Approximate Degree of AND-OR trees

Pritish Kamath
pritish@mit.edu

Prashant Vasudevan
prashvas@mit.edu

December 12, 2014

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 with error ε for every $\mathbf{x} \in \{1, -1\}^n$. In this project, we explored questions related to approximate degree of AND-OR trees. In particular, we started with the broad goal of addressing the following conjecture (due to Aaronson [Aar14]) : Every read-once AND-OR formula with N leaves has approximate degree $\Omega(\sqrt{N})$.

We present some partial progress towards the above conjecture for constant depth non-regular AND-OR trees.

1 Introduction

Approximate degree is an important measure of the complexity of a boolean function. It captures whether a function can be point-wise approximated by low-degree polynomials with real coefficients. In quantum computing, the measure of approximate degree has been used to prove tight lower bounds on quantum query complexity of several functions. Lower bounds on approximate degree have a lot of applications, related to circuit lower bounds and communication complexity. Also, upper bounds on approximate degree of functions underly the best known algorithms for PAC learning DNFs and read-once formulas.

AND-OR trees is a very general class of functions, which form the skeleton of formulas. Approximate degree of AND-OR trees have received a lot of interest and after several years of developments, tight lower bounds on approximate degree of certain special sub-class of AND-OR trees are known. From a beautiful result of Reichardt [Rei11], it is known that the quantum query complexity of any AND-OR tree with N leaves is $\Theta(\sqrt{N})$. This immediately gives an upper bound of $O(N^{1/2})$ on the approximate degree of any AND-OR tree. Aaronson [Aar14] conjectured that this bound might be optimal.

Prior work on lower bounding the approximate degree of AND-OR trees has shown tight lower bounds for "regular" AND-OR trees † of depth-2, and *nearly* optimal lower bounds for "regular" AND-OR trees of constant depth. In this work, we try to prove lower bounds for "non-regular" AND-OR trees. Our contributions are as follows,

(i) We give a lower bound of $\Omega\left(\left(\frac{N}{2^d \log^{2d-3} N}\right)^{1/2}\right)$ on the approximate degree of any (*non-regular*) depth-d AND-OR formula. This is a generalization of the lower bound of $\Omega\left(\left(\frac{N}{\log^{d-2} N}\right)^{1/2}\right)$ on approximate degree of *regular* AND-OR formulas, due to Bun-Thaler [BT13b].

(ii) The above result gives an $\Omega\left((N/\log N)^{1/2}\right)$ lower bound on approximate degree of depth-2 formulas. We introduce the notion of *approximate weighted degree*, and suggest an approach which can potentially give a lower bound of $\Omega\left(N^{1/2}\right)$ on the approximate degree of depth-2 formulas. We point out connections between approximate weighted degree and Ambainis' result on quantum search with variable times [Amb06].

†By "regular", we mean that all gates in the same layer have the same fan-in

2 Preliminaries

Definition 1. *A polynomial $p : \mathbb{R}^n \to \mathbb{R}$ is said to ε-approximate a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ if $\|f - p\|_\infty \triangleq \max_{\mathbf{x}\in\{-1,1\}^n} |f(\mathbf{x}) - p(\mathbf{x})| \le \varepsilon$.*

Definition 2. *For any function $f : \{1,-1\}^n \to \{1,-1\}$, $\widetilde{\deg}_\varepsilon(f)$ is the smallest degree of a polynomial that ε-approximates it. For ease of notation, we use $\widetilde{\deg}(f)$ to denote $\widetilde{\deg}_{1/2}(f)$.*

Definition 3. *For any $S \subseteq [n]$, we define the polynomial $\chi_S : \mathbb{R}^n \to \mathbb{R}$ as $\chi_S(x_1,\ldots,x_n) = \prod_{i\in S} x_i$.*

Note that we can write any multilinear polynomial $p : \mathbb{R}^n \to \mathbb{R}$ as $p(\mathbf{x}) = \sum_{S\subseteq[n]} c_S\chi_S(\mathbf{x})$ for some coefficients $\{c_S\}$. We use $\mathbf{x}$ or $\mathbf{y}$ to denote a collection of variables, and x_i or x_{ij} to denote individual variables.

3 Prior Work

Our work mainly builds upon prior work on approximate degree lower bounds for constant depth regular AND-OR trees.

