Hardness Amplification and the Approximate Degree of Constant-Depth Circuits

Mark Bun[1(✉)] and Justin Thaler[2]

[1] Harvard University, Cambridge, Massachusetts
mbun@seas.harvard.edu
[2] Yahoo! Labs, New York, USA
jthaler@fas.harvard.edu

Abstract. We establish a generic form of hardness amplification for the approximability of constant-depth Boolean circuits by polynomials. Specifically, we show that if a Boolean circuit cannot be pointwise approximated by low-degree polynomials to within constant error in a certain one-sided sense, then an OR of disjoint copies of that circuit cannot be pointwise approximated even with very high error. As our main application, we show that for every sequence of degrees $d(n)$, there is an explicit depth-three circuit $F : \{-1,1\}^n \to \{-1,1\}$ of polynomial-size such that any degree-d polynomial cannot pointwise approximate F to error better than $1 - \exp(-\tilde{\Omega}(nd^{-3/2}))$. As a consequence of our main result, we obtain an $\exp(-\tilde{\Omega}(n^{2/5}))$ upper bound on the the discrepancy of a function in AC^0, and an $\exp(\tilde{\Omega}(n^{2/5}))$ lower bound on the threshold weight of AC^0, improving over the previous best results of $\exp(-\Omega(n^{1/3}))$ and $\exp(\Omega(n^{1/3}))$ respectively.

Our techniques also yield a new lower bound of $\Omega(n^{1/2}/\log^{(d-2)/2}(n))$ on the approximate degree of the AND-OR tree of depth d, which is tight up to polylogarithmic factors for any constant d, as well as new bounds for read-once DNF formulas. In turn, these results imply new lower bounds on the communication and circuit complexity of these classes, and demonstrate strong limitations on existing PAC learning algorithms.

1 Introduction

The ε-approximate degree of a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\widetilde{\deg}_\varepsilon(f)$, is the minimum degree of a real polynomial that approximates f to error ε in the ℓ_∞ norm. Approximate degree has pervasive applications in theoretical

The full version of this paper is available at http://arxiv.org/abs/1311.1616.

Supported by an NDSEG Fellowship and NSF grant CNS-1237235.

Parts of this work were done while the author was a graduate student at Harvard University, and a Research Fellow at the Simons Institute for the Theory of Computing. This work was supported by an NSF Graduate Research Fellowship, NSF grants CNS-1011840 and CCF-0915922, and a Research Fellowship from the Simons Institute for the Theory of Computing.

M.M. Halldórsson et al. (Eds.): ICALP 2015, Part I, LNCS 9134, pp. 268–280, 2015.
DOI: 10.1007/978-3-662-47672-7_22

computer science. For example, lower bounds on approximate degree underly many tight lower bounds on quantum query complexity (e.g., [2,3,5,31]), and have been used to resolve several long-standing open questions in communication complexity [27]. Meanwhile, upper bounds on approximate degree underly many of the fastest known learning algorithms, including PAC learning DNF and read-once formulas [4,14], agnostically learning disjunctions [12], and PAC learning in the presence of irrelevant information [15,24].

Despite the range and importance of these applications, large gaps remain in our understanding of approximate degree. The approximate degree of any *symmetric* Boolean function has been understood since Paturi's 1992 paper [22], but once we move beyond symmetric functions, few general results are known.

In this paper, we perform a careful study of the approximate degree of constant-depth Boolean circuits. In particular, we establish a generic form of hardness amplification for the pointwise approximation of small depth circuits by low-degree polynomials: we show that if a Boolean circuit f cannot be pointwise approximated to within constant error in a certain one-sided sense by polynomials of a given degree, then the circuit F obtained by taking an OR of disjoint copies of f cannot be approximated even with error exponentially close to 1. Notice that if f is computed by a circuit of polynomial size and constant depth, then so is F.

Our proof extends a recent line of work [8,18,25,33] that seeks to prove approximate degree lower bounds by constructing explicit *dual polynomials*, which are dual solutions to a linear program that captures the approximate degree of any function. Specifically, we show that given a dual polynomial demonstrating that f cannot be approximated to within constant error, we can construct a dual polynomial demonstrating that F cannot be approximated even with error exponentially close to 1.

As the main application of our hardness amplification technique, for any $d > 0$ we exhibit an explicit function $F : \{-1,1\}^n \to \{-1,1\}$ computed by a polynomial size circuit of depth three for which any degree-d polynomial cannot pointwise approximate F to error $1 - \exp(-\tilde{\Omega}(nd^{-3/2}))$. We then use this result to obtain new bounds on two quantities that play central roles in learning theory, communication complexity, and circuit complexity: *discrepancy* and *threshold weight*. Specifically, we prove a new upper bound of $\exp(-\tilde{\Omega}(n^{2/5}))$ for the discrepancy of a function in AC^0, and a new lower bound of $\exp(\tilde{\Omega}(n^{2/5}))$ for the threshold weight of AC^0. As a second application, our hardness amplification result allows us to resolve, up to polylogarithmic factors, the approximate degree of AND-OR trees of arbitrary constant depth. Finally, our techniques also yield new lower bounds for read-once DNF formulas.

