Dual Lower Bounds for Approximate Degree and Markov-Bernstein Inequalities

Mark Bun[*]   Justin Thaler[†]

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 error 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, 14, 24, 35, 38]. The same lower bound was recently obtained independently by Sherstov using related techniques [31]. 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 [40]. 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 [6, 7, 23, 34], quantum computing (where it has been used to prove lower bounds on quantum query complexity, e.g. [2, 4, 17]), communication complexity [9, 11, 20, 36, 37, 39, 40], and computational learning theory (where approximate degree upper bounds underly the best known algorithms for PAC learning DNF formulas and agnostically learning disjunctions) [16, 18].

In this paper, we seek to advance our understanding of this fundamental complexity measure. We focus on proving approximate degree lower bounds 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. [9, 11, 20, 37, 39, 40]). See the survey of Sherstov [32] 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 perhaps 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 [40] 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 [26]. Our solution builds on work of Špalek [40], 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 from approximation theory. These inequalities bound the derivative of a polynomial in terms of its degree. Combined with the well-known symmetrization technique (see e.g. [1, 23]), Markov-type inequalities have traditionally been the primary tool used to prove approximate degree lower bounds on Boolean functions (e.g. [2, 3, 24, 38]). 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 [8, 21, 22], 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, and we let $\mathbf{1}_n \in \{-1, 1\}^n$ denotes the n-dimensional all-ones vector. 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 $S \subseteq \{1, \ldots, n\}$ and $x \in \{-1, 1\}^n$, let $\chi_S(x) = \prod_{i \in S} x_i$. Given a Boolean function f, let $p(x) = \sum_{|S| \leq d} c_S \chi_S(x)$ be a polynomial of degree d that minimizes $\|p - f\|_\infty$, where the coefficients

c_S are real numbers. Then p is an optimum of the following linear program.

$$
\begin{array}{ll}
\min & \varepsilon \\
\text{such that} & \left| f(x) - \sum_{|S| \le d} c_S \chi_S(x) \right| \le \varepsilon \quad \text{for each } x \in \{-1,1\}^n \\
& c_S \in \mathbb{R} \hspace{4.5cm} \text{for each } |S| \le d \\
& \varepsilon \ge 0
\end{array}
$$

The dual LP is as follows.

$$
\begin{array}{ll}
\max & \sum_{x \in \{-1,1\}^n} \phi(x) f(x) \\
\text{such that} & \sum_{x \in \{-1,1\}^n} |\phi(x)| = 1 \\
& \sum_{x \in \{-1,1\}^n} \phi(x) \chi_S(x) = 0 \quad \text{for each } |S| \le d \\
& \phi(x) \in \mathbb{R} \hspace{3.3cm} \text{for each } x \in \{-1,1\}^n
\end{array}
$$

Strong LP-duality yields the following well-known dual characterization of approximate degree (cf. [37]).

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

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

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 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 [24]. This was subsequently improved to $\Omega(\sqrt{N \log N})$ by Shi [38], and improved further to $\Omega(N^{2/3})$ by Ambainis [3]. Most recently, Sherstov proved an $\Omega(N^{3/4})$ lower bound in [35], which was the best lower bound prior to our work. The best upper bound is $O(N)$ due to Høyer, Mosca, and de Wolf [14], which matches our new lower bound.

By refining Sherstov's analysis in [35], 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 [33]. In particular, this implies that the approximate degree of the two-level AND-OR tree is $\Theta(N)$.

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

Independent work by Sherstov. Independently of our work, Sherstov [31] 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 [31] 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 [31] is short (barely more than a page), 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 [35], which roughly showed that approximate degree increases multiplicatively under function composition. Specifically, Sherstov showed the following.

Proposition 3 ([35, 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 $\varepsilon, \delta > 0$,*

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

Sherstov's proof of Proposition 3 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 $(\varepsilon - 4\delta\text{bs}(F))$-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 $\text{bs}(F)$ grows.

Proposition 3 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 $\text{bs}(\text{AND}_M) = M$. 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 3 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 $F(f, \ldots, f)$. The details of our construction follow.

3.2   Proof of Theorem 2

Nisan and Szegedy [24] proved the now well-known result that for any constant $0 < \varepsilon < 1$, $\deg_\varepsilon(\text{AND}_n) = \deg_\varepsilon(\text{OR}_n) = \Theta(\sqrt{n})$. Let $\Psi : \{-1, 1\}^M \to \mathbb{R}$ be a dual witness for the $(1/3)$-approximate degree of AND_M whose existence is guaranteed by Theorem 1. There is some $\varepsilon > 1/3$

and $d = \Theta(\sqrt{M})$ such that Ψ satisfies:

$$\sum_{x \in \{-1,1\}^M} \Psi(x)\text{AND}_M(x) = \varepsilon, \tag{4}$$

$$\sum_{x \in \{-1,1\}^M} |\Psi(x)| = 1, \tag{5}$$

$$\sum_{x \in \{-1,1\}^M} \Psi(x)\chi_S(x) = 0 \text{ for each } |S| \leq d. \tag{6}$$

Likewise, let ψ be the dual witness for the $(1 - (\varepsilon - 1/3)/4)$-approximate degree of OR_N. By Theorem 1, there is some $\delta < (\varepsilon - 1/3)/4$ and some $d' = \Theta(\sqrt{N})$ such that ψ satisfies:

$$\sum_{x \in \{-1,1\}^N} \psi(x)\text{OR}_N(x) = 1 - \delta, \tag{7}$$

$$\sum_{x \in \{-1,1\}^N} |\psi(x)| = 1, \tag{8}$$

$$\sum_{x \in \{-1,1\}^N} \psi(x)\chi_S(x) = 0 \text{ for each } |S| \leq d'. \tag{9}$$

We will also make use of the following easy lemma, which tells us the precise values of $\psi(\mathbf{1}_N)$ and $\Psi(-\mathbf{1}_M)$. This is essentially a restatement of a result due to Gavinsky and Sherstov [12].

Lemma 4

$$1 - \delta = \sum_{x \in \{-1,1\}^N} \psi(x)\,\text{OR}_N(x) = 2\psi(\mathbf{1}_N). \tag{10}$$

In particular, $\psi(\mathbf{1}_N) > 0$. Similarly,

$$\varepsilon = \sum_{x \in \{-1,1\}^M} \Psi(x)\,\text{AND}_M(x) = -2\Psi(-\mathbf{1}_M). \tag{11}$$

In particular, $\Psi(-\mathbf{1}_M) < 0$.

Proof of Lemma 4: The first part follows because $\sum_{x \in \{-1,1\}^N} \psi(x)\text{OR}_N(x) = 2\psi(\mathbf{1}) - \sum_{x \in \{-1,1\}^N} \psi(x)$. The second term on the right-hand side is zero because ψ is orthogonal to all polynomials of degree at most d, and in particular ψ is orthogonal to the constant function. The proof for the second part is similar. ∎

As in Sherstov's proof of Proposition 3, we define $\zeta : \left(\{-1,1\}^N\right)^M \to \mathbb{R}$ by

$$\zeta(x_1, \ldots, x_M) := 2^M \Psi(\ldots, \widetilde{\text{sgn}}(\psi(x_i)), \ldots) \prod_{i=1}^M |\psi(x_i)|, \tag{12}$$

where $x_i = (x_{i,1}, \ldots, x_{i,N})$.

By Theorem 1, in order 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 show that

$$\sum_{(x_1,\ldots,x_M)\in(\{-1,1\}^N)^M} \zeta(x_1,\ldots,x_M)\text{AND-OR}_N^M(x_1,\ldots,x_M) \geq 1/3. \tag{13}$$

$$\sum_{(x_1,\ldots,x_M)\in(\{-1,1\}^N)^M} |\zeta(x_1,\ldots,x_M)| = 1. \tag{14}$$

$$\sum_{(x_1,\ldots,x_M)\in(\{-1,1\}^N)^M} \zeta(x_1,\ldots,x_M)\chi_S(x_1,\ldots,x_M) = 0 \text{ for each } |S| \leq d\cdot d'. \tag{15}$$

Eq. (15) is proved exactly as in [35]; we provide Sherstov's argument in Appendix A.2 for completeness. We now argue that Expression (13) and Eq. (14) hold as well.

Proof of Eq. (14). Let μ be the distribution on $(\{-1,1\}^N)^M$ given by $\mu(x_1,\ldots,x_M) = \prod_{i=1}^M |\psi(x_i)|$. Since ψ is orthogonal to the constant polynomial, it has expected value 0, and hence the string $(\ldots,\widetilde{\text{sgn}}(\psi(x_i)),\ldots)$ is distributed uniformly in $\{-1,1\}^M$ when one samples $(x_1,\ldots,x_M)$ according to μ. Thus,

$$\sum_{(x_1,\ldots,x_M)\in(\{-1,1\}^N)^M} |\zeta(x_1,\ldots,x_M)| = \sum_{z\in\{-1,1\}^M} |\Psi(z)| = 1$$

by Eq. (5), proving Eq. (14). ∎

Proof of Expression (13). Using the same distribution μ as in the proof of Eq. (14), observe that

$$\sum_{(x_1,\ldots,x_M)\in(\{-1,1\}^N)^M} \zeta(x_1,\ldots,x_M)\text{AND-OR}_N^M(x_1,\ldots,x_M)$$

$$= 2^M \mathbf{E}_\mu[\Psi(\ldots,\widetilde{\text{sgn}}(\psi(x_i)),\ldots)\text{AND}_M(\ldots,\text{OR}_N(x_i),\ldots)]$$

$$= \sum_{z\in\{-1,1\}^M} \Psi(z)\left(\sum_{(x_1,\ldots,x_M)\in(\{-1,1\}^N)^M} \text{AND}_M(\ldots,\text{OR}_N(x_i),\ldots)\,\mu(x_1,\ldots,x_M|z)\right), \tag{16}$$

where $\mu(\mathbf{x}|z)$ denotes the probability of $\mathbf{x}$ under μ, conditioned on $(\ldots,\widetilde{\text{sgn}}(\psi(x_i)),\ldots) = z$.

