Dual Lower Bounds for Approximate Degree and Markov-Bernstein Inequalities*

Mark Bun** and Justin Thaler***

School of Engineering and Applied Sciences
Harvard University, Cambridge, MA
{mbun,jthaler}@seas.harvard.edu

Abstract. The ε-*approximate degree* of a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ is the minimum degree of a real polynomial that approximates f to within ε in the ℓ_∞ norm. We prove several lower bounds on this important complexity measure by explicitly constructing solutions to the dual of an appropriate linear program. Our first result resolves the ε-approximate degree of the two-level AND-OR tree for any constant $\varepsilon > 0$. We show that this quantity is $\Theta(\sqrt{n})$, closing a line of incrementally larger lower bounds [3, 11, 21, 30, 32]. The same lower bound was recently obtained independently by Sherstov using related techniques [25]. Our second result gives an explicit *dual polynomial* that witnesses a tight lower bound for the approximate degree of any symmetric Boolean function, addressing a question of Špalek [34]. Our final contribution is to reprove several Markov-type inequalities from approximation theory by constructing explicit dual solutions to natural linear programs. These inequalities underly the proofs of many of the best-known approximate degree lower bounds, and have important uses throughout theoretical computer science.

1 Introduction

Approximate degree is an important measure of the complexity of a Boolean function. It captures whether a function can be approximated by a low-degree polynomial with real coefficients in the ℓ_∞ norm, and it has many applications in theoretical computer science. The study of approximate degree has enabled progress in circuit complexity [7, 8, 19, 29], quantum computing (where it has been used to prove lower bounds on quantum query complexity, e.g. [2, 5, 14]), communication complexity [4, 10, 17, 27, 31, 33, 34], and computational learning theory (where approximate degree upper bounds underly the best known algorithms for PAC learning DNF formulas and agnostically learning disjunctions) [13, 15].

In this paper, we seek to advance our understanding of this fundamental complexity measure. We focus on proving approximate degree lower bounds

* The full version of this paper is available at http://arxiv.org/abs/1302.6191

** Supported in part by NSF grant CNS-1237235.

*** Supported by an NSF Graduate Research Fellowship and NSF grants CNS-1011840 and CCF-0915922.

F.V. Fomin et al. (Eds.): ICALP 2013, Part I, LNCS 7965, pp. 303–314, 2013.

by specifying explicit *dual polynomials*, which are dual solutions to a certain linear program capturing the approximate degree of any function. These polynomials act as certificates of the high approximate degree of a function. Their construction is of interest because these dual objects have been used recently to resolve several long-standing open problems in communication complexity (e.g. [4, 10, 17, 27, 33, 34]). See the survey of Sherstov [26] for an excellent overview of this body of literature.

Our Contributions. Our first result resolves the approximate degree of the function $f(x) = \wedge_{i=1}^{N} \vee_{j=1}^{N} x_{ij}$, showing this quantity is $\Theta(N)$. Known as the two-level AND-OR tree, f is the simplest function whose approximate degree was not previously characterized. A series of works spanning nearly two decades proved incrementally larger lower bounds on the approximate degree of this function, and this question was recently re-posed by Aaronson in a tutorial at FOCS 2008 [1]. Our proof not only yields a tight lower bound, but it specifies an explicit dual polynomial for the high approximate degree of f, answering a question of Špalek [34] in the affirmative.

Our second result gives an explicit dual polynomial witnessing the high approximate degree of any *symmetric* Boolean function, recovering a well-known result of Paturi [22]. Our solution builds on the work of Špalek [34], who gave an explicit dual polynomial for the OR function, and addresses an open question from that work.

Our final contribution is to reprove several classical Markov-type inequalities of approximation theory using simpler ideas from linear programming. These inequalities bound the derivative of a polynomial in terms of its degree. Combined with the well-known symmetrization technique (see e.g. [1, 19]), Markov-type inequalties have traditionally been the primary tool used to prove approximate degree lower bounds on Boolean functions (e.g. [2, 3, 21, 32]). Our proofs of these inequalities specify explicit dual solutions to a natural linear program (that differs from the one used to prove our first two results). While these inequalities have been known for over a century [9, 18], to the best of our knowledge our proof technique is novel, and we believe it sheds new light on these results.