2 Hardness Amplification

Recall that the ε-approximate degree of a Boolean function f is the minimum degree of a real polynomial that pointwise approximates f to error ε. Another fundamental measure of the complexity of f is its *threshold degree*, denoted

$\deg_\pm(f)$. The threshold degree of f is the least degree of a real polynomial that agrees in sign with f at all Boolean inputs.

Central to our results is a measure of the complexity of a Boolean function that we call *one-sided approximate degree*. This quantity, which we denote by $\widetilde{\deg}_\varepsilon(f)$, is an intermediate complexity measure that lies between ε-approximate degree and threshold degree. Unlike approximate degree and threshold degree, one-sided approximate degree treats inputs in $f^{-1}(1)$ and inputs in $f^{-1}(-1)$ asymmetrically.

More specifically, $\widetilde{\deg}_\varepsilon(f)$ captures the least degree of a *one-sided approximation* for f. Here, a one-sided approximation p for f is a polynomial that approximates f to error at most ε at all points $x \in f^{-1}(1)$, and satisfies the threshold condition $p(x) \leq -1 + \varepsilon$ at all points $x \in f^{-1}(-1)$. Notice that $\widetilde{\deg}_\varepsilon(f)$ is always *at most* $\widetilde{\deg}_\varepsilon(f)$, but can be smaller. Similarly, $\widetilde{\deg}_\varepsilon(f)$ is always *at least* $\deg_\pm(f)$, but can be larger.

One-sided approximate degree is the complexity measure that we amplify for constant-depth circuits: given a depth k circuit f on m variables that has one-sided approximate degree greater than d, we show how to generically transform f into a depth $k+1$ circuit F on $t \cdot m$ variables such that F cannot be pointwise approximated by degree d polynomials even to error $1 - 2^{-t}$.[1]

Theorem 1. *Suppose $f : \{-1,1\}^m \to \{-1,1\}$ has one-sided approximate degree $\widetilde{\deg}_{1/2}(f) > d$. Denote by $F : \{-1,1\}^{m \cdot t} \to \{-1,1\}$ the block-wise composition $\mathrm{OR}_t(f, \ldots, f)$, where OR_t denotes the OR function on t variables. Then F cannot be pointwise approximated by degree-d polynomials to within error $1 - 2^{-t}$ by degree-d polynomials. That is, the $(1 - 2^{-t})$-approximate degree of F is greater than d.*

Remark: Theorem 1 demonstrates that one-sided approximate degree admits a form of hardness amplification within AC^0, which does not generally hold for the ordinary approximate degree. Indeed, Theorem 1 fails badly if the condition $\widetilde{\deg}_{1/2}(f) > d$ is replaced with the weaker condition $\widetilde{\deg}_{1/2}(f) > d$ (in fact, $f = \mathrm{OR}_m$ is a counter-example).

A *dual formulation* of one-sided approximate degree was previously exploited by Gavinsky and Sherstov to separate the multi-party communication versions of NP and co-NP [9], as well as by the current authors [8] and independently by Sherstov [25] to resolve the approximate degree of the two-level AND-OR tree. In this paper, we introduce the primal formulation of one-sided approximate degree, which allows us to express Theorem 1 as a hardness amplification result. We also argue for the importance of one-sided approximate degree as a complexity measure in its own right.

Prior Work on Hardness Amplification for Approximate Degree. For the purposes of this discussion, we informally consider a hardness amplification result for approximate degree to be any statement of the following form:

[1] Follow-up work by Sherstov [26] has established a lower bound on the *threshold degree* of F. Specifically, he has shown that there is some constant c such that $\deg_\pm(F) > \min\{ct, d\}$. See Section 6 for further discussion of this result.

Fix two functions $f : \{-1,1\}^m \to \{-1,1\}$ and $g : \{-1,1\}^t \to \{-1,1\}$. Then the composed function $g(f,\ldots,f) : \{-1,1\}^{m\cdot t} \to \{-1,1\}$ is strictly harder to approximate in the ℓ_∞ norm by low-degree polynomials than is the function f.

We think of such a result as establishing that application of the outer function g to t disjoint copies of f amplifies the hardness of f. Here we consider polynomial degree to be a resource, and "harder to approximate" can refer either to the amount of resources required for the approximation, to the error of the approximation, or to a combination of the two.

Two particular kinds of hardness amplification results for approximate degree have received particular attention. *Direct-sum* theorems focus on amplifying the degree required to obtain an approximation, but do not focus on amplifying the error. For example, a typical direct-sum theorem identifies conditions on f and g that guarantee that $\widetilde{\deg}_\varepsilon(g(f,\ldots,f)) \geq \widetilde{\deg}_\varepsilon(g) \cdot \widetilde{\deg}_\varepsilon(f)$. In contrast, a *direct-product* theorem focuses on amplifying both the error and the minimum degree required to achieve this error. An *XOR lemma* is a special case of either type of theorem where the combining function g is the XOR function. Ideally, an XOR lemma of the direct-product form establishes that there exists a sufficiently small constant $\delta > 0$ such that $\widetilde{\deg}_{1-2^{-\delta t}}(\mathrm{XOR}_t(f,\ldots,f)) \geq t \cdot \widetilde{\deg}_{1/3}(f)$. That is, an XOR lemma establishes that approximating the XOR of t disjoint copies of f requires a t-fold blowup in degree relative to f, even if one allows error exponentially close to 1.

