Electronic Colloquium on Computational Complexity, Report No. 150 (2014)

Lower Bounds for the Approximate Degree of Block-Composed Functions

Justin Thaler*

Abstract

We describe a new hardness amplification result for point-wise approximation of Boolean functions by low-degree polynomials. Specifically, for any function f on N bits, define $F(x_1,\ldots,x_M) = \mathrm{OMB}(f(x_1),\ldots,f(x_M))$ to be the function on $M \cdot N$ bits obtained by block-composing f with a specific DNF known as ODD-MAX-BIT. We show that, if f requires large degree to approximate to error $2/3$ in a certain one-sided sense (captured by a complexity measure known as *positive one-sided approximate degree*), then F requires large degree to approximate even to error $1-2^{-M}$. This generalizes a result of Beigel [5], who proved an identical result for the special case $f = \mathrm{OR}$.

Unlike related prior work, our result implies strong approximate degree lower bounds even for many functions F that have low *threshold degree*. Our proof is constructive: we exhibit a solution to the dual of an appropriate linear program capturing the approximate degree of any function.

As an application, we give an explicit AC^0 function with *margin complexity* $\exp\left(\tilde{\Omega}(n^{2/5})\right)$ and *dimension complexity* $n^{O(\log n)}$. The previous best separation was due to Buhrman et al. [6], who gave an AC^0 function with margin complexity $\exp\left(\Omega(n^{1/3})\right)$ and dimension complexity $\mathrm{poly}(n)$.

*Yahoo! Labs.

ISSN 1433-8092

1 Introduction

Approximate degree and threshold degree are two measures of Boolean function complexity that capture the difficulty of point-wise approximation by low-degree polynomials. The ε-approximate degree of a function f, denoted $\widetilde{\deg}_\varepsilon(f)$, is the least degree of a real polynomial that point-wise approximates f to error ε. The threshold degree of f, denoted $\deg_\pm(f)$, is the least degree of a real polynomial that sign-represents f at all points.

Approximate degree has found a diverse array of algorithmic and complexity-theoretic applications. On the complexity side, approximate degree lower bounds underlie many tight lower bounds on quantum query complexity [1, 2, 4, 21, 35], and have proven instrumental in resolving a host of long-standing open problems in communication and circuit complexity [6, 10–12, 14, 23, 27, 28, 33–36, 38]. On the algorithms side, upper bounds on these complexity measures underlie the fastest known learning algorithms in a number of important models, including the PAC, agnostic, and mistake-bounded models [16, 19, 20, 29]. They also yield fast algorithms for private data release [9, 41].

Despite these applications, our understanding of approximate and threshold degree remains limited. While tight upper and lower bounds are known for some specific functions, including symmetric functions [13, 26, 32] and certain read-once formulae, few general results are known, and characterizing the approximate and threshold degrees of many simple functions remains open. However, a handful of recent works has established various forms of "hardness amplification" for approximate degree [7, 8, 22, 30, 37, 39]. Roughly speaking, these results show how to take a function f which is hard to approximate by low-degree polynomials in a weak sense, and turn f into a related function F that is hard to approximate by low-degree polynomials in a much stronger sense.

Our Contributions. We extend this recent line of work by establishing a new, generic form of hardness amplification for approximate degree. Unlike prior work, our result implies strong lower bounds even for many functions F that have low threshold degree (e.g., halfspaces). In contrast, analogous hardness amplification results [7, 8, 22, 30, 37, 39] apply only to functions with polynomially large threshold degree. As the main application of our technique, we exploit the aforementioned property to obtain an improved separation between the *margin* and *dimension complexities* of an AC^0 function.

We prove our results by constructing explicit *dual polynomials*, which are dual solutions to an appropriate linear program capturing the approximate degree of any function. This "method of dual polynomials" has proven to be a powerful technique for establishing lower bounds on approximate degree. Our construction departs qualitatively from earlier applications of the method, and we believe it to be of interest in its own right. In addition to implying approximate degree lower bounds, dual polynomials have been used to resolve several long-standing open problems in communication complexity, and they yield explicit distributions under which various communication problems are hard [12, 14, 27, 34–36, 38].

1.1 Overview of Our Results

Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Our hardness amplification method relies heavily on a complexity measure known as *one-sided approximate degree*, or, more precisely, its "positive" and "negative" variants, denoted $\widetilde{\deg}_{+,\varepsilon}(f)$ and $\widetilde{\deg}_{-,\varepsilon}(f)$ respectively. These are intermediate complexity measures that lie between ε-approximate degree and threshold degree, and they have played a central role in recent

prior work on hardness amplification for approximate degree [7,8,30,37].[1] Unlike the latter two complexity measures, $\widetilde{\deg}_{+,\varepsilon}(f)$ and $\widetilde{\deg}_{-,\varepsilon}(f)$ treat inputs in $f^{-1}(+1)$ and inputs in $f^{-1}(-1)$ asymmetrically.

In more detail, a polynomial p is said to be a positive one-sided ε-approximation for a Boolean function f if $|p(x)-f(x)| \leq \varepsilon$ for all $x \in f^{-1}(-1)$, and $p(x) \geq 1-\varepsilon$ for all $x \in f^{-1}(+1)$. The positive one-sided ε-approximate degree of f is the least degree of a positive one-sided ε-approximation for f. Negative one-sided ε-approximate degree is defined analogously. (Appendix A contains formal definitions.) Notice that $\widetilde{\deg}_{+,\varepsilon}(f)$ and $\widetilde{\deg}_{-,\varepsilon}(f)$ are always at most $\widetilde{\deg}_{\varepsilon}(f)$, but can be much smaller. Similarly, $\widetilde{\deg}_{+,\varepsilon}(f)$ and $\widetilde{\deg}_{-,\varepsilon}(f)$ are always at least $\deg_{\pm}(f)$, but can be much larger.

Let $\mathrm{OMB} : \{-1,1\}^n \to \{-1,1\}$ denote a specific polynomial size DNF formula known as ODD-MAX-BIT, defined as follows. On input $x = (x_1,\ldots,x_n)$, let i^* denote the largest index such that $x_{i^*} = -1$, and let $i^* = 0$ if no such index exists. We define

$$\mathrm{OMB}(x_1,\ldots,x_n) = \begin{cases} -1 \text{ if } i^* \text{ is odd} \\ 1 \text{ otherwise} \end{cases}$$

When appropriate, we also use subscripts after function symbols to indicate the number of variables over which the function is defined. Thus, OMB_M denotes the OMB function on M inputs.

For any function $f : \{-1,1\}^N$, define $F : \left(\{-1,1\}^N\right)^M \to \{-1,1\}$ to be the block-composition of OMB_M with f, i.e., $F = \mathrm{OMB}_M(f,\ldots,f)$. Our hardness amplification result establishes that if $\widetilde{\deg}_{+,\varepsilon}(f)$ is large for some ε bounded away from 1, then $\widetilde{\deg}_{+,\varepsilon}(F)$ is large even for ε exponentially close to 1.

Theorem 1. *If $\widetilde{\deg}_{+,2/3}(f) \geq d$, then $\widetilde{\deg}_{+,\varepsilon}(F) \geq d$ for $\varepsilon = 1-2^{-M}$.*

Theorem 1 is tight whenever f has a $(1/3)$-approximation q of degree d satisfying $q(x) = f(x)$ for all $x \in f^{-1}(-1)$. This is the case for many important functions, including $f = \mathrm{OR}_N$ (see Section 1.2.2 and Remark 6), and $f = \overline{\mathrm{ED}}_N$, where $\overline{\mathrm{ED}}_N$ is a function arising in the proof of Theorem 2 below (see Remark 21).

An Application: Improved Margin-Dimension Gap for $\mathbf{AC^0}$. Margin complexity and dimension complexity are combinatorial quantities that play central roles in learning theory, communication complexity, and circuit complexity. For example, margin complexity is known to be essentially equivalent to *discrepancy* [24], which in turn characterizes the communication complexity class PP^{cc}. Meanwhile, dimension complexity characterizes the communication complexity class UPP^{cc}.

The communication complexity measures $\mathrm{PP}^{cc}(f)$ and $\mathrm{UPP}^{cc}(f)$ both capture the difficulty of computing f to small-bias. $\mathrm{UPP}^{cc}(f)$ is the minimum communication cost of any randomized protocol that computes f with strictly positive bias. $\mathrm{PP}^{cc}(f)$ is similar, but defines the cost of a protocol to be the sum of the communication cost and the logarithm of the reciprocal of the protocol's bias. We define PP^{cc} and UPP^{cc} formally in Appendix B, where we also discuss applications of these communication models to circuit complexity and learning theory.

Both UPP^{cc} and PP^{cc} were introduced in 1986 by Babai et al. [3], and determining whether these classes were equal was open until 2008, when Buhrman et al. [6] and Sherstov [31] independently resolved

[1] Strictly speaking, the terms positive and negative one-sided approximate degree were introduced by Kanade and Thaler [17], who gave applications of these complexity measures to learning theory. Earlier works on hardness amplification for pointwise approximation by polynomials only used negative one-sided approximate degree, and referred to this complexity measure without qualification as one-sided approximate degree [8, 30]. For our purposes, the distinction between positive and negative one-sided approximate degree is crucial.

the problem. Sherstov established the existence of a function F with PP^{cc} communication complexity $\mathrm{PP}^{cc}(F) = \Omega(n)$, and UPP^{cc} communication complexity $\mathrm{UPP}^{cc}(F) = O(\log n)$ [31, Theorem 1.2]. However, this function is not in AC^0. Buhrman et al. exhibited an explicit function F' in AC^0 satisfying $\mathrm{PP}^{cc}(F') = \Omega(n^{1/3})$, and $\mathrm{UPP}^{cc}(F') = O(\log n)$.

Prior to our work, the result of Burhman et al. was the best known separation between the PP^{cc} and UPP^{cc} communication complexity of an AC^0 function (equivalently, between the margin complexity and dimension complexity of an AC^0 function). We use Theorem 1 to improve on their separation, as formalized in the following theorem.