Let $A_1 = \{x\in\{-1,1\}^N : \psi(x) \geq 0, \text{OR}_N(x) = -1\}$ and $A_{-1} = \{x\in\{-1,1\}^N : \psi(x) < 0, \text{OR}_N(x) = 1\}$, so $A_1\cup A_{-1}$ is the set of all inputs x where the sign of $\psi(x)$ disagrees with $\text{OR}_N(x)$. Notice that $\sum_{x\in A_1\cup A_{-1}} |\psi(x)| = \delta/2$ because ψ has correlation $1-\delta$ with OR_N.

As noted in [35], for any given $z\in\{-1,1\}^M$, the following two random variables are identically distributed:

- The string $(\ldots,\text{OR}_N(x_i),\ldots)$ when one chooses $(\ldots,x_i,\ldots)$ from the conditional distribution $\mu(\cdot|z)$.

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

Thus, Expression (16) equals

$$\sum_{z\in\{-1,1\}^M} \Psi(z)\cdot\mathbf{E}[\mathrm{AND}_M(\ldots,y_iz_i,\ldots)], \tag{17}$$

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

We first argue that the term corresponding to $z=-\mathbf{1}_M$ contributes $\Psi(z)$ to Expression (17). By Eq. (10) of Lemma 4, if $\mathrm{OR}_N(x)=1$ (i.e., if $x=\mathbf{1}_N$), then $\widetilde{\mathrm{sgn}}(\psi(x))=1$. This implies that A_{-1} is empty; that is, if $\widetilde{\mathrm{sgn}}(\psi(x))=-1$, 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 Part 2 of Lemma 4, $\widetilde{\mathrm{sgn}}(\Psi(-\mathbf{1}_M))=-1$, and thus the term corresponding to $z=-\mathbf{1}_M$ contributes $-\Psi(z)$ to Expression (17) as claimed.

All $z\ne-\mathbf{1}_M$ can be handled as in Sherstov's proof of Proposition 3, because AND_M has low block sensitivity at these inputs. To formalize this, we invoke the following proposition, whose proof we provide in Appendix A.1 for completeness.

Proposition 5 ([35]) *Let $F:\{-1,1\}^M\to\{-1,1\}$ be a given Boolean function. Let $y\in\{-1,1\}^M$ be a random string whose ith bit is set to -1 with probability at most $\alpha\in[0,1]$, and to $+1$ otherwise, independently for each i. Then for every $z\in\{-1,1\}^M$,*

$$\mathbf{P}_y[F(z_1,\ldots,z_M)\ne F(z_1y_1,\ldots,z_My_M)]\le 2\alpha\,\mathrm{bs}_z(F).$$

In particular, since $\mathrm{bs}_z(\mathrm{AND}_M)=1$ for all $z\ne-\mathbf{1}_M$, Proposition 5 implies that for all $z\ne-\mathbf{1}_M$, and $F=\mathrm{AND}_M$, $\mathbf{P}_y[F(z_1,\ldots,z_M)=F(z_1y_1,\ldots,z_ky_k)]\ge 1-2\delta$.

Recall that the term corresponding to $z=-\mathbf{1}_M$ contributes $-\Psi(-\mathbf{1}_M)$ to the sum, we obtain the following lower bound on Expression (17).

$$\sum_{z\in\{-1,1\}^M}\Psi(z)\cdot\mathbf{E}[\mathrm{AND}_M(\ldots,y_iz_i,\ldots)]\ge-\Psi(-\mathbf{1}_M)+\left(\sum_{z\ne-\mathbf{1}_M}\Psi(z)\mathrm{AND}_M(z)\right)-4\delta\left(\sum_{z\ne-\mathbf{1}_M}|\Psi(z)|\right)$$

$$\ge\left(\sum_{z\in\{-1,1\}^M}\Psi(z)\mathrm{AND}_M(z)\right)-4\delta=\varepsilon-4\delta>1/3.$$

∎

This completes the proof of Theorem 2.

∎

Remark 6 *Špalek [40] 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 to any symmetric function in Section 4 below). In Appendix B we show how to generalize Špalek's argument in a different way to handle any constant $\varepsilon\in(0,1)$ (due to reuse of notation, Appendix B is intended to be read after Section 4). With these dual polynomials in hand, the dual solution ζ given in our proof is completely explicit. This answers a question of Špalek [40, 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 [26] 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 [40], building on work of Szegedy, exhibited an explicit 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 [40, 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. The simplest symmetric functions are the t-threshold functions:

$$\tau_t(x) = \begin{cases} -1 & \text{if } |x| \geq t \\ 1 & \text{otherwise.} \end{cases}$$

Important special cases include $\mathrm{OR} = \tau_1$, $\mathrm{AND} = \tau_n$, and the majority function $\mathrm{MAJ} = \tau_{\lceil n/2 \rceil}$. 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 Theorem 1 as it applies to symmetric functions. Following the notation in [40], 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 Theorem 1 and it becomes convenient to work with the following reformulation of Theorem 1.

Corollary 7 *A symmetric function $f : \{-1, 1\}^n \to \{-1, 1\}$ has ε-approximate degree greater than d if and only if 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.

We exploit the following method for constructing polynomials of pure high degree d. Let ψ be a multivariate polynomial of degree $n - d$, and let $\chi_{[n]}(x)$ denote the parity function on n variables. Consider the function $\phi(x) = \psi(x)\chi_{[n]}(x)$, i.e., ϕ is obtained by multiplying ψ by the parity function. It is straightforward to check that ϕ has pure high degree d. Notice that if ψ is symmetric, then so is ϕ, and the corresponding univariate polynomials satisfy $\Phi(k) = \Psi(k) \cdot (-1)^k$. Therefore, to show that a symmetric function f with a "jump" at t has approximate degree greater than d, it is enough to exhibit an $(n - d)$-degree univariate polynomial Ψ such that $(-1)^i \Psi(i)$ has high correlation with its associated univariate function F.

We are now in a position to state the lower bound that we will prove in this section. Paturi [26] 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 8 ([26, 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 the Jackson theorems from approximation theory. He proved the lower bound by combining symmetrization with an appeal to the Markov-Bernstein inequality (see Section 5) – however, 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 9 *Given f and F as above, let $1 \leq t \leq n$ be an integer with $F(t - 1) \neq F(t)$. Then $\widetilde{\deg}(f) = \Omega(\sqrt{t(n - t + 1)})$.*

In particular, the approximate degree of the symmetric t-threshold function is $\Omega(\sqrt{t(n - t + 1)})$. This special case serves as a useful model for understanding our construction.

4.2 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 $S = \{k^2 : k^2 \leq n\}$ 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

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

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 $|Q \cdot \text{OR}|/\|Q\|_1$ 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:

Step 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 $j = t$.

Step 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 $t - 1, t, t + 1$.

Step 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 Section 4.3, we carry out this line of attack in the case where $t = \Omega(n)$. This partial result also gives the right intuition for general t, although the details are somewhat more complicated. Namely, in Step 3, we may need to rely on the alternative points t and $t + 2$ to contribute high positive correlation between F and Q, rather than inputs $t - 1$ and t.

---

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

10

4.3 A Dual Polynomial for MAJ

We first construct a dual polynomial that witnesses an $\Omega(t)$ lower bound for symmetric functions having a "jump" at $t \le n/2$. Notice that this bound matches Proposition 9 if $t = \Omega(n)$, but is weaker otherwise (e.g. in the case of OR). By setting $t = \lceil \frac{n}{2} \rceil$, we can write down down a clean dual polynomial for the majority function MAJ. This case is illustrative, as one can view Špalek's dual polynomial for OR and our dual polynomial for MAJ as two ends of a spectrum, with our general construction interpolating between the two extremes.

Proposition 10 *Let* $f : \{-1, 1\}^n \to \{-1, 1\}$ *be a Boolean function with associated univariate function F. If $1 \le t \le n/2$ such that $F(t-1) \ne F(t)$, then $\widetilde{\deg}(f) = \Omega(t)$.*

Proof: We follow the proof outline given in the previous section. Define the set

$$S = \{t \pm 4\ell : 0 \le \ell \le t/4\}.$$

Note that $|S| = \Omega(t)$. We claim that $\pi_S(i) := \prod_{j \in S, j \ne i} |j - i|$ is minimized at $i = t$. Notice that translating all points in S by a constant does not affect $\pi_S(i)$, and scaling all points in S by a constant does not affect $\operatorname{argmin}_i \pi_S(i)$. Thus, it is enough to show that $\pi_{S^*}(i)$ is minimized at $i = 0$ for the set $S^* = \{\pm\ell : \ell \le t\}$. In this case, $\pi_{S^*}(i)$ takes the simple form $(t-i)!(t+i)!$, and we see that

$$\frac{\pi_{S^*}(0)}{\pi_{S^*}(i)} = \frac{(t!)^2}{(t-i)!(t+i)!} = \frac{t}{t+|i|} \cdot \frac{t-1}{t+|i|-1} \cdot \ldots \cdot \frac{t-|i|+1}{t+1}$$

is a product of terms smaller than 1, so $\pi_{S^*}(i)$ is indeed minimized at $i = 0$.

With Step 1 completed, we let $T = S \cup \{t-1, t+1\}$ and define the polynomial

$$P(x) = (-1)^s \frac{4^{2h}(h!)^2}{n!} \prod_{j \in [n] \setminus T} (x - j),$$

where $h = \lfloor t/4 \rfloor$ and s is a sign bit to be determined later. The normalization is chosen so that $\binom{n}{t}|P(t)| = 1$. We divide by both $(x - (t-1))$ and $(x - (t+1))$ to ensure that the rate of decay of $P(x)$ is at least quadratic as x moves away from t. This will ultimately allow us to show that most of the ℓ_1 mass of P comes from the points $x = t-1$ and $x = t$.