O'Donnell and Servedio [21] proved an XOR lemma for *threshold degree*, establishing that $\mathrm{XOR}_t(f,\ldots,f)$ has threshold degree t times the threshold degree of f. In later work, Sherstov [33] proved a direct sum result for approximate degree that holds whenever the combining function g has low block-sensitivity. His techniques also capture O'Donnell and Servedio's XOR lemma for threshold degree as a special case. In [31], Sherstov proved a number of hardness amplification results for approximate degree. Most notably, he proved an optimal XOR lemma, as well as a direct-sum theorem that holds whenever the combining function has close to maximal approximate degree (i.e., approximate degree $\Omega(t)$). Sherstov used his XOR lemma to prove direct product theorems for quantum query complexity, and in subsequent work [32], to show direct product theorems for the multiparty communication of set disjointness.

Comparison to Prior Work. In this paper, we are interested in establishing approximate degree lower bounds for constant-depth circuits over the basis $\{\mathrm{AND}, \mathrm{OR}, \mathrm{NOT}\}$. For this purpose, it is essential to consider combining functions (such as OR, see Theorem 1) that are themselves in AC^0, ruling out the use of XOR as a combining function. Our hardness amplification result (Theorem 1) is orthogonal to direct-sum theorems: direct-sum theorems focus on amplifying degree but not error, while Theorem 1 focuses on amplifying error but not degree. Curiously, Theorem 1 is nonetheless a critical ingredient in our proof of a direct-sum type theorem for AND-OR trees of constant depth (Theorem 3).

Proof Idea. As discussed in the introduction, our proof of Theorem 1 relies on a dual characterization of one-sided approximate degree (see the full version of this

work). Specifically, for any m-variate Boolean function f satisfying $\widetilde{\mathrm{odeg}}_{1/2}(f) > d$, there exists a dual object $\psi : \{-1,1\}^m \to \mathbb{R}$ that witnesses this fact — we refer to ψ as a "dual polynomial" for f. The dual polynomial ψ satisfies three important properties: (1) ψ has high correlation with f, (2) ψ has zero correlation with all polynomials of degree at most d, and (3) $\psi(x)$ agrees in sign with $f(x)$ for all $x \in f^{-1}(-1)$. We refer to the second property by saying ψ has *pure high degree* d, and we refer to the third property by saying that ψ has *one-sided error*.

Our proof proceeds by taking a dual witness ψ to the high one-sided approximate degree of f, and a certain dual witness Ψ for the function OR_t, and combining them to obtain a dual witness ζ for the fact that $\widetilde{\deg}_{1-2^{-t}}(\mathrm{OR}_t(f,\ldots,f)) > d$. Our analysis of the combined dual witness crucially exploits two properties: first, that ψ has one-sided error and second, that the vector whose entries are all equal to -1 has very large (in fact, maximal) Hamming distance from the unique input in $\mathrm{OR}_t^{-1}(1)$.

Our method of combining the two dual witnesses was first introduced by Sherstov [33, Theorem 3.3] and independently by Lee [18]. This method was also used by the present authors in [8] to resolve the approximate degree of the two-level AND-OR tree, and by Sherstov [31] to prove direct sum and direct product theorems for polynomial approximation. However, as discussed above, prior work used this method of combining dual witnesses exclusively to amplify the *degree* in the resulting lower bound; in contrast, we use the combining method in the proof of Theorem 1 to amplify the *error* in the resulting lower bound.

From a technical perspective, the primary novelty in the proof of Theorem 1 lies in our choice of an appropriate (and simple) dual witness Ψ for OR_t, and the subsequent analysis of the correlation of the combined witness ζ with $\mathrm{OR}_t(f,\ldots,f)$. By our choice of Ψ, we are able to show that ζ has correlation with $\mathrm{OR}_t(f,\ldots,f)$ that is *exponentially* close to 1, yielding a lower bound even on the degree of approximations with very high error.

3 Lower Bounds For AC^0

3.1 A New One-Sided Approximate Degree Lower Bound for AC^0

Our ultimate goal is to use Theorem 1 to construct a function F in AC^0 that is hard to approximate by low-degree polynomials even with error exponentially close to 1. However, in order to apply Theorem 1, we must first identify an AC^0 function f such that $\widetilde{\mathrm{odeg}}_{1/2}(f)$ is large.

To this end, we identify fairly general conditions guaranteeing that the one-sided approximate degree of a function is *equal* to its approximate degree, up to a logarithmic factor. To express our result, let $[N] = \{1,\ldots,N\}$, and let m, N, R be a triple of positive integers such that $R \geq N$, and $m = N \cdot \log_2 R$. In most cases, we will take $R = N$. We specifically consider Boolean functions f on $\{-1,1\}^m$ that interpret their input x as the values of a function g_x mapping $[N] \to [R]$. That is, we break x up into N blocks each of length $\log_2 R$, and regard each block x_i as the binary representation of $g_x(i)$. Hence, we think of

f as computing some *property* ϕ_f of functions $g_x : [N] \to [R]$. We say that a property ϕ is *symmetric* if for all $g : [N] \to [R]$, all permutations σ on $[R]$, and all permutations π on $[N]$, it holds that $\phi(g) = \phi(\sigma \circ g \circ \pi)$.

