2015 IEEE 56th Annual Symposium on Foundations of Computer Science

The Power of Asymmetry in Constant-Depth Circuits*

Alexander A. Sherstov†
Computer Science Department
University of California, Los Angeles
Los Angeles, CA 90095 USA
sherstov@cs.ucla.edu

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 a novel approximation-theoretic technique of independent interest, which exploits asymmetry in circuits to prove their hardness for polynomials.

Keywords

Constant-depth circuits; threshold degree; polynomial representations of Boolean functions; polynomial threshold functions; polynomial approximation; communication complexity; computational learning theory

I. 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 real 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. The formal study of this complexity measure began in 1969 with the pioneering work of Minsky and Papert [20]. Motivated by applications to neural networks, the authors of [20] 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 found applications far beyond artificial intelligence. In theoretical computer science, sign-representing polynomials have played a key role in a variety of contexts, from circuit lower bounds [18], [19] and size-depth trade-offs [23], [35] to the closure of PP under intersection [7].

The notion of threshold degree has been especially influential in the study of AC^0, the class of circuits of constant depth and polynomial size with $\wedge, \vee, \neg$ gates. Aspnes et al. [4] used sign-representing polynomials to give an entirely different proof of classic lower bounds for AC^0. In communication complexity, the notion of threshold degree was critical to constructing an AC^0 circuit with exponentially small discrepancy and hence maximum communication complexity in nearly every model [25], [26]. That discrepancy result was used in [25] to show the optimality of Allender's classic simulation of AC^0 by majority circuits, solving the open problem [18] on the relation between the two circuit classes. A more sophisticated application of threshold degree was the first exponential lower bound on the sign-rank of AC^0 circuits [24], obtained twenty-two years after the problem was posed by Babai et al. [5]. Subsequent work [12], [6], [27], [33] resolved other questions in communication complexity and circuit complexity related to constant-depth circuits by generalizing the threshold degree method of [25], [26].

Sign-representing polynomials have also enabled algorithmic breakthroughs in the study of constant-depth circuits. One such example is the current fastest algorithm for learning DNF formulas, due to Klivans and Servedio [16], with running time $\exp(\tilde{O}(n^{1/3}))$. The authors of [16] 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

* The full version of this paper with complete proofs is available online [34] as an ECCC technical report.

† Supported by NSF CAREER award CCF-1149018 and an Alfred P. Sloan Foundation Research Fellowship.

0272-5428/15 $31.00 © 2015 IEEE
DOI 10.1109/FOCS.2015.34

Boolean formulas, obtained by O'Donnell and Servedio [22] for formulas of constant depth and by Ambainis et al. [3] for formulas of 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.

A. Our results

A longstanding open problem in the area is to determine the maximum threshold degree of an AC^0 circuit. This problem is motivated by algorithmic and complexity-theoretic applications [16], [22], [17], [24], [10], in addition to being a natural question in its own right. The progress to date is summarized in Table I.1. In their seminal monograph, Minsky and Papert [20] 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 [16] obtained an upper bound of $O(n^{1/3}\log n)$ on the threshold degree of polynomial-size DNF formulas 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. Until recently, the only progress on this question was due to O'Donnell and Servedio [22], who proved a threshold degree lower bound of $\Omega(n^{1/3}\log^{2(d-2)/3} n)$ for circuits of depth d. The authors of [22] formally posed the challenge of obtaining a polynomial improvement on Minsky and Papert's lower bound. Such an improvement was obtained last year in [32], with a threshold degree lower bound of $\Omega(n^{(d-1)/(2d-1)})$ for circuits of depth d. In particular, the result in [32] subsumes all previous lower bounds, with a strict improvement starting at depth $d = 3$. The main contribution of this paper is a polynomially stronger lower bound for every depth $d \geqslant 3$. For depth 3, we obtain:

THEOREM I.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 $O(n^{(2^{d-1}-1)/(2^d-1)})$ on the threshold degree of formulas of depth d and size n, due to O'Donnell and Servedio [22]. Prior to our work, that upper bound was only known to be tight for $d = 1$ and $d = 2$, by the classic results of Minsky and Papert [20]. Theorem I.1 suggests that O'Donnell and Servedio's upper bound is tight for all d, a fascinating possibility.

A comment is in order on the admittedly unusual formula f in Theorem I.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 bound in Theorem I.1. In particular, the previous lower bound of $\Omega(n^{2/5})$ for depth-3 formulas, obtained

Depth	Threshold degree	Reference
2	$\Omega(n^{1/3})$	Minsky and Papert [20]
d	$\Omega(n^{1/3}\log^{2(d-2)/3} n)$	O'Donnell and Servedio [22]
d	$\Omega(n^{\frac{d-1}{2d-1}})$	Sherstov [32]
3	$\Omega(n^{3/7})$	This paper
4	$\Omega(\sqrt{n})$	This paper

Table I.1: Lower bounds on the threshold degree of $\wedge, \vee, \neg$-circuits of polynomial size and constant depth. In all bounds, n denotes the number of variables.

in [32], was shown in that paper to be tight for formulas of the form $f(x) = \bigvee_{i=1}^{n_1} \bigwedge_{j=1}^{n_2} \bigvee_{k=1}^{n_3} x_{i,j,k}$ for any n_1, n_2, n_3. Asymmetry plays the same critical role in all of our constructions, as we will shortly elaborate 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 I.1:

THEOREM I.2 (Main result, $d \geqslant 4$). *There is an explicitly given $\wedge, \vee$-circuit $f: \{0,1\}^n \to \{0,1\}$ of depth 4 and polynomial size such that*

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

As we discuss in Remark VIII.5 at the end of the paper, one can use O'Donnell and Servedio's technique [22] to improve the lower bound of Theorem I.2 by an arbitrary polylogarithmic factor, at the expense of increasing the circuit depth by a constant. Theorem I.2 solves a recent open problem due to Bun and Thaler [10] and Thaler [11], 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 I.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 I.2 above and restated in technical detail as Theorem VIII.4 at the end of the paper, gives an $\wedge, \vee$-circuit $f: \{0,1\}^n \to \{0,1\}$ of depth 4 and polynomial size with discrepancy $\exp(-\Omega(\sqrt{n}))$ and threshold weight $\exp(\Omega(\sqrt{n}))$. The best previous bounds [32] 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, where $d \geqslant 2$ stands for the depth of the circuit. This passage from threshold degree to discrepancy and threshold weight uses the pattern matrix method [26], a by-now standard technique. We refer the interested reader to [32, Section 8] for details, including the definitions of discrepancy and threshold weight.