Write the ℓ_1 contribution due to the point r as

$$\binom{n}{r}|P(r)| = \binom{n}{r} \frac{4^{2h}(h!)^2}{n!} \frac{\prod_{j \in [n] \setminus \{r\}} |r - j|}{\prod_{j \in T \setminus \{r\}} |r - j|} = \frac{4^{2h}(h!)^2}{\prod_{j \in T \setminus \{r\}} |r - j|}.$$

For $r = t \pm 1$ this becomes

$$\frac{4^{2h}(h!)^2}{2 \prod_{\ell=1}^{h} (4\ell - 1)(4\ell + 1)} = \frac{1}{2} \prod_{\ell=1}^{h} \left(1 + \frac{1}{16\ell^2 - 1}\right)$$

$$\le \frac{1}{2} \exp\left(\sum_{\ell=1}^{h} \frac{1}{15\ell^2}\right)$$

$$\le \frac{1}{2} e^{\pi^2/90} < 1,$$

where the first inequality holds because $1 + x \leq e^x$ for all $x \geq 0$. This shows that the ℓ_1 contributions of the points $t-1$ and $t+1$ are equal, and not too large:

$$\binom{n}{t-1} |P(t-1)| = \binom{n}{t+1} |P(t+1)| < 1.$$

Now we analyze the remaining summands, and show that their total contribution is much smaller than 1. Recall that the choice $i = t$ minimizes $\pi_S(i)$, and that $\pi_S(t) = 4^{2h}(h!)^2$. Therefore,

$$\binom{n}{t+4\ell} |P(t+4\ell)| = \frac{4^{2h}(h!)^2}{\prod_{j \in T \setminus \{t+4\ell\}} |t + 4\ell - j|} \leq \frac{1}{|4\ell+1||4\ell-1|} \leq \frac{1}{15\ell^2}.$$

We can use this quadratic decay to bound the total ℓ_1 mass of the points outside of $\{t-1, t, t+1\}$:

$$\sum_{j \in S \setminus \{t\}} \binom{n}{j} |P(j)| \leq \sum_{\ell=-h}^{h} \frac{1}{15\ell^2} \leq \frac{2}{15} \cdot \frac{\pi^2}{6} < \frac{1}{4}.$$

For the final part of our construction, we multiply P by parity to get $Q(i) = (-1)^i P(i)$. Since $P(t-1)$ and $P(t)$ have the same sign, $Q(t-1)$ and $Q(t)$ have opposite signs. Since $F(t-1)$ and $F(t)$ also have opposite signs, we can choose $s \in \{-1, 1\}$ to ensure that

$$Q \cdot F > \binom{n}{t-1} |P(t-1)| + \binom{n}{t} |P(t)| - \binom{n}{t-1} |P(t+1)| - \sum_{j \in S \setminus \{t\}} \binom{n}{j} |P(j)|$$

$$\geq 1 - \frac{1}{4} = \frac{3}{4}.$$

As the total ℓ_1 mass $\|P\|_1$ is at most $3 + \frac{1}{4}$, we get that $(Q \cdot F)/\|Q\|_1 > \frac{3}{13}$. By Corollary 7, the $\frac{3}{13}$-approximate degree of f is $\Omega(t)$. $\blacksquare$

4.4   General Symmetric Boolean Functions

We now show how to generalize our dual polynomial for MAJ and Špalek's dual polynomial for OR to handle arbitrary symmetric functions. Recall that we are given a Boolean function f, an associated univariate function F, and a number t such that $F(t-1) \neq F(t)$. As our goal is to show that $\widetilde{\deg}(F) = \Omega(\sqrt{t(n-t+1)})$, we may without loss of generality assume that $t \leq n/2$ throughout this section. As the case of $t = 1$ is handled by Špalek's construction, we may also assume $t \geq 2$ to improve the constants in our analysis.

As a first attempt at defining a suitable set S for use in constructing a dual polynomial for f, we consider the set $S' = \{tk^2 : k^2 \leq n/t\}$. Fact 24 in Appendix C implies that $\prod_{i \in S', i \neq j} |j - i|$ is minimized at t. Unfortunately, the set S' is too small – it has size only $\Theta(\sqrt{n/t})$ instead of $\Theta\left(\sqrt{t(n-t+1)}\right)$. The trick is to notice that the distance between any two points in S' is at least t. Therefore, we should be able to interlace $\Theta(t)$ translated copies of S', and still have the desired product minimized near t. The following lemma gives the details, but its proof is rather technical and deferred to Appendix C.

Lemma 11 *Let*

$$S = \{tk^2 + 4\ell : 1 \le k \le \sqrt{(n-t+1)/t}, 0 \le \ell \le ct\} \cup \{t - 4\ell : 0 \le \ell \le ct\}$$

where $c \le 1/32$. Then for $i \in S$, the product

$$\pi_S(i) := \prod_{\substack{i' \in S \\ i' \ne i}} |i - i'|$$

is minimized for some i^* with $(1 - 4c)t \le i^* \le (1 + 4c)t$.

Thus the product of differences is minimized somewhere in a $\Theta(t)$-sized neighborhood of t. The exact location depends delicately on n and t. However, since the product $\pi_S(i)$ is invariant under translations of S, we can assume that the minimizer i^* is one of the points $t - 1, t$, or $t + 1$.

For intuition, observe that one can view the set S of Lemma 11 as interpolating between the set for OR used by Špalek, and the set used to prove our $\Omega(t)$ lower bound in Proposition 23. Notice that S contains all points of the form $t \pm 4\ell$, plus additional points corresponding to perfect squares when $t = o(n)$.

Let S be the set in Lemma 11 (or a translate thereof), and suppose the corresponding product is minimized at i^*. Let $T = S \cup \{i^* - 1, i^* + 1\}$. Define the polynomial

$$P(x) = (-1)^s \frac{\pi_S(i^*)}{n!} \prod_{j \in [n] \setminus T} (x - j),$$

where the sign bit s is to be determined later. The choice of normalization is so that $\binom{n}{i^*}|P(i^*)| = 1$. Our goal is now to show that the ℓ_1 mass of P is concentrated at the points $i^* - 1, i^*$ and $i^* + 1$. The following lemma shows that the contribution of a point x to the ℓ_1 mass of P decays at a quadratic rate as x moves away from i^*. This is precisely because we include both points $i^* - 1$ and $i^* + 1$ in T. Full details are in Appendix C.

Lemma 12 *Let $r = tk^2 + 4\ell$ for $k \ge 2$ and $0 \le \ell \le ct$. Then $\binom{n}{r}|P(r)| \le 1/(t^2(k^2 - 2)^2)$. If $v = i^* + 4\ell$, then $\binom{n}{v}|P(v)| \le 1/(16\ell^2 - 1)$.*

Since the sum of the inverse squares of the integers is bounded by a constant, the total contribution of these points to $\|P\|_1$ is dominated by the mass contributed by $P(i^*)$.

Lemma 13

$$\sum_{j \in T \setminus \{i^* - 1, i^*, i^* + 1\}} \binom{n}{j}|P(j)| \le \frac{2}{5}.$$

This bound allows us to sketch a proof of Proposition 9 in full generality.

Proof sketch of Proposition 9: We consider three cases based on which of $\binom{n}{i^* - 1}|P(i^* - 1)|$, $\binom{n}{i^*}|P(i^*)| = 1$, and $\binom{n}{i^* + 1}|P(i^* + 1)|$ is the smallest. The relationship between these terms determines how we choose the location of i^* relative to the "jump" at t. We set i^* so that after multiplying P by parity to obtain a polynomial Q, the larger two of these terms contribute positively to $Q \cdot F$. They will hence dominate the (possibly negative) correlation due to the smallest

term, as well as the contribution of size at most $2/5$ due to remaining points in T. Ultimately, we show that $(Q \cdot F)/\|Q\|_1 \le \frac{1}{14}$, which gives the asserted lower bound by Corollary 7. The calculations for each of these cases are analogous to those in the proof of Proposition 23 and given in Appendix C.

∎

4.5 On Complementary Slackness

In this section, we give some additional intuition, based on complementary slackness, that helps to explain the structure of the dual polynomial exhibited in Section 4.4. The discussion that follows is deliberately informal and is meant to complement the formal argument given in Section 4.4.

We illustrate the idea by considering the symmetric t-threshold function f, which evaluates to -1 on inputs of Hamming weight at least t and evaluates to 1 on all other inputs. In [30], Sherstov gives an explicit, asymptotically optimal polynomial p for approximating f in the ℓ_∞ norm. If this polynomial p were in fact an *exactly* optimal solution to the primal linear program of Section 2.1, then complementary slackness (cf. [27, pg. 95]) would imply that the optimal dual polynomial ϕ is supported on the points corresponding to the constraints made tight by the primal optimal polynomial p. That is, it would hold that $\phi(x) = 0$ except for those $x \in \{-1, 1\}^n$ for which $|p(x) - f(x)| = \varepsilon$. We will refer to such values of x as *maximum-error points of p*.

While it is not clear whether Sherstov's polynomial p is exactly optimal, our dual polynomial is still approximately consistent with the conditions obtained by applying complementary slackness to p. Sherstov's construction of p works by taking a Chebyshev polynomial of degree $\Theta(\sqrt{n/t})$, shifting and scaling it, and then composing it with a Chebyshev polynomial of degree $\Theta(t)$. This is reminiscent of our dual solution, which interlaces $\Theta(t)$ copies of a set of size $\Theta(\sqrt{n/t})$. In general, it is difficult to determine the precise maximum-error points of Sherstov's polynomial p. However, our dual polynomial can be viewed as placing nonzero weight on close *approximations* to the maximum-error points of p. We explain this viewpoint below.

Let $T_d : \mathbb{R} \to \mathbb{R}$ denote the degree-d Chebyshev polynomial of the first kind. It is well-known that the extreme points of T_d are the degree-d *Chebyshev nodes*, which take the form $\cos(k\pi/d)$ for $0 \le k \le d$. Truncating the Taylor expansion of $\cos(x) = 1 - x^2/2 + \ldots$ after the quadratic term, one sees that for $d = \sqrt{n}$, $\cos(k\pi/d) \approx 1 - (ck^2/d^2) = 1 - ck^2/n$ for some constant c.

It is known [24] that an appropriately shifted-and-scaled Chebyshev polynomial Q_d of degree $d = \Theta(\sqrt{n})$ itself yields an asymptotically optimal approximation $Q_d(\sum_{i=1}^n x_i/n)$ to the OR function. Recall from Section 4.2 that Špalek's dual polynomial for the OR function [40] only places nonzero weight on inputs of Hamming weight equal to a perfect square (or equal to two). We can therefore view Špalek's dual polynomial for the OR function as placing nonzero weight only on points whose Hamming weight closely approximates a constant multiple of a Chebyshev node (Appendix B shows that the dual polynomial for the OR function is robust to scaling the non-vanishing Hamming weight values by constants, i.e., it suffices to place nonzero weight only on inputs of Hamming weight ck^2 or 1 for any constant $c \ge 2$).