Theorem 2. *Let $f : \{-1,1\}^m \to \{-1,1\}$ be a Boolean function corresponding to a symmetric property ϕ_f of functions $g_x : [N] \to [R]$. Suppose that for every pair $x, y \in f^{-1}(-1)$, there is a pair of permutations σ on $[R]$ and π on $[N]$ such that $g_x = \sigma \circ g_y \circ \pi$. Then $\widetilde{\deg}_\varepsilon(f) \geq \frac{1}{\log_2 R} \cdot \widetilde{\deg}_\varepsilon(f)$ for all $\varepsilon > 0$.*

Proof Idea. It is enough to show that any one-sided ε-approximation p to f can be transformed into an actual ε-approximation r to f in a manner that does not increase the degree by too much (i.e., in a manner guaranteeing that $\deg(r) \leq (\log_2 R) \deg(p)$).

Our transformation from p to r consists of two steps. In the first step, we turn p into a "symmetric" polynomial $p^{\mathrm{sym}}(x) := \mathbb{E}_{y \sim x}[p(y)]$ where $y \sim x$ if $g_y = \sigma \circ g_x \circ \pi$ for some permutations σ on $[R]$ and π on $[N]$. It follows from work of Ambainis [3] that the map $p \mapsto p^{\mathrm{sym}}$ increases the degree of p by a factor of at most $\log_2 R$. In the second step, we argue that there is an affine transformation r of p^{sym} that is an actual ε-approximation to f, completing the construction.

The existence of the affine transformation r of p^{sym} follows from two observations: (1) if p is a one-sided approximation for f, then so is p^{sym} (this holds because ϕ_f is symmetric), and (2) p^{sym} takes on a constant value v on $f^{-1}(-1)$, i.e., $p^{\mathrm{sym}}(x) = v$ for all $x \in f^{-1}(-1)$ (this holds because $x \sim y$ for every pair of inputs $x, y \in f^{-1}(-1)$). Thus even if p^{sym} poorly approximates f on $f^{-1}(-1)$, we can still obtain a good approximation r by applying an affine transformation to the range of p^{sym} that maps v to -1 and moves all values closer to 1.

In our primary application of Theorem 2, we let $f : \{-1,1\}^m \to \{-1,1\}$ be the Element Distinctness function. Aaronson and Shi [2] showed that the approximate degree of Element Distinctness is $\Omega((m/\log m)^{2/3})$. Element Distinctness is computed by a CNF of polynomial size, and Aaronson and Shi's result remains essentially the best-known lower bound for the approximate degree of a function in AC^0. Theorem 2 applies to Element Distinctness, yielding the following corollary.

Corollary 1. *Let $f : \{-1,1\}^m \to \{-1,1\}$ denote the* Element Distinctness *function. Then $\widetilde{\deg}(f) = \tilde{\Omega}(m^{2/3})$.*

The best known lower bound on the one-sided approximate degree of an AC^0 function that followed from prior work was $\Omega(m^{1/2})$ (which holds for the AND function [9,20]). Section 6 describes some further implications of Theorem 2.

3.2 Accuracy-Degree Tradeoff Lower Bounds for AC^0

By Corollary 1, we can apply Theorem 1 to Element Distinctness to obtain a depth-three Boolean circuit F with $t \cdot m$ inputs such that $\widetilde{\deg}_\varepsilon(F) = \tilde{\Omega}(m^{2/3})$,

for $\varepsilon = 1 - 2^{-t}$. By choosing t and m appropriately, we obtain a depth-three circuit on $n = t \cdot m$ variables of size $\mathrm{poly}(n)$ such that any degree-d polynomial cannot pointwise approximate F to error better than $1 - \exp(-\tilde{\Omega}(nd^{-3/2}))$.

Corollary 2. *For every $d > 0$, there is a depth-3 Boolean circuit $F : \{-1,1\}^n \to \{-1,1\}$ of size $\mathrm{poly}(n)$ such that any degree-d polynomial cannot pointwise approximate F to error better than $1 - \exp(-\tilde{\Omega}(nd^{-3/2}))$. In particular, there is a depth-3 circuit F such that any polynomial of degree at most $n^{2/5}$ cannot pointwise approximate F to error better than $1 - \exp(-\tilde{\Omega}(n^{2/5}))$.*

3.3 Discrepancy Upper Bound

Discrepancy is a central quantity in communication complexity and circuit complexity. For instance, upper bounds on the discrepancy of a function f immediately yield lower bounds on the cost of small-bias communication protocols for computing f (The full version of this work has details). The first exponentially small discrepancy upper bounds for AC^0 were proved by Burhman et al. [7] and Sherstov [29,30], who exhibited constant-depth circuits with discrepancy $\exp(-\Omega(n^{1/3}))$. We improve the best-known upper bound to $\exp(-\tilde{\Omega}(n^{2/5}))$.

Table 1. Comparison of our new discrepancy bound for AC^0 to prior work. The circuit depth column lists the depth of the circuit used to exhibit the bound.