Closing a long line of work Sherstov and Bun-Thaler independently proved a tight lower bound on the approximate degree of depth-2 regular AND-OR trees.

Theorem 1 ([BT13a, She13]). *If $f = \bigwedge_{i=1}^m \bigvee_{j=1}^n x_{ij}$, then $\widetilde{\deg}(f) = \Omega(\sqrt{mn})$.*

Building upon this work, Bun and Thaler proved close-to-optimal lower bounds on approximate degree of constant depth regular AND-OR trees, by introducing a new technique of hardness amplification.

Theorem 2 ([BT13b]). *If f is a regular AND-OR tree of depth d, with N leaves, then*

$$\widetilde{\deg}(f) \ge \Omega\left(\sqrt{\frac{N}{\log^{(d-2)} N}}\right)$$

4 Dual Functions

In this section, we present the technique of dual functions that was used to prove most of the results mentioned earlier. Much of the exposition here is taken from [BT13a] and sources cited therein.

4.1 Functions

Given a function $f : \{-1,1\}^n \to \{-1,1\}$, suppose $p : \{-1,1\}^n \to \mathbb{R}$, is a polynomial of degree d that ε-approximates f (that is, $\|f - p\|_\infty \le \varepsilon$). Without loss of generality, we may assume that p is multilinear - we only care about its values on $\{-1,1\}^n$. So we can write p as $p(\mathbf{x}) = \sum_{S:|S|\le d} c_S\chi_S(\mathbf{x})$,

Then, it is straightforward to see that the value of the following linear program, with c_S as variables, is at most ε (because the objective at p is at most ε).

$$\begin{aligned} \text{Min}: &\ \delta \\ \text{Constraints}: &\ \forall \mathbf{x}\in\{-1,1\}^n : \Big|\, f(\mathbf{x}) - \sum_{S:|S|\le d} c_S\chi_S(\mathbf{x}) \,\Big| \le \delta \end{aligned}$$

The dual of this linear program is as follows, in variables $\phi(\mathbf{x})$, for each $\mathbf{x}\in\{-1,1\}^n$.

$$\begin{aligned} \text{Max}: &\ \sum_{\mathbf{x}} \phi(\mathbf{x}) f(\mathbf{x}) \\ \text{Constraints}: &\ \sum_{\mathbf{x}} |\phi(\mathbf{x})| = 1 \\ &\ \sum_{\mathbf{x}} \phi(\mathbf{x})\chi_S(\mathbf{x}) = 0 \quad \forall S : |S| \le d \end{aligned}$$

By weak duality, and by the observation about the primal above, if there is indeed a polynomial that approximates f to within ε, then the value of this dual is less than ε. The contrapositive says that if we can find a function $\phi : \{-1,1\}^n \to \mathbb{R}$ such that the objective of the dual at the point ϕ is more than ε, then there is no polynomial of degree d that ε-approximates f.

The converse of this statement is implied by strong duality, which gives us the following theorem.

Theorem 3. *For any $f : \{-1,1\}^n \to \{-1,1\}$, $\widetilde{\deg}_\varepsilon(f) \geq d$ iff there is a function $\phi : \{-1,1\}^n \to \mathbb{R}$ such that:*

$$\sum_{\mathbf{x}\in\{-1,1\}^n} |\phi(\mathbf{x})| = 1$$
$$\sum_{\mathbf{x}\in\{-1,1\}^n} \phi(\mathbf{x})f(\mathbf{x}) > \varepsilon$$
$$\sum_{\mathbf{x}\in\{-1,1\}^n} \phi(\mathbf{x})\chi_S(\mathbf{x}) = 0 \ \ \forall S : |S| < d$$

We call such a function a *dual function* for f that witnesses $\widetilde{\deg}_\varepsilon(f) \geq d$.

4.2 Distributional view of dual functions

An equivalent, sometimes more intuitive, view of the above is in terms of dual distributions.

Suppose $\widetilde{\deg}_\varepsilon(f) \geq d$, and ϕ is a dual function witnessing this. One of the implications of this is that ϕ does not correlate with any polynomial of degree $< d$. In particular, it doesn't correlate with the constant polynomial 1. This implies that $\sum_{\mathbf{x}} \phi(\mathbf{x}) = 0$. We already have $\sum_{\mathbf{x}} |\phi(\mathbf{x})| = 1$, and these two together give imply that: $\sum_{\mathbf{x}:\phi(\mathbf{x})<0} |\phi(\mathbf{x})| = \sum_{\mathbf{x}:\phi(\mathbf{x})>0} |\phi(\mathbf{x})| = \frac{1}{2}$.