Theorem 2. *There is an explicit function F' computed by a polynomial size circuit of constant depth satisfying: $\mathrm{UPP}^{cc}(F') = O(\log^2 n)$ and $\mathrm{PP}^{cc}(F') = \tilde{\Omega}(n^{2/5})$, where the $\tilde{\Omega}$ notation hides logarithmic factors.*

At the core of our proof of Theorem 2 is the identification of an AC^0 function F such that $\widetilde{\deg}_{\varepsilon}(F)$ is "large" even for exponentially close to 1, yet $\deg_{\pm}(F)$ is "small". It is already well-known that this behavior is exhibited by many halfspaces — in fact, crucial to the PP^{cc} vs. UPP^{cc} separation achieved by Burhman et al. [6] is the fact that OMB itself is a halfspace (i.e., $\deg_{\pm}(\mathrm{OMB}_n) = 1$), and yet $\widetilde{\deg}_{\varepsilon}(\mathrm{OMB}_n) = \Omega(n^{1/3})$, even for $\varepsilon = 1 - 2^{-n^{1/3}}$ (see Section 1.2.2).

We improve over the result of Burhman et al. by considering the function $F = \mathrm{OMB}_{n^{2/5}}(\overline{\mathrm{ED}}_{n^{3/5}}, \ldots, \overline{\mathrm{ED}}_{n^{3/5}})$. Prior work has shown that $\overline{\mathrm{ED}}_N$ satisfies $\widetilde{\deg}_{+,2/3}(\overline{\mathrm{ED}}_N) = \tilde{\Omega}(N^{2/3})$ [8], so Theorem 1 implies that $\widetilde{\deg}_{+,\varepsilon}(F) = \tilde{\Omega}(n^{2/5})$ even for $\varepsilon = 1 - 2^{-n^{2/5}}$. Yet we show that $\deg_{\pm}(F) = O(\log n)$. The key property that we use to establish this threshold degree upper bound is that $\overline{\mathrm{ED}}_N$ can be sign-represented by a polynomial p of degree $O(\log N)$ such that p is *exactly* correct on all inputs x such that $\overline{\mathrm{ED}}_N(x) = +1$.

1.2 Technical Comparison to Prior Work

1.2.1 The Method of Dual Polynomials

A dual witness to the statement $\widetilde{\deg}_{\varepsilon}(f) \geq d$ is a non-zero real-valued function $\psi : \{-1,1\}^N \to \mathbb{R}$ satisfying two conditions: (a) $\sum_{x \in \{-1,1\}^N} \psi(x) \cdot f(x) \geq \varepsilon \cdot C$, where $C = \sum_{x \in \{-1,1\}^N} |\psi(x)|$, and (b) ψ has zero correlation with all polynomials of degree at most d. We refer to Property (a) by saying that ψ is *ε-correlated* with f. We refer to Property (b) by saying that ψ has *pure high degree d*. We refer to ψ as a *dual polynomial* for f.

A dual witness to the statement that $\widetilde{\deg}_{+,\varepsilon}(f) \geq d$ must satisfy an additional correlation condition, namely: (c) $\phi(x)$ agrees in sign with $f(x)$ for all $x \in f^{-1}(+1)$. We refer to Property (c) by saying that ϕ has *positive one-sided error*. (See Appendix A for details of the duality theory.)

We prove Theorem 1 by showing the following: given a dual polynomial ψ_{in} witnessing the assumed $\widetilde{\deg}_{+,2/3}$ lower bound on the inner function f, one can construct an explicit dual polynomial ψ_{comb} witnessing the claimed lower bound on the composed function $F = \mathrm{OMB}(f, \ldots, f)$.

1.2.2 Prior Work on the Approximate Degree of OMB

Beigel [5] proved that for any $d > 0$, there is an $\varepsilon \in 1 - 2^{-\Omega(n/d^2)}$ such that $\widetilde{\deg}_{\varepsilon}(\mathrm{OMB}_n) \geq d$, and used this result to give an oracle separating the (Turing Machine) complexity class PP from P^{NP}. Note that $\mathrm{OMB}_M(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ is a sub-function of $\mathrm{OMB}_{M \cdot (2N)}$. Moreover, it is known that $\widetilde{\deg}_{+,2/3}(\mathrm{OR}_N) = \Omega(N^{1/2})$ [8, 15, 25]. Hence, Theorem 1 can be viewed as a substantial generalization of Beigel's result: we recover Beigel's lower bound as a special case of Theorem 1 by letting $f = \mathrm{OR}_{d^2}$. Unlike Beigel's proof,

which used a non-constructive symmetrization technique, our proof of Theorem 1 constructs an explicit dual polynomial witnessing the lower bound.

For any $\varepsilon > 0$, Klivans and Servedio [20] gave an optimal ε-approximating polynomial for the function OMB, showing that Beigel's lower bound (and hence also our Theorem 1) is asymptotically tight for all $d > 0$.

1.2.3 Earlier Constructions of Dual Polynomials for Block-Composed Functions

Given functions g_M, f_N, Sherstov [39] and Lee [22] independently described a powerful method for constructing a dual polynomial for the composed function $F = g_M(f_N, \ldots, f_N) : \{-1,1\}^{M \cdot N} \to \{-1,1\}$. This method takes a dual polynomial ψ_{in} for f_N, and a dual polynomial ψ_{out} for g, and combines them to obtain a dual polynomial ψ_{comb} for the composed function F.

Specifically, denoting an $(M \cdot N)$-bit input as $(x_1, \ldots, x_M) \in \left(\{-1,1\}^N\right)^M$, Sherstov and Lee defined

$$\psi_{\text{comb}}(x_1, \ldots, x_M) = \psi_{\text{out}}\left(\widetilde{\text{sgn}}\left(\psi_{\text{in}}(x_1)\right), \ldots, \widetilde{\text{sgn}}\left(\psi_{\text{in}}(x_M)\right)\right) \cdot \prod_{i=1}^{M} |\psi_{\text{in}}(x_i)|. \tag{1}$$

Here, $\widetilde{\text{sgn}} : \mathbb{R} \to \{-1,0,1\}$ denotes the function satisfying $\widetilde{\text{sgn}}(t) = 1$ if $t > 0$, $\widetilde{\text{sgn}}(t) = -1$ if $t < 0$, and $\widetilde{\text{sgn}}(0) = 0$.

Recall that for ψ_{comb} to witness a good lower bound for the approximate degree of F, it must be well-correlated with F (Property (a) of Section 1.2.1), and it must have large pure high degree (Property (b) of Section 1.2.1). Sherstov and Lee showed that the pure high degree of ψ_{comb} is multiplicative in the pure high degrees of ψ_{in} and ψ_{out}. That is, if ψ_{in} has pure high degree d_1, and ψ_{out} has pure high degree d_2, then ψ_{comb} has pure high degree $d_1 \cdot d_2$. And while ψ_{comb} is not *in general* well-correlated with the composed function F, several important examples have been identified in which this is the case, as we now explain.

Sherstov [37] and independently Bun and Thaler [7] used the combining technique of Eq. (1) to resolve the (1/3)-approximate degree of the two-level AND-OR tree. Subsequent work by Bun and Thaler [8] used Eq. (1) to establish a hardness amplification result that looks similar to our Theorem 1. Specifically, Bun and Thaler proved:

Theorem 3 (Bun and Thaler [8])**.** *Suppose* $\widetilde{deg}_{-,2/3}(f) \geq d$. *Then* $\widetilde{deg}_{-,\varepsilon}(OR_M(f, \ldots, f)) \geq d$, *for* $\varepsilon = 1 - 2^{-M}$.

Theorem 3 is identical to our Theorem 1, but for two differences: first, in our Theorem 1, the outer function in the composition is OMB, while in Theorem 3 it is OR. Second, the hypothesis in Theorem 1 is that the inner function f satisfies $\widetilde{\deg}_{+,2/3}(f) \geq d$, while the assumption in Theorem 3 is that $\widetilde{\deg}_{-,2/3}(f) \geq d$. These differences are crucial for obtaining a hardness amplification result that applies to functions with low threshold degree. Indeed, subsequent work by Sherstov refined Bun and Thaler's construction to yield a threshold degree lower bound, rather than a $\widetilde{\deg}_{-,\varepsilon}$ lower bound [30].

Theorem 4 (Sherstov [30])**.** *Suppose* $\widetilde{deg}_{-,2/3}(f) \geq d$. *Then* $\deg_{\pm}(OR_M(f, \ldots, f)) \geq \min\{d, cM\}$ *for some constant* $c > 0$.

Sherstov's proof of Theorem 4 also draws heavily on Eq. (1): he constructs a dual witness of the form $\psi_{\text{comb}} + \psi_{\text{fix}}$, where ψ_{comb} is the dual witness constructed by Bun and Thaler using Eq. (1) to prove Theorem 3, and ψ_{fix} is used "zero out" ψ_{comb} on points x such that $0 \neq \widetilde{\text{sgn}}(\psi_{\text{comb}}(x)) \neq \widetilde{\text{sgn}}(\text{OR}_M(f, \ldots, f))$. This ensures that $\psi_{\text{comb}} + \psi_{\text{fix}}$ is perfectly correlated with F.

Sherstov used Theorem 4 to give a depth three circuit with threshold degree $\tilde{\Omega}(n^{2/5})$. He also established the following result, which yields a polynomially stronger lower bound for depth $k > 3$.

Theorem 5 (Sherstov [30])**.** *For any $k \geq 2$, there is a depth k (read-once) Boolean circuit computing a function F satisfying* $\deg_{\pm}(F) = \Omega(n^{(k-1)/(2k-1)})$.

Sherstov's proof of Theorem 5 relies on an elaborate inductive construction of a dual polynomial that is also reminiscent of Eq. (1).

1.2.4 Complementary Slackness and the Need for New Techniques

In this section, we explain why any dual witness establishing Theorem 1 must qualitatively depart from the dual witnesses constructed in prior work (cf. Section 1.2.3). In brief, we first argue that the dual witnesses constructed in prior work are implicitly tailored to show optimality of a specific technique for approximating block-composed functions. We then explain that this technique is far from optimal for the functions to which Theorem 1 applies.