Reference	Discrepancy Bound	Circuit Depth
Sherstov [30]	$\exp(-\Omega(n^{1/5}))$	3
Buhrman et al. [7]	$\exp(-\Omega(n^{1/3}))$	3
Sherstov [29]	$\exp(-\Omega(n^{1/3}))$	3
This work	$\exp(-\tilde{\Omega}(n^{2/5}))$	4

Our result relies on a powerful technique developed by Sherstov [29], known as the pattern-matrix method. This technique allows one to automatically translate lower bounds on the ε-approximate degree of a Boolean function F into upper bounds on the *discrepancy* of a related function F' as long as ε is exponentially close to one. By applying the pattern-matrix method to Corollary 2, we obtain the following result.

Corollary 3. *There is a depth-4 Boolean circuit $F' : \{-1,1\}^n \to \{-1,1\}$ with discrepancy $\exp(-\tilde{\Omega}(n^{2/5}))$.*

3.4 Threshold Weight Lower Bound

A *polynomial threshold function* (PTF) for a Boolean function f is a multilinear polynomial p with integer coefficients that agrees in sign with f on all Boolean inputs. The *weight* of an n-variate polynomial p is the sum of the absolute value of its coefficients. The *degree-d threshold weight* of a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $W(f,d)$, refers to the least weight of a degree-d

PTF for f. We let $W(f)$ denote the quantity $W(f,n)$, i.e., the least weight of any threshold function for f regardless of its degree. As discussed in the full version of this work, threshold weight has important applications in learning theory.

Threshold weight is closely related to ε-approximate degree when ε is very close to 1. This allows us to translate Corollary 2 into a lower bound on the degree-d threshold weight of AC^0.

Corollary 4. *For every $d > 0$, there is a depth-3 Boolean circuit $F : \{-1,1\}^n \to \{-1,1\}$ of size $\mathrm{poly}(n)$ such that $W(F,d) \geq \exp(\tilde{\Omega}(nd^{-3/2}))$. In particular, $W(F, n^{2/5}) = \exp(\tilde{\Omega}(n^{2/5}))$.*

A result of Krause [16] allows us to extend our new degree-d threshold weight lower bound for F into a *degree independent* threshold weight lower bound for a related function F'. The previous best lower bound on the threshold weight of AC^0 was $\exp(\Omega(n^{1/3}))$, due to Krause and Pudlák [17].

Corollary 5. *There is a depth-4 Boolean circuit $F' : \{-1,1\}^n \to \{-1,1\}$ satisfying $W(F') = \exp(\tilde{\Omega}(n^{2/5}))$.*

Moreover, while the threshold weight bound of Corollary 5 is stated for polynomial threshold functions over $\{-1,1\}^n$, we show that the same threshold weight lower bound also holds for polynomials over $\{0,1\}^n$.

4 Approximate Degree Lower Bounds for AND-OR Trees

The d-level AND-OR tree on n variables is a function described by a read-once circuit of depth d consisting of alternating layers of AND gates and OR gates. We assume for simplicity that all gates have fan-in $n^{1/d}$. For example, the two-level AND-OR tree is a read-once CNF in which all gates have fan-in $n^{1/2}$.

Until recently, the approximate degree of AND-OR trees of depth two or greater had resisted characterization, despite 19 years of attention [3,8,10,20,25,33,34]. The case of of depth two was reposed as a challenge problem by Aaronson in 2008 [1], as it captured the limitations of existing lower bound techniques. This case was resolved last year by the current authors [8], and independently by Sherstov [25], who proved a lower bound of $\Omega(\sqrt{n})$, matching an upper bound of Høyer, Mosca, and de Wolf [10]. However, the case of depth three or greater remained open. To our knowledge, the best known lower bound for $d \geq 3$ was $\Omega(n^{1/4+1/2d})$, which follows by combining the depth-two lower bound [8,25] with an earlier direct-sum theorem of Sherstov [33, Theorem 3.1].

By combining the techniques of our earlier work [8] with our hardness amplification result (Theorem 1), we improve this lower bound to $\Omega(n^{1/2}/\log^{(d-2)/2}(n))$ for any constant $d \geq 2$. A line of work on quantum query algorithms [4,10,23] established an upper bound of $O(n^{1/2})$ for AND-OR trees of any depth, demonstrating that our result is optimal up to polylogarithmic factors.

Theorem 3. *Let $\text{AND-OR}_{d,n}$ denote the d-level AND-OR tree on n variables. Then $\widetilde{\deg}(\text{AND-OR}_{d,n}) = \Omega(n^{1/2}/\log^{(d-2)/2} n)$ for any constant $d \geq 2$.*

Proof Idea. To introduce our proof technique, we first describe the method used in [8] to construct an optimal dual polynomial in the case $d = 2$, and we identify why this method breaks down when trying to extend to the case $d = 3$. We then explain how to use our hardness amplification result (Theorem 1) to construct a different dual polynomial that does extend to the case $d = 3$.

Let M denote the fan-in of all gates in OR-AND$_{2,M^2}$. In our earlier work [8], we constructed a dual polynomial for OR-AND$_{2,M^2}$ as follows. It is known that there is a dual polynomial γ_1 witnessing the fact that $\widetilde{\text{odeg}}(\text{AND}_M) = \Omega(M^{1/2})$, and a dual polynomial γ_2 witnessing the fact that $\widetilde{\deg}(\text{OR}_M) = \Omega(M^{1/2})$. We then combined the dual witnesses γ_1 and γ_2, using the same "combining" technique as in the proof of Theorem 1, to obtain a dual witness $\gamma_3 : \{-1, 1\}^{M^2} \to \mathbb{R}$ for the high approximate degree of OR-AND$_{2,M^2}$.