2 Preliminaries

We work with Boolean functions $f : \{-1,1\}^n \to \{-1,1\}$ under the standard convention that 1 corresponds to logical false, and -1 corresponds to logical true. We let $\|f\|_\infty = \max_{x \in \{-1,1\}} |f(x)|$ denote the ℓ_∞ norm of f. The ε-approximate degree of a function $f : \{-1,1\}^n \to \{-1,1\}$, denoted $\deg_\varepsilon(f)$, 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 $x \in \{-1,1\}^n$. We use $\widetilde{\deg}(f)$ to denote $\deg_{1/3}(f)$, and use this to refer to the *approximate degree* of a function without qualification. The choice of $1/3$ is arbitrary, as $\widetilde{\deg}(f)$ is related to $\deg_\varepsilon(f)$ by a constant factor for any constant $\varepsilon \in (0,1)$. We let OR_n and AND_n denote the OR function and AND function on n variables respectively. Define $\widetilde{\mathrm{sgn}}(x) = -1$ if $x < 0$ and 1 otherwise.

In addition to approximate degree, *block sensitivity* is also an important measure of the complexity of a Boolean function. We introduce this measure because functions with low block sensitivity are an "easy case" in the analysis of Theorem 2 below. The block sensitivity $\mathrm{bs}_x(f)$ of a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ at the point x is the maximum number of pairwise disjoint subsets $S_1, S_2, S_3, \cdots \subseteq \{1,2,\ldots,n\}$ such that $f(x) \neq f(x^{S_1}) = f(x^{S_2}) = f(x^{S_3}) = \ldots$ Here, x^S denotes the vector obtained from x by negating each entry whose index is in S. The block sensitivity $\mathrm{bs}(f)$ of f is the maximum of $\mathrm{bs}_x(f)$ over all $x \in \{-1,1\}^n$.

2.1 A Dual Characterization of Approximate Degree

For a subset $S \subset \{1,\ldots,n\}$ and $x \in \{-1,1\}^n$, let $\chi_S(x) = \prod_{i \in S} x_i$. Strong LP-duality yields the following well-known dual characterization of approximate degree (cf. [27]).

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

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

$$\sum_{x \in \{-1,1\}^n} |\phi(x)| = 1, \tag{2}$$

and

$$\sum_{x \in \{-1,1\}^n} \phi(x)\chi_S(x) = 0 \text{ for each } |S| \leq d. \tag{3}$$

If ϕ satisfies Eq. (3), 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.

3 A Dual Polynomial for the AND-OR Tree

Define $\text{AND-OR}_N^M : \{-1,1\}^{MN} \to \{-1,1\}$ by $f(x) = \wedge_{i=1}^M \vee_{j=1}^N x_{ij}$. AND-OR_N^N is known as the two-level AND-OR tree, and its approximate degree has resisted characterization for close to two decades. Nisan and Szegedy proved an $\Omega(N^{1/2})$ lower bound on $\widetilde{\deg}(\text{AND-OR}_N^N)$ in [21]. This was subsequently improved to $\Omega(\sqrt{N \log N})$ by Shi [32], and improved further to $\Omega(N^{2/3})$ by Ambainis [3]. Most recently, Sherstov proved an $\Omega(N^{3/4})$ lower bound in [30], which was the best lower bound prior to our work. The best upper bound is $O(N)$ due to Høyer, Mosca, and de Wolf [11], which matches our new lower bound.

By refining Sherstov's analysis in [30], we will show that $\widetilde{\deg}(\text{AND-OR}_N^M) = \Omega(\sqrt{MN})$, which matches an upper bound implied by a result of Sherstov [28]. In particular, this implies that the approximate degree of the two-level AND-OR tree is $\Theta(N)$.

Theorem 2. $\widetilde{\deg}(\text{AND-OR}_N^M) = \Theta(\sqrt{MN})$.

Independent Work by Sherstov. Independently of our work, Sherstov [25] has discovered the same $\Omega(\sqrt{MN})$ lower bound on $\widetilde{\deg}(\text{AND-OR}_N^M)$. Both his proof and ours exploit the fact that the OR function has a dual polynomial with one-sided error. Our proof proceeds by constructing an explicit dual polynomial for AND-OR_N^M, by combining a dual polynomial for OR_N with a dual polynomial for AND_M. In contrast, Sherstov mixes the primal and dual views: his proof combines a dual polynomial for OR_N with an approximating polynomial p for AND-OR_N^M to construct an approximating polynomial q for AND_M. The proof in [25] shows that q has much lower degree than p, so the desired lower bound on the degree of p follows from known lower bounds on the degree of q.

The proof of [25] is shorter, while our proof has the benefit of yielding an explicit dual polynomial witnessing the lower bound.

3.1 Proof Outline

Our proof is a refinement of a result of Sherstov [30], which roughly showed that approximate degree increases multiplicatively under function composition. Specifically, Sherstov showed the following.