B. Proof overview

At first glance, Theorems I.1 and I.2 seem unrelated. In reality, they are corollaries to a more general result that we prove. A key notion here is that of *one-sided approximate degree*, defined for a Boolean function $f: \{0,1\}^n \to \{0,1\}$ 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 a small constant, with $\epsilon = 1/3$ being the canonical setting. One-sided approximate degree has played an important role in the area [16], [14], [9], [28], [10], [32], with applications to both complexity theory and algorithms. It is related to threshold degree in a straightforward way: if p is a one-sided approximant for a given function, then $p - \frac{1}{2}$ is a sign-representing polynomial for the same function. 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 related function with proportionately high threshold degree. This transformation allows us to obtain our threshold degree lower bounds by working in the significantly simpler setting of one-sided approximation. Quantitatively, our hardness amplification theorem transforms any given circuit $f: \{0,1\}^n \to \{0,1\}$ with one-sided approximate degree n^{α} in a black-box manner into a polynomially larger circuit $F: \{0,1\}^N \to \{0,1\}$ with threshold degree $\deg_{\pm}(F) = \Omega(N^{\beta})$, where $\beta = \beta(\alpha)$ is the monotonically increasing function given by

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

THEOREM I.3 (Hardness amplification). *Let $f: \{0,1\}^n \to \{0,1\}$ be a given Boolean function, with approximate degree n^{α}. Then there is an explicitly given function $F: \{0,1\}^N \to \{0,1\}$ of the form*

$$F = \mathrm{OR}_{m_1} \circ ((\mathrm{AND}_{m_2} \circ \neg f) \wedge (\mathrm{AND}_{m_3} \circ \mathrm{OR}_{m_4}))$$

with threshold degree

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

where $\beta = \beta(\alpha)$ is given by (I.1).

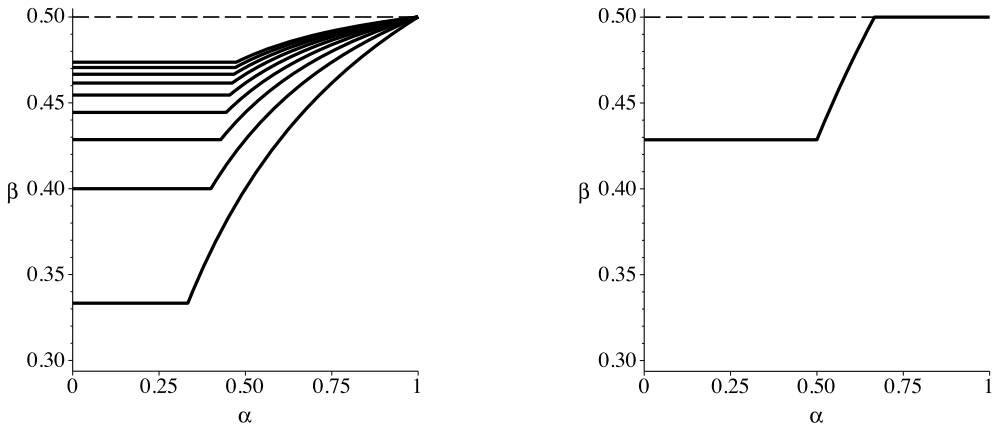

Figure I.1: Transforming a circuit $f\colon\{0,1\}^n \to \{0,1\}$ with one-sided approximate degree n^α into a polynomially larger circuit $F\colon\{0,1\}^N \to \{0,1\}$ with threshold degree $\Omega(N^\beta)$. The graphs plot the dependence $\beta = \beta(\alpha)$ in previous work and this paper. *Left:* the best previous construction [32], 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.

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, the transformed function F is 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 I.3 and known lower bounds on the one-sided approximate degree of constant-depth circuits. Specifically, we obtain Theorem I.1 by applying our hardness amplification result to the function $f = \mathrm{NOR}_n$, whose one-sided approximate degree is $\Theta(\sqrt{n})$. To obtain Theorem I.2, we instead let f be a certain polynomial-size CNF formula with one-sided approximate degree $\Omega(n^{2/3})$.

We find Theorem I.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 [32, 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 I.3 always corresponds to an increase in circuit depth by 2. Figure I.1 compares the previous hardness amplification result from [32] with Theorem I.3 in this paper, plotting the dependence $\beta = \beta(\alpha)$ in the two cases. As the figure shows, we strictly improve on previous work for $d = 1$ and $d = 2$, 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 I.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 as well as new intermediate notions of approximation.

1) By hypothesis, the original function f does not have a low-degree one-sided approximating polynomial. In the notation of Theorem I.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.
2) 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.
3) 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 role of asymmetry in this step.

4) Finally, we invoke a result from previous work [32] that characterizes the threshold degree of a disjunction of functions in terms of the one-sided rational approximation of the individual functions.

Steps 1, 2, and 3 in this program correspond to Sections VI, VII, and V, respectively. These components are put together in Section VIII, which completes the proof. We provide additional details and intuition at each stage of the proof, and conclude the paper by discussing the potential of our technique to give stronger bounds.