Recall that we say a dual witness has *pure high degree* d if it has zero correlation with every polynomial of degree at most d. It followed from earlier work [33] that γ_3 has pure high degree equal to the product of the pure high degrees of γ_1 and γ_2, yielding an $\Omega(M)$ lower bound on the pure high degree of γ_3. The new ingredient of the analysis in [8] was to use the one-sided error of the "inner" dual witness γ_1 to argue that γ_3 also had good correlation with OR-AND$_{2,M^2}$.

Extending to Depth Three. Let $M = n^{1/3}$ denote the fan-in of all gates in AND-OR$_{3,n}$. To construct a dual witness for AND-OR$_{3,n}$ = AND$_M$ (OR-AND$_{2,M^2}$, ..., OR-AND$_{2,M^2}$), it is natural to try the following approach. Let γ_4 be a dual polynomial witnessing the fact that the approximate degree of AND$_M$ $= \Omega(\sqrt{M})$. Then we can combine γ_3 and γ_4 as above to obtain a dual function γ_5.

The difficulty in establishing that γ_5 is a dual witness to the high approximate degree of AND-OR$_{3,n}$ is in showing that γ_5 has good correlation with AND-OR$_3$. In our earlier work, we showed γ_3 has large correlation with OR-AND$_{2,n}$ by exploiting the fact that the inner dual witness γ_1 had one-sided error, i.e., $\gamma_1(y)$ agrees in sign with AND$_M$ whenever $y \in \text{AND}_M^{-1}(-1)$. However, γ_3 itself does not satisfy an analogous property: there are inputs $x_i \in \text{OR-AND}_{2,M^2}^{-1}(-1)$ such that $\gamma_3(x_i) > 0$, *and* there are inputs $x_i \in \text{OR-AND}_{2,M^2}^{-1}(1)$ such that $\gamma_3(x_i) < 0$.

To circumvent this issue, we use a different inner dual witness γ_3' in place of γ_3. Our construction of γ_3' utilizes our hardness amplification analysis to achieve the following: while γ_3' has error "on both sides", the error from the "wrong side" is very small. The hardness amplification step causes γ_3' to have pure high degree that is lower than that of the dual witness γ_3 constructed in [8] by a $\sqrt{\log n}$ factor. However, the hardness amplification step permits us to prove the desired lower bound on the correlation of γ_5 with AND-OR$_{3,n}$. The proof for the general case, which is quite technical, appears in the full version of this work.

5 Lower Bounds for Read-Once DNFs and CNFs

Our techniques also yield new lower bounds on the approximate degree and degree-d threshold weight of read-once DNF and CNF formulas. Before stating our results, we discuss relevant prior work.

In their seminal work on perceptrons, Minsky and Papert exhibited a read-once DNF $f : \{-1,1\}^n \to \{-1,1\}$ with *threshold degree* $\Omega(n^{1/3})$ [19]. That is, a real polynomial requires degree $\Omega(n^{1/3})$ just to agree with f in sign. However, to our knowledge no non-trivial lower bound on the degree-d threshold *weight* of read-once DNFs was known for any $d = \omega(n^{1/3})$.

In an influential result, Beigel [6] exhibited a polynomial-size (read-many) DNF called ODD-MAX-BIT satisfying the following: there is some constant $\delta > 0$ such that $\widetilde{\deg}_{1-2^{-\delta n/d^2}}(\text{ODD-MAX-BIT}) > d$, and hence also $W(\text{ODD-MAX-BIT}, d) = \exp(\Omega(n/d^2))$. Motivated by applications in computational learning theory, Klivans and Servedio showed that Beigel's lower bound is essentially tight for $d < n^{1/3}$ [15]. Very recently, Servedio, Tan, and Thaler showed an alternative lower bound on the degree-d threshold weight of ODD-MAX-BIT. Specifically, they showed that $W(\text{ODD-MAX-BIT}, d) = \exp(\Omega(\sqrt{n/d}))$ [24]. The lower bound of Servedio et al. improves over Beigel's for any $d > n^{1/3}$, and is essentially tight in this regime (i.e., when $d > n^{1/3}$).

While ODD-MAX-BIT is a relatively simple DNF (in fact, it is a *decision list*), it is not a read-once DNF. Our results extend the lower bounds of Servedio et al. and Beigel from decision lists to read-once DNFs and CNFs. In the statement of the results below, we restrict ourselves to DNFs, as the case of CNFs is entirely analogous.

5.1 Extending Servedio et al.'s Lower Bound to Read-Once DNFs

In order to extend the lower bound of Servedio et al. to read-once DNFs and CNFs, we extend our hardness amplification techniques from one-sided approximate degree to a new quantity we call *degree-d one-sided non-constant approximate weight*. This quantity captures the least L_1 *weight* (excluding the constant term) of a polynomial of degree at most d that is a one-sided approximation of f. We denote the degree-d one-sided approximate weight of a Boolean function f by $W^*_\varepsilon(f, d)$, where ε is an error parameter. We prove the following analog of Theorem 1.