Proposition 1 ([30, Theorem 3.3]). *Let $F : \{-1,1\}^M \to \{-1,1\}$ and $f : \{-1,1\}^N \to \{-1,1\}$ be given functions. Then for all $\varepsilon, \delta > 0$,*

$$\deg_{\varepsilon - 4\delta \,\mathrm{bs}(F)}(F(f,\ldots,f)) \geq \deg_\varepsilon(F) \deg_{1-\delta}(f).$$

Sherstov's proof of Proposition 1 proceeds by taking a dual witness Ψ to the high ε-approximate degree of F, and combining it with a dual witness ψ to the high $(1-\delta)$-approximate degree of f to obtain a dual witness ζ for the high $(\varepsilon - 4\delta \mathrm{bs}(F))$-approximate degree of $F(f,\ldots,f)$. His proof proceeds in two steps: he first shows that ζ has pure high degree at least $\deg_\varepsilon(F)\deg_{1-\delta}(f)$, and then he lower bounds the correlation of ζ with $F(f,\ldots,f)$. The latter step of this analysis yields a lower bound on the correlation of ζ with $F(f,\ldots,f)$ that deteriorates rapidly as the block sensitivity $\mathrm{bs}(F)$ grows.

Proposition 1 itself does not yield a tight lower bound for $\widetilde{\deg}(\text{AND-OR}_N^M)$, because the function AND_M has maximum block sensitivity $\mathrm{bs}(\text{AND}_M) = M$. We address this by refining the second step of Sherstov's analysis in the case where $F = \text{AND}_M$ and $f = \text{OR}_N$. We leverage two facts. First, although the block sensitivity of AND_M is high, it is only high at one input, namely the all-true input. At all other inputs, AND_M has low block sensitivity and the analysis of Proposition 1 is tight. Second, we use the fact that any dual witness to the high approximate degree of OR_N has one-sided error. Namely, if $\psi(x) < 0$ for such a dual witness ψ, then we know that $\psi(x)$ agrees in sign with $\text{OR}_N(x)$. This property allows us to handle the all-true input to AND_M separately: we use it to show that despite the high block-sensitivity of AND_M at the all-true input y, this input nonetheless contributes positively to the correlation between ζ and $F(f,\ldots,f)$.

3.2 Proof of Thm. 2

As in Sherstov's proof of Proposition 1, we define $\zeta : \left(\{-1,1\}^N\right)^M \to \mathbb{R}$ by

$$\zeta(x_1,\ldots,x_M) := 2^M \Psi(\ldots,\widetilde{\operatorname{sgn}}(\psi(x_i)),\ldots)\prod_{i=1}^{M}|\psi(x_i)|, \tag{4}$$

where Ψ and ψ are dual witnesses to the high ε-approximate degree of AND_M and $(1-\delta)$-approximate degree of OR_N, respectively, for suitable ε and δ, and $x_i = (x_{i,1},\ldots,x_{i,N})$. To show that ζ is a dual witness for the fact that the $(1/3)$-approximate degree of AND-OR_N^M is $\Omega(\sqrt{MN})$, it suffices to check that ζ satisfies the conditions of Thm. 1. The only place where our analysis differs from that of Sherstov's is in verifying Expression (1), i.e. that

$$\sum_{(x_1,\ldots,x_M)\in(\{-1,1\}^N)^M} \zeta(x_1,\ldots,x_M)\text{AND-OR}_N^M(x_1,\ldots,x_M) > 1/3. \tag{5}$$

Let $A_1 = \{x \in \{-1,1\}^N : \psi(x) \geq 0, \mathrm{OR}_N(x) = -1\}$ and $A_{-1} = \{x \in \{-1,1\}^N : \psi(x) < 0, \mathrm{OR}_N(x) = 1\}$, so $A_1 \cup A_{-1}$ is the set of all inputs x where the sign of $\psi(x)$ disagrees with $\mathrm{OR}_N(x)$. Notice that $\sum_{x\in A_1\cup A_{-1}}|\psi(x)| < \delta/2$ because ψ has correlation $1-\delta$ with f. A sequence of manipulations found in the full version of this paper shows that the left-hand side of (5) equals

$$\sum_{z\in\{-1,1\}^M} \Psi(z)\cdot\mathbf{E}[\mathrm{AND}_M(\ldots,y_iz_i,\ldots)], \tag{6}$$

where $y \in \{-1,1\}^M$ is a random string whose ith bit independently takes on value -1 with probability $2\sum_{x\in A_{z_i}}|\psi(x)| < \delta$.

