Electronic Colloquium on Computational Complexity, Report No. 197 (2018)

Approximate degree of AND via Fourier analysis

Andrej Bogdanov*

Abstract

We give a new proof that the approximate degree of the AND function over n inputs is $\Omega(\sqrt{n})$. Our proof extends to the notion of weighted degree, resolving a conjecture of Kamath and Vasudevan. As a consequence we confirm that the approximate degree of any read-once depth-2 De Morgan formula is the square root of the formula size up to constant. This generalizes a result of Sherstov (TOC 2013) and Bun and Thaler (Inf. Comput. 2015).

1 Introduction

The approximate degree of a Boolean function [NS94] is the smallest degree of a real-valued polynomial that approximates it pointwise. Nisan and Szegedy showed that it is polynomially related to a host of complexity measures including exact multilinear degree, sensitivity, deterministic query complexity, and randomized query complexity. Beals et al. [BBC+01] added quantum query complexity to the list, initiating a fruitful framework for proving optimality of various quantum algorithms. More recent works study approximate degree as a complexity measure in its own right, with focus on "low" complexity classes like symmetric functions, small De Morgan formulas, and bounded-depth AND-OR circuits [Pat92, She13, She18, BT15, BT17, BKT18].

Many of these works rely on Nisan and Szegedy's $\Omega(\sqrt{n})$ lower bound for the approximate degree of the n-bit AND function. Their proof employs a symmetrization argument [MP69], reducing the problem to a question about univariate polynomial approximation over the reals to which tools from approximation theory are applied.

We give a new proof of this via Fourier analysis over the Boolean cube. Our proof generalizes to the notion of *weighted degree* in which different variables make different contributions to the degree of the approximating polynomial. In contrast, it is unclear if symmetrization arguments can be extended to the weighted setting.

Kamath and Vasudevan [KV14] conjectured our main result and showed it implies that the approximate degree of any depth-two read-once De Morgan formula of size N is $\Theta(\sqrt{N})$. Previously Sherstov [She13] and Bun and Thaler [BT15] showed that this is true under the

*`andrejb@cse.cuhk.edu.hk`. Department of Computer Science and Engineering and Institute of Theoretical Computer Science and Communications, Chinese University of Hong Kong.

ISSN 1433-8092

assumption the formula is regular, that is all terms have the same number of variables. While the degree of any size-N read-once AND-OR formulas is upper bounded by $O(\sqrt{N})$ [Rei11], to the best of our knowledge a matching lower bound is known for only one other instance—the complete binary AND-OR tree.

Notation and definitions. We use $\langle w, x\rangle = \sum w_i x_i$ for inner product over the reals, $\|w\|_p = (\sum w_i^p)^{1/p}$ for the p-norm of a vector/function, and $\ominus$ for symmetric set difference. For a vector w of n non-negative weights, the *weighted degree* $\deg_w p$ of a function $p\colon \{-1,1\}^n \to \mathbb{R}$ is the maximum value of $w(S) = \sum_{i\in S} w_i$ taken over all monomials $\chi_S(x) = \prod_{i\in S} x_i$ that appear in the Fourier expansion of p, namely

$$\deg_w p = \max\{w(S)\colon \hat{p}(S) \neq 0\}, \qquad \text{where } p = \sum_{S\subseteq[n]} \hat{p}(S)\cdot \chi_S.$$

2 Main Theorem

Let $\delta\colon \{-1,1\}^n \to \{0,1\}$ be the point function

$$\delta(x) = \begin{cases} 1, & \text{if } x = 1^n, \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 1. *For every $p\colon \{-1,1\}^n \to \mathbb{R}$,*

$$\|p - \delta\|_\infty \geq \sqrt{\frac{\Pr_{x\sim\{-1,1\}^n}[\langle w, x\rangle > \deg_w p]}{2}}.$$

Claim 2. $\deg_w p \cdot q \leq \deg_w p + \deg_w q$.