This ensures that the following pair of functions, defined from $\{-1,1\}^n \to \mathbb{R}$, are, in fact, probability distributions.

$$\begin{aligned} \mu_1(\mathbf{x}) &= \begin{cases} 2\phi(\mathbf{x}) & \text{if } \phi(\mathbf{x}) \geq 0 \\ 0 & \text{if } \phi(\mathbf{x}) < 0 \end{cases} \\ \mu_{-1}(\mathbf{x}) &= \begin{cases} 0 & \text{if } \phi(\mathbf{x}) \geq 0 \\ -2\phi(\mathbf{x}) & \text{if } \phi(\mathbf{x}) < 0 \end{cases} \end{aligned}$$

Observe, now, that for any function g, $\sum_{\mathbf{x}} \phi(\mathbf{x})g(\mathbf{x}) = \frac{1}{2}\left(\mathbb{E}_{\mathbf{x}\sim\mu_1}[g(x)] - \mathbb{E}_{\mathbf{x}\sim\mu_{-1}}[g(x)]\right)$. Hence the correlation properties of the dual function translate directly to properties of expectations under these distributions.

In the other direction, given any pair of disjoint distributions μ_1, μ_{-1}, we can define a function $\phi(\mathbf{x}) = \frac{1}{2}(\mu_1(\mathbf{x}) - \mu_{-1}(\mathbf{x}))$ that inherits as properties of its correlation with other functions whatever properties the difference in expected values under the distributions had. (The requirement of disjointness of the supports of μ_1 and μ_{-1} may be traded for some stronger requirement on the expectations, but we will make do with this for now.)

This process lets us translate the above theorem into the language of these *dual distributions* as follows.

Theorem 4. *For any $f : \{-1,1\}^n \to \{-1,1\}$, $\widetilde{\deg}_\varepsilon(f) > d$ iff there is a pair of disjoint distributions (μ_1, μ_{-1}) on $\{-1,1\}^n$ such that:*

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_1}[f(\mathbf{x})] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_{-1}}[f(\mathbf{x})] &> 2\varepsilon \\ \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_1}[\chi_S(\mathbf{x})] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_{-1}}[\chi_S(\mathbf{x})] &= 0 \quad \textit{for all } S : |S| \leq d \end{aligned}$$

Thus, intuitively, we show that a function f cannot be approximated by polynomials of a certain degree d, by demonstrating two distributions such that f can *distinguish* between these distributions with 2ε advantage, while no polynomial of degree d can distinguish them at all.

4.3 Application

One simple example of an approximate degree lower bound is of PARITY. We know that $\widetilde{\deg}_\varepsilon(\text{PARITY}) = n$ for any $\varepsilon < 1$. This can be shown by exhibiting the dual function which is PARITY itself! This is easy to see if one considers that $\text{PARITY}(\mathbf{x}) = \chi_{[n]}(\mathbf{x}) = \prod_i x_i$.

The case with most other functions turns out to not be so straightforward, though - constructing dual functions even for simple functions like OR (done in [Spa08]) seems to be somewhat non-trivial. Note that we know, even prior to the above work, by applying symmetrization lemma and Markov's theorem, that the ε-approximate degree for any constant ε for both AND_n and OR_n is $\Omega(\sqrt{n})$, and hence that the corresponding dual functions exist.

But constructing dual functions (or distributions) seems to be more or less the strongest approach to proving approximate degree lower bounds, at least in the case of AND-OR trees, seeing as both [BT13a] and [She13], which bound the approximate degree of depth-2 AND-OR trees, and [BT13b], which extends this to constant-depth trees, proceed essentially by either demonstrating the existence of or constructing the necessary dual functions.

To give an idea of how such feats might be achieved, we present the following short sketch of the technique used in [She13] to prove theorem 1.