All $z \neq -\mathbf{1}_M$ can be handled as in Sherstov's proof of Proposition 1, because AND_M has low block sensitivity at these inputs. These inputs contribute a total of at least $\varepsilon - 4\delta - |\Psi(-\mathbf{1}_M)|$ to Expression (6). We only need to argue that the term corresponding to $z = -\mathbf{1}_M$ contributes $|\Psi(-\mathbf{1}_M)|$ to the correlation. In the full version, we argue that any dual witness for the OR_N function has one-sided error [12]. That is, if $\mathrm{OR}(x) = 1$ (i.e. if $x = \mathbf{1}_N$), then $\widetilde{\operatorname{sgn}}(\psi(x)) = 1$. This implies that A_{-1} is empty; that is, if $\widetilde{\operatorname{sgn}}(\psi(x)) = -1$, then it must be the case that $\mathrm{OR}_N(x) = -1$. Therefore, for $z = -\mathbf{1}_M$, the y_i's are all -1 with probability 1, and hence $\mathbf{E}_y[\mathrm{AND}_M(\ldots,y_iz_i,\ldots)] = \mathrm{AND}_M(-\mathbf{1}_M) = -1$. By the one-sided error of any dual witness for AND_M, $\widetilde{\operatorname{sgn}}(\Psi(-\mathbf{1}_M)) = -1$, and thus the term corresponding to $z = -\mathbf{1}_M$ contributes $-\Psi(z) = |\Psi(z)|$ to Expression (6) as claimed. □

Remark 1. Špalek [34] has exhibited an explicit dual witness showing that the ε-approximate degree of both the AND function and the OR function is $\Omega(\sqrt{n})$, for $\varepsilon = 1/14$ (in fact, we generalize Špalek's construction in the next section to any symmetric function). It is relatively straightforward to modify his construction to handle any constant $\varepsilon \in (0,1)$. With these dual polynomials in hand, the dual solution ζ we construct in our proof is completely explicit. This answers a question of Špalek [34, Section 4] in the affirmative.

4 Dual Polynomials for Symmetric Boolean Functions

In this section, we construct a dual polynomial witnessing a tight lower bound on the approximate degree of any symmetric function. The lower bound we recover was first proved by Paturi [22] via a symmetrization argument combined with the classical Markov-Bernstein inequality from approximation theory (see Section 5). Paturi also provided a matching upper bound. Špalek [34], building on work of Szegedy, presented a dual witness to the $\Omega(\sqrt{n})$-approximate degree of the OR function and asked whether one could construct an analogous dual polynomial for the symmetric t-threshold function [34, Section 4]. We accomplish this in the more general case of arbitrary symmetric functions by extending the ideas underlying Špalek's dual polynomial for OR.

4.1 Symmetric Functions

For a vector $x \in \{-1,1\}^n$, let $|x| = \frac{1}{2}(n - (x_1 + \cdots + x_n))$ denote the number of -1's in x. A Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ is symmetric if $f(x) = f(y)$ whenever $|x| = |y|$. That is, the value of f depends only on the number of inputs that are set to -1.

Let $[n] = \{0, 1, \ldots, n\}$. To each symmetric function f, we can associate a unique univariate function $F : [n] \to \{-1,1\}$ by taking $F(|x|) = f(x)$. Throughout this section, we follow the convention that lower case letters refer to multivariate functions, while upper case letters refer to their univariate counterparts.

We now discuss the dual characterization of approximate degree established in Thm. 1, as it applies to symmetric functions. Following the notation in [34], the standard inner product $p \cdot q = \sum_{x \in \{-1,1\}^n} p(x)q(x)$ on symmetric functions p, q induces an inner product on the associated univariate functions:

$$P \cdot Q := \sum_{i=0}^{n} \binom{n}{i} P(i)Q(i).$$

We refer to this as the *correlation* between P and Q. Similarly, the ℓ_1-norm $\|p\|_1 = \sum_{x \in \{-1,1\}^n} |p(x)|$ induces a norm $\|P\|_1 = \sum_{i=0}^{n} \binom{n}{i} P(i)$. These definitions carry over verbatim when f is real-valued instead of Boolean-valued.

If f is symmetric, we can restrict our attention to symmetric ϕ in the statement of Thm. 1, and it becomes convenient to work with the following reformulation of Thm. 1.

Corollary 1. *A symmetric function $f : \{-1,1\}^n \to \{-1,1\}$ has ε-approximate degree greater than d iff there exists a symmetric function $\phi : \{-1,1\}^n \to \mathbb{R}$ with pure high degree d such that*

$$\frac{\Phi \cdot F}{\|\Phi\|_1} = \frac{\phi \cdot f}{\|\phi\|_1} > \varepsilon.$$

(Here, F and Φ are the univariate function associated to f and ϕ, respectively).