II. Preliminaries

A. 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, the notation $\log x$ 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 $e_1, e_2, \ldots, e_n$ correspond to the standard basis for $\mathbb{R}^n$. For a linear subspace S, we let $S^\perp$ 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 $x \in X$.

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 $x \in \{0,1\}^n$.

B. 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 all 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 n^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.

C. 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 $\lambda_1 f_1 + \lambda_2 f_2 + \cdots + \lambda_k f_k$, where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are nonnegative and sum to 1. The *convex hull* of $F \subseteq \mathbb{R}^X$, denoted $\operatorname{conv} F$, is the set of all convex combinations of functions in F.

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

D. 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, an 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 that uniform approximation. Formally, we have $\deg_\pm(f) = \lim_{\epsilon \nearrow 1/2} \deg_\epsilon(f)$. In particular, $\deg_\pm(f) \leqslant \deg_\epsilon(f)$ for $0 \leqslant \epsilon < 1/2$.

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 $f \colon \{0,1\}^n \to \mathbb{R}$ 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 ϵ-approximate degree are invariant under negation:

$$\deg_{\pm}(f) = \deg_{\pm}(\neg f), \tag{II.1}$$

$$\deg_{\epsilon}(f) = \deg_{\epsilon}(\neg f) \tag{II.2}$$

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

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

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 II.1 makes it clear, among other things, that the canonical constant $\epsilon = 1/3$ in the definition of one-sided approximate degree could have been 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 separately and compose the resulting approximants. For this approach to work, the approximating polynomial for f needs to be robust to noise in the inputs, i.e., it needs to approximate f not only on the Boolean hypercube but also on any perturbation of a Boolean vector. The following result from [30] gives an optimal procedure for making any polynomial robust to noise in the inputs.

THEOREM II.2 (Sherstov). *Let $p: \{0,1\}^n \to [-1,1]$ be a given polynomial. Then for every $\delta > 0$, there is a polynomial $p_{\text{robust}}: \mathbb{R}^n \to \mathbb{R}$ of degree $O(\deg p + \log \frac{1}{\delta})$ such that*

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

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

E. 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 [21], was one of the first results in this line of work.

THEOREM II.3 (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. [8] and de Wolf [13] generalized Nisan and Szegedy's theorem to an arbitrary error parameter ϵ. For our purposes, only the upper bound is needed.

THEOREM II.4 (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 [10], [32]. The following tight lower bound is due to Ambainis [2].

THEOREM II.5 (Ambainis).

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

Bun and Thaler [10] 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. As a consequence, Bun and Thaler [10] obtained a

polynomial-size CNF formula in n variables with one-sided approximate degree $\Omega(n/\log n)^{2/3}$. We are able to improve on this bound by a logarithmic factor:

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

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

for a suitable (explicitly given) mapping $\phi\colon \{0,1\}^{6\lceil \log n \rceil} \to \{0,1\}$. *Then*

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

Moreover, F *is computable by a CNF formula of polynomial size.*

The proof of this theorem is available in the full version of the paper [34, Section 3].

III. One-sided rational approximation

We now review one-sided rational approximation of Boolean functions, studied recently in [32]. Let $f\colon X \to \{0,1\}$ be a Boolean function of interest, $d_0, d_1 \geqslant 0$ given reals. Following [32], 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

1) $|p_1| < \epsilon p_0$ on $f^{-1}(0)$,
2) $|p_0| < \epsilon p_1$ on $f^{-1}(1)$,
3) $\deg p_0 \leqslant d_0$,
4) $\deg p_1 \leqslant d_1$.

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 $f^{-1}(1)$:

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

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

$$R(\mathrm{OR}_n, 0, 1) = 0. \tag{III.2}$$

An analogous argument shows that

$$R(\mathrm{AND}_n, 1, 0) = 0. \tag{III.3}$$

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,

$$\lim_{\epsilon \searrow 0} \left\| \mathrm{AND}_n - \frac{\epsilon}{\epsilon + \sum (1 - x_i)} \right\|_\infty = 0, \tag{III.4}$$

$$\lim_{\epsilon \searrow 0} \left\| \mathrm{OR}_n - \frac{\sum x_i}{\epsilon + \sum x_i} \right\|_\infty = 0. \tag{III.5}$$

These results on rational approximation should be contrasted with Theorem II.3, which states that approximating the AND_n function even in the one-sided sense requires a polynomial of degree $\Omega(\sqrt{n})$.

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

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

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

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 $f^{-1}(1)$. ∎

A. 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. [7] 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 [15], [29], [31]. The tightest and most recent version is as follows [32], stated in the terminology of this paper.

THEOREM III.2 (cf. Beigel et al.). *Let $f: X \to \{0,1\}$ be given. Then for all ℓ,*

$$\deg_\pm(\mathrm{OR}_\ell \circ f) \leqslant 2 \min_{d_0,d_1} \left\{ \ell d_0 + d_1 : R(f,d_0,d_1) < \frac{1}{\sqrt[4]{2\ell}} \right\}. \tag{III.6}$$

In particular,

$$\deg_\pm(\mathrm{OR}_\ell \circ f) \leqslant 2 \min_{d_0,d_1} \left\{ \ell d_0 + d_1 : R(f,d_0,d_1) < \frac{1}{2} \right\} \left\lceil \frac{1+\log \ell}{4} \right\rceil. \tag{III.7}$$

It was recently shown in [32] that Theorem III.2 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 [32, Theorem 6.7].