We start with the dual distributions witnessing $\widetilde{\deg}_\varepsilon(\mathrm{AND}_m) \geq d$ and $\widetilde{\deg}_{1-\delta}(\mathrm{OR}_n) \geq d'$ given as $(\mu_1^{\mathrm{AND}}, \mu_{-1}^{\mathrm{AND}})$ and $(\mu_1^{\mathrm{OR}}, \mu_{-1}^{\mathrm{OR}})$ respectively (ε and δ to be chosen later). We now construct the dual distributions for $\mathrm{AND}_m \circ \mathrm{OR}_n$, namely $\mu_b^{\text{AND-OR}}$ (for $b \in \{1, -1\}$), which is sampled as follows:

1. First sample $(z_1, \ldots, z_m) \sim \mu_b^{\mathrm{AND}}$.
2. For each i, sample $(x_{i1}, \ldots, x_{in}) \sim \mu_{z_i}^{\mathrm{OR}}$.

To show that this dual distribution, indeed proves the desired lower bound on approximate degree, we need to show that it satisfies the two conditions stated in Theorem 4. We leave out the full proof of condition (i) here, but using the *one-sidedness* of the dual function of OR_n, it follows that under the distributions defined as above,

$$\mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_1^{\text{AND-OR}}} \mathrm{AND}_m \circ \mathrm{OR}_n(\mathbf{x}) - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_{-1}^{\text{AND-OR}}} \mathrm{AND}_m \circ \mathrm{OR}_n(\mathbf{x}) \geq 2(\varepsilon - \delta)$$

[We suggest the reader to refer [She13] for complete details, although the presentation there is slightly different from ours].

We sketch the proof of condition (ii), since it will be relevant later for our notions of approximate weighted degree. For any χ_S where $|S| < d \cdot d'$, we have the following (where $S_i = S \cap \{x_{i1}, x_{i2}, \ldots, x_{in}\}$):

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_b^{\text{AND-OR}}}[\chi_S(\mathbf{x})] &= \mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_b^{\mathrm{AND}}}\left[\prod_{i=1}^{m} \mathop{\mathbb{E}}_{\mathbf{x}_i\sim\mu_{z_i}^{\mathrm{OR}}} \chi_{S_i}(\mathbf{x}_i)\right] \\
&= \mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_b^{\mathrm{AND}}}\left[\prod_{i=1}^{m}\left(\frac{(1+z_i)}{2} \mathop{\mathbb{E}}_{\mathbf{x}_i\sim\mu_1^{\mathrm{OR}}} \chi_{S_i}(\mathbf{x}_i) + \frac{(1-z_i)}{2} \mathop{\mathbb{E}}_{\mathbf{x}_i\sim\mu_{-1}^{\mathrm{OR}}} \chi_{S_i}(\mathbf{x}_i)\right)\right]
\end{aligned}$$

Now, if for some i it happens that $|S_i| < d'$, then the corresponding term in the product above would be a constant, as the expected value under μ_1^{OR} and μ_{-1}^{OR} would be the same. As $|S| < d \cdot d'$, less than d terms can survive this. Each of these terms contribute a degree of 1 in z to the product, leading to a polynomial in $\mathbf{z}$ of degree strictly less than d, which has the same expected value under μ_1^{AND} and μ_{-1}^{AND} as the approximate degree of AND_m is at least d. This tells us that any polynomial of degree less than $d \cdot d'$ has the same expected value under either distribution, thus getting us the second condition in Theorem 4. Thus, we get that $\widetilde{\deg}_{\varepsilon-\delta}(\mathrm{AND}_m \circ \mathrm{OR}_n) \geq d \cdot d' = \Omega(\sqrt{mn})$.

5 Our Results

We try to prove lower bounds for non-regular AND-OR trees. Our first result is a close-to-optimal lower bound for constant depth AND-OR trees (see Theorem 5). We prove this by showing that it is possible to find a "large enough" regular AND-OR tree embedded inside any non-regular AND-OR tree, and the lower bound then follows from a direct reduction to Theorems 1 and 2.

Our second result (rather, a *pseudo*-result) is an approach to prove an optimal lower bound on approximate degree of depth-2 AND-OR trees. We first generalize the notion of approximate degree to define *approximate weighted degree* (denoted by $\widetilde{\text{wt-deg}}$: see Definition 5). We then show an upper bound on the $\widetilde{\text{wt-deg}}(\mathrm{AND}_m)$, and conjecture the optimality of our upper bound (Conjecture 1). We show that this conjecture in fact implies tight lower bounds on the approximate degree of non-regular depth-2 AND-OR trees.