We clarify that the pure high degree of a multivariate polynomial ϕ does not correspond to the smallest degree of a monomial in the associated univariate function Φ. When we talk about the pure high degree of a univariate polynomial Φ, we mean the pure high degree of its corresponding multilinear polynomial ϕ. It is straightforward to check that if ψ is a multivariate polynomial of degree $n-d$, then multiplying ψ by the parity function yields a univariate function $\Phi(k) := \Psi(k) \cdot (-1)^k$ with pure high degree d.

We are now in a position to state the lower bound that we will prove in this section. Paturi [22] completely characterized the approximate degree of a symmetric Boolean function by the location of the layer t closest to the center of the Boolean hypercube such that $F(t-1) \neq F(t)$.

Theorem 3 ([22], Theorem 4). *Given a nonconstant symmetric Boolean function f with associated univariate function F, let $\Gamma(f) = \min\{|2t-n-1| : F(t-1) \neq F(t), 1 \leq k \leq n\}$. Then $\widetilde{\deg}(f) = \Theta(\sqrt{n(n-\Gamma(f))})$.*

Paturi proved the upper bound non-explicitly by appealing to Jackson theorems from approximation theory. He proved the lower bound by combining symmetrization with an appeal to the Markov-Bernstein inequality (see Section 5) – his proof does not yield an explicit dual polynomial. We construct an explicit dual polynomial to prove the following proposition, which is easily seen to imply Paturi's lower bound.

Proposition 2. *Given f and F as above, let $1 \leq t \leq n$ be an integer with $F(t-1) \neq F(t)$. Then $\widetilde{\deg}(f) = \Omega(\sqrt{t(n-t+1)})$.*

Proof Outline. We start with an intuitive discussion of Špalek's construction of a dual polynomial for OR, with the goal of elucidating how we extend the construction to arbitrary symmetric functions. Consider the perfect squares $S = \{k^2 : 0 \leq k^2 \leq n\}$ and the univariate polynomial

$$R(x) = \frac{1}{n!} \prod_{i \in [n] \setminus S} (x-i).$$

This polynomial is supported on S, and for all $k \in S$,

$$\binom{n}{k^2} |R(k^2)| = \binom{n}{k^2} \cdot \frac{1}{n!} \cdot \frac{\prod_{\substack{i \in [n] \\ i \neq k^2}} |k^2 - i|}{\prod_{\substack{i \in S \\ i \neq k^2}} |k^2 - i|} = \frac{1}{\prod_{\substack{i \in S \\ i \neq k^2}} |k^2 - i|}.$$

Note the remarkable cancellation in the final equality. This quotient is maximized at $k = 1$. In other words, the threshold point $t = 1$ makes the largest contribution to the ℓ_1 mass of R. Moreover, one can check that $R(0)$ is only a constant factor smaller than $R(1)$.

Špalek exploits this distribution of the ℓ_1 mass by considering the polynomial $P(x) = R(x)/(x-2)$. The values of $P(x)$ are related to $R(x)$ by a constant multiple for $x = 0, 1$, but $P(k)$ decays as $|P(k^2)| \approx |R(k^2)|/k^2$ for larger values.

This decay is fast enough that a *constant fraction* of the ℓ_1 mass of P comes from the point $P(0)$.[1] Now P is an $(n - \Omega(\sqrt{n}))$-degree univariate polynomial, so we just need to show that $Q(i) = (-1)^i P(i)$ has high correlation with OR. We can write

$$Q \cdot \mathrm{OR} = 2Q(0) - Q \cdot \mathbf{1} = 2Q(0),$$

since the multilinear polynomial associated to Q has pure high degree $\Omega(\sqrt{n})$, and therefore has zero correlation with constant functions. Because a constant fraction of the ℓ_1 mass of Q comes from $Q(0)$, it follows that $|Q \cdot \mathrm{OR}|/\|Q\|_1$ is bounded below by a constant. By perhaps changing the sign of Q, we get a good dual polynomial for OR.