Approximating Block-Composed Functions via "Robustification". Sherstov [40] provided a generic technique forapproximating block-composed functions. Specifically, he showed that for any polynomial $p : \{-1,1\}^M \to [-1,1]$, and every $\delta > 0$, there is a polynomial $p_{\text{robust}} : \mathbb{R}^M \to \mathbb{R}$ of degree $O(\deg(p) + \log(1/\delta))$ that is robust to noise in the sense that $|p(y) - p_{\text{robust}}(y + \mathbf{e})| < \delta$ for all $y \in \{-1,1\}^M$ and $\mathbf{e} \in [1/3, 1/3]^M$. Hence, given functions $g = g_M, f = f_N$, one can obtain an $(\varepsilon + \delta)$-approximating polynomial for the block-composition $g(f, \ldots, f)$ as follows: let p be an ε-approximating polynomial for g, and q a $(1/3)$-approximating polynomial for f. Then the block composition $p^* := p_{\text{robust}}(q, \ldots, q)$ is an $(\varepsilon + \delta)$-approximating polynomial for $g(f, \ldots, f)$. Notice that the degree of p^* is at most the product of the degrees of p_{robust} and q.

This generic construction yields asymptotically optimal ε-approximating polynomials for essentially all block-composed functions considered in prior work on hardness amplification. Indeed, this holds for the two-level AND-OR tree when $\varepsilon = 1/3$ [7, 37], as well as for the functions considered in Theorems 3, 4, and 5, for ε exponentially close to 1 (see e.g. [30, Theorem 1.2]).

Showing Robustification Is Optimal (Except When It's Not). Intuitively, the dual witness ψ_{comb} constructed via Eq. (1) is specifically tailored to show optimality of the above generic technique for approximating block-composed functions. Indeed, ψ_{comb} "almost" obeys complementary slackness with respect to p^* in the following sense.

Suppose that p_{robust} achieved *exactly* optimal error ε among all degree d polynomial approximations to the outer function g. Then p_{robust} yields an optimal solution to the relevant linear program capturing the ε-approximate degree of g (cf. Appendix A). Complementary slackness states that there is an optimal dual solution (i.e., a weighting of the constraints from the primal linear program) which places non-zero weight only on constraints that are made tight by the primal optimum. In our context, this means that there is an optimal dual polynomial ψ_{out} for g such that $\psi_{\text{out}}(y) \neq 0$ only for "maximal error points" $y \in \{-1,1\}^M$, i.e., points y satisfying $|p_{\text{robust}}(y) - g(y)| = \varepsilon$. Let ψ_{in} be any dual polynomial for the inner function f, and suppose ψ_{out} is combined with ψ_{in} as per Eq. (1) to obtain a dual polynomial ψ_{comb} for $g(f, \ldots, f)$.

If ψ_{in} were *perfectly* correlated with f, then one can check that $\psi_{\text{comb}}(x) \neq 0$ only for $x = (x_1, \ldots, x_M) \in \left(\{-1,1\}^N\right)^M$ such that

$$\begin{aligned} &|p_{\text{robust}}(q(x_1), \ldots, q(x_M)) - g(f(x_1), \ldots, f(x_M))| = \\ &|p_{\text{robust}}(f(x_1), \ldots, f(x_M)) - g(f(x_1), \ldots, f(x_M))| \pm \delta \geq \varepsilon - \delta \approx \varepsilon. \end{aligned}$$

Put another way, ψ_{comb} places non-zero weight only on points on which $p_{\text{robust}}(q, \ldots, q)$ achieves "nearly maximal error" of at least $\varepsilon - \delta$. This is what we mean when we say that ψ_{comb} "almost" satisfies complementary slackness with respect to the primal solution corresponding to $p_{\text{robust}}(q, \ldots, q)$.

In general, ψ_{in} will not be perfectly correlated with f, but the analyses of ψ_{comb} from prior work identify settings in which ψ_{comb} still places "most" of its weight on points x such that $|p_{\text{robust}}(q,\ldots,q) - g(f,\ldots,f)| = \varepsilon \pm \delta \approx \varepsilon$.

When Robustification Is Sub-Optimal. In contrast to these earlier results, Theorem 1 applies to functions for which $p^* := p_{\text{robust}}(q,\ldots,q)$ is not an optimal approximating polynomial. To see this, recall from Section 1.2.2 that, for any $\varepsilon > 0$, Klivans and Servedio [20] gave an optimal ε-approximating polynomial for the function $\text{OMB}_{M\cdot(2N)}$, which contains the function $\text{OMB}_M(\text{OR}_N,\ldots,\text{OR}_N)$ as a subfunction. It can be seen that the approximating polynomial exhibited by Klivans and Servedio is of the form $p(q,\ldots,q)$, where p is a *non-robust* ε-approximating polynomial for OMB_M (for some $\varepsilon = 1 - 2^{-\Theta(M)}$), and q is a $(1/3)$-approximating polynomial for OR_N.

Since $p_{\text{robust}}(q,\ldots,q)$ is not an optimal approximating polynomial for $\text{OMB}_M(\text{OR}_N,\ldots,\text{OR}_N)$, we do not expect there to be any dual witness obeying complementary slackness with respect to $p_{\text{robust}}(q,\ldots,q)$. Accordingly, the dual witness ψ_{comb} that we construct to prove Theorem 1 departs from Eq. (1).

Remark 6. *One reason that Klivans and Servedio [20] do not need to use a robust approximating polynomial for the outer function* OMB_M *is that they use an inner approximation q for the inner function f that is* exactly *correct for inputs in $f^{-1}(+1)$. Hence, they can use an outer approximation p that is robust only to highly restricted noise vectors. Namely, for any input x, p needs to be robust only to noise vectors* $\mathbf{e}$ *such that $\mathbf{e}_i = 0$ on all coordinates i such that $x_i = +1$.*

1.2.5 Roadmap for the Rest of the Paper

For completeness, we collect formal definitions of approximate degree and its one-sided variants, along with their dual characterizations, in Appendix A. We introduce notation and establish preliminary lemmas in Section 2. Section 3 provides an intuitive overview of the dual witness we construct to prove Theorem 1, before providing proof details. Section 4 proves Theorem 2.

2 Notation and Preliminary Facts

Given a set $T \subseteq \{-1,1\}^N$, we let $\mathbb{I}_T$ denote the indicator vector of T; that is, $\mathbb{I}_T(x) = 1$ if $x \in T$, and $\mathbb{I}_T(x) = 0$ otherwise. Given a dual polynomial $\psi : \{-1,1\}^N \to \mathbb{R}$, we define the L_1*-weight* of T under ψ to be $W_\psi(T) = \sum_{x\in T} |\psi(x)|$. We refer to $W_\psi(\{-1,1\}^N)$ as the L_1*-norm* of ψ.

We define the function $\widetilde{\text{sgn}} : \mathbb{R} \to \{-1,0,1\}$ via:

$$\widetilde{\text{sgn}}(t) = \begin{cases} 1 \text{ if } t > 0 \\ -1 \text{ if } t < 0 \\ 0 \text{ otherwise.} \end{cases}$$

We say that a dual polynomial ψ for a function f *makes an error* on input x if $0 \neq \widetilde{\text{sgn}}(\psi(x)) \neq \widetilde{\text{sgn}}(f(x))$.

Crucial to our proof are the following two facts that provide methods of combining multiple dual witnesses while preserving their pure high degree.

Fact 7. *If $\psi_1, \psi_2 : \left(\{-1,1\}^N\right)^M \to \{-1,1\}$ both have pure high degree d, then so does $\psi_1 + \psi_2$.*

Proof. Let $g : \left(\{-1,1\}^N\right)^M \to \{-1,1\}$ be any polynomial of degree at most d. Then

$$\sum_{x\in(\{-1,1\}^N)^M} (\psi_1(x)+\psi_2(x))\, g(x) = \left(\sum_{x\in(\{-1,1\}^N)^M} \psi_1(x)g(x)\right) + \left(\sum_{x\in(\{-1,1\}^N)^M} \psi_2(x)g(x)\right) = 0+0 = 0.$$

□

Fact 8. *Suppose that $\psi_1,\ldots,\psi_M : \{-1,1\}^N \to \{-1,1\}$ are each defined over disjoint sets of variables, and there is some i such that ψ_i has pure high degree d. Then so does the function $\psi : \left(\{-1,1\}^N\right)^M \to \{-1,1\}$ defined via $\psi(x_1,\ldots,x_M) = \prod_{i=1}^M \psi_i(x_i)$.*

Proof. Let $g : \left(\{-1,1\}^N\right)^M \to \mathbb{R}$ be any polynomial of degree at most d. Letting $x = (x_1,\ldots,x_M) \in \left(\{-1,1\}^N\right)^M$, we assume without loss of generality that $g = \prod_{j=1}^M g_j(x_i)$, where $\deg(g_j) \leq d$ for all j (the general case follows from this special case by linearity). We must show that ψ is uncorrelated with g. To see this, note that:

$$\sum_{x\in(\{-1,1\}^N)^M} \psi(x)\cdot g(x) = \sum_{x_1,\ldots,x_M\in\{-1,1\}^N} \psi(x_1,\ldots,x_M)\cdot g(x_1,\ldots,x_M) = \sum_{x_1,\ldots,x_M\in\{-1,1\}^N} \prod_{j=1}^M (\psi_j(x_j)\cdot g_j(x_j))$$

$$= \left(\sum_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_M\in\{-1,1\}^N} \prod_{j=1,j\neq i}^M \psi_i(x_i)\cdot g_j(x_i)\right)\left(\sum_{x_i\in\{-1,1\}^N} \psi_i(x_i)\cdot g_i(x_i)\right)$$

$$= \left(\sum_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_M\in\{-1,1\}^N} \prod_{j=1,j\neq i}^M \psi_j(x_j)\cdot g_j(x_j)\right)\cdot 0 = 0$$

Here, the second equality holds by definition of ψ, and the fourth because ψ_i has pure high degree d and $\deg(g_i) \leq \deg(g) \leq d$.

□

3 Proof of Theorem 1

This section proves Theorem 1, which we restate here for the reader's convenience. Recall from the introduction that for any Boolean function $f : \{-1,1\}^N \to \{-1,1\}$, F denotes the function $\mathrm{OMB}_M(f,\ldots,f)$ that maps $\{-1,1\}^{M\cdot N}$ to $\{-1,1\}$.