Proof. It suffices to prove the claim when p and q are monomials, i.e., $p = \chi_S$ and $q = \chi_T$. Then

$$\deg_w \chi_S \cdot \chi_T = \deg_w \chi_{S\ominus T} = w(S\ominus T) \leq w(S) + w(T) = \deg_w \chi_S + \deg_w \chi_T. \qquad \square$$

Claim 3. *There exists a probability mass function $\mathcal{D}$ over $\{-1,1\}^n$ such that (1) $\mathcal{D}$ is the square of a polynomial d of weighted degree at most $\|w\|_1/2$ and (2) $\mathcal{D}(1^n) \geq 1/2$.*

Proof. Let $\mathcal{H}$ be the set of subsets S of $\{1,\ldots,n\}$ of weight at most $\|w\|_1/2$ and

$$\mathcal{D}(x) = \frac{1}{Z}\,\mathbb{E}_{S\sim\mathcal{H}}[\chi_S(x)]^2,$$

where the choice of S is uniform over $\mathcal{H}$, and Z is a normalizing constant. Property (1) holds by definition. To verify property (2), observe that the expectation evaluates to 1 when $x = 1^n$. It remains to verify that $Z \leq 2$. Since $Z = \sum_{x\in\{-1,1\}^n} \mathbb{E}_S[\chi_S(x)]^2$, we can write

$$Z = \sum_{x\in\{-1,1\}^n} \mathbb{E}_{S,T}[\chi_S(x)\chi_T(x)] = \sum_{x\in\{-1,1\}^n} \mathbb{E}_{S,T}[\chi_{S\ominus T}(x)] = \mathbb{E}_{S,T} \sum_{x\in\{-1,1\}^n} \chi_{S\ominus T}(x).$$

For fixed S and T, the value of the last sum is 2^n when $S = T$ and zero otherwise. Therefore

$$Z = 2^n \cdot \Pr_{S,T \sim \mathcal{H}}[S = T].$$

The set $\mathcal{H}$ contains at least half the subsets of $\{1, \ldots, n\}$ because $w(S) + w(\overline{S}) = \|w\|_1$, so at least one among every complementing pair must be in $\mathcal{H}$. Therefore the collision probability of $\mathcal{H}$ is at least 2^{-n+1}, and $Z \leq 2$ as desired. $\square$

Proof of Theorem 1. Let $\mathcal{D}\colon \{-1,1\}^n \to \mathbb{R}$ be the distribution from Claim 3. Then

$$\begin{aligned}
\|p - \delta\|_\infty^2 &\geq \mathbb{E}_{x \sim \mathcal{D}}\big[(p(x) - \delta(x))^2\big] \\
&= \sum_{x \in \{-1,1\}^n} \mathcal{D}(x) \cdot \big(p(x) - \delta(x)\big)^2 \\
&= \sum_{x \in \{-1,1\}^n} \big(d(x)p(x) - d(x)\delta(x)\big)^2 \\
&= \sum_{x \in \{-1,1\}^n} \big(d(x)p(x) - d(1^n)\delta(x)\big)^2 \\
&\geq \frac{1}{2} \sum_{x \in \{-1,1\}^n} \left(\frac{d(x)p(x)}{d(1^n)} - \delta(x)\right)^2,
\end{aligned}$$

The last inequality follows from part (1) of Claim 3. Let $q = d \cdot p / d(1^n)$. By Parseval's identity

$$\sum_{x \in \{-1,1\}^n} \big(q(x) - \delta(x)\big)^2 = 2^n \sum_{T \subseteq [n]} \big(\hat{q}(T) - \hat{\delta}(T)\big)^2 \geq 2^n \sum_{T\colon w(T) > \deg_w q} \hat{\delta}(T)^2,$$

because q has no coefficients of weight exceeding $\deg_w q$. The Fourier transform of δ is $\hat{\delta}(T) = 2^{-n}$ for all T so

$$\begin{aligned}
\|p - \delta\|_\infty^2 &\geq \frac{1}{2} \cdot 2^n \cdot \sum_{T\colon w(T) > \deg_w q} 2^{-2n} \\
&= \frac{1}{2} \Pr_{\text{random } T \subseteq [n]}[w(T) > \deg_w q] \\
&= \frac{1}{2} \Pr_{x \sim \{-1,1\}^n}[\langle w, x \rangle + \|w\|_1/2 > \deg_w q]. \qquad (1)
\end{aligned}$$

By Claim 2 and part (2) of Claim 3,

$$\deg_w q \leq \deg_w d + \deg_w p \leq \frac{\|w\|_1}{2} + \deg_w p.$$

Plugging $\deg_w q$ into (1) and simplifying gives the desired inequality. $\square$

Consequences

When $w_1 = \dots w_n = 1$ the weighted degree is the standard polynomial degree, and we recover the Nisan-Szegedy lower bound on the approximate degree of the AND function.