THEOREM III.3 (Sherstov). *Let $d_0, d_1 \geqslant 0$ be integers, $f: X \to \{0,1\}$ a given Boolean function. If $R(f,d_0,d_1) > \epsilon$, then*

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

B. 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 [32, Theorem 6.4].

THEOREM III.4 (Sherstov). *Let $f: X \to \{0,1\}$ be a given Boolean function, $d_0, d_1 \geqslant 0$. Then for every $\epsilon > 0$, the nonexistence of polynomials p_0, p_1 such that*

1) $|p_1| < \epsilon p_0$ *on* $f^{-1}(0)$,
2) $|p_0| < \epsilon p_1$ *on* $f^{-1}(1)$,
3) $\deg p_0 \leqslant d_0$,
4) $\deg p_1 \leqslant d_1$,

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

5) $\phi_0 \geqslant \epsilon |\phi_1|$ *on* $f^{-1}(0)$,
6) $\phi_1 \geqslant \epsilon |\phi_0|$ *on* $f^{-1}(1)$,
7) $\deg p \leqslant d_0 \implies \langle \phi_0, p \rangle = 0$,
8) $\deg p \leqslant d_1 \implies \langle \phi_1, p \rangle = 0$,
9) $\phi_0 \not\equiv 0$,
10) $\phi_1 \not\equiv 0$.

As a corollary, we have the following dual characterization of one-sided rational approximation [32, Corollary 6.5].

COROLLARY III.5 (Sherstov). *Let $f: X \to \{0,1\}$ be a given function, $R(f,d_0,d_1) > 0$. Then $R(f,d_0,d_1)$ is the supremum over all $\epsilon > 0$ for which there exist $\phi_0, \phi_1: X \to \mathbb{R}$ with*

1) $\phi_0 \geqslant \epsilon |\phi_1|$ *on* $f^{-1}(0)$,
2) $\phi_1 \geqslant \epsilon |\phi_0|$ *on* $f^{-1}(1)$,
3) $\deg p \leqslant d_0 \implies \langle \phi_0, p \rangle = 0$,
4) $\deg p \leqslant d_1 \implies \langle \phi_1, p \rangle = 0$,
5) $\phi_0 \not\equiv 0$,
6) $\phi_1 \not\equiv 0$.

IV. 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 $f: X \to \{0,1\}$. 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

1) $|p_1| < \epsilon p_0$ on $f^{-1}(0)$,
2) $p_0 \in (\frac{\epsilon}{\Delta} p_1, \epsilon p_1)$ on $f^{-1}(1)$,
3) $\deg p_0 \leqslant d_0$,
4) $\deg p_1 \leqslant d_1$.

A moment's reflection shows that the feasibility of 1)–4) is monotonic in ϵ, in the sense that increasing ϵ can only make it easier to satisfy 1)–4). As a result, $R_\Delta(f, d_0, d_1)$ is always well-defined and ranges in $[0, 1]$. The quotient of polynomials in the above definition obeys

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

A. Relation to one-sided approximation

It is helpful to contrast (IV.1) with its counterpart (III.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 IV.1. *Let $f: X \to \{0,1\}$ be given. Then for all $d_0, d_1 \geqslant 0$,*

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

$$R(f, d_0, d_1) \geqslant R(f, d_0, d_1)^2 \geqslant \lim_{\Delta \to +\infty} R_\Delta(f, 2d_0, 2d_1). \tag{IV.3}$$

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

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

trivially imply

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

which proves (IV.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

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

we arrive at

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

which yields $\lim_{\Delta \to +\infty} R_\Delta(f, 2d_0, 2d_1) \leqslant R(f, d_0, d_1)^2$. This directly implies (IV.3) since $R(f, d_0, d_1) \in [0, 1]$ by definition. ∎

B. 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 III.4).

THEOREM IV.2. *Let $f: X \to \{0,1\}$ 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*

1) $|p_1| < \epsilon p_0$ *on* $f^{-1}(0)$,
2) $p_0 \in (\frac{\epsilon}{\Delta} p_1, \epsilon p_1)$ *on* $f^{-1}(1)$,
3) $\deg p_0 \leqslant d_0$,
4) $\deg p_1 \leqslant d_1$

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

6) $\phi_0 \geqslant \epsilon |\phi_1|$ *on* $f^{-1}(0)$,
7) $\phi_1 \geqslant \epsilon \max\{-\phi_0, -\frac{1}{\Delta}\phi_0\}$ *on* $f^{-1}(1)$,
8) $\deg p \leqslant d_0 \implies \langle \phi_0, p \rangle = 0$,
9) $\deg p \leqslant d_1 \implies \langle \phi_1, p \rangle = 0$,
10) $\phi_0 \not\equiv 0$,
11) $\phi_1 \not\equiv 0$.

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 1) and 2) can be rewritten as

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

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

$$\epsilon^{1-f} \mu - (-\epsilon)^{1-f} \lambda \in P_0^{\perp}, \tag{IV.4}$$
$$(-\tfrac{\epsilon}{\Delta})^f \mu - (-\epsilon)^f \lambda \in P_1^{\perp}. \tag{IV.5}$$

By basic arithmetic, the existence of such μ and λ is in turn equivalent to the existence of $\phi_0, \phi_1: X \to \mathbb{R}$, not both identically zero, that obey 6)–9), where we identify ϕ_0 and ϕ_1 with the left-hand side of (IV.4) and (IV.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 6)–7) 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 IV.3. *Let $f: X \to \{0,1\}$ 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 $\phi_0, \phi_1: X \to \mathbb{R}$ with*

1) $\phi_0 \geqslant \epsilon |\phi_1|$ *on* $f^{-1}(0)$,
2) $\phi_1 \geqslant \epsilon \max\{-\phi_0, -\frac{1}{\Delta}\phi_0\}$ *on* $f^{-1}(1)$,
3) $\deg p \leqslant d_0 \implies \langle \phi_0, p \rangle = 0$,
4) $\deg p \leqslant d_1 \implies \langle \phi_1, p \rangle = 0$,
5) $\phi_0 \not\equiv 0$,
6) $\phi_1 \not\equiv 0$.