Moving to the general case and eliding many details, Sherstov approximates the symmetric t-threshold function f with a polynomial p roughly of the form

$$p(x) = T_t\left(q\left(\sum_{i=1}^n x_i/n\right)\right),$$

14

where q is a shifted-and-scaled version of the Chebyshev polynomial of degree $\sqrt{n/t}$. Following the intuition above that the degree-d Chebyshev nodes are approximated by points of the form $1 - ck^2/d^2$ for some constant c, the inner polynomial $q(\sum_{i=1}^{n} x_i/n)$ in Sherstov's construction hits its extreme points at inputs of Hamming weight close to ctk^2 for the non-negative integers $k \leq \sqrt{n/t}$. Moreover, q alternates between -1 and $+1$ at its extreme points. Thus, as the Hamming weight of the input x increases from ctk^2 to $ct(k+1)^2$, the inner polynomial $q(\sum_{i=1}^{n} x_i/n)$ passes through all t maximum-error points of the *outer* polynomial T_t.

Note also that the degree-d Chebyshev nodes $\cos(k\pi/d) \approx 1 - k^2/d^2$ are clustered near the endpoints of the interval $[-1, 1]$ rather than the middle of the interval, so most of the maximum-error points of the composed polynomial in fact fall very close to inputs of Hamming weight ctk^2.

To see how these maximum-error points correspond to the support of our dual polynomial ϕ for the t-threshold function, recall that there is some constant c' such ϕ takes nonzero values only on inputs with Hamming weight in the set

$$S = \{tk^2 + 4\ell : 1 \leq k \leq \sqrt{(n-t+1)/t}, 0 \leq \ell \leq c't\} \cup \{t - 4\ell : 0 \leq \ell \leq c't\}.$$

Roughly speaking, our dual witness thus takes nonzero values only on inputs of Hamming weight very close to tk^2 for each $k \leq \sqrt{n/t}$ (i.e., for each Hamming weight of the form tk^2, our dual witness takes nonzero values on t distinct Hamming weights in the vicinity of tk^2), just as as predicted above.

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, 7, 16, 19, 24, 26, 28, 37]).

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 $x \in [-1, 1]$.

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

One might initially be concerned that our goal is to bound $|p'(x)|$, while the above LP only yields an upper bound on $p'(x)$. But for any polynomial p satisfying $\|p\|_{[-1,1]} \leq 1$ whose derivative is negative, $-p$ is a feasible solution to the above LP achieving value $|p'(x)|$. Thus, the value of the above LP indeed equals $\sup_{p \in B_n} |p'(0)|$, where B denotes the set of all degree n polynomials p satisfying $\|p'\|_{[-1,1]} \leq 1$.

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.

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

The dual to PRIMAL can be written as

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

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 make use of a number of trigonometric identities that can all be established by elementary methods. These identities are presented in Appendix D.

Proposition 14 *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 $e_1 = (0, 1, 0, 0, \ldots, 0)^T$.

Before proving the proposition, we explain its consequences. Note that y is clearly nonnegative, and thus is the unique feasible solution for DUAL. Therefore it is the dual-optimal solution, and exactly recovers the Markov-Bernstein inequality at $x = 0$:

Corollary 15 *Let p be a polynomial of degree $n = 2m + 1$ with $\|p\|_{[-1,1]} \le 1$. Then $p'(0) \le n$.*

Proof: Let y be as in Proposition 14. This is the unique feasible point for DUAL. By Lemma 30 in Appendix D,

$$\sum_{j=0}^{n-1} \sec^2 \left(\frac{j\pi}{n} \right) = n^2,$$

so we immediately see that $\sum_{j=0}^{n} y_j = n$. By weak LP duality, the value of PRIMAL is at most n.
∎

While we have recovered the Markov-Bernstein inequality only for odd-degree polynomials at the point $x = 0$, a simple "shift-and-scale" argument recovers the asymptotic bound for any x bounded away from the endpoints $\{-1, 1\}$.

Corollary 16 *Let p be a polynomial of degree n with $\|p\|_{[-1,1]} \le 1$. Then for any $x_0 \in (-1, 1)$, $|p'(x_0)| \le \frac{n+1}{1-|x_0|} \|p\|_{[-1,1]}$. In particular, for any constant $\varepsilon \in (0, 1)$, $\|p'\|_{[-1+\varepsilon, 1-\varepsilon]} = O(n)\|p\|_{[-1,1]}$.*

Proof: Assume without loss of generality that $x_0 \in [0, 1)$ – an identical argument holds if $x_0 \in (-1, 0]$. By Corollary 15, $|q'(0)| \le (n+1)\|q\|_{[-1,1]}$ for any polynomial q of degree at most n. Define the degree-n polynomial $q(x) = p((1-x_0)x + x_0)$. Since $(1-x_0)x + x_0 \in [-1, 1]$ for every $x \in [-1, 1]$, we have $\|q\|_{[-1,1]} \le \|p\|_{[-1,1]}$. Moreover, $q'(x) = p'((1-x_0)x + x_0)(1-x_0)$, so $q'(0) = p'(x_0)(1-x_0)$. Therefore,

$$|p'(x_0)| = \frac{|q'(0)|}{1 - |x_0|} \le \frac{n+1}{1-|x_0|}\|p\|_{[-1,1]}.$$

∎

We remark that the full Markov-Bernstein inequality guarantees that $|p'(x)| \le \frac{n}{\sqrt{1-x^2}}\|p\|_{[-1,1]}$, 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. Moreover, we can recover the Markov-Bernstein inequality near ± 1 by considering a different linear program (cf. Subsection 5.2).

Proof of Proposition 14: We write

$$(Ay)_i = \frac{(-1)^m}{2n} + \frac{(-1)^{i+m+1}}{2n} + \frac{1}{n}\sum_{j=1}^{n-1}(-1)^{j+m}\cos^{i-2}\left(\frac{j\pi}{n}\right). \tag{18}$$

Our goal is to show that $(Ay)_i = 1$ for $i = 1$, and $(Ay)_i = 0$ for all other i. The case where i is even is easy. Since $\cos(\pi - \theta) = -\cos\theta$, the terms in the sum naturally pair up. Specifically,

$$(-1)^{j+m}\cos^{i-2}\left(\frac{j\pi}{n}\right) + (-1)^{(n-j)+m}\cos^{i-2}\left(\frac{(n-j)\pi}{n}\right) = 0,$$

so the sum in Eq. (18) is clearly zero.

Now suppose i is odd and larger than 1. Then Lemma 27 in Appendix D implies that $(Ay)_i = 0$. All that remains is the case of $i = 1$. We write the sum explicitly as

$$(Ay)_1 = \frac{(-1)^m}{n} + \frac{1}{n}\sum_{j=1}^{n-1}(-1)^{j+m}\sec\left(\frac{j\pi}{n}\right).$$

By Lemma 29 in Appendix D, this evaluates to 1. ∎

5.2 Proving the Markov-Bernstein Inequality at $x = 1$

A similar strategy allows us to bound the derivative of a degree-n polynomial p at the point $x = 1$. We can expand $p(x)$ around 1 as $p(x) = c_n(x-1)^n + c_{n-1}(x-1)^{n-1} + \cdots + c_1(x-1) + c_0$. Then $p'(1) = c_1$. A modest update to PRIMAL captures the problem of maximizing $p'(1)$ subject to boundedness constraints at the Chebyshev nodes.

$$
\begin{aligned}
\max \quad & c_1 \\
\text{such that} \quad & \sum_{i=0}^{n} c_i(x-1)^i \leq 1, \ \forall x = \cos(j\pi/n), \ 0 \leq j \leq n, \ j \text{ even} \\
- \ & \sum_{i=0}^{n} c_i(x-1)^i \leq 1, \ \forall x = \cos(j\pi/n), \ 0 \leq j \leq n, \ j \text{ odd}
\end{aligned}
$$

The dual linear program takes the form

$$
\begin{aligned}
\min \quad & \sum_{i=0}^{n} y_i \\
\text{such that} \quad & By = e_1 \\
& y_j \geq 0 \ \forall j \in \{0, \ldots, n\}
\end{aligned}
$$

where $B_{0,0} = 1$ and $B_{ij} = (-1)^j(\cos(j\pi/n) - 1)^i$ otherwise. The determinant of B is, up to sign, a Vandermonde determinant, and in particular is nonzero. Thus, $By = e_1$ has a unique solution. Again, we can write down this solution explicitly.

Proposition 17 *Let n be a natural number, and define the $(n+1) \times (n+1)$ matrix B as above. Then*

$$
y = \left(\frac{2n^2 + 1}{6}, \csc^2\left(\frac{\pi}{2n}\right), \csc^2\left(\frac{\pi}{n}\right), \ldots, \csc^2\left(\frac{(n-1)\pi}{2n}\right), \frac{1}{2} \right)
$$

is the unique solution to $By = e_1$.

Proof: We just need to show that $By = e_1$. First, if $i \neq 0$ then

$$
(By)_i = \frac{(-1)^n(-2)^i}{2} + \sum_{j=1}^{n-1} (-1)^j \csc^2\left(\frac{j\pi}{2n}\right) \left(\cos\left(\frac{j\pi}{n}\right) - 1\right)^i.
$$

Using the half-angle identity $\sin^2(\theta/2) = (1 - \cos\theta)/2$, this becomes

$$
(By)_i = (-1)^{i+n} 2^{i-1} + (-1)^i 2^i \sum_{j=1}^{n-1} (-1)^j \sin^{2i-2}\left(\frac{j\pi}{2n}\right).
$$

If $i = 1$, then the sine terms are identically 1 so $(By)_1$ evaluates to 1. If $i > 1$, then by Lemma 28 in Appendix D, the sum of sine terms evaluates to $\frac{1}{2}(-1)^n - (-1)^n = -\frac{1}{2}(-1)^n$. Therefore, $(By)_i = 0$ for all $i > 1$.