A natural approach to extend Špalek's argument to symmetric functions with a "jump" at t is the following:

1) Find a set S with $|S| = \Omega(\sqrt{t(n-t+1)})$ such that the maximum contribution to the ℓ_1 norm of $R(x) = \frac{1}{n!}\prod_{i \in [n] \setminus S}(x - i)$ comes from the point $x = t$. Equivalently,
$$\binom{n}{j}|R(j)| = \frac{1}{\prod_{\substack{i \in S \\ i \neq j}} |j - i|}$$
is maximized at $j = t$.
2) Define a polynomial $P(x) = R(x)/(x - (t-1))(x - (t+1))$. Dividing $R(x)$ by the factor $(x - t - 1)$ is analogous to Špalek's division of $R(x)$ by $(x - 2)$. We also divide by $(x - t + 1)$ because we will ultimately need our polynomial $P(x)$ to decay faster than Špalek's by a factor of $|x - t|$ as x moves away from the threshold. By dividing by both $(x - t - 1)$ and $(x - t + 1)$, we ensure that most of the ℓ_1 mass of P is concentrated at the points $t - 1, t, t + 1$.
3) Obtain Q by multiplying P by parity, and observe that $Q(t-1)$ and $Q(t)$ have opposite signs. Since $F(t-1)$ and $F(t)$ also have opposite signs, we can ensure that both $t - 1$ and t contribute positive correlation. Suppose these two points contribute a $1/2 + \varepsilon$ constant fraction of the ℓ_1-norm of Q. Then even in the worst case where the remaining points all contribute negative correlation, $Q \cdot F$ is still at least a 2ε fraction of $\|Q\|_1$ and we have a good dual polynomial. Notice that the pure high degree of Q is $|S| + 2$, yielding the desired lower bound.

In the case where $t = \Omega(n)$, we can use the set

$$S = \{t \pm 4\ell : 0 \leq \ell \leq t/4\},$$

yielding a remarkably clean dual polynomial for the majority function. This partial result also gives the right intuition for general t, although the details are somewhat more complicated and spelled out in the full version of this paper. In general, the set S interpolates between the set for OR used by Špalek, and the set described above for linear t. In particular, S contains all points of the form $t \pm 4\ell$, plus additional points corresponding to perfect squares when $t = o(n)$.

[1] It is also necessary to check that $P(2)$ is only a constant factor larger than $P(0)$.

5 A Constructive Proof of Markov-Bernstein Inequalities

The Markov-Bernstein inequality for polynomials with real coefficients asserts that

$$p'(x) \leq \min\left\{\frac{n}{\sqrt{1-x^2}}, n^2\right\} \|p\|_{[-1,1]}, x \in (-1,1)$$

for every real polynomial of degree at most n. Here, and in what follows,

$$\|p\|_{[-1,1]} := \sup_{y \in [-1,1]} |p(y)|.$$

This inequality has found numerous uses in theoretical computer science, especially in conjunction with symmetrization as a method for bounding the ε-approximate degree of various functions (e.g. [2, 8, 13, 16, 21, 22, 27]).

We prove a number of important special cases of this inequality based on linear programming duality. Our proofs are constructive in that we exhibit explicit dual solutions to a linear program bounding the derivative of a constrained polynomial.

The special cases of the Markov-Bernstein inequality that we prove are sufficient for many applications in theoretical computer science. The dual solutions we exhibit are remarkably clean, and we believe that they shed new light on these classical inequalities.

5.1 Proving the Markov-Bernstein Inequality at $x = 0$

The following linear program with uncountably many constraints captures the problem of finding a polynomial $p(x) = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$ with real-valued coefficients that maximizes $|p'(0)|$ subject to the constraint that $\|p\|_{[-1,1]} \leq 1$. Below the variables are $c_0, \ldots c_n$, and there is a constraint for every $x \in [-1, 1]$.

$$\begin{array}{rl} & \max c_1 \\ \text{such that} & \sum_{i=0}^{n} c_i x^i \leq 1, \ \forall x \in [-1,1] \\ & -\sum_{i=0}^{n} c_i x^i \leq 1, \ \forall x \in [-1,1] \end{array}$$

We will actually upper bound the value of the following LP, which is obtained from the above by throwing away all but finitely many constraints. Not coincidentally, the constraints that we keep are those that are tight for the primal solution corresponding to the Chebyshev polynomials of the first kind. Throughout this section, we refer to this LP as Primal.

$$\begin{array}{rl} & \max c_1 \\ \text{such that} & \sum_{i=0}^{n} c_i x^i \leq 1, \ \forall x = \cos(k\pi/n), \ k \in \{0, 2, \ldots, n-1\} \\ & -\sum_{i=0}^{n} c_i x^i \leq 1, \ \forall x = \cos(k\pi/n), \ k \in \{1, 3, \ldots, n\} \end{array}$$

The dual to Primal can be written as

$$\begin{aligned} &\min \textstyle\sum_{i=0}^{n} y_i \\ &\text{such that } Ay = e \\ &\qquad y_j \geq 0 \ \forall j \in \{0, \ldots, n\} \end{aligned}$$

where $A_{ij} = (-1)^j \cos^i(j\pi/n)$ and $e = (0, 1, 0, 0, 0, \ldots, 0)^T$. We refer to this linear program as DUAL.