Theorem 1. *If $\widetilde{deg}_{+,2/3}(f) \geq d$, then $\widetilde{deg}_{+,\varepsilon}(F) \geq d$ for $\varepsilon = 1-2^{-M}$.*

Proof. Let ψ_{in} denote a dual witness for the fact that $\widetilde{\deg}_{+,2/3}(f) \geq d$, normalized to ensure that its L_1-norm is 1. Recall from Section 1.2.1 that ψ_{in} satisfies three properties: (a) ψ_{in} has pure high degree at least d, (b) ψ_{in} has correlation $\varepsilon' \geq 2/3$ with f, and (c) $\psi_{\text{in}}(x_i) \geq 0$ for all $x_i \in f^{-1}(+1)$. Let E denote the set of all $x_i \in \{-1,1\}^N$ on which $\psi_{\text{in}}(x_i)$ is in error, i.e., $0 \neq \widetilde{\text{sgn}}(\psi_{\text{in}}(x_i)) \neq \widetilde{\text{sgn}}(f(x_i))$.

Proof Overview. For any vector $x = (x_1,\ldots,x_M) \in \left(\{-1,1\}^N\right)^M$, we think of x_M as the "most significant" block in x, because if $f(x_M) = -1$, then F evaluates to -1 regardless of the values of the other blocks $x_1,\ldots,x_{M-1}$. Similarly, we think of x_1 as the "least significant block" of x.

We think of our dual witness ψ_{comb} as being constructed iteratively. The first iteration will create a dual witness $\psi^{(1)}$ that "uses" the least significant block x_1 to "achieve" pure high degree at least d. However, $\psi^{(1)}$ will only have correlation ε' with F, and hence it will make errors if $\varepsilon' < 1$. The second iteration creates a dual witness $\psi^{(2)}_{\text{comb}} = \psi^{(1)} + \psi^{(2)}$, where $\psi^{(2)}$ is a correction term that zeros out there errors of $\psi^{(1)}$. Moreover, $\psi^{(2)}$ will use the second block x_2 to achieve pure high degree at least d. By Fact 7, this ensures that $\psi^{(2)}_{\text{comb}}$ also has pure high degree at least d.

If $\psi^{(2)}$ zeroed out all of the errors of $\psi^{(1)}$ without introducing any new errors, then $\psi^{(2)}_{\text{comb}}$ would have perfect correlation with F, and we would be done. Unfortunately, $\psi^{(2)}$ does introduce new errors. But we have made tangible progress: we show that the number of errors $\psi^{(2)}$ makes, relative to $\psi^{(1)}$, falls by a factor of $W_{\psi_{\text{in}}}(f^{-1}(+1))/W_{\psi_{\text{in}}}(E) = \varepsilon'/(1-\varepsilon')$. Since $\varepsilon' \geq 2/3$, we conclude that $\varepsilon'/(1-\varepsilon') \geq 2$, and hence that $\psi^{(2)}$ makes at most half as many errors as $\psi^{(1)}$.

In general, the ith iteration adds in a correction term $\psi^{(i)}$ that zeros out all of the errors of the dual witness $\psi^{(i-1)}_{\text{comb}}$ constructed in the previous iteration. $\psi^{(i)}$ will use the ith input block x_i to achieve pure high degree at least d, and will introduce at most a $W_{\psi_{\text{in}}}(E)/W_{\psi_{\text{in}}}(f^{-1}(+1)) \leq 1/2$ fraction of the errors made by $\psi^{(i-1)}$. At the end of iteration M, we have constructed a dual witness $\psi_{\text{comb}} := \sum_{i=1}^{M} \psi^{(i)}$ that makes only a $\left(W_{\psi_{\text{in}}}(E)/W_{\psi_{\text{in}}}(f^{-1}(+1))\right)^M = ((1-\varepsilon')/\varepsilon')^M \leq 2^{-M}$ fraction of the errors made by $\psi^{(1)}$, and we are done.

Proof Details.

Properties of ψ_{in}. Throughout, we let $Q^-, Q^+ \subseteq \{-1,1\}^N$ denote the set of inputs x_i for which $\psi_{\text{in}}(x_i) < 0$ and $\psi_{\text{in}}(x_i) > 0$ respectively. We assume $d \geq 1$, as otherwise Theorem 1 holds trivially. We make use of the following simple facts about $\mathbb{I}_{Q^+}$ and $\mathbb{I}_{Q^-}$.

Fact 9. $\sum_{x_i \in \{-1,1\}^N} \mathbb{I}_{Q^-}(x_i) \cdot |\psi_{in}(x_i)| = \sum_{x_i \in \{-1,1\}^N} \mathbb{I}_{Q^+}(x_i) \cdot |\psi_{in}(x_i)| = 1/2$.

Proof. Since ψ_{in} witnesses the fact that $\widetilde{\deg}_{+,1/2}(f) \geq d$, ψ_{in} has pure high degree at least $d \geq 1$. In particular, ψ_{in} is uncorrelated with any constant function. Hence, $\sum_{x_i \in \{-1,1\}^N} \psi_{\text{in}}(x_i) = 0$. Since $\sum_{x_i \in \{-1,1\}^N} |\psi_{\text{in}}(x_i)| = 1$, it follows that $\sum_{x_i \in \{-1,1\}^N : x_i \in Q^+} |\psi_{\text{in}}(x_i)| = \sum_{x_i \in \{-1,1\}^N : x_i \in Q^-} |\psi_{\text{in}}(x_i)| = 1/2$, which is equivalent to the statement we wished to prove. □

A crucial implication of Property (c) is that if ψ_{in} outputs a negative value on input x_i, we can "trust" that $f(x_i) = -1$, as formalized in the next fact.

Fact 10. *For all $x_i \in Q^-$, it holds that $f(x_i) = -1$. Equivalently, $E \subseteq f^{-1}(-1)$, or in other words $E \cap f^{-1}(+1) = \emptyset$.*

The following two facts relate the correlation of ψ_{in} with f to the L_1-weight of the sets E and $f^{-1}(+1)$ under ψ_{in}.

Fact 11. $W_{\psi_{in}}(E) = (1-\varepsilon')/2$.

Proof. By Property (a), $\varepsilon' = \sum_{x_i \in \{-1,1\}^N} \psi_{\text{in}}(x_i) \cdot f(x_i) = 1 - 2\sum_{x_i \in E} |\psi_{\text{in}}(x_i)|$. □

Fact 12. $W_{\psi_{in}}(f^{-1}(+1)) = \varepsilon'/2$.

Proof. This holds by the following sequence of equalities:

$$1/2 = \sum_{x_i \in Q^+} |\psi_{\text{in}}(x_i)| = \sum_{x_i \in E} |\psi_{\text{in}}(x_i)| + \sum_{x_i \in f^{-1}(+1)} |\psi_{\text{in}}(x_i)| = \left(1/2 - \varepsilon'/2\right) + \sum_{x_i \in f^{-1}(+1)} |\psi_{\text{in}}(x_i)|.$$

Here, the first equality holds by Fact 9, the second because ψ_{in} satisfies Property (c), and the third by Fact 11. □

Construction of ψ_{comb}. The dual witness we construct is:

$$\psi_{\text{comb}}(x_1,\ldots,x_M) = \sum_{i=1}^{M} \psi^{(i)}, \text{ where} \tag{2}$$

$$\psi^{(i)} = (-1)^{i-1}\cdot(2/\varepsilon')^{M-1}\left(\prod_{j<i}\mathbb{I}_E(x_j)\cdot|\psi_{\text{in}}(x_j)|\right)\cdot\psi_{\text{in}}(x_i)\cdot\left(\prod_{j=i+1}^{M}\mathbb{I}_{f^{-1}(+1)}(x_j)\cdot|\psi_{\text{in}}(x_j)|\right). \tag{3}$$

Recall that, to show that ψ_{comb} witnesses the fact that $\widetilde{\deg}_{+,\varepsilon}(F)\geq d$ for $\varepsilon = 1-2^{-M}$, it suffices to show that ψ_{comb} satisfies three properties (cf. Lemma 25 in Appendix A): (a) it must have pure high degree at least d, (b) it must satisfy $\sum_{x\in(\{-1,1\}^N)^M}\psi_{\text{comb}}(x)\cdot F(x)\geq C\cdot\varepsilon$, where $C=\sum_{(x\in\{-1,1\}^N)^M}|\psi_{\text{comb}}(x)|$, and (c) it must have positive one-sided error. We establish each in turn below, in Propositions 13, 14, and 17.

Proposition 13. *ψ_{comb} has pure high degree at least d.*

Proof. Since ψ_{in} has pure high degree at least d, Fact 8 implies that each term $\psi^{(i)}$ in the sum within Eq. (2) also has pure high degree at least d. The lemma then follows by Fact 7. □

Proposition 14. *$\sum_{x\in(\{-1,1\}^N)^M}\psi_{comb}(x)\cdot F(x)\geq C\cdot\varepsilon$.*

The proof of Proposition 14 will make use of the following two lemmas.

Lemma 15. *$C\geq 1/2$.*

Proof. Consider the set $S=\{(x_1,\ldots,x_M): x_1\in Q^-\text{ and }x_2,\ldots,x_M\in f^{-1}(+1)\}$. We claim that the weight, $W_{\psi_{\text{comb}}}(S)$, that ψ_{comb} places on the set S is $1/2$. The lemma clearly follows.

To see this, fix $x=(x_1,\ldots,x_M)\in S$. We first note that for all $i\geq 2$, $\psi^{(i)}(x)=0$. Indeed, $Q^-\cap E=\emptyset$ (cf. Fact 10), and hence $\mathbb{I}_E(x_1)=0$. Thus, it is immediate from Eq. (3) that $\psi^{(i)}(x)=0$ for $i\geq 2$.