V. 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 III.3, which gives a clean characterization of the threshold degree of every such composition. Specifically, that theorem shows that the threshold degree of $\mathrm{OR}_\ell \circ F$ is large if and only if F does not have a low-degree one-sided rational approximant with constant error. Quantitatively,

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

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.

A. 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 (V.1) 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 $0 \leqslant \epsilon \leqslant 1/3$ 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

$$\tilde{f} \cdot \tilde{g}. \tag{V.2}$$

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 D and denominator degree d. Another, incomparable approach is to appeal to DeMorgan's law and approximate $f \wedge g$ by

$$\frac{1}{\frac{1}{\tilde{f}^2} + \frac{1}{\tilde{g}^2}}, \tag{V.3}$$

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 $f \wedge g$.

B. The lower bound

Given the restricted form of (V.2) and (V.3), 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 V.1. *Let* $f: X \to \{0,1\}$ *and* $g: Y \to \{0,1\}$ *be given functions,* $0 < \epsilon \leqslant 1$, *and* $\Delta > 1$. *Assume that there exist* $\phi_0, \phi_1', \phi_1'': X \to \mathbb{R}$ *such that*

$$\phi_0 \geqslant \epsilon |\phi_1'| \text{ on } f^{-1}(0), \tag{V.4}$$
$$\phi_0 \geqslant \epsilon |\phi_1''| \text{ on } f^{-1}(0), \tag{V.5}$$
$$\phi_1' \geqslant \epsilon \max\{-\phi_0, -\tfrac{1}{\Delta}\phi_0\} \text{ on } f^{-1}(1), \tag{V.6}$$
$$\phi_1'' \geqslant \epsilon |\phi_0| \text{ on } f^{-1}(1), \tag{V.7}$$
$$\deg p \leqslant d \implies \langle \phi_0, p \rangle = 0, \tag{V.8}$$
$$\deg p \leqslant D \implies \langle \phi_1', p \rangle = 0, \tag{V.9}$$
$$\deg p \leqslant d \implies \langle \phi_1'', p \rangle = 0, \tag{V.10}$$
$$\phi_0 \not\equiv 0, \tag{V.11}$$
$$\phi_1' \not\equiv 0, \tag{V.12}$$
$$\phi_1'' \not\equiv 0. \tag{V.13}$$

Assume furthermore that

$$R(g, d, D) > \frac{\epsilon}{\sqrt{\Delta}}. \tag{V.14}$$

Then

$$R\left(f \wedge g, \frac{d}{2}, \frac{D}{2}\right) \geqslant \frac{\epsilon}{\sqrt{2}}. \tag{V.15}$$

The statement of Theorem V.1 is admittedly technical but its intuitive content is satisfying and easy to explain. The conditions (V.4), (V.6), (V.8), (V.9), (V.11), (V.12), (V.14) can be summarized as

$$\left.\begin{aligned} R_\Delta(f, d, D) &> \frac{\epsilon}{\sqrt{\Delta'}} \\ R_{\Delta'}(g, d, D) &> \frac{\epsilon}{\sqrt{\Delta}} \end{aligned}\right\} \text{ for every } \Delta' > 1, \tag{V.16}$$

by the dual characterization of hybrid rational approximation (Theorem IV.2 and Corollary IV.3). The remaining conditions (V.5), (V.7), (V.8), (V.10), (V.11), (V.13) are equivalent to

$$R(f, d, d) \geqslant \epsilon, \tag{V.17}$$

by the dual characterization of one-sided rational approximation (Theorem III.4 and Corollary III.5). The hypothesis (V.16) rules out an approximant for $f \wedge g$ of the form (V.2), whereas the hypothesis (V.17) rules out an approximant of the form (V.3). The theorem states, informally, 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 of $\Delta', \Delta'' > 1$. 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 (V.2).

The proof of Theorem V.1 is available in the full version of this paper [34, Section 6.2].

VI. 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 VI.1, is a hardness amplification result that implies

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 the hybrid *rational* approximation of related functions.

THEOREM VI.1. *Let* $f: X \to \{0,1\}$ *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 $\Phi_0, \Phi_1', \Phi_1'': X^{cn} \to \mathbb{R}$ *such that:*

1) $\Phi_1' \geqslant (1-\epsilon)\max\{-\Phi_0, -2^{-n}\Phi_0\}$ *on* $F^{-1}(1)$,
2) $\Phi_1'' \geqslant (1-\epsilon)|\Phi_0|$ *on* $F^{-1}(1)$,
3) $\Phi_0 \geqslant (1-\epsilon)\max\{|\Phi_1'|, |\Phi_1''|\}$ *on* $F^{-1}(0)$,
4) $\langle \Phi_1', P\rangle = 0$ *whenever* $\deg P \leqslant \frac{1}{c}\deg^+_{1/3}(\neg f)\sqrt{n}$,
5) $\langle \Phi_0, P\rangle = \langle \Phi_1'', P\rangle = 0$ *whenever* $\deg P \leqslant \min\{\frac{1}{c}\deg^+_{1/3}(\neg f), \frac{1}{2}n\}$,
6) $\Phi_0 \not\equiv 0$,
7) $\Phi_1' \not\equiv 0$,
8) $\Phi_1'' \not\equiv 0$.

The conclusion of Theorem VI.1 is easiest to understand in terms of the dual characterization of one-sided and hybrid rational approximation (Corollaries III.5 and IV.3, respectively). Specifically, properties 1)–8) are equivalent 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 VI.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 VI.1 is an adaptation of a recent hardness amplification result in [32, 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 several 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 f's one-sided approximate degree, and the end result is the triple of dual objects $\Phi_0, \Phi_1', \Phi_1''$ in the theorem statement. The intermediate building blocks in our proof are all borrowed from [32], but we are able to combine them in a way that is considerably simpler and more intuitive.