Finally, we need to show that $(By)_0 = 0$. We expand

$$
(By)_0 = \frac{2n^2 + 1}{6} + \frac{1}{2}(-1)^n + \sum_{j=1}^{n-1} (-1)^j \csc^2\left(\frac{j\pi}{2n}\right).
$$

By Lemma 32, this evaluates to 0. $\blacksquare$

Corollary 18 *If p is a polynomial of degree n with $\|p\|_{[-1,1]} \leq 1$, then $p'(1) \leq n^2$.*

Proof: Let y be as in Proposition 17. Notice that $y_j \geq 0$ for all $j \in \{0, \ldots, n\}$. Combined with Proposition 17, it is clear that y is dual-feasible. By Lemma 31,

$$\sum_{j=0}^{n} y_j = \frac{2n^2 + 1}{6} + \frac{1}{2} + \sum_{j=1}^{n-1} \csc^2 \left(\frac{j\pi}{2n} \right)$$

$$= \frac{n^2}{3} + \frac{2}{3} + \frac{4n^2 - 4}{6} = n^2.$$

∎

By combining Corollary 18 with a shifting and scaling argument similar to the one used to prove Corollary 16, we recover an asymptotic statement of Markov's inequality for the first derivative of a constrained polynomial.

Corollary 19 *If p is a polynomial of degree n, then for all $x_0 \in [-1,1]$ with $x_0 \neq 0$, $|p'(x_0)| \leq \frac{n^2}{|x_0|} \|p\|_{[-1,1]}$. Thus, for any constant $\varepsilon \in (0,1)$, $\|p'\|_{[-1,-\varepsilon] \cup [\varepsilon,1]} = O(n^2) \|p\|_{[-1,1]}$.*

Proof: The argument is the same as in the proof of Corollary 16, except we instead use the auxiliary polynomial $q(x) = p(|x_0|x)$.

∎

Combining this with Corollary 16, we recover an asymptotically tight version of Markov's inequality for the whole interval $[-1,1]$.

Corollary 20 *If p is a polynomial of degree n, then for all $x \in [-1,1]$, $|p'(x)| \leq O(n^2) \|p\|_{[-1,1]}$.*

5.3 Markov's inequality for higher derivatives

In 1892, V. Markov proved the following generalization of the Markov-Bernstein inequality to higher derivatives. Let p be a real polynomial of degree at most n, and let T_n be the nth Chebyshev polynomial of the first kind. Then

$$|p^{(k)}(x)| \leq T_n^{(k)}(1) \|p\|_{[-1,1]}$$

for every $x \in [-1,1]$. We use complementary slackness to prove an important special case of this inequality, namely that $p^{(k)}(1) \leq T_n^{(k)}(1) \|p\|_{[-1,1]}$.

While A. A. Markov's inequality for the first derivative has a short proof (see [10] for a proof using tools from approximation theory), the generalization to higher derivatives is considered a deep theorem [29]. The shortest known proof of this theorem proceeds in two steps [29, Section 3.1]. In the first step, it is shown that among all points $x \in [-1,1]$, the quantity $\sup_{p \in B} |p^{(k)}(x)|$ is maximized at $x = 1$, where again B is the set of degree n polynomials p with real coefficients such that $\|p\|_{[-1,1]} \leq 1$. In the second step, it is shown that $p^{(k)}(1) \leq T_n^{(k)}(1) \|p\|_{[-1,1]}$. It is this second step that we prove here using complementary slackness.

The following lemma, found in [13], relates the determinant of a Vandermonde matrix that skips a row to the determinant of an ordinary Vandermonde matrix. For each integer $0 \leq k \leq n$, we define the elementary symmetric polynomial

$$e_k(x_1, \ldots, x_n) = \sum_{1 \leq j_1 < j_2 < \cdots < j_k \leq n} x_{j_1} x_{j_2} \ldots x_{j_k}.$$

Lemma 21 *Let $0 \leq k \leq n$. Then*

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & \ldots & x_n \\ \vdots & \vdots & & \vdots \\ x_1^{k-1} & x_2^{k-1} & \ldots & x_n^{k-1} \\ x_1^{k+1} & x_2^{k+1} & \ldots & x_n^{k+1} \\ \vdots & \vdots & & \vdots \\ x_1^n & x_2^n & \ldots & x_n^n \end{vmatrix} = e_{n-k}(x_1, x_2, \ldots, x_n) \begin{vmatrix} 1 & 1 & 1 & 1 \\ x_1 & x_2 & \ldots & x_n \\ \vdots & \vdots & & \vdots \\ x_1^k & x_2^k & \ldots & x_n^k \\ \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2^{n-1} & \ldots & x_n^{n-1} \end{vmatrix}$$

Proposition 22 *Let p be a polynomial of degree n with $|p(x)| \leq 1$ for $x \in [-1, 1]$. Then*

$$p^{(k)}(1) \leq T_n^{(k)}(1)$$

where $T_n(x)$ is the n-th Chebyshev polynomial of the first kind.

Proof: This is obvious if $k = 0$, since $T_n(1) = 1$, so we assume $k > 0$. Recall the expansion $p(x) = c_n(x-1)^n + c_{n-1}(x-1)^{n-1} + \cdots + c_1(x-1) + c_0$. Then the k-th derivative of p at 1 is simply $k!c_k$. We consider the linear program

$$\begin{aligned} \max \quad & k!c_k \\ \text{such that} \quad & \textstyle\sum_{i=0}^n c_i(x-1)^i \leq 1, \quad \forall x = \cos(j\pi/n), \ 0 \leq j \leq n, \ j \text{ even} \\ & -\textstyle\sum_{i=0}^n c_i(x-1)^i \leq 1, \quad \forall x = \cos(j\pi/n), \ 0 \leq j \leq n, \ j \text{ odd} \end{aligned}$$

and its dual

$$\begin{aligned} \min \quad & \textstyle\sum_{i=0}^n y_i \\ \text{such that} \quad & By = k!e_k \\ & y_j \geq 0 \ \forall j \in \{0, \ldots, n\} \end{aligned}$$

where $B_{0,0} = 1$ and $B_{ij} = (-1)^j(\cos(j\pi/n) - 1)^i$ otherwise. Notice that all primal constraints are tight for the primal solution corresponding to T_n, the degree n Chebyshev polynomial of the first kind.

The determinant of B is, up to sign, a Vandermonde determinant, and in particular is nonzero. Thus, $By = k!e_k$ has a unique solution. If we can show that this solution has positive entries, complementary slackness (cf. [27, pg. 95]) implies that T_n is a primal optimal solution, and the result will follow.

We now use Cramer's rule to investigate the solution to $By = k!e_k$. Recall that Cramer's rule tells us that entry y_j is given by $\det B_j / \det B$ where the matrix B_j is obtained from B by replacing its jth column with $k!e_k$. Using the formula for the Vandermonde determinant, $\det B$ is given by

$$(-1)^{\lfloor (n+1)/2 \rfloor} \prod_{0 \leq j < j' \leq n} \left(\cos\left(\frac{j'\pi}{n}\right) - \cos\left(\frac{j\pi}{n}\right) \right).$$

Since $\cos(x)$ is a decreasing function on the interval $[0, \pi]$, all the terms in the product are negative. Therefore, the sign of $\det B$ is $(-1)^{\lfloor (n+1)/2 \rfloor + \binom{n}{2}}$.

20

For convenience, let $\alpha_j = \cos(j\pi/n) - 1$. Consider the numerator of Cramer's rule for entry y_j. This is the determinant of the matrix B_j,

$$\begin{pmatrix} 1 & -1 & \cdots & (-1)^{j-1} & 0 & (-1)^{j+1} & \cdots & (-1)^n \\ 0 & -\alpha_1 & \cdots & (-1)^{j-1}\alpha_{j-1} & 0 & (-1)^{j+1}\alpha_{j+1} & \cdots & (-1)^n(-2) \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & -\alpha_1^k & \cdots & (-1)^{j-1}\alpha_{j-1}^k & k! & (-1)^{j+1}\alpha_{j+1}^k & \cdots & (-1)^n(-2)^k \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & -\alpha_1^n & \cdots & (-1)^{j-1}\alpha_{j-1}^n & 0 & (-1)^{j+1}\alpha_{j+1}^n & \cdots & (-1)^n(-2)^n \end{pmatrix}.$$

Taking the cofactor expansion along the replaced column, and factoring out -1 from each of the appropriate columns gives

$$k!(-1)^{\lfloor (n+1)/2 \rfloor + j} \cdot (-1)^{j+k} \begin{vmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 \\ 0 & \cdots & \alpha_{j-1} & \alpha_{j+1} & \cdots & -2 \\ \vdots & & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & \alpha_{j-1}^{k-1} & \alpha_{j+1}^{k-1} & \cdots & (-2)^{k-1} \\ 0 & \cdots & \alpha_{j-1}^{k+1} & \alpha_{j+1}^{k+1} & \cdots & (-2)^{k+1} \\ \vdots & & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & \alpha_{j-1}^n & \alpha_{j+1}^n & \cdots & (-2)^n \end{vmatrix}.$$

The matrix satisfies the conditions of Lemma 21, so we can write this as

$$k!(-1)^{\lfloor (n+1)/2 \rfloor + k} e_{n-k}(\alpha_0, \ldots, \alpha_{j-1}, \alpha_{j+1}, \ldots, \alpha_n) \prod_{\substack{0 \le i < i' \le n \\ i,i' \ne j}} (\alpha_{i'} - \alpha_i).$$

There are $\binom{n-1}{2}$ strictly negative terms in the product, and as long as $k > 0$, e_{n-k} has sign $(-1)^{n-k}$. So the sign of the whole product is $(-1)^{\lfloor (n+1)/2 \rfloor + n + \binom{n-1}{2}}$. Dividing by the sign of $\det B$, we get $(-1)^{n + \binom{n-1}{2} - \binom{n}{2}} = 1$. ∎

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 known 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, 5], as does the approximate degree of approximate majority (see [25, 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

---

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

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] S. Aaronson. The polynomial method in quantum and classical computing. In *Proc. of Foundations of Computer Science* (FOCS), page 3, 2008. Slides available at www.scottaaronson.com/talks/polymeth.ppt.

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

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

[4] R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bound by polynomials. *J. ACM*, 48(4): 778–797, 2001.

[5] P. Beame, W. Machmouchi. The quantum query complexity of AC^0. *Quantum Information & Computation* 12(7–8): 670–676, 2012.