Our goal is to prove that PRIMAL has value at most n. For odd n, it is well-known that this value is achieved by the coefficients of $(-1)^{(n-1)/2} T_n(x)$, the degree n Chebyshev polynomial of the first kind. Our knowledge of this primal-optimal solution informed our search for a dual-optimal solution, but our proof makes no explicit reference to the Chebyshev polynomials, and we do not need to invoke strong LP duality; weak duality suffices. Our arguments rely on a number of trigonometric identities that can all be established by elementary methods.

Proposition 3. *Let $n = 2m + 1$ be odd. Define the $(n + 1) \times (n + 1)$ matrix A by $A_{ij} = (-1)^{j+m} \cos^i(j\pi/n)$ for $0 \leq i, j \leq n$. Then*

$$y = \frac{1}{n}(1/2, \sec^2(\pi/n), \sec^2(2\pi/n), \ldots, \sec^2((n-1)\pi/n), 1/2)^T$$

is the unique solution to $Ay = e_1$, where $e_1 = (0, 1, 0, 0, \ldots, 0)^T$.

Note that y is clearly nonnegative, and thus is the unique feasible solution for DUAL. Therefore it is the dual-optimal solution, and as the entries of y sum to n, it exactly recovers the Markov-Bernstein inequality at $x = 0$:

Corollary 2. *Let p be a polynomial of degree $n = 2m + 1$ with $\|p\|_{[-1,1]} \leq 1$. Then $p'(0) \leq n$.*

While we have recovered the Markov-Bernstein inequality only for odd-degree polynomials at zero, a simple "shift-and-scale" argument recovers the asymptotic bound for any x bounded away from the endpoints $\{-1, 1\}$.

Corollary 3. *Let p be a polynomial of degree n with $\|p\|_{[-1,1]} \leq 1$. Then for any $x_0 \in (-1, 1)$, $|p'(x_0)| \leq \frac{n+1}{1-|x_0|}\|p\|_{[-1,1]}$. In particular, for any constant $\varepsilon \in (0, 1)$, $\|p'\|_{[-1+\varepsilon, 1-\varepsilon]} = O(n)\|p\|_{[-1,1]}$.*

We remark that the full Markov-Bernstein inequality guarantees that $|p'(x)| \leq \frac{n}{\sqrt{1-x^2}}\|p\|_{[-1,1]}$, which has quadratically better dependence on the distance from x to ± 1. However, for x bounded away from ± 1 our bound is asymptotically tight and sufficient for many applications in theoretical computer science, such as proving that the approximate degree of the Majority function on n variables is $\Omega(n)$. Moreover, we can recover the Markov-Bernstein inequality near ± 1 by considering a different linear program. We omit the details from this extended abstract for brevity.

6 Conclusion

The approximate degree is a fundamental measure of the complexity of a Boolean function, with pervasive applications throughout theoretical computer science. We have sought to advance our understanding of this complexity measure by resolving the approximate degree of the AND-OR tree, and reproving old lower bounds through the construction of explicit dual witnesses. Nonetheless, few general results on approximate degree are known, and our understanding of the approximate degree of fundamental classes of functions remains incomplete. For example, the approximate degree of AC^0 remains open [2, 6], as does the approximate degree of approximate majority (see [20, Page 11]).[2]

Resolving these open questions may require moving beyond traditional symmetrization-based arguments, which transform a polynomial p on n variables into a polynomial q on $m < n$ variables in such a way that $\widetilde{\deg}(q) \leq \widetilde{\deg}(p)$, before obtaining a lower bound on $\widetilde{\deg}(q)$. Symmetrization necessarily "throws away" information about p; in contrast, the method of constructing dual polynomials appears to be a very powerful and complete way of reasoning about approximate degree. Can progress be made on these open problems by directly constructing good dual polynomials?