Corollary 4. *For every degree-d polynomial p,*

$$\|p - \delta\|_\infty \geq \sqrt{\frac{1}{2^{n+1}} \sum_{t<n/2-d} \binom{n}{t}}.$$

The expression on the right is lower bounded by the larger of $1/2 - O(d/\sqrt{n})$ and $2^{-O(d^2/n)}$. In the large d regime, this matches the best-known lower bound asymptotically and is tight up to polylogarithmic factors in the exponent [KLS96]. For small d the correct bound is $1/2 - \Theta(d^2/n)$ [BT15], so Corollary 4 is not tight.

The second corollary is a tight lower bound on the weighted approximate degree of AND. The tightness up to constant follows from a quantum algorithm of Ambainis [Amb10].

Corollary 5. *For every w and p, if $\deg_w p \leq \sqrt{1-\varepsilon} \cdot \|w\|_2$ then $\|p - \delta\|_\infty = \Omega(\varepsilon)$.*

Proof. Let $X = w_1 x_1 + \cdots + w_n x_n$ where $x \sim \{-1, 1\}^n$ is uniform over the Boolean cube. Then $\mathbb{E}[X^2] = \|w\|_2^2$ and $\mathbb{E}[X^4] = \sum w_i^4 + 3 \sum w_i^2 w_j^2 \leq 3\,\mathbb{E}[X^2]^2$. By the Paley-Zygmund inequality,

$$\Pr\bigl[|X| > \sqrt{1-\varepsilon} \cdot \|w\|_2\bigr] = \Pr\bigl[X^2 > (1-\varepsilon) \cdot \|w\|_2^2\bigr] \geq \frac{\varepsilon^2}{3}.$$

Since X is a symmetric random variable, X exceeds $\sqrt{1-\varepsilon} \cdot \|w\|_2$ with probability at least $\varepsilon^2/6$. Plugging into Theorem 1 we obtain that $\|p - \delta\|_\infty \geq \varepsilon/\sqrt{12}$. □

3 Approximate degree of depth-two read-once De Morgan formulas

Kamath and Vasudevan [KV14] showed that Corollary 5 implies the following lower bound on functions of the form $f(x) = AND(OR(x_1), \dots, OR(x_n))$, where the OR terms are disjoint.

Theorem 6. *There is a universal constant c such that for every p of degree at most $c\sqrt{N}$, $\|p - f\|_\infty \geq 1/3$, where N is the number of variables in f.*

Sherstov [She13] and Bun and Thaler [BT15] proved this under the restrictive assumption that the formula is regular, namely $x_1, \dots, x_n$ are of equal size. Kamath and Vasudevan showed that the result for regular formulas implies a weaker bound of $\Omega(\sqrt{N/\log N})$ for the general case.

We give the proof of Theorem 6 for completeness. Two distributions over $\{0, 1\}^n$ are indistinguishable by $S \subseteq [n]$ if their projections on S are identical.

Proof. By the duality between polynomial approximation and bounded indistinguishability [BIVW16] and standard amplification of distinguishing advantage, Corollaries 4 and 5 imply the following for sufficiently small constants c_1, c_2.

1. For every w there exists a pair of distributions ν_0, ν_1 over $\{0,1\}^n$ that are indistinguishable by any subset of weight at most $c_2\|w\|_2$, but $\mathbb{E}_{Y\sim\nu_0}[AND(Y)] = 0$ and $\mathbb{E}_{Y\sim\nu_1}[AND(Y)] \geq 2/3$.

2. For every m there exists a pair of distributions μ_0, μ_1 over $\{0,1\}^m$ that are indistinguishable by any subset of size at most $c_1\sqrt{m}$, but $\mathbb{E}_{X\sim\mu_0}[OR(X)] \leq 1/3$ and $\mathbb{E}_{X\sim\mu_1}[OR(X)] = 1$.