The proof of Theorem VI.1 is available in the full version of this paper [34, Section 7].

VII. 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 the function $\mathrm{AND}_n \circ \mathrm{OR}_r$ 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 VII.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 II.4, there is a polynomial $p: \{0,1\}^n \to [0,1]$ of degree $O(\sqrt{nr})$ with

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

Theorem II.2 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 $p_{\text{robust}}\colon \mathbb{R}^n \to \mathbb{R}$ of degree $O(\sqrt{nr})$ with

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

Again by Theorem II.4, there is a degree-$O(\sqrt{r})$ polynomial q with $\|\mathrm{OR}_r - q\|_\infty \leqslant 1/3$. By (VII.1) and (VII.2), the composed polynomial $p_{\text{robust}} \circ q$ satisfies $\|\mathrm{AND}_n \circ \mathrm{OR}_r - p_{\text{robust}} \circ q\|_\infty \leqslant 2^{-r-1} + 2^{-\sqrt{nr}-1} \leqslant 2^{-r}$. In particular, $p_{\text{robust}} \circ q$ is a one-sided approximant for $\mathrm{AND}_n \circ \mathrm{OR}_r$ with error 2^{-r}. This completes the proof since $p_{\text{robust}} \circ q$ has degree $\deg(p_{\text{robust}}) \deg(q) = O(r\sqrt{n}) = O(\min\{r\sqrt{n}, n\})$.

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 VII.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: 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 VII.2 (Lower bound). *There is an absolute constant $c > 0$ such that*

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

The proof of Theorem VII.2 is available in the full version of this paper [34, Section 8].

VIII. MAIN RESULTS

The technical centerpiece of this paper is a hardness amplification technique 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 $F = \mathrm{OR}_\ell \circ ((\mathrm{AND}_k \circ f) \wedge g)$, where g is an auxiliary function to which we refer as the "amplifier." If the original function has one-sided approximate degree $\deg^+_{1/3}(f) \geqslant n^\alpha$, then the transformed function has threshold degree $\deg_\pm(F) = \Omega(N^\beta)$ for some monotonically growing exponent $\beta = \beta(\alpha)$ that depends on g. We formalize our technique in this generality in Section VIII-A. In Section VIII-B, we specialize the amplifier g to be a read-once formula of depth 2 and prove that the resulting construction achieves

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

As corollaries, we obtain our main lower bounds on the threshold degree of AC^0 circuits and read-once formulas, by choosing f accordingly in each case. In Sections VIII-C, we prove matching upper bounds. In the concluding Section VIII-D, we discuss the limitations of our technique and propose directions for future work.

A. 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 III and IV, the composition theorem from Section V, and the lower bound on hybrid rational approximation from Section VI.

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

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

Then

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

and

$$\deg_\pm(\mathrm{OR}_\ell \circ ((\mathrm{AND}_k \circ \neg f) \wedge g)) \geqslant c\min\{\ell k, \ell(d+1), \ell\deg^+_{1/3}(f), \sqrt{k}\deg^+_{1/3}(f), D\} \tag{VIII.4}$$

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

Proof: The lower bound (VIII.4) on the threshold degree is a direct consequence of (VIII.3) and Theorem III.3. Thus, it suffices to prove (VIII.3).

For some absolute integer constant $c_0 \geqslant 1$ and all $k \geqslant c_0$, Theorem VI.1 gives functions $\Phi_0, \Phi_1', \Phi_1'' \colon (\operatorname{dom} f)^k \to \mathbb{R}$ such that

$$\Phi_0 \geqslant \frac{1}{2}|\Phi_1'| \text{ on } (\mathrm{AND}_k \circ \neg f)^{-1}(0), \tag{VIII.5}$$

$$\Phi_0 \geqslant \frac{1}{2}|\Phi_1''| \text{ on } (\mathrm{AND}_k \circ \neg f)^{-1}(0), \tag{VIII.6}$$

$$\Phi_1' \geqslant \frac{1}{2}\max\{-\Phi_0, -2^{-k/c_0}\Phi_0\} \text{ on } (\mathrm{AND}_k \circ \neg f)^{-1}(1), \tag{VIII.7}$$

$$\Phi_1'' \geqslant \frac{1}{2}|\Phi_0| \text{ on } (\mathrm{AND}_k \circ \neg f)^{-1}(1), \tag{VIII.8}$$

$$\langle \Phi_0, P \rangle = 0 \text{ whenever } \deg P \leqslant \frac{1}{2c_0}\min\{\deg^+_{1/3}(f), k\}, \tag{VIII.9}$$

$$\langle \Phi_1', P \rangle = 0 \text{ whenever } \deg P \leqslant \frac{1}{c_0\sqrt{c_0}}\deg^+_{1/3}(f)\sqrt{k}, \tag{VIII.10}$$

$$\langle \Phi_1'', P \rangle = 0 \text{ whenever } \deg P \leqslant \frac{1}{2c_0}\min\{\deg^+_{1/3}(f), k\}, \tag{VIII.11}$$

$$\Phi_0, \Phi_1', \Phi_1'' \not\equiv 0. \tag{VIII.12}$$

On the other hand, it follows from (VIII.2) and the error-reduction property of rational approximation (Proposition III.1) that

$$R\left(g, \frac{d}{c_0}, \frac{D}{c_0}\right) > 2^{-k/c_0}. \tag{VIII.13}$$

Applying Theorem V.1 to (VIII.5)–(VIII.13) with parameters $\epsilon = 1/2$ and $\Delta = 2^{-k/c_0}$, we infer that

$$R\left((\mathrm{AND}_k \circ \neg f) \wedge g, \frac{1}{2c_0}\min\{k, d, \deg^+_{1/3}(f)\}, \frac{1}{c_0\sqrt{c_0}}\min\{D, \deg^+_{1/3}(f)\sqrt{k}\}\right) \geqslant \frac{1}{2\sqrt{2}} \tag{VIII.14}$$

for all $k \geqslant c_0$.