References

1. Aaronson, S.: The polynomial method in quantum and classical computing. In: Proc. of Foundations of Computer Science (FOCS), p. 3 (2008), Slides available at `http://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 of Computing 1(1), 37–46 (2005)
4. Chattopadhyay, A., Ada, A.: Multiparty communication complexity of disjointness. Electronic Colloquium on Computational Complexity (ECCC) 15(002) (2008)
5. Beals, R., Buhrman, H., Cleve, R., Mosca, M., de Wolf, R.: Quantum lower bound by polynomials. Journal of the ACM 48(4), 778–797 (2001)
6. Beame, P., Machmouchi, W.: The quantum query complexity of AC^0. Quantum Information & Computation 12(7-8), 670–676 (2012)
7. Beigel, R.: The polynomial method in circuit complexity. In: Proc. of the Conference on Structure in Complexity Theory, pp. 82–95 (1993)
8. Beigel, R.: Perceptrons, PP, and the polynomial hierarchy. In: Computational Complexity, vol. 4, pp. 339–349 (1994)
9. Bernstein, S.N.: On the V. A. Markov theorem. Trudy Leningr. Industr. In-ta, no 5, razdel fiz-matem nauk, 1 (1938)
10. Buhrman, H., Vereshchagin, N.K., de Wolf, R.: On computation and communication with small bias. In: Proc. of the Conference on Computational Complexity, pp. 24–32 (2007)
11. Høyer, P., Mosca, M., de Wolf, R.: Quantum search on bounded-error inputs. In: Baeten, J.C.M., Lenstra, J.K., Parrow, J., Woeginger, G.J. (eds.) ICALP 2003. LNCS, vol. 2719, pp. 291–299. Springer, Heidelberg (2003)

[2] This open problem is due to Srikanth Srinivasan.

12. Gavinsky, D., Sherstov, A.A.: A separation of NP and coNP in multiparty communication complexity. Theory of Computing 6(1), 227–245 (2010)
13. Kalai, A., Klivans, A.R., Mansour, Y., Servedio, R.A.: Agnostically learning halfspaces. SIAM Journal on Computing 37(6), 1777–1805 (2008)
14. Klauck, H., Špalek, R., de Wolf, R.: Quantum and classical strong direct product theorems and optimal time-space tradeoffs. SIAM Journal on Computing 36(5), 1472–1493 (2007)
15. 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)
16. Klivans, A.R., Sherstov, A.A.: Lower bounds for agnostic learning via approximate rank. Computational Complexity 19(4), 581–604 (2010)
17. Lee, T., Shraibman, A.: Disjointness is hard in the multi-party number-on-the-forehead model. In: Proc. of the Conference on Computational Complexity, pp. 81–91 (2008)
18. Markov, V.: On functions which deviate least from zero in a given interval, St. Petersburg (1892) (Russian)
19. Minsky, M.L., Papert, S.A.: Perceptions: An Introduction to Computational Geometry. MIT Press, Cambridge (1969)
20. Open problems in analysis of Boolean functions. Compiled for the Simons Symposium. CoRR, abs/1204.6447, February 5-11 (2012)
21. Nisan, N., Szegedy, M.: On the degree of boolean functions as real polynomials. Computational Complexity 4, 301–313 (1994)
22. Paturi, R.: On the degree of polynomials that approximate symmetric Boolean functions (Preliminary Version). In: Proc. of the Symp. on Theory of Computing (STOC), pp. 468–474 (1992)
23. Schrijver, A.: Theory of Linear and Integer Programming. John Wiley & Sons, New York (1986)
24. Sherstov, A.A.: Approximate inclusion-exclusion for arbitrary symmetric functions. Computational Complexity 18(2), 219–247 (2009)
25. Sherstov, A.A.: Approximating the AND-OR tree. Electronic Colloquium on Computational Complexity (ECCC) 20(023) (2013)
26. Sherstov, A.A.: Communication lower bounds using dual polynomials. Bulletin of the EATCS 95, 59–93 (2008)
27. Sherstov, A.A.: The pattern matrix method. SIAM J. Comput. 40(6), 1969–2000 (2011)
28. Sherstov, A.A.: Making polynomials robust to noise. In: Proceedings of Symp. Theory of Computing, pp. 747–758 (2012)
29. Sherstov, A.A.: Separating AC^0 from depth-2 majority circuits. SIAM Journal on Computing 28(6), 2113–2129 (2009)
30. Sherstov, A.A.: The intersection of two halfspaces has high threshold degree. In: Proc. of Foundations of Computer Science (FOCS), pp. 343–362 (2009); To appear in SIAM J. Comput. (special issue for FOCS 2009)
31. Sherstov, A.A.: The multiparty communication complexity of set disjointness. In: Proceedings of Symp. Theory of Computing, pp. 525–548 (2012)
32. Shi, Y.: Approximating linear restrictions of Boolean functions. Manuscript (2002), `http://web.eecs.umich.edu/~shiyy/mypapers/linear02-j.ps`
33. Shi, Y., Zhu, Y.: Quantum communication complexity of block-composed functions. Quantum Information & Computation 9(5), 444–460 (2009)
34. Špalek, R.: A dual polynomial for OR. Manuscript (2008), `http://arxiv.org/abs/0803.4516`