So it suffices to show that $\sum_{x\in S}-\psi^{(1)}(x)\geq 1/2$. This follows from the following calculation:

$$\sum_{x\in S}-\psi^{(1)}(x) = (2/\varepsilon')^{M-1}\cdot\left(\sum_{x_1\in Q^-}-\psi_{\text{in}}(x_1)\right)\cdot\left(\prod_{j=2}^{M}\left(\sum_{x_j\in\{-1,1\}^N}\mathbb{I}_{f^{-1}(+1)}(x_j)\cdot|\psi_{\text{in}}(x_j)|\right)\right)$$
$$= (2/\varepsilon')^{M-1}\cdot(1/2)\cdot\prod_{j=2}^{M}(\varepsilon'/2) = 1/2,$$

where the first equality holds by Eq. (3), and the second holds by Facts 9 and 12. □

Lemma 16. *Let $E_{comb}\subseteq\left(\{-1,1\}^N\right)^M$ denote the set of inputs on which ψ_{comb} makes an error, i.e., $0\neq\widetilde{\text{sgn}}(\psi_{comb}(x))\neq\widetilde{\text{sgn}}(F(x))$. Let $E^M\subseteq\left(\{-1,1\}^N\right)^M$ denote $\{(x_1,\ldots,x_M): x_i\in E\text{ for all }i\}$. Then $E_{comb}=E^M$.*

Proof. We first show that $E^M \subseteq E_{\text{comb}}$ before showing that $E_{\text{comb}} \subseteq E^M$. Suppose that $x = (x_1, \ldots, x_M) \in E^M$. Fact 10 states that $E \subseteq f^{-1}(-1)$, and hence $\mathbb{I}_{f^{-1}(+1)}(x_M) = 0$. It is then immediate from Eq. (3) that $\psi^{(i)}(x) = 0$ for all $i < M$. Meanwhile, by Eq. (3) it holds that

$$\widetilde{\text{sgn}}(\psi^{(M)}(x)) = (-1)^{M-1} \cdot \widetilde{\text{sgn}}(\psi_{\text{in}}(x_M)) = (-1)^{M-1}.$$

Here, we used the fact that $\widetilde{\text{sgn}}(\psi_{\text{in}}(x_M)) > 0$ if $x_M \in E$. (To see this, note that since $x_M \in E$, it holds that

$$0 \neq \widetilde{\text{sgn}}(\psi_{\text{in}}(x_M)) \neq f(x_M) = -1,$$

where the final equality holds because $E \subseteq f^{-1}(-1)$.) At the same time, $F(x) = \text{OMB}_M(-1, -1, \ldots, -1) = (-1)^M$. Thus, $x \in E_{\text{comb}}$ as claimed.

Fix any $x = (x_1, \ldots, x_M) \in \left(\{-1, 1\}^N\right)^M$ such that there exists an $i \in \{1, \ldots, M\}$ satisfying $x_i \notin E$. To show that $E_{\text{comb}} \subseteq E^M$, we must show that $x \notin E_{\text{comb}}$. To this end, let i^* be the smallest coordinate such that $x_{i^*} \notin E$. It is clear that $\psi_{\text{comb}}(x) = 0$ if $\psi_{\text{in}}(x_i) = 0$ for any $i \in [M]$, and hence $x \notin E_{\text{comb}}$. So assume throughout that $\psi_{\text{in}}(x_i) \neq 0$ for all i. The proof proceeds via a case analysis.

- Case 1: There exists a $j > i^*$ such that $x_j \notin f^{-1}(+1)$. In this case, $\mathbb{I}_{f^{-1}(+1)}(x_j) = 0$, so it is immediate from Eq. (3) that $\psi^{(k)}(x) = 0$ for all $k < j$. Meanwhile, since $\mathbb{I}_E(x_{i^*}) = 0$, it is immediate from Eq. (3) that $\psi^{(k)}(x) = 0$ for all $k \geq j$. Thus, $\psi_{\text{comb}}(x) = \sum_{k=0}^{M} \psi^{(k)}(x) = 0$, implying that $x \notin E_{\text{comb}}$.

- Case 2: $i^* = 1$, and $x_j \in f^{-1}(+1)$ for all $j > i^*$. In this case, it is clear by Eq. (3) that

$$\widetilde{\text{sgn}}(\psi^{(1)}(x)) = (-1)^0 \cdot \widetilde{\text{sgn}}(\psi_{\text{in}}(x_1)) = \widetilde{\text{sgn}}(\psi_{\text{in}}(x_1)) = \widetilde{\text{sgn}}(f(x_1)) = F(x_1, \ldots, x_M). \tag{4}$$

  Here, the third equality holds because $x_1 \notin E$, and the fourth equality exploits the fact that if $x_j \in f^{-1}(+1)$ for all $j > 1$, then $F(x) = f(x_1)$.

  Meanwhile, since $x_1 \notin E$, it holds that $\mathbb{I}_E(x_1) = 0$, and so it is clear by Eq. (3) that $\psi^{(k)}(x) = 0$ for all $k \geq 2$. Combining this with Eq. (4), we conclude that $\widetilde{\text{sgn}}(\psi_{\text{comb}}(x)) = \widetilde{\text{sgn}}(\psi^{(1)}(x)) = F(x_1, \ldots, x_M)$. Thus, $x \notin E_{\text{comb}}$.

- Case 3: $i^* \geq 2$, and $x_j \in f^{-1}(+1)$ for all $j > i^*$. First, we argue that $\psi^{(k)} = 0$ for all $k < i^* - 1$. Indeed, for all such k, $x_{k+1} \in E \subseteq f^{-1}(-1)$ (cf. Fact 10), and so it holds that $\mathbb{I}_{f^{-1}(+1)}(x_{k+1}) = 0$. Hence, it is immediate from Eq. (3) that $\psi^{(k)}(x) = 0$.

  Next, we argue that $\psi^{(k)} = 0$ for all $k \geq i^* + 1$. Indeed, $x_{i^*} \notin E$, so $\mathbb{I}_E(x_{i^*}) = 0$. It is then immediate from Eq. (3) that $\psi^{(k)}(x) = 0$ for all $k \geq i^* + 1$.

  Finally, we claim that either $\psi^{(i^*-1)}(x) + \psi^{(i^*)}(x) = 0$ or $\widetilde{\text{sgn}}(\psi^{(i^*-1)}(x) + \psi^{(i^*)}(x)) = F(x)$. This follows from the following calculation.

  - Case 3a: Suppose $x_{i^*} \notin f^{-1}(+1)$, i.e., that $\mathbb{I}_{f^{-1}(+1)}(x_{i^*}) = 0$. Then is clear from Eq. (3) that $\psi^{(i^*-1)}(x) = 0$. Meanwhile, since $x_{i^*} \notin E$, it is clear from Eq. (3) that

$$\widetilde{\text{sgn}}(\psi^{(i^*)}(x)) = (-1)^{i^*-1} \cdot \widetilde{\text{sgn}}(\psi_{\text{in}}(x_{i^*})) = (-1)^{i^*-1} \cdot f(x_{i^*}) = F(x),$$

    where the final equality exploits the fact that if $x_j \in f^{-1}(+1)$ for all $j > i^*$, and $x_{i^*-1} \in E \subseteq f^{-1}(-1)$ (Fact 10), then $F(x) = (-1)^{i^*-1} \cdot f(x_{i^*})$.

– Case 3b: Suppose $x_{i^*} \in f^{-1}(+1)$. We claim that it holds that $\psi^{(i^*-1)}(x) = -\psi^{(i^*)}(x)$. To see this, note that in this case

$$\psi^{(i^*-1)}(x) = (-1)^{i^*-2} \cdot (2/\varepsilon')^{M-1} \cdot \psi_{\text{in}}(x_{i^*-1}) \cdot \prod_{j \neq i^*-1} |\psi_{\text{in}}(x_j)|, \text{ and} \tag{5}$$

$$\psi^{(i^*)}(x) = (-1)^{i^*-1} \cdot (2/\varepsilon')^{M-1} \cdot \psi_{\text{in}}(x_{i^*}) \cdot \prod_{j \neq i^*} |\psi_{\text{in}}(x_j)|. \tag{6}$$

Both of the above quantities are clearly equal in absolute value, but it remains to show that $\psi^{(i^*-1)}(x) = -\psi^{(i^*)}(x)$. Since $x_{i^*-1} \in E \subseteq f^{-1}(-1)$ (Fact 10), it holds that $\widetilde{\text{sgn}}(\psi_{\text{in}}(x_{i^*-1})) = +1$. Meanwhile, since $x_{i^*} \notin E$, $\widetilde{\text{sgn}}(\psi_{\text{in}}(x_{i^*})) = f(x_{i^*}) = +1$. Hence, $\widetilde{\text{sgn}}(\psi^{(i^*-1)}(x)) = (-1)^{i^*-2}$, while $\widetilde{\text{sgn}}(\psi^{(i^*)}(x)) = (-1)^{i^*-1}$, completing the proof.

Combining all of the above, we conclude that $\psi_{\text{comb}}(x) = \sum_{j=1}^{M} \psi_{\text{comb}}^{(j)}(x) = \psi_{\text{comb}}^{(i^*-1)}(x) + \psi_{\text{comb}}^{(i^*)}(x)$, and the latter expression is either equal to 0 or agrees in sign with $F(x)$. Thus, $x \notin E_{\text{comb}}$. This completes the proof of Lemma 16.