We are now in a position to prove that (VIII.3) holds with $c = 1/\max\{6c_0, 3c_0\sqrt{c_0}\}$. For $k \geqslant c_0$, the bound follows directly from (VIII.14) and the error-reduction property (Proposition III.1). To prove validity in the complementary case $k < c_0$, observe that the left-hand side of (VIII.3) is trivially bounded from below by $R((\mathrm{AND}_k \circ \neg f) \wedge g, cd, cD)$. The latter quantity is easily seen to be large:

$$\begin{aligned}
R((\mathrm{AND}_k \circ \neg f) \wedge g, cd, cD) &\geqslant R(g, cd, cD) && \text{since } f \not\equiv 0 \\
&\geqslant R\left(g, \frac{d}{2k}, \frac{D}{2k}\right) && \text{since } c \leqslant \frac{1}{2c_0} < \frac{1}{2k} \\
&\geqslant R(g, d, D)^{1/(2k)} && \text{by Proposition III.1} \\
&> \frac{1}{\sqrt{2}} && \text{by (VIII.2).}
\end{aligned}$$

■

B. Results using depth-2 amplifiers

We now establish the main results of our paper by invoking Theorem VIII.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 [22] obtained an upper bound of $O(N^{(2^{d-1}-1)/(2^d-1)})$ on the threshold degree of any $\wedge, \vee$-formula of depth d and size N. This bound was only known to be tight for $d \leqslant 2$, by the classic results of Minsky and Papert [20]. We are able 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 [32], was polynomially weaker: $\Omega(N^{2/5})$.

THEOREM VIII.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 I.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 VIII.2. Specifically, we showed in [32] that all formulas of the form $\mathrm{OR}_{N_1} \circ \mathrm{AND}_{N_2} \circ \mathrm{OR}_{N_3}$ on $N = N_1 N_2 N_3$ variables have threshold degree $O(N^{2/5} \log N)$.

Proof of Theorem VIII.2*:* Theorem VII.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 II.3 states that

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

for $f = \mathrm{NOR}_n$. Appealing to Theorem VIII.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 $f\colon \{0,1\}^n \to \{0,1\}$ with given 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 $F = \mathrm{OR}_\ell \circ ((\mathrm{AND}_k \circ f) \wedge g)$ for some parameters k and ℓ and some read-once formula g of depth 2. In particular, the transformation $f \mapsto F$ preserves membership in AC^0 as well as the read-once property. The result that we are about to state subsumes Theorem VIII.2 on the threshold degree of constant-depth formulas and settles Theorem I.3 from the Introduction.

THEOREM VIII.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 VIII.2, and we focus on the complementary case $\alpha \geqslant 1/2$. Theorem VII.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 VIII.1 with parameters $d = cn^{\alpha}$, $D = cn^{\alpha}\sqrt{n^{\alpha}}$, $k = n^{\alpha}$, and $\ell = \sqrt{n^{\alpha}}$, we obtain $\deg_{\pm}(F) = \Omega(n^{\alpha}\sqrt{n^{\alpha}}) = \Omega(\min\{N^{3\alpha/(3\alpha+2)}, \sqrt{N}\})$. ∎

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 I.2.

THEOREM VIII.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 $\phi\colon \{0,1\}^{6\lceil \log n\rceil} \to \{0,1\}$ *is as constructed in Theorem* II.6. *Then*

$$\deg_{\pm}(F) = \Omega(\sqrt{N}). \tag{VIII.15}$$

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

Proof: Recall from Theorem II.6 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, (VIII.15) follows directly from Theorem VIII.3. Theorem II.6 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. ∎

REMARK VIII.5. The above lower bound on the threshold degree of AC^0 can be strengthened by an arbitrary polylogarithmic factor at the expense of increasing the circuit depth by a constant. In more detail, O'Donnell and Servedio [22] proved that $\deg_{\pm}(F \circ \mathrm{XOR}_k) = k \deg_{\pm}(F)$ 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 $\log^c n$ by composing the function in Theorem VIII.4 with $\mathrm{XOR}_{\log^c n}$.

C. Tightness for degree-2 amplifiers

In Theorem VIII.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 VIII.4, is tight up to a logarithmic factor.

THEOREM VIII.6. *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 $N = 2\ell k n$ *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 VIII.4. In particular, Theorem VIII.6 shows that the threshold degree lower bound in our main result (Theorem VIII.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 VIII.6 shows that the square root barrier in our hardness amplification technique (Theorem VIII.3) is inherent due to the possibility of a large gap between the one-sided approximate degree of a function and that of its negation.

The proof of Theorem VIII.6 is available in the full version of this paper [34, Section 9.3].

D. Tightness for arbitrary amplifiers

In this final section, we explore the limitations of Theorem VIII.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 $N \geqslant \ell k n$ variables. It is clear from the statement of Theorem VIII.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,

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

To summarize, Theorem VIII.1 cannot give a threshold degree lower bound asymptotically superior to

$$\max\{N^{1/3}, N^{3\alpha/(3\alpha+2)}\} \tag{VIII.16}$$

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

Recall that Theorem VIII.3 in this paper actually achieves (VIII.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 (VIII.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 (VIII.16) for all α, it is natural to wonder how to go *beyond* this barrier. In other words, given an $\wedge, \vee$-circuit $f\colon\{0,1\}^n \to \{0,1\}$ with one-sided approximate degree n^α, we would like to construct a related $\wedge, \vee$-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.

References

[1] S. Aaronson and Y. Shi, "Quantum lower bounds for the collision and the element distinctness problems," *J. ACM*, vol. 51, no. 4, pp. 595–605, 2004.