[6] R. Beigel. The polynomial method in circuit complexity. In *Proc. of the Conference on Structure in Complexity Theory*, pages 82–95, 1993.

[7] R. Beigel. Perceptrons, PP, and the polynomial hierarchy. In *Computational Complexity*, 4: 339–349, 1994.

[8] S. N. Bernstein. On the V. A. Markov theorem. Trudy Leningr. Industr. In-ta, no 5, razdel fiz-matem nauk, 1 (1938).

[9] H. Buhrman, N. K. Vereshchagin, and R. de Wolf. On computation and communication with small bias. In *Proc. of the Conference on Computational Complexity* (CCC), pages 24–32, 2007.

[10] N. L. Carothers. A short course on approximation theory. Lecture notes available online at http://personal.bgsu.edu/ carother/Approx.html.

[11] A. Chattopadhyay and A. Ada. Multiparty communication complexity of disjointness. In *Electronic Colloquium on Computational Complexity* (ECCC), 08(002), 2008.

[12] D. Gavinsky and A. A. Sherstov. A separation of NP and coNP in multiparty communication complexity. *Theory of Computing*, 6(1): 227–245, 2010.

[13] E. R. Heineman. Generalized Vandermonde determinants. *Trans. of the AMS*, 31(3): 464–476, 1929.

[14] P. Høyer, M. Mosca, and R. de Wolf. Quantum search on bounded-error inputs. In *Proc. of International Colloquium on Automata, Languages, and Programming* (ICALP), pages 291–299, 2003.

[15] L. B. W. Jolley. *Summation of Series*. Dover Publications, second edition, 1961.

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

[17] H. Klauck, R. Špalek, and R. de Wolf. Quantum and classical strong direct product theorems and optimal time-space tradeoffs. SIAM J. Comput., 36(5): 1472–1493, 2007.

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

[19] A. R. Klivans and A. A. Sherstov. Lower bounds for agnostic learning via approximate rank. *Computational Complexity*, 19(4): 581–604, 2010.

[20] T. Lee and A. Shraibman. Disjointness is hard in the multi-party number-on-the-forehead model. In *Proc. of the Conference on Computational Complexity* (CCC), pages 81–91, 2008.

[21] A. Markov. On a question by D. I. Mendeleev. *Zapiski Imper. Akad. Nauk*, 62: 1–24, 1890.

[22] V. Markov. On functions which deviate least from zero in a given interval, St. Petersburg, 1892 (Russian).

[23] M. L. Minsky and S. A. Papert. Perceptions: An Introduction to Computational Geometry. MIT Press, Cambridge, MA., 1969.

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

[25] Open problems in analysis of Boolean functions. Compiled for the Simons Symposium, February 5–11, 2012. *CoRR*, abs/1204.6447, 2012.

[26] R. Paturi. On the degree of polynomials that approximate symmetric Boolean functions (Preliminary Version). In *Proc. of the Symposium on Theory of Computing* (STOC), pages 468–474, 1992.

[27] A. Schrijver. Theory of Linear and Integer Programming. John Wiley & Sons, New York, NY., 1986.

[28] R. A. Servedio, L.-Y. Tan, and J. Thaler. Attribute-efficient learning and weight-degree tradeoffs for polynomial threshold functions. *J. of Machine Learning Research - Proc. Track*, 23: 14.1-14.19, 2012.

[29] A. Shadrin. Twelve proofs of the Markov inequality, *Approximation Theory: A volume dedicated to Borislav Bojanov*, (D. K. Dimitrov et al, Eds), Marin Drinov Acad. Publ. House, Sofia, pages 233–299, 2004.

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

[31] A. A. Sherstov. Approximating the AND-OR tree. *Theory of Computing*, 9(20): 653–663, 2013.

[32] A. A. Sherstov. Communication lower bounds using dual polynomials. *Bulletin of the EATCS*, 95: 59–93, 2008.

[33] A. A. Sherstov. Making polynomials robust to noise. *Proc. of the Symposium on Theory of Computing* (STOC), pages 747–758, 2012.

[34] A. A. Sherstov. Separating AC^0 from depth-2 majority circuits. *SIAM J. Comput.*, 28(6): 2113-2129, 2009.

[35] A. A. Sherstov. The intersection of two halfspaces has high threshold degree. *Proc. of Foundations of Computer Science* (FOCS), pages 343–362, 2009. To appear in *SIAM J. Comput. (special issue for FOCS 2009)*

[36] A. A. Sherstov. The multiparty communication complexity of set disjointness. *Proc. of the Symposium on Theory of Computing* (STOC), pages 525–548, 2012.

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

[38] Y. Shi. Approximating linear restrictions of Boolean functions. Manuscript, 2002. Available online at `web.eecs.umich.edu/~shiyy/mypapers/linear02-j.ps`.

[39] Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation* 9(5): 444–460, 2009.

[40] R. Špalek. A dual polynomial for OR. Manuscript, 2008. Available online at: `http://arxiv.org/abs/0803.4516`

[41] Wolfram Research. The Wolfram functions site, August 2003. `http://functions.wolfram.com/01.11.23.0001.01`.

A   Final Details of Theorem 2

A.1   Proof of Proposition 5

Let $r = \lfloor 1/\alpha \rfloor$. Then

$$\mathbf{P}_y[F(z) \neq F(z_1 y_1, \ldots, z_M y_M)] \leq \mathbf{P}_y[F(z) \neq F(z_1 w_1, \ldots, z_M w_M) \text{ for some } w \preceq y] \qquad (19)$$

where $w \preceq y$ if $\{i : w_i = -1\} \subseteq \{i : y_i = -1\}$. By monotonicity, it suffices to bound the right hand side under the assumption that each bit of y takes the value -1 independently with probability exactly $1/r$.

Consider a matrix $Y \in \{-1, 1\}^{r \times M}$ where each column is chosen independently at random from the r vectors having a -1 in one slot and a $+1$ in all the others. Let $y^1, y^2, \ldots, y^r$ denote the rows of Y. While these rows are not independent, each is individually a random string whose ith bit

24

independently takes the value -1 with probability $1/r$. Thus the right-hand side of Expression (19) equals

$$\frac{1}{r}\sum_{j=1}^{r}\mathbf{P}_Y[F(z)\neq F(z_1w_1,\ldots,z_Mw_M) \text{ for some } w \preceq y^j]$$

$$=\frac{1}{r}\mathbf{E}_Y\left[\#\{j : F(z) \neq F(z_1w_1,\ldots,z_Mw_M) \text{ for some } w \preceq y^j\}\right]$$

The latter count has at most $\mathrm{bs}_z(F)$ nonzero terms because $y^1,\ldots,y^j$ are the characteristic vectors of disjoint sets. The asserted inequality follows because $1/r = 1/\lfloor 1/\alpha \rfloor \leq 2\alpha$. ∎

A.2   Proof of Equation 15

We prove that the polynomial ζ defined in Eq. (12) satisfies Eq. (15), reproduced here for convenience.

$$\sum_{(x_1,\ldots,x_M)\in(\{-1,1\}^N)^M}\zeta(x_1,\ldots,x_M)\chi_S(x_1,\ldots,x_M) = 0 \text{ for each } |S| \leq d \cdot d'. \qquad (15)$$

To prove Eq. (15), notice that since Ψ is orthogonal on $\{-1,1\}^M$ to all polynomials of degree at most d, we have the Fourier representation

$$\Psi(z) = \sum_{\substack{T\subseteq\{1,\ldots,M\}\\|T|>d}}\hat{\Psi}(T)\chi_T(z)$$

for some reals $\hat{\Psi}(T)$. We can thus write

$$\zeta(x_1,\ldots,x_M) = 2^M\sum_{|T|>d}\hat{\Psi}(T)\prod_{i\in T}\psi(x_i)\prod_{i\notin T}|\psi(x_i)|.$$

Given a subset $S \subseteq \{1,\ldots,M\} \times \{1,\ldots,N\}$ with $|S| \leq d \cdot d'$, partition $S = (\{1\} \times S_1) \cup \cdots \cup (\{M\} \times S_M)$ where each $S_i \subseteq \{1,\ldots,N\}$. Then

$$\sum_{(x_1,\ldots,x_M)\in(\{-1,1\}^N)^M}\zeta(x_1,\ldots,x_M)\chi_S(x_1,\ldots,x_M)$$

$$= 2^M\sum_{|T|>d}\hat{\Psi}(T)\prod_{i\in T}\underbrace{\left(\sum_{x_i\in\{-1,1\}^N}\psi(x_i)\chi_{S_i}(x_i)\right)}\prod_{i\notin T}\left(\sum_{x_i\in\{-1,1\}^N}|\psi(x_i)|\chi_{S_i}(x_i)\right).$$

Since $|S| \leq d \cdot d'$, by the pigeonhole principle, $|S_i| \leq d'$ for at least $M - d$ indices $i \in \{1,\ldots,M\}$. Thus for each set T, at least one of the underbraced factors is zero, as χ_{S_i} is orthogonal to ψ whenever $|S_i| \leq d'$.

B   A Dual Polynomial for the ε-Approximate Degree of OR

Špalek [40] constructed an explicit dual witness for the fact that the OR function on n variables has $(1/14)$-approximate degree $\Omega(\sqrt{n})$. We extend his argument to exhibit a dual witness that shows that OR has ε-approximate degree $\Omega(\sqrt{n})$ for *any* constant $\varepsilon \in (0,1)$.

Proposition 23 *Let $\varepsilon \in (0,1)$. Then OR_n has approximate degree $\Omega(\sqrt{n(1-\varepsilon)})$.*

Proof: As in Section 4, we associate with each symmetric function p a univariate function P and vice versa. Let $c = \max\{2, \lceil 8/(1-\varepsilon)\rceil\}$. Let $m = \lfloor\sqrt{n/c}\rfloor$ and define the set

$$T = \{1\} \cup \{ck^2 : 0 \le k \le m\}.$$

Note that $|T| = \Omega(\sqrt{n/c})$. Define the polynomial

$$P(x) = (-1)^s \frac{c^{2m}(m!)^2}{n!} \prod_{j \in [n]\setminus T} (x-j),$$

where s is a sign bit to be determined later. It is easy to check that $|P(0)| = 1$.