□

Proof of Proposition 14. Note that

$$\sum_{x \in (\{-1,1\}^N)^M} \psi_{\text{comb}}(x) \cdot F(x) = \sum_{x \in (\{-1,1\}^N)^M} |\psi_{\text{comb}}(x)| - 2 \sum_{x \in E_{\text{comb}}} |\psi_{\text{comb}}(x)| = C - 2 \sum_{x \in E_{\text{comb}}} |\psi_{\text{comb}}(x)|, \tag{7}$$

where we recall from Lemma 16 that $E_{\text{comb}} = E^M$ is the set of points on which ψ_{comb} makes an error. Observe that for each j:

$$\sum_{x \in E^M} \psi^{(j)}(x) \quad \leq (2/\varepsilon')^{M-1} \prod_{i=1}^{M} \left(\sum_{x_i \in E} |\psi_{\text{in}}(x_i)| \right) \leq (2/\varepsilon')^{M-1} \cdot \prod_{i=1}^{M} \big((1-\varepsilon')/2\big) \leq 3^{M-1}/6^M < 2^{-M-1}. \tag{8}$$

Here, the first equality holds because, for all $x \in E^M$ and $j < M$, $\psi^{(j)}(x) = 0$; this follows by combining Eq. (3) with the fact that $E \cap f^{-1}(+1) = \emptyset$ (Fact 10) (see also the $E^M \subseteq E_{\text{comb}}$ direction in the proof of Lemma 16). The second inequality holds by Fact 11, and the third holds because $\varepsilon' \geq 2/3$. Combining Lemma 15 with Eq. (7) and Eq. (8), we conclude that $\sum_{x \in (\{-1,1\}^N)^M} \psi_{\text{comb}}(x) \cdot F(x) \geq C - 2^{-M-1} \geq C(1 - 2^{-M})$, completing the proof. □

Proposition 17. *$\psi_{comb}(x) \geq 0$ for all $x \in F^{-1}(+1)$.*

Proof. Assume without loss of generality that M is odd. Lemma 16 implies that the set E_{comb} on which ψ_{comb} makes an error is equal to E^M. Since $E \subseteq f^{-1}(-1)$ (cf. Fact 10), it is obvious from the definition of F that $E^M \subseteq F^{-1}(-1)$. It follows that ψ_{comb} makes no errors on $F^{-1}(+1)$, implying the proposition. □

Theorem 1 follows by combining Propositions 13, 14, and 17 and the dual characterization of $\widetilde{\deg}_{+,\varepsilon}$ (cf. Lemma 25 in Appendix A). □

4 Proof of Theorem 2

For convenience, we restate Theorem 2, which gives an improved separation between the PP^{cc} and UPP^{cc} communication complexities of an AC^0 function F' (equivalently, between the margin and dimension complexities of F'). As stated in the introduction, the previous best separation for AC^0 was due to Buhrman et al. [6], who identified a function F' with $\mathrm{UPP}^{cc}(F') = O(\log n)$ and $\mathrm{PP}^{cc}(F') = \Omega(n^{1/3})$.

Theorem 2. *There is an explicit function F' computed by a polynomial size circuit of constant depth satisfying: $\mathrm{UPP}^{cc}(F') = O(\log^2 n)$ and $\mathrm{PP}^{cc}(F') = \tilde{\Omega}(n^{2/5})$, where the $\tilde{\Omega}$ notation hides logarithmic factors.*

Proof. Before describing the function F', we first introduce the concept of discrepancy.

Discrepancy. Consider a Boolean function $F : X \times Y \to \{-1,1\}$, and let $M^{(F)}$ be its communication matrix $M^{(F)} = [F(x,y)]_{x\in X, y\in Y}$. A combinatorial rectangle of $X \times Y$ is a set of the form $A \times B$ with $A \subseteq X$ and $B \subseteq Y$. For a distribution μ over $X \times Y$, the *discrepancy* of F with respect to μ is defined to be the maximum over all rectangles R of the *bias* of F on R. That is:

$$\mathrm{disc}_\mu(F) = \max_R \left| \sum_{(x,y)\in R} \mu(x,y)F(x,y) \right|.$$

The discrepancy of F, $\mathrm{disc}(F)$, is defined to be $\min_\mu \mathrm{disc}_\mu(F)$. It is known that discrepancy characterizes the communication model PP^{cc}, in the sense that $\mathrm{PP}^{cc}(F) = \Theta(\log(1/\mathrm{disc}(F)) + \log\log(|X|\cdot|Y|))$ [18].

Sherstov's pattern matrix method [34] shows how to generically transform an AC^0 function F such that $\widetilde{\deg}_\varepsilon(F)$ is large, into another AC^0 function with low discrepancy, as long as ε is exponentially close to 1.

Lemma 18 ([34], adapted from Corollary 1.2 and Theorem 7.3)**.** *Let $F : \{-1,1\}^n \to \{-1,1\}$ be given, and define the communication problem $F' : \{-1,1\}^{4n} \times \{-1,1\}^{4n} \to \{-1,1\}$ by*

$$F'(x,y) = F(\ldots, \vee_{j=1}^4 (x_{i,j} \wedge y_{i,j}), \ldots).$$

Suppose that $\widetilde{deg}_\varepsilon(F) \geq d$ for $\varepsilon = 1 - 2^{-d}$. Then $disc(F')^2 \leq 2n \cdot 2^{-d}$.

A Function F' with Small Discrepancy. Bun and Thaler [8, Corollary 3], building on work of Aaronson and Shi [1], exhibit a function known as $\mathrm{ED}_N : \{-1,1\}^N \to \{-1,1\}$ (short for ELEMENT DISTINCTNESS) that is computed by a polynomial size CNF formula, and satisfies $\widetilde{\deg}_{-,1/3}(\mathrm{ED}_N) = \Omega((N/\log N)^{2/3})$. Specifically, ED_N is defined as follows: Fix an $R = \Theta(N)$ that is a power of 2, and let $N = m \cdot \log_2 R$ for some $m = \Theta(N/\log N)$. ED_N takes N bits as input, and interprets its input as m blocks $(x_1, \ldots, x_m)$ with each block consisting of $\log_2 R$ bits. Each block is interpreted as a number in the range $[R]$, and ED_N evaluates to -1 on x if and only if all m numbers are distinct.

It is easy to see that for any function f, $\widetilde{\deg}_{+,\varepsilon}(f) = \widetilde{\deg}_{-,\varepsilon}(\overline{f})$, where $\overline{f}$ denotes the negation of f. Hence, Bun and Thaler's result implies the following:

Lemma 19. *There is a function, $\overline{\mathrm{ED}}_N : \{-1,1\}^N \to \{-1,1\}$, computed by a DNF formula of polynomial size, such that $\widetilde{deg}_{+,2/3}(\overline{\mathrm{ED}}_N) = \Omega((N/\log N)^{2/3})$.*

Fix an $n > 0$. Let $F : \{-1,1\}^n \to \{-1,1\}$ be defined via: $F = \mathrm{OMB}_M(\overline{\mathrm{ED}}_N, \ldots, \overline{\mathrm{ED}}_N)$, where $M = n^{2/5}$ and $N = n^{3/5}$. Clearly F is computed by a polynomial size circuit of depth four. Theorem 1, combined with Lemma 19, implies that

$$\widetilde{\deg}_{+,\varepsilon}(F) = \tilde{\Omega}(n^{2/5}), \text{ for some } \varepsilon = 1 - 2^{-\tilde{\Omega}(n^{2/5})}. \tag{9}$$

Combining Eq. (9) with Lemma 18, we obtain a function F' (computed by a polynomial size circuit of depth 6) satisfying $\mathrm{disc}(F') \leq 2^{-\tilde{\Omega}(n^{2/5})}$. Since $\mathrm{PP}^{cc}(F') = \Theta(\log(1/\mathrm{disc}(F')) + \log\log(|X|\cdot|Y|))$, it follows that $\mathrm{PP}^{cc}(F') = \tilde{\Omega}(n^{2/5})$. Thus, to complete the proof of Theorem 2, it suffices to show that $\mathrm{UPP}^{cc}(F') = O(\log^2 n)$.

Bounding $\mathrm{UPP}^{cc}(F')$. The following lemma contains the core of the argument.

Lemma 20. *$\deg_{\pm}(F) = d$ for some $d = O(\log n)$.*

Proof. Given two inputs $z_i, z_j \in \{-1,1\}^{\log_2 R}$, let $\mathrm{EQ}(z_i, z_j)$ denote the function that evaluates to 1 if $z_i = z_j$, and evaluates to 0 otherwise. Trivially, $\mathrm{EQ}(z_i, z_j)$ is exactly computed by a polynomial of degree at most $2\log_2 R$.

Let $z = (z_1, \ldots, z_m) \in \left(\{-1,1\}^{\log_2 R}\right)^m = \{-1,1\}^N$ denote an input to $\overline{\mathrm{ED}}$. Define

$$q(z) := \sum_{i,j \in [m], i \neq j} \mathrm{EQ}(z_i, z_j).$$

Let $K = \binom{m}{2}$. Notice that q satisfies the following two properties.

- Property 1: If $\overline{\mathrm{ED}}_N(z) = 1$, then $q(z) = 0$, because $z_i \neq z_j$ for all $i \neq j$.
- Property 2: If $\overline{\mathrm{ED}}_N(z) = -1$, then $q(z) \in \{1, \ldots, K\}$, because there is at least one pair $i \neq j$ such that $z_i = z_j$.

Let $x = (x_1, \ldots, x_M) \in \left(\{-1,1\}^N\right)^M$ denote an input to the function $F = \mathrm{OMB}_M(\overline{\mathrm{ED}}_N, \ldots, \overline{\mathrm{ED}}_N)$. Define

$$p(x) = 1/2 + \sum_{i=1}^{M} (-1)^i K^{2\cdot(i-1)} \cdot q(x_i). \tag{10}$$

Notice that $\deg(p) \leq \deg(q) \leq \deg(\mathrm{EQ}) \leq 2\log R$. We claim that $\widetilde{\mathrm{sgn}}(p(x)) = F(x)$ for all x. To see this, first consider any x such that $\overline{\mathrm{ED}}_N(x_i) = 1$ for all i. Then $F(x) = 1$, and Property 1 above implies that $p(x) = 1/2$, so $\widetilde{\mathrm{sgn}}(p(x)) = F(x)$ in this case.

Now consider any x such that $\overline{\mathrm{ED}}_N(x_i) = -1$ for some i. Let i^* be the largest such i. Suppose without loss of generality that i^* is odd (the analysis in the case that i^* is even is analogous). Then $F(x) = -1$, so we need only show that $p(x) < 0$.

Notice that term i^* in the sum within Eq. (10) equals

$$(-1)^{i^*} \cdot K^{2\cdot(i^*-1)} \cdot q(x_{i^*}) \in \{-K^{2\cdot(i^*-1)}, -2\cdot K^{2\cdot(i^*-1)}, \ldots, -K^{2\cdot i^*-1}\}, \tag{11}$$

where we have exploited Property 2 above, as well as the fact that i^* is odd.

For all $j > i^*$, term j in the sum within Eq. (10) equals

$$(-K)^{2\cdot(j-1)} \cdot q(x_j) = 0, \tag{12}$$

where we have exploited Property 1 above.