Theorem 4. *Fix $d > 0$. Let $f : \{-1,1\}^m \to \{-1,1\}$, and suppose $W^*_{3/4}(f, d) > w$. Let $F : \{-1,1\}^{m \cdot t} \to \{-1,1\}$ denote the function $\text{OR}_t(f, \ldots, f)$. Then any degree-$d$ polynomial that approximates F to error $1 - 2^{-t}$ requires weight $2^{-5t}w$.*

Adapting a proof of Servedio et al., we can show that $W^*_{3/4}(\text{AND}_m, d) \geq 2^{\Omega(m/d)}$. By applying Theorem 4 with $f = \text{AND}_m$, along with standard manipulations, we are able to extend the lower bound of Servedio et al. to read-once CNFs and DNFs.

Corollary 6. *For each $d = o(n/\log^4 n)$, there is a read-once DNF F satisfying $W(F, d) = \exp(\Omega(\sqrt{n/d}))$.*

In particular, there is a read-once DNF that cannot be computed by any PTF of $\text{poly}(n)$ weight, unless the degree is $\tilde{\Omega}(n)$.

5.2 Extending Beigel's Lower Bound to Read-Once DNFs

It is known that $\widetilde{\text{odeg}}(\text{AND}_m) = \Omega(m^{1/2})$. By applying Theorem 1 with $f = \text{AND}_m$, we obtain the following result.

Corollary 7. *There is an (explicit) read-once DNF* $F : \{-1,1\}^n \to \{-1,1\}$ *with* $\widetilde{\deg}_{1-2^{-n/d^2}}(F) = \Omega(d)$.

We remark that for $d < n^{1/3}$, Corollary 7 is subsumed by Minsky and Papert's seminal result that exhibited a read-once DNF F with threshold degree $\Omega(n^{1/3})$ [19]. However, for $d > n^{1/3}$, it is not subsumed by Minsky and Papert's result, nor by Corollary 6. Indeed, Corollary 6 yields a lower bound on the degree-d threshold weight of read-once DNFs, but not a lower bound on the *approximate-degree* of read-once DNFs.

6 Discussion

Subsequent Work by Sherstov. In 1969, Minsky and Papert gave a lower bound of $\Omega(n^{1/3})$ on the threshold degree of an explicit read-once DNF formula. Klivans and Servedio [14] proved their lower bound to be tight within a logarithmic factor for DNFs of polynomial size, but it remained a well-known open question to give a threshold degree lower bound of $\Omega(n^{1/3+\delta})$ for a function in AC^0; the only progress prior to our work was due to O'Donnell and Servedio [21], who established an $\Omega(n^{1/3}\log^k n)$ lower bound for any constant $k > 0$.

Let f denote the Element Distinctness function on $n^{3/5}$ variables. In an earlier version of this work, we conjectured that the function $F = \text{OR}_{n^{2/5}}(f,\ldots,f)$ appearing in Corollary 2 in fact satisfies $\deg_{\pm}(f) = \tilde{\Omega}(n^{2/5})$, and observed that this would yield the first polynomial improvement on Minsky and Papert's lower bound. Sherstov [26, Theorem 7.1] has recently proved our conjecture. His proof, short and elegant, extends our dual witness construction in the proof of Theorem 1 to establish a different form of hardness amplification, from one-sided approximate degree to threshold degree. Specifically, he shows that if a Boolean function f has one-sided approximate degree d, then the block-wise composition $\text{OR}_t(f,\ldots,f)$ has threshold degree at least $\min\{ct, d\}$ for some constant c. This result is incomparable to our Theorem 1 when $t \leq d$, but when $t \gg d$, Sherstov's result is a substantial strengthening of Theorem 1.

In the same work, Sherstov has also proven a much stronger and more difficult result: for any $k > 2$, he gives a read-once formula of depth k with threshold degree $\Omega(n^{(k-1)/(2k-1)})$. Notice that for any constant $\delta > 0$, this yields an AC^0 function with threshold degree $\Omega(n^{1/2-\delta})$. This in turn yields an improvement of our discrepancy upper bound (Corollary 3) for AC^0 to $\exp(-\Omega(n^{1/2-\delta}))$, and of our threshold weight lower bound (Corollary 5) to $\exp(\Omega(n^{1/2-\delta}))$.

Subsequent Work by Kanade and Thaler. Existing applications of one-sided approximate degree [8,9,25,26] have all been of a negative nature (proving communication or circuit lower bounds, establishing limitations on PAC learning

algorithms, etc.). Kanade and Thaler [13] have identified a positive (algorithmic) application of one-sided approximate degree. Specifically, they show that one-sided approximate degree upper bounds imply fast algorithms in the reliable agnostic learning framework of Kalai et al. [11]. This framework captures learning tasks in which one type of error (such as false negative errors) is costlier than other types. Kanade and Thaler use this result to give the first sub-exponential time algorithms for distribution-independent reliable learning of several fundamental concept classes.

In light of these developments, we are optimistic that the notion of one-sided approximate degree will continue to enable progress on questions within the analysis of Boolean functions and computational complexity theory.

Acknowledgments. We are grateful to Sasha Sherstov, Robert Špalek, Li-Yang Tan, and the anonymous reviewers for valuable feedback on earlier versions of this manuscript.