The ℓ_1 contribution due to the r'th layer of the Boolean hypercube is

$$\binom{n}{r}|P(r)| = \binom{n}{r}\frac{c^m(m!)^2}{n!} \frac{\prod_{j\in[n]\setminus\{r\}} |r-j|}{\prod_{j\in T\setminus\{r\}} |r-j|} = \frac{c^m(m!)^2}{\prod_{j\in T\setminus\{r\}} |r-j|}.$$

For $r = 1$ the right hand side evaluates to

$$\frac{c^m(m!)^2}{\prod_{i=1}^m (ci^2-1)} = \prod_{i=1}^m \frac{i^2}{i^2-1/c} \ge 1.$$

Thus, the total ℓ_1 contribution of the inputs of Hamming weight 1 is at least 1.

For $r = ck^2$ where $k > 0$, we get

$$\frac{c^m(m!)^2}{(ck^2-1)\prod_{i\in[m]\setminus\{k\}} |ci^2 - ck^2|} = \frac{(m!)^2}{(ck^2-1)\prod_{i\in[m]\setminus\{k\}} (i+k)|i-k|}$$

$$= \frac{2(m!)^2}{(ck^2-1)(m+k)!(m-k)!}$$

$$\le \frac{2}{ck^2-1}$$

where the last inequality follows as in [40] because

$$\frac{(m!)^2}{(m+k)!(m-k)!} = \frac{m}{m+k}\cdot\frac{m-1}{m+k-1}\cdot\ldots\cdot\frac{m-k+1}{m+1}$$

is a product of factors that are each smaller than 1. This shows that the total ℓ_1 contribution of the Hamming layers excluding 0 and 1 is at most

$$\sum_{k=1}^m \frac{2}{ck^2-1} < \sum_{k=1}^\infty \frac{4}{ck^2} < \frac{8}{c}.$$

For the final part of our construction, we let $Q(i) = (-1)^i P(i)$. Then the multilinear polynomial corresponding to Q has pure high degree $\Omega(\sqrt{n(1-\varepsilon)})$. Since $P(0)$ and $P(1)$ have the same sign, $Q(0)$ and $Q(1)$ have opposite signs. Since $\mathrm{OR}(0)$ and $\mathrm{OR}(1)$ also have opposite signs, we can choose $s \in \{-1, 1\}$ to ensure that

$$Q \cdot \mathrm{OR} \geq \binom{n}{0}|P(0)| + \binom{n}{1}|P(1)| - \sum_{j \in S \setminus \{0,1\}} \binom{n}{j}|P(j)|$$

$$\geq 1 + n|P(1)| - \frac{8}{c}.$$

As the total ℓ_1 mass $\|Q\|_1$ of Q is at most $1 + n|P(1)| + 8/c$, we see that

$$\frac{Q \cdot \mathrm{OR}}{\|Q\|_1} \geq \frac{1 + n|P(1)| - 8/c}{1 + n|P(1)| + 8/c} = 1 - \frac{16}{c + cn|P(1)| + 8} \geq 1 - \frac{16}{2c + 8}.$$

Since $c > 8/(1-\varepsilon)$, the right hand side is at least ε. Corollary 7 then implies that the ε-approximate degree of OR is $\Omega(\sqrt{n(1-\varepsilon)})$. ∎

C Dual Polynomials for Symmetric Functions

Proof of Lemma 11: Fix an ℓ such that $\ell \in [\lfloor ct \rfloor]$ and let $i(k) = tk^2 + 4\ell$. It is enough to show that $\prod_{i' \in S, i' \neq i} |i - i'|$ is minimized at $k = 1$. We can expand this product as

$$\prod_{\substack{i' \in S \\ i' \neq i}} |i - i'| = \prod_{m=0}^{\lfloor ct \rfloor} \left((tk^2 + 4\ell - (t - 4m)) \prod_{\substack{j=1 \\ j \neq k}}^{\lfloor \sqrt{(n-t+1)/t} \rfloor} |tk^2 + 4\ell - (tj^2 + 4m)| \right) \times \prod_{\substack{m=0 \\ m \neq \ell}}^{\lfloor ct \rfloor} |4\ell - 4m|.$$

Cancelling the factor independent of k and considering each index m separately, we just need to show that for any fixed $0 \leq \ell, m \leq ct$, the product

$$(tk^2 + 4\ell - (t - 4m)) \prod_{j \neq k} |tk^2 + 4\ell - (tj^2 + 4m)|$$

as a function of $k \geq 1$ is minimized at $k = 1$. Divide each factor by t to obtain

$$\left(k^2 - 1 + \frac{4(\ell + m)}{t} \right) \prod_{j \neq k} |k^2 - j^2| \left(1 + \frac{4(\ell - m)}{t(k^2 - j^2)} \right). \tag{20}$$

We first obtain a lower bound for this expression when $k \geq 2$. Consider the following two facts.

Fact 24 *Let $k \leq m$ be nonnegative integers. Then*

$$\prod_{\substack{j \in [m] \\ j \neq k}} |k^2 - j^2| \geq \prod_{\substack{j \in [m] \\ j \neq 1}} |1 - j^2|. \tag{21}$$

In other words, this product of differences of squares is minimized at $k = 1$.

27

Proof of Fact 24 This is clear if $k = 0$, so suppose $k \geq 2$. Then the left-hand side of Expression (21) can be written as

$$\prod_{\substack{j \in [m] \\ j \neq k}} |k^2 - j^2| = \prod_{\substack{j \in [m] \\ j \neq k}} (k+j)|k-j| = \frac{(m+k)!}{2k(k-1)!} \cdot k!(m-k)! = \frac{1}{2}(m+k)!(m-k)!.$$

Taking the ratio of the left-hand side of Expression (21) to the right gives us

$$\frac{(m+k)!(m-k)!}{(m+1)!(m-1)!} = \frac{(m+k)(m+k-1)\ldots(m+2)}{(m-1)(m-2)\ldots(m-k+1)}$$

which is a product of numbers that are all at least 1. ∎

Fact 25 *Let k be a nonnegative integer. Then*

$$\sum_{\substack{j \in \mathbb{Z} \\ j \neq k}} \frac{1}{|j^2 - k^2|} \leq \frac{\pi^2}{3}.$$

Proof of Fact 25 First suppose $j > k$. Then

$$j^2 - k^2 = (j-k)^2 + 2jk - 2k^2 > (j-k)^2.$$

Thus

$$\sum_{j>k} \frac{1}{j^2 - k^2} < \sum_{j>k} \frac{1}{(j-k)^2} = \frac{\pi^2}{6}.$$

A similar argument holds for $j < k$. ∎

Combining the two facts, Expression (20) is at least

$$(k^2 - 1)\prod_{j \neq k}\left(1 - \frac{4c}{|k^2 - j^2|}\right)\prod_{j \neq 1}|1 - j^2| \geq \left(1 - \sum_{j \neq k}\frac{4c}{|k^2 - j^2|}\right)\prod_{j \neq 1}|1 - j^2|$$

$$\geq \left(1 - \frac{4c\pi^2}{3}\right)\prod_{j \neq 1}|1 - j^2|$$

whenever $k \geq 2$. On the other hand, setting $k = 1$ in Expression 20 gives us at most

$$8c\prod_{j \neq 1}|1 - j^2|\left(1 + \frac{4c}{|1 - j^2|}\right) \leq 8c\exp\left(\sum_{j \neq 1}\frac{4c}{|1 - j^2|}\right)\prod_{j \neq 1}|1 - j^2|$$

$$\leq 8c\exp\left(\frac{4c\pi^2}{3}\right)\prod_{j \neq 1}|1 - j^2|$$

28

which is easily verified to be smaller than our lower bound for the $k \geq 2$ case if $c \leq 1/32$. ■

Proof of Lemma 12: Write

$$|P(r)| = \frac{\pi_S(i^*)}{n!} \frac{\prod_{j \in [n] \setminus \{r\}} |r - j|}{|r - (i^* - 1)||r - (i^* + 1)| \prod_{j \in S \setminus \{r\}} |r - j|}$$

$$\leq \frac{1}{n!} \frac{r!(n-r)!}{|r - (i^* - 1)||r - (i^* + 1)|} \qquad \text{by definition of } i^*$$

$$\leq \frac{1}{\binom{n}{r}} \frac{1}{t^2(k^2 - (1 + 4c + 1/t))^2}.$$

The bound follows since $c \leq 1/32$ and $t \geq 2$. The calculation for v is similar. ■

Proof of Lemma 13: Using the bounds from the previous lemma, as well as the facts that $c \leq 1/32$ and $t \geq 2$, the left hand side is at most

$$\sum_{\ell \neq 0} \frac{1}{16\ell^2 - 1} + \sum_{k \geq 2} \sum_{\ell=0}^{\lfloor ct \rfloor} \frac{1}{t^2(k^2 - 2)^2} \leq 2 \sum_{\ell=1}^{\infty} \frac{1}{15\ell^2} + \frac{ct+1}{t^2} \sum_{k=2}^{\infty} \frac{1}{(k^2 - 2)^2}$$

$$\leq \frac{\pi^2}{45} + \frac{17}{64} \sum_{k=2}^{\infty} \frac{4}{k^4}$$

$$= \frac{\pi^2}{45} + \frac{17}{16} \left(\frac{\pi^4}{90} - 1 \right)$$

$$\leq \frac{2}{5}.$$

■

Proof of Proposition 9: By symmetry, we can assume that $t \leq n/2$. Moreover, we may assume that t is the largest such integer with $F(t-1) \neq F(t)$. We have already handled a few special cases: The case of $t = 1$ corresponds to Špalek's construction for the OR function [40], and the case of $t = \Omega(n)$ follows from Proposition 23. We can therefore assume that $2 \leq t \leq n/4$. We now consider three separate cases based on which of the terms $\binom{n}{i^*-1}|P(i^* - 1)|, 1, \binom{n}{i^*+1}|P(i^* + 1)|$ is the smallest. In all of these cases, we will show that we can construct a polynomial Q such that $(Q \cdot F)/\|Q\|_1 \geq 1/14$.

Case 1: $\binom{n}{i^*-1}|P(i^* - 1)|, 1 \geq \binom{n}{i^*+1}|P(i^* + 1)|$.