Finally, we can bound the sum of the first $i^* - 1$ terms in Eq. (10) via:

$$\sum_{j=1}^{i^*-1} (-1)^j \cdot K^{2\cdot(j-1)} \cdot q(x_j) \leq \sum_{j=1}^{i^*-1} K^{2\cdot(j-1)} \cdot K = \sum_{j=1}^{i^*-1} K^{2\cdot j-1} \leq K^{2i^*-2} - 1. \tag{13}$$

Here, the inequality holds by Property 2 above.

Combining Equations Eq. (11), Eq. (12), and Eq. (13), we conclude that

$$p(x) \leq 1/2 - K^{2i^*-2} + (K^{2i^*-2} - 1) = -1/2,$$

completing the proof. □

Recall from Lemma 18 that in the communication problem corresponding to F', Alice has input $x \in \{-1,1\}^{4n}$, Bob has input $y \in \{-1,1\}^{4n}$, and the goal is to output $F(\dots, \vee_{j=1}^{4}(x_{i,j} \wedge y_{i,j}), \dots)$. We use Lemma 20 in a standard way to give a simple UPP^{cc} protocol P of cost $O(\log^2 n)$ that computes the function F'.

Suppose that $p(x) = \sum_{S \subseteq \{-1,1\}^n, |S| \leq d} c_S \chi_S(x)$. Alice picks an S at random, with probability proportional to $|c_S|$. Alice then sends Bob the set S, along with the values $\{x_{i,j} : i \in S, 1 \leq j \leq 4\}$. Notice that the total communication required is $\log \binom{n}{d} + 4 \cdot d = O(\log^2 n)$ bits. Bob uses this information to compute $\chi_S(\dots, \vee_{j=1}^{4}(x_{i,j} \wedge y_{i,j}), \dots)$, and outputs $\widetilde{\text{sgn}}\left(c_S \cdot \chi_S(\dots, \vee_{j=1}^{4}(x_{i,j} \wedge y_{i,j}), \dots)\right)$.

It is easy to see that Bob outputs 1 with probability $1/2 + \frac{p(x)}{2\sum_{S \subseteq \{-1,1\}^n, |S| \leq d} |c_S|}$. Since $p(x)$ sign-represents f (cf. Lemma 20), this implies that P computes f with positive bias, and hence P is a UPP^{cc} protocol for f achieving communication cost $O(\log^2 n)$. □

Remark 21. *Theorem 1 is tight up to logarithmic factors for the function $F = \text{OMB}(\overline{\text{ED}}_N, \dots, \overline{\text{ED}}_N)$ appearing in the proof of Theorem 2, by the following analysis.*

Since $\widetilde{deg}_{+,2/3}(\overline{\text{ED}}_N) = d$ for some $d = \Omega((N/\log N)^{2/3})$ (cf. Lemma 19), Theorem 1 states that $\widetilde{deg}_{+,\varepsilon}(F) \geq d$ for $\varepsilon = 1 - 2^{-M}$. Meanwhile, if M and N are polynomially related, the proof of Lemma 20 can easily be modified to demonstrate not just that $\deg_{\pm}(F) = O(\log n)$, but that $\widetilde{deg}_{\varepsilon}(F) = O(\log N)$ for some $\varepsilon = 1 - K^{-\Theta(M)} = 1 - 2^{\tilde{O}(M)}$, where $K = \tilde{\Theta}(N^2)$ is the parameter appearing in the proof of Lemma 20.

This analysis also reveals that the approximate degree of F experiences a "sharp threshold" as the error parameter ε approaches 1 from below: while F can be approximated to error $1 - 2^{\tilde{\Theta}(M)}$ using degree $O(\log N)$, degree $\Omega((N/\log N)^{2/3})$ is necessary to approximate F to error $1 - 2^{-M}$.

5 Future Directions

Our analysis naturally suggests several directions for future work. The primary question is to determine what is the "right" analog of Theorem 1 when the hypothesis that $\widetilde{\deg}_{+,2/3}(f) \geq d$ is replaced with the hypothesis that $\widetilde{\deg}_{-,2/3}(f) \geq d$. We conjecture that the following bound holds:

Conjecture 22. *Suppose that $f\colon \{-1,1\}^N \to \{-1,1\}$ satisfies $\widetilde{deg}_{-,2/3}(f) \geq d$. Then letting $F = \text{OMB}_M(f, \dots, f)$, it holds that $\deg_{\pm}(F) = \Omega(\min\{d \cdot M^{1/3}, M\})$.*

Recall that Bun and Thaler [8] proved that $\widetilde{\deg}_{-,2/3}(\text{ED}_N) = \Omega((N/\log N)^{2/3})$ (cf. Lemma 19). Thus, we obtain the following special case of Conjecture 22, which we highlight separately.

Conjecture 23. *Let $F = \text{OMB}_{n^{1/2}}(\text{ED}_{n^{1/2}}, \dots, \text{ED}_{n^{1/2}})$. Then $\deg_{\pm}(F) = \tilde{\Omega}(n^{1/2})$.*

A proof of Conjecture 23 would yield a polynomial improvement over the current best threshold degree lower bound for an AC^0 function, which is $\Omega(n^{(k-1)/(2k-1)})$ for any constant depth $k \geq 2$ [30] (cf. Theorem

5). On the other hand, disproving Conjecture 23 would likely require the development of new techniques for constructing low-degree threshold representations for block-composed functions.

It would also be interesting to determine whether block-composition with OMB is still an effective form of hardness amplification if the hypothesis that $\widetilde{\deg}_{+,2/3}(f) \geq d$ from Theorem 1 is replaced with the weaker hypothesis that $\widetilde{\deg}_{2/3}(f) \geq d$. Is this enough to guarantee that $\widetilde{\deg}_{\varepsilon}(\mathrm{OMB}_M(f,\ldots,f)) \geq d$, for some $\varepsilon = 1 - 2^{-\Omega(M)}$?

Acknowledgements The author is grateful to Mark Bun for insightful comments on an earlier version of this manuscript.

References

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

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

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

[4] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, 2001.

[5] Richard Beigel. Perceptrons, pp, and the polynomial hierarchy. *Computational Complexity*, 4:339–349, 1994.

[6] Harry Buhrman, Nikolai K. Vereshchagin, and Ronald de Wolf. On computation and communication with small bias. In *22nd Annual IEEE Conference on Computational Complexity (CCC 2007), 13-16 June 2007, San Diego, California, USA*, pages 24–32. IEEE Computer Society, 2007.

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

[8] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. *Electronic Colloquium on Computational Complexity (ECCC)*, 20:151, 2013.

[9] Karthekeyan Chandrasekaran, Justin Thaler, Jonathan Ullman, and Andrew Wan. Faster private release of marginals on small databases. *CoRR*, abs/1304.3754, 2013.

[10] Arkadev Chattopadhyay and Anil Ada. Multiparty communication complexity of disjointness. *Electronic Colloquium on Computational Complexity (ECCC)*, 15(002), 2008.

[11] Matei David and Toniann Pitassi. Separating NOF communication complexity classes RP and NP. *Electronic Colloquium on Computational Complexity (ECCC)*, 15(014), 2008.

[12] Matei David, Toniann Pitassi, and Emanuele Viola. Improved separations between nondeterministic and randomized multiparty communication. *TOCT*, 1(2), 2009.

[13] Ronald de Wolf. A note on quantum algorithms and the minimal degree of ε-error polynomials for symmetric functions. *Quantum Information & Computation*, 8(10):943–950, 2010.

[14] Dmitry Gavinsky and Alexander A. Sherstov. A separation of NP and conp in multiparty communication complexity. *Theory of Computing*, 6(1):227–245, 2010.

[15] Dmitry Gavinsky and Alexander A. Sherstov. A separation of np and conp in multiparty communication complexity. *Theory of Computing*, 6(1):227–245, 2010.

[16] Adam Tauman Kalai, Adam R. Klivans, Yishay Mansour, and Rocco A. Servedio. Agnostically learning halfspaces. *SIAM J. Comput.*, 37(6):1777–1805, 2008.

[17] Varun Kanade and Justin Thaler. Distribution-independent reliable learning. In Maria-Florina Balcan and Csaba Szepesvári, editors, *Proceedings of The 27th Conference on Learning Theory, COLT 2014, Barcelona, Spain, June 13-15, 2014*, volume 35 of *JMLR Proceedings*, pages 3–24. JMLR.org, 2014.

[18] Hartmut Klauck. Lower bounds for quantum communication complexity. *SIAM J. Comput.*, 37(1):20–46, 2007.

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

[20] Adam R. Klivans and Rocco A. Servedio. Toward attribute efficient learning of decision lists and parities. *Journal of Machine Learning Research*, 7:587–602, 2006.

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

[22] Troy Lee. A note on the sign degree of formulas. *CoRR*, abs/0909.4607, 2009.

[23] Troy Lee and Adi Shraibman. Disjointness is hard in the multiparty number-on-the-forehead model. *Computational Complexity*, 18(2):309–336, 2009.

[24] Nati Linial and Adi Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Struct. Algorithms*, 34(3):368–394, 2009.

[25] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[26] Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions (preliminary version). In S. Rao Kosaraju, Mike Fellows, Avi Wigderson, and John A. Ellis, editors, *Proceedings of the 24th Annual ACM Symposium on Theory of Computing, May 4-6, 1992, Victoria, British Columbia, Canada*, pages 468–474. ACM, 1992.

[27] Anup Rao and Amir Yehudayoff. Simplified lower bounds on the multiparty communication complexity of disjointness. *Electronic Colloquium on Computational Complexity (ECCC)*, 21:60, 2014.

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

[29] Rocco A. Servedio, Li-Yang Tan, and Justin Thaler. Attribute-efficient learning and weight-degree tradeoffs for polynomial threshold functions. In Shie Mannor, Nathan Srebro, and Robert C. Williamson, editors, *COLT*, volume 23 of *JMLR Proceedings*, pages 14.1–14.19. JMLR.org, 2012.

[30] A. A. Sherstov. Breaking the Minsky-Papert barrier for constant-depth circuits. In *STOC*, 2014.

[31] Alexander A. Sherstov. Halfspace matrices. *Computational Complexity*, 17(2):149–178, 2008.