Let m_t be the size of the t-th term of f and set $w_t = \sqrt{m_t}$. Given $b \in \{0,1\}$ sample $X^b \in \{0,1\}^N$ as follows. First sample $Y \in \{0,1\}^n$ from ν_b. Then for each bit Y_t, sample $X_t \in \{0,1\}^{m_t}$ from μ_{Y_t} with length parameter $m = m_t$. Set $X^b = X_0X_1\dots X_n$.

First we argue that X^0 and X^1 are distinguishable by f. When $b = 0$, there always exists a term t for which $Y_t = 0$, so $\Pr[OR(X_t) = 1] \leq 1/3$. Therefore $\Pr[f(X^0) = 0] \leq 1/3$. When $b = 1$, with probability at least $2/3$ all bits of Y are ones, in which case $f(X^1)$ evaluates to 1. Therefore $\Pr[f(X^1) = 1] \geq 2/3$. It follows that $\mathbb{E}[f(X^1)] - \mathbb{E}[f(X^0)] \geq 1/3$.

Next we argue that X^0 and X^1 are indistinguishable by any subset S of $c\sqrt{N}$ inputs. Let $T \subseteq [n]$ be the set of terms t that intersect S in more than $c_1\sqrt{m_t}$ variables. Then the weight of T is at most

$$w(T) = \sum_{t\in T} w_t = \sum_{t\in T} \sqrt{m_t} < \frac{|S|}{c_1}.$$

On the other hand, if X^0 and X^1 are distinguishable by S, then T must have weight more than $c_2\|w\|_2 = c_2\sqrt{N}$. It follows that $|S| > c_1c_2\sqrt{N}$.

In conclusion, X^0 and X^1 are indistinguishable by any subset of size $c_1c_2\sqrt{N}$, but are distinguishable by f with advantage 1/3. By duality, the 1/3-approximate degree of f is at least $c_1c_2\sqrt{N}$. □

References

[Amb10] Andris Ambainis. Quantum search with variable times. *Theory Comput. Syst.*, 47(3):786–807, 2010.

[BBC+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *J. ACM*, 48(4):778–797, July 2001.

[BIVW16] Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. Bounded indistinguishability and the complexity of recovering secrets. In *Advances in Cryptology - CRYPTO 2016 - 36th Annual International Cryptology Conference*, pages 593–618, 2016.

[BKT18] Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: tight quantum query bounds via dual polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 297–310, 2018.

[BT15] Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and Markov-Bernstein inequalities. *Inf. Comput.*, 243(C):2–25, August 2015.

[BT17] Mark Bun and Justin Thaler. A nearly optimal lower bound on the approximate degree of AC^0. In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 1–12, 2017.

[KLS96] Jeff Kahn, Nathan Linial, and Alex Samorodnitsky. Inclusion-exclusion: Exact and approximate. *Combinatorica*, 16(4):465–477, Dec 1996.

[KV14] Pritish Kamath and Prashant Vasudevan. Approximate degree of AND-OR trees, 2014. Manuscript available at `https://www.scottaaronson.com/showcase3/kamath-pritish-vasudevan-prashant.pdf`.

[MP69] Marvin Minsky and Seymour Papert. *Perceptrons*. MIT Press, Cambridge, MA, 1969.

[NS94] Noam Nisan and Mario Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.

[Pat92] Ramamohan Paturi. On the degree of polynomials that approximate symmetric boolean functions. In *Symposium on the Theory of Computing (STOC)*, pages 468–474, 1992.

[Rei11] Ben W. Reichardt. Reflections for quantum query algorithms. In *Proceedings of the Twenty-second Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '11, pages 560–569, Philadelphia, PA, USA, 2011. Society for Industrial and Applied Mathematics.

[She13] Alexander A. Sherstov. Approximating the AND-OR tree. *Theory of Computing*, 9:653–663, 2013.

[She18] Alexander A. Sherstov. Algorithmic polynomials. In *Symposium on the Theory of Computing (STOC)*, pages 311–324. ACM, 2018.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il