Recall that by translating S by at most $4ct$ (thereby keeping it a subset of $[n]$), we can assume that $i^* = t$. Let $Q(i) = (-1)^i P(i)$. Then the multilinear polynomial associated to Q has pure high degree $|T| = \Omega(\sqrt{t(n - t + 1)})$. The ℓ_1 norm of Q is

$$\|Q\|_1 = \sum_{i \in T} \binom{n}{i} |Q(i)|$$

$$\leq \binom{n}{t-1} |Q(t-1)| + 1 + \binom{n}{t+1} |Q(t+1)| + \frac{2}{5} \qquad \text{by Lemma 13}$$

$$\leq 2 \binom{n}{t-1} |Q(t-1)| + \frac{7}{5}.$$

29

Choose the sign bit s in the definition of P so that $Q(t) = F(t)$. Since $P(t-1)$ has the same sign as $P(t)$, it holds that $\widetilde{\operatorname{sgn}}(Q(t-1)) = \widetilde{\operatorname{sgn}}(F(t-1))$. Therefore,

$$Q \cdot F = \binom{n}{t-1}|Q(t-1)| + 1 + \sum_{i \in T \setminus \{t-1,t\}} \binom{n}{i} F(i)Q(i)$$

$$\geq \binom{n}{t-1}|Q(t-1)| + 1 - \binom{n}{t+1}|Q(t+1)| - \frac{2}{5}$$

$$\geq \frac{1}{2}\binom{n}{t-1}|Q(t-1)| + \frac{1}{2} - \frac{2}{5}$$

$$= \frac{1}{2}\binom{n}{t-1}|Q(t-1)| + \frac{1}{10}.$$

Using the fact that $(A+B)/(C+D) \geq \min(A/C, B/D)$ for positive A, B, C, D,

$$\frac{Q \cdot F}{\|Q\|_1} \geq \frac{1}{14}.$$

Case 2: $1, \binom{n}{i^*+1}|P(i^*+1)| \geq \binom{n}{i^*-1}|P(i^*-1)|$.

This time, translate S so that $i^* = t - 1$. We remark that under this translation we still have $T \subseteq [n]$, since we assumed $t \geq 2$. The remainder of the analysis is identical to Case 1, interchanging the roles of $t-1$ and $t+1$.

Case 3: $\binom{n}{i^*-1}|P(i^*-1)|, \binom{n}{i^*+1}|P(i^*+1)| \geq 1$.

Translate S so that $i^* = t + 1$, and choose s so that $Q(t) = (-1)^{i^*-1}P(i^*-1) = F(t)$. Observe that $F(t+2) = F(t)$, since we chose $t \leq n/4$ to be the largest such integer with $F(t-1) \neq F(t)$. Then $Q(t+2) = (-1)^{i^*+1}P(i^*+1)$ has the same sign as $F(t+2)$. The ℓ_1 norm calculation follows as in Case 1 to give

$$\|Q\|_1 \leq \binom{n}{t}|Q(t)| + \binom{n}{t+2}|Q(t+2)| + \frac{7}{5} \leq 2\binom{n}{t}|Q(t)| + 2\binom{n}{t+2}|Q(t+2)|.$$

The correlation with F is

$$Q \cdot F = \binom{n}{t}|Q(t)| + \binom{n}{t+2}|Q(t+2)| + \sum_{i \in T \setminus \{t,t+2\}} \binom{n}{i} F(i)Q(i)$$

$$\geq \binom{n}{t}|Q(t)| + \binom{n}{t+2}|Q(t+2)| - \frac{7}{5}$$

$$\geq \frac{1}{2}\binom{n}{t}|Q(t)| + \frac{1}{2}\binom{n}{t+2}|Q(t+2)|,$$

so $(Q \cdot F)/\|Q\|_1 \geq 1/4$. ∎

D Index of Trigonometric Identities

Lemma 26 ([15, No. 429])

$$\sum_{j=0}^{n}(-1)^j \cos(j\theta) = \frac{1}{2} + (-1)^n \frac{\cos((n+1/2)\theta)}{2\cos(\theta/2)}.$$

Lemma 27 *Let $i < n$ be odd natural numbers. Then*

$$\sum_{j=0}^{n}(-1)^j \cos^i\left(\frac{j\pi}{n}\right) = 1.$$

Proof: For odd i, consider the well-known power reduction formula

$$\cos^i \theta = 2^{1-i} \sum_{k=0}^{(i-1)/2} \binom{i}{k} \cos((i-2k)\theta).$$

Letting $\theta = \pi/n$ and applying the previous lemma,

$$\sum_{j=0}^{n}(-1)^j \cos^i\left(\frac{j\pi}{n}\right) = 2^{1-i} \sum_{k=0}^{(i-1)/2} \binom{i}{k} \sum_{j=0}^{n}(-1)^j \cos\left(\frac{(i-2k)j\pi}{n}\right)$$

$$= 2^{1-i} \sum_{k=0}^{(i-1)/2} \binom{i}{k}\left(\frac{1}{2} + (-1)^n \frac{\cos((i-2k)\pi/2n + (i-2k)\pi)}{2\cos((i-2k)\pi/2n)}\right).$$

$$= 2^{1-i} \sum_{k=0}^{(i-1)/2} \binom{i}{k} = 1.$$

∎

Lemma 28 *Let $2i < n$. Then*

$$\sum_{j=0}^{n}(-1)^j \sin^{2i}\left(\frac{j\pi}{2n}\right) = \frac{1}{2}(-1)^n.$$

Proof: Consider the power reduction formula

$$\sin^{2i}(\theta) = 2^{-2i}\binom{2i}{i} + 2^{1-2i}\sum_{k=0}^{i-1}(-1)^{i-k}\binom{2i}{k}\cos((2i-2k)\theta).$$

Let $\theta = \pi/2n$ and apply Lemma 26:

$$\sum_{j=0}^{n}(-1)^j \sin^{2i}\left(\frac{j\pi}{2n}\right) = \frac{1}{2}(1+(-1)^n)2^{-2i}\binom{2i}{i} + 2^{1-2i}\sum_{k=0}^{i-1}(-1)^{i-k}\binom{2i}{k}\sum_{j=0}^{n}(-1)^j\cos\left(\frac{(i-k)j\pi}{n}\right)$$

$$= (1+(-1)^n)2^{-1-2i}\binom{2i}{i} + 2^{1-2i}\sum_{k=0}^{i-1}(-1)^{i-k}\binom{2i}{k}\left(\frac{1}{2}+(-1)^n\frac{\cos((i-k)\pi/2n+(i-k)\pi)}{2\cos((i-k)\pi/2n)}\right)$$

$$= (1+(-1)^n)2^{-1-2i}\binom{2i}{i} + 2^{1-2i}\sum_{k=0}^{i-1}(-1)^{i-k}\binom{2i}{k}\left(\frac{1}{2}+\frac{1}{2}(-1)^{n+i-k}\right)$$

$$= (1+(-1)^n)2^{-1-2i}\binom{2i}{i} + 2^{-2i}\sum_{k=0}^{i-1}(-1)^{i-k}\binom{2i}{k} + (-1)^n 2^{-2i}\sum_{k=0}^{i-1}\binom{2i}{k}$$

Using the identity

$$\sum_{k=0}^{2i}(-1)^k\binom{2i}{k} = 0$$

and the symmetry of the binomial coefficients, the first sum evaluates to $-\frac{1}{2}\binom{2i}{i}$. Therefore,

$$\sum_{j=0}^{n}(-1)^j\sin^{2i}\left(\frac{j\pi}{2n}\right) = (1+(-1)^n)2^{-1-2i}\binom{2i}{i} - 2^{-1-2i}\binom{2i}{i} + (-1)^n 2^{-2i}\left(2^{2i-1} - \frac{1}{2}\binom{2i}{i}\right)$$

$$= \frac{1}{2}(-1)^n.$$

$\blacksquare$

Lemma 29 *Let $n = 2m+1$ be odd. Then*

$$\sum_{k=0}^{n}(-1)^k\sec\left(\frac{k\pi}{n}\right) = (-1)^m n + 1.$$

Proof: This follows from the identity [41]

$$\sum_{k=0}^{m}\sec\left(\frac{2k\pi}{2m+1}\right) = \frac{1}{2}(-1)^m(2m+1) + \frac{1}{2},$$

and the observation that $\sec(2k\pi/n) = -\sec((n-2k)\pi/n)$.

$\blacksquare$

Lemma 30 *Let n be odd. Then*

$$\sum_{k=0}^{n-1}\sec^2\left(\frac{k\pi}{n}\right) = n^2.$$

Proof: We start with the identity [15, No. 445]

$$\sum_{k=0}^{n-1} \tan^2\left(\theta + \frac{k\pi}{n}\right) = n^2 \cot\left(\frac{n\pi}{2} + n\theta\right) + n(n-1).$$

Letting $\theta = 0$, this evaluates to $n(n-1)$ as long as n is odd. Substituting $\tan^2(k\pi/n) = \sec^2(k\pi/n) - 1$ into the left-hand side gives the identity. ∎

Lemma 31 ([15, No. 440])

$$\sum_{j=1}^{n-1} \csc^2\left(\frac{j\pi}{2n}\right) = \frac{4n^2 - 4}{6}.$$

Lemma 32

$$\sum_{j=1}^{n-1} (-1)^j \csc^2\left(\frac{j\pi}{2n}\right) = -\frac{n^2}{3} - \frac{1}{6} - \frac{1}{2}(-1)^n.$$

Proof: Consider the identity [15, Nos. 441, 442]

$$\sum_{\substack{j=1 \\ j \text{ odd}}}^{n-1} \csc^2\left(\frac{j\pi}{2n}\right) = \frac{n^2}{2} + \frac{1}{4}((-1)^n - 1).$$

Let $\theta = \pi/2n$ and subtract two copies of the second identity from the identity in . Then we get

$$\sum_{j=1}^{n-1} (-1)^j \csc^2\left(\frac{j\pi}{2n}\right) = \frac{4n^2 - 4}{6} - n^2 - \frac{1}{2}((-1)^n - 1) = -\frac{n^2}{3} - \frac{1}{6} - \frac{1}{2}(-1)^n.$$

∎