[32] Alexander A. Sherstov. Approximate inclusion-exclusion for arbitrary symmetric functions. *Computational Complexity*, 18(2):219–247, 2009.

[33] Alexander A. Sherstov. Separatingac^0 from depth-2 majority circuits. *SIAM J. Comput.*, 38(6):2113–2129, 2009.

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

[35] Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. In Lance Fortnow and Salil P. Vadhan, editors, *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 41–50. ACM, 2011.

[36] Alexander A. Sherstov. The multiparty communication complexity of set disjointness. In Howard J. Karloff and Toniann Pitassi, editors, *Proceedings of the 44th Symposium on Theory of Computing Conference, STOC 2012, New York, NY, USA, May 19 - 22, 2012*, pages 525–548. ACM, 2012.

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

[38] Alexander A. Sherstov. Communication lower bounds using directional derivatives. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *Symposium on Theory of Computing Conference, STOC'13, Palo Alto, CA, USA, June 1-4, 2013*, pages 921–930. ACM, 2013.

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

[40] Alexander A. Sherstov. Making polynomials robust to noise. *Theory of Computing*, 9:593–615, 2013.

[41] Justin Thaler, Jonathan Ullman, and Salil P. Vadhan. Faster algorithms for privately releasing marginals. In Artur Czumaj, Kurt Mehlhorn, Andrew M. Pitts, and Roger Wattenhofer, editors, *Automata, Languages, and Programming - 39th International Colloquium, ICALP 2012, Warwick, UK, July 9-13, 2012, Proceedings, Part I*, volume 7391 of *Lecture Notes in Computer Science*, pages 810–821. Springer, 2012.

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

A Polynomial Approximations and their Dual Characterizations

The presentation in this section borrows heavily from our earlier work [8].

A.1 Approximate Degree

The ε-approximate degree of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\widetilde{\deg}_\varepsilon(f)$, is the minimum (total) degree of any real polynomial p such that $\|p-f\|_\infty \leq \varepsilon$, i.e., $|p(x)-f(x)| \leq \varepsilon$ for all $x \in \{-1,1\}^n$. Any polynomial p satisfying $\|p-f\|_\infty \leq \varepsilon$ is called an *ε-approximation* for f. By convention, $\widetilde{\deg}(f)$ denotes $\widetilde{\deg}_{1/3}(f)$, and this quantity is referred to with qualification as the *approximate degree* of a function. The choice of $1/3$ is arbitrary, as $\widetilde{\deg}(f)$ is related to $\widetilde{\deg}_\varepsilon(f)$ by a constant factor for any constant $\varepsilon \in (0,1)$.

Given a Boolean function f, let p be a real polynomial that minimizes $\|p-f\|_\infty$ among all polynomials of degree at most d. Since we work over $x \in \{-1,1\}^n$, we may assume without loss of generality that p is multilinear with the representation $p(x) = \sum_{|S| \leq d} c_S \chi_S(x)$ where the coefficients c_S are real numbers. Then p is an optimum of the following linear program.

min	ε	
such that	$\left\lvert f(x) - \sum_{\lvert S\rvert \leq d} c_S \chi_S(x)\right\rvert \leq \varepsilon$	for each $x \in \{-1,1\}^n$
	$c_S \in \mathbb{R}$	for each $\lvert S\rvert \leq d$
	$\varepsilon \geq 0$	

The dual LP is as follows.

max	$\sum_{x \in \{-1,1\}^n} \phi(x) f(x)$	
such that	$\sum_{x \in \{-1,1\}^n} \lvert\phi(x)\rvert = 1$	
	$\sum_{x \in \{-1,1\}^n} \phi(x) \chi_S(x) = 0$	for each $\lvert S\rvert \leq d$
	$\phi(x) \in \mathbb{R}$	for each $x \in \{-1,1\}^n$

Strong LP-duality thus yields the following well-known dual characterization of approximate degree (cf. [34]).

Lemma 24. *Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Then $\widetilde{\deg}_\varepsilon(f) > d$ if and only if there is a polynomial $\phi : \{-1,1\}^n \to \mathbb{R}$ such that*

$$\sum_{x \in \{-1,1\}^n} f(x)\phi(x) > \varepsilon, \tag{14}$$

$$\sum_{x \in \{-1,1\}^n} |\phi(x)| = 1, \tag{15}$$

and

$$\sum_{x \in \{-1,1\}^n} \phi(x)\chi_S(x) = 0 \text{ for each } |S| \leq d. \tag{16}$$

If ϕ satisfies Eq. (16), we say ϕ has *pure high degree d*. We refer to any feasible solution ϕ to the dual LP as a *dual polynomial* for f.

A.2 Positive One-Sided Approximate Degree

Positive one-sided ε-approximate degree, denoted $\widetilde{\deg}_{+,\varepsilon}(f)$, is the least degree of a real polynomial p with that is an *positive one-sided ε-approximation* to f, meaning

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

2. $p(x) \geq 1-\varepsilon$ for all $x \in f^{-1}(+1)$.

That is, we require p to be very accurate on inputs in $f^{-1}(-1)$, but only require "one-sided accuracy" on inputs in $f^{-1}(+1)$. The primal and dual LPs change in a simple but crucial way if we look at one-sided approximate degree rather than approximate degree. Let $p(x) = \sum_{|S| \leq d} c_S \chi_S(x)$ be a polynomial of degree d for which the positive one-sided ε-approximate degree of f is attained. Then p is an optimum of the following linear program.

min	ε	
such that	$\left\lvert f(x) - \sum_{\lvert S\rvert \leq d} c_S \chi_S(x)\right\rvert \leq \varepsilon$	for each $x \in f^{-1}(-1)$
	$\sum_{\lvert S\rvert \leq d} c_S \chi_S(x) \geq 1-\varepsilon$	for each $x \in f^{-1}(+1)$
	$c_S \in \mathbb{R}$	for each $\lvert S\rvert \leq d$
	$\varepsilon \geq 0$	

The dual LP is as follows.

max	$\sum_{x \in \{-1,1\}^n} \phi(x) f(x)$	
such that	$\sum_{x \in \{-1,1\}^n} \lvert\phi(x)\rvert = 1$	
	$\sum_{x \in \{-1,1\}^n} \phi(x) \chi_S(x) = 0$	for each $\lvert S\rvert \leq d$
	$\phi(x) \geq 0$ for each $x \in f^{-1}(+1)$	
	$\phi(x) \in \mathbb{R}$	for each $x \in \{-1,1\}^n$

We again appeal to strong LP-duality for the following dual characterization of positive one-sided approximate degree.

Lemma 25. *Fix any constant $C > 0$. Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Then $\widetilde{deg}_{+,\varepsilon}(f) > d$ if and only if there is a polynomial $\phi : \{-1,1\}^n \to \mathbb{R}$ such that*

$$\sum_{x \in \{-1,1\}^n} f(x)\phi(x) > C \cdot \varepsilon, \tag{17}$$

$$\sum_{x \in \{-1,1\}^n} |\phi(x)| = C, \tag{18}$$

$$\sum_{x \in \{-1,1\}^n} \phi(x)\chi_S(x) = 0 \text{ for each } |S| \leq d, \tag{19}$$

and

$$\phi(x) \geq 0 \text{ for each } x \in f^{-1}(+1). \tag{20}$$

Observe that a feasible solution ϕ to this dual LP is a feasible solution to the dual LP for approximate degree, with the additional constraint that $\phi(x)$ agrees in sign with $f(x)$ whenever $x \in f^{-1}(+1)$. We refer to any such feasible solution ϕ as a dual polynomial for f with *positive one-sided error*.

A.3 Negative One-Sided Approximate Degree

Negative one-sided ε-approximate degree, denoted $\widetilde{\deg}_{-,\varepsilon}(f)$, is defined analogously to positive one-sided ε-approximate degree. Specifically, it equals the least degree of a real polynomial p with that is an *negative one-sided ε-approximation* to f, meaning

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

2. $p(x) \leq -1+\varepsilon$ for all $x \in f^{-1}(-1)$.

Negative one-sided approximate degree has a dual characterization analogous to Lemma 25. However, we do not make use of this dual characterization in this work, and therefore omit the details for brevity.

B Communication Complexity Models

Let $f\colon X \times Y \to \{-1,1\}$ be a function. Consider a two-party communication problem in which Alice is given an input $x \in X$, Bob is given an input $y \in Y$, and their goal is to output $f(x,y)$ with probability at least $1/2+\beta$ for some bias $\beta > 0$. Alice and Bob each have access to an arbitrarily long sequence of private random bits, and the cost $C(P)$ of a protocol P is the worst-case number of bits they must exchange over all inputs $(x,y) \in X \times Y$. Babai et al. [3] defined the PP and UPP communication models to capture the complexity of computing f with small bias. The PP communication complexity of f, denoted by $\mathrm{PP}^{cc}(f)$, is the minimum value of $C(P)+\log(1/\beta(P))$ over all protocols P that compute f with positive bias. The UPP communication complexity of f, denoted by $\mathrm{UPP}^{cc}(f)$, is the minimum value of $C(P)$ over all protocols P that compute f with positive bias.

Applications. Both PP^{cc} and UPP^{cc} have important applications in learning theory and circuit complexity. On the circuit complexity side, lower bounds on $\mathrm{PP}^{cc}(f)$ imply corresponding lower bounds on the size of majority-of-threshold circuits computing f (see, e.g., [33]). Meanwhile, lower bounds on $\mathrm{UPP}^{cc}(f)$ imply a corresponding lower bound on the size of threshold-of-majority circuits computing f (see e.g. [28]). On the learning theory side, upper bounds on UPP communication complexity imply fast algorithms for distribution-independent PAC learning [42]. For example, the fastest known algorithm for PAC learning DNF formulae — a challenge problem posed in Valiant's original paper [42] on the PAC model — is due to Klivans and Servedio [19], and follows from an upper bound on the threshold degree of DNFs (which in turn implies an upper bound on the UPP communication complexity of DNFs). Meanwhile, upper bounds on PP communication complexity imply efficient algorithms in the online mistake-bounded learning model (see, e.g., [20]).

ECCC ISSN 1433-8092

http://eccc.hpi-web.de