References

1. Aaronson, S.: The polynomial method in quantum and classical computing. In: FOCS (2008). (Slides available at www.scottaaronson.com/talks/polymeth.ppt)
2. Aaronson, S., Shi, Y.: Quantum lower bounds for the collision and the element distinctness problems. Journal of the ACM **51**(4), 595–605 (2004)
3. Ambainis, A.: Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. Theory Comput. **1**(1), 37–46 (2005)
4. Ambainis, A., Childs, A.M., Reichardt, B., Špalek, R., Zhang, S.: Any AND-OR formula of size N can be evaluated in time $N^{1/2+o(1)}$ on a quantum computer. SIAM J. Comput. **39**(6), 2513–2530 (2010)
5. Beals, R., Buhrman, H., Cleve, R., Mosca, M., de Wolf, R.: Quantum lower bound by polynomials. J. ACM **48**(4), 778–797 (2001)
6. Beigel, R.: Perceptrons, PP, and the polynomial hierarchy. Computational Complexity **4**, 339–349 (1994)
7. Buhrman, H., Vereshchagin, N.K., de Wolf, R.: On computation and communication with small bias. CCC, pp. 24–32 (2007)
8. Bun, M., Thaler, J.: Dual lower bounds for approximate degree and markov-bernstein inequalities. In: Fomin, F.V., Freivalds, R., Kwiatkowska, M., Peleg, D. (eds.) ICALP 2013, Part I. LNCS, vol. 7965, pp. 303–314. Springer, Heidelberg (2013)
9. Gavinsky, D., Sherstov, A.A.: A separation of NP and coNP in multiparty communication complexity. Theory of Computing **6**(1), 227–245 (2010)
10. Høyer, P., Mosca, M., de Wolf, R.: Quantum search on bounded-error inputs. In: ICALP, pp. 291–299 (2003)
11. Kalai, A., Kanade, V., Mansour, Y.: Reliable agnostic learning. J. Comput. Syst. Sci. **78**(5), 1481–1495 (2012)
12. Kalai, A., Klivans, A., Mansour, Y., Servedio, R.: Agnostically learning halfspaces. SIAM Journal on Computing **37**(6), 1777–1805 (2008)
13. Kanade, V., Thaler, J.: Distribution-independent reliable learning. In: COLT (2014)

14. Klivans, A.R., Servedio, R.A.: Learning DNF in time $2^{\tilde{O}(n^{1/3})}$. J. of Comput. and System Sci. **68**(2), 303–318 (2004)
15. Klivans, A.R., Servedio, R.A.: Toward attribute efficient learning of decision lists and parities. Journal of Machine Learning Research **7**, 587–602 (2006)
16. Krause, M.: On the computational power of Boolean decision lists. Computational Complexity **14**(4), 362–375 (2005)
17. Krause, M., Pudlák, P.: On the computational power of depth-2 circuits with threshold and modulo gates. Theor. Comput. Sci. **174**(1–2), 137–156 (1997)
18. Lee, T.: A note on the sign degree of formulas (2009). CoRR abs/0909.4607
19. Minsky, M.L., Papert, S.A.: Perceptions: An Introduction to Computational Geometry. MIT Press, Cambridge (1969)
20. Nisan, N., Szegedy, M.: On the degree of boolean functions as real polynomials. Computational Complexity **4**, 301–313 (1994)
21. O'Donnell, R., Servedio, R.: New degree bounds for polynomial threshold functions. Combinatorica **30**(3), 327–358 (2010)
22. Paturi, R.: On the degree of polynomials that approximate symmetric Boolean functions (Preliminary Version). STOC, pp. 468–474 (1992)
23. Reichardt, B.: Reflections for quantum query algorithms. In: SODA (2011)
24. Servedio, R.A., Tan, L.-Y., Thaler, J.: Attribute-Efficient learning and weight-degree tradeoffs for polynomial threshold functions. COLT **23**, 14.1–14.19 (2012)
25. Sherstov, A.A.: Approximating the AND-OR Tree. Theory of Computing (2013)
26. Sherstov, A.A.: Breaking the Minsky-Papert Barrier for constant-depth circuits. STOC (2014)
27. Sherstov, A.A.: Communication lower bounds using dual polynomials. Bulletin of the EATCS **95**, 59–93 (2008)
28. Sherstov, A.A.: Optimal bounds for sign-representing the intersection of two halfspaces by polynomials. STOC, pp. 523–532 (2010)
29. Sherstov, A.A.: The pattern matrix method. SIAM J. Comput. **40**(6), 1969–2000 (2011)
30. Sherstov, A.A.: Separating AC^0 from depth-2 majority circuits. SIAM J. Comput. **28**(6), 2113–2129 (2009)
31. Sherstov, A.A.: Strong direct product theorems for quantum communication and query complexity. SIAM J. Comput. **41**(5), 1122–1165 (2012)
32. Sherstov, A.A.: The multiparty communication complexity of set disjointness. STOC, pp. 525–524 (2012)
33. Sherstov, A.A.: The intersection of two halfspaces has high threshold degree. FOCS, pp. 343–362 (2009). (To appear in SIAM J. Comput. (special issue for FOCS 2009))
34. Shi, Y.: Approximating linear restrictions of Boolean functions. Manuscript (2002) web.eecs.umich.edu/shiyy/mypapers/linear02-j.ps