[2] A. Ambainis, "Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range," *Theory of Computing*, vol. 1, no. 1, pp. 37–46, 2005.

[3] A. Ambainis, A. M. Childs, B. Reichardt, R. Špalek, and S. Zhang, "Any AND-OR formula of size N can be evaluated in time $N^{1/2+o(1)}$ on a quantum computer," in *Proceedings of the Forty-Eighth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2007, pp. 363–372.

[4] J. Aspnes, R. Beigel, M. L. Furst, and S. Rudich, "The expressive power of voting polynomials," *Combinatorica*, vol. 14, no. 2, pp. 135–148, 1994.

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

[6] P. Beame and T. Huynh, "Multiparty communication complexity and threshold circuit size of AC^0," *SIAM J. Comput.*, vol. 41, no. 3, pp. 484–518, 2012.

[7] R. Beigel, N. Reingold, and D. A. Spielman, "PP is closed under intersection," *J. Comput. Syst. Sci.*, vol. 50, no. 2, pp. 191–202, 1995.

[8] 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.

[9] M. Bun and J. Thaler, "Dual lower bounds for approximate degree and Markov-Bernstein inequalities," in *Proceedings of the Fortieth International Colloquium on Automata, Languages and Programming* (ICALP), 2013, pp. 303–314.

[10] ——, "Hardness amplification and the approximate degree of constant-depth circuits," in *Electronic Colloquium on Computational Complexity (ECCC)*, 2013, report TR13-151.

[11] ——, "Lower bounds for the approximate degree of block-composed functions," in *Electronic Colloquium on Computational Complexity (ECCC)*, 2014, report TR14-150.

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

[13] R. de Wolf, "A note on quantum algorithms and the minimal degree of ϵ-error polynomials for symmetric functions," *Quantum Information and Computation*, vol. 8, no. 10, pp. 943–950, 2008.

[14] D. Gavinsky and A. A. Sherstov, "A separation of NP and coNP in multiparty communication complexity," *Theory of Computing*, vol. 6, no. 10, pp. 227–245, 2010.

[15] A. R. Klivans, R. O'Donnell, and R. A. Servedio, "Learning intersections and thresholds of halfspaces," *J. Comput. Syst. Sci.*, vol. 68, no. 4, pp. 808–840, 2004.

[16] A. R. Klivans and R. A. Servedio, "Learning DNF in time $2^{\tilde{O}(n^{1/3})}$," *J. Comput. Syst. Sci.*, vol. 68, no. 2, pp. 303–318, 2004.

[17] ——, "Toward attribute efficient learning of decision lists and parities," *J. Machine Learning Research*, vol. 7, pp. 587–602, 2006.

[18] M. Krause and P. Pudlák, "On the computational power of depth-2 circuits with threshold and modulo gates," *Theor. Comput. Sci.*, vol. 174, no. 1–2, pp. 137–156, 1997.

[19] ——, "Computing Boolean functions by polynomials and threshold circuits," *Comput. Complex.*, vol. 7, no. 4, pp. 346–370, 1998.

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

[21] N. Nisan and M. Szegedy, "On the degree of Boolean functions as real polynomials," *Computational Complexity*, vol. 4, pp. 301–313, 1994.

[22] R. O'Donnell and R. A. Servedio, "New degree bounds for polynomial threshold functions," *Combinatorica*, vol. 30, no. 3, pp. 327–358, 2010.

[23] R. Paturi and M. E. Saks, "Approximating threshold circuits by rational functions," *Inf. Comput.*, vol. 112, no. 2, pp. 257–272, 1994.

[24] A. A. Razborov and A. A. Sherstov, "The sign-rank of AC^0," *SIAM J. Comput.*, vol. 39, no. 5, pp. 1833–1855, 2010, preliminary version in *Proceedings of the Forty-Ninth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2008.

[25] A. A. Sherstov, "Separating AC^0 from depth-2 majority circuits," *SIAM J. Comput.*, vol. 38, no. 6, pp. 2113–2129, 2009, preliminary version in *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing* (STOC), 2007.

[26] ——, "The pattern matrix method," *SIAM J. Comput.*, vol. 40, no. 6, pp. 1969–2000, 2011, preliminary version in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing* (STOC), 2008.

[27] ——, "The multiparty communication complexity of set disjointness," in *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), 2012, pp. 525–544.

[28] ——, "Approximating the AND-OR tree," *Theory of Computing*, vol. 9, no. 20, pp. 653–663, 2013.

[29] ——, "The intersection of two halfspaces has high threshold degree," *SIAM J. Comput.*, vol. 42, no. 6, pp. 2329–2374, 2013, preliminary version in *Proceedings of the Fiftieth Annual IEEE Symposium on Foundations of Computer Science* (FOCS), 2009.

[30] ——, "Making polynomials robust to noise," *Theory of Computing*, vol. 9, pp. 593–615, 2013, preliminary version in *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing* (STOC), 2012.

[31] ——, "Optimal bounds for sign-representing the intersection of two halfspaces by polynomials," *Combinatorica*, vol. 33, no. 1, pp. 73–96, 2013, preliminary version in *Proceedings of the Forty-Second Annual ACM Symposium on Theory of Computing* (STOC), 2010.

[32] ——, "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, full version available as ECCC Report TR14-009, January 2014.

[33] ——, "Communication lower bounds using directional derivatives," *J. ACM*, vol. 61, no. 6, pp. 1–71, 2014, preliminary version in *Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing* (STOC), 2013.

[34] ——, "The power of asymmetry in constant-depth circuits," in *Electronic Colloquium on Computational Complexity (ECCC)*, 2015, technical report.

[35] K.-Y. Siu, V. P. Roychowdhury, and T. Kailath, "Rational approximation techniques for analysis of neural networks," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 455–466, 1994.