We also suggest an approach to prove this conjecture which goes via obtaining a certain generalization of robust polynomials constructed by Sherstov [She09]. We don't know if these generalized robust polynomials exist or not, and we leave it as an open conjecture (Conjecture 2). We feel that this conjecture may be of independent interest.

We present all our results in this section. All the main proofs have been shifted to the Appendix for better clarity of reading.

5.1 Lower bounds by direct reduction

Theorem 5. *For f being computed by any (non-regular) AND-OR tree of depth $d \geq 2$ with N leaves,*

$$\widetilde{\deg}(f) \geq \Omega\left(\sqrt{\frac{N}{2^d \log^{(2d-3)} N}}\right)$$

Proof. See Appendix. □

5.2 Lower bounds via weighted approximate degree

Definition 4 (Weighted Degree)**.** *Fix a* weight *function $T : [n] \to \mathbb{N}$. The weighted degree of the monomial $m = \prod_{i=1}^n x_i^{e_i}$ is equal to $\text{wt-}\deg^T(m) = \sum_{i=1}^n T(i)e_i$. The weighted degree of any polynomial $p \in \mathbb{F}[x_1, \cdots, x_n]$ is the largest weighted degree of any monomial in p. For ease of notation, we use $\text{wt-}\deg(p)$ to denote $\text{wt-}\deg^T(p)$ (whenever the weight function is obvious from context).*

Intuitively, weighted degree is a generalization of degree where each variable has a different "cost".

Definition 5 (ε-Approximate Weighted Degree)**.** *For a weight function $T : [n] \to \mathbb{N}$, and any function $f : \{1,-1\}^n \to \{1,-1\}$, $\widetilde{\text{wt-}\deg}_\varepsilon^T(f)$ is the minimum value of $\text{wt-}\deg^T(p)$ where the polynomial $p : \mathbb{R}^n \to \mathbb{R}$ is such that for all $\mathbf{x} \in \{1,-1\}^n$, $|f(\mathbf{x})-p(\mathbf{x})| \leq \varepsilon$. For ease of notation, we use $\widetilde{\text{wt-}\deg}(f)$ to denote $\widetilde{\text{wt-}\deg}_{1/2}^T(f)$ (whenever the weight function is obvious from context).*

Our main motivation is to understand the approximate weighted degree of the AND function. By symmetry, this is same as the approximate weighted degree of the OR function (since $\mathrm{OR}(\mathbf{x}) = -\mathrm{AND}(-\mathbf{x})$). We now show an upper bound on the approximate weighted degree of OR using a connection to Ambainis' result on quantum search with variable times [Amb06].

Lemma 1. *Suppose we have an input $\mathbf{x} = (x_1, \cdots, x_n) \in \{1,-1\}^n$, where querying x_i takes time $T(i)$. Let $Q_\varepsilon^T(f)$ be the minimum time taken by any quantum algorithm to compute $f(x_1, \cdots, x_n)$ with error probability less than ε. Then, $\widetilde{\text{wt-}\deg}_{2\varepsilon}^T(f) \leq 2 \cdot Q_\varepsilon^T(f)$.*

Proof. See Appendix. □

From Ambainis' quantum search with variable times [Amb06], we know that $Q_{0.25}^T(\mathrm{OR}_n) = \Theta\left(\left(\sum_{i=1}^n T(i)^2\right)^{1/2}\right)$. Combining this with the above lemma, we get the following theorem,

Theorem 6. *For any weight function $T : [n] \to \mathbb{N}$,*

$$\widetilde{\text{wt-}\deg}(\mathrm{AND}_n) = \widetilde{\text{wt-}\deg}(\mathrm{OR}_n) \leq O\left(\left(\sum_{i=1}^n T(i)^2\right)^{1/2}\right)$$

We conjecture that this upper bound is in fact optimal.

Conjecture 1. *For any weight function $T : [n] \to \mathbb{N}$, $\widetilde{\text{wt-}\deg}\,(\mathrm{AND}_m) \geq \Omega\left(\left(\sum_{i=1}^n T(i)^2\right)^{1/2}\right)$*

We now show tight lower bounds for any depth-2 AND-OR tree, assuming Conjecture 1 to be true.

Theorem 7. *Let $f(\mathbf{x}) = \bigwedge_{i=1}^{m} \bigvee_{j=1}^{n_i} x_{ij}$. Suppose that for weight function $T : [m] \to \mathbb{N}$ given by $T(i) = c\sqrt{n_i}$ (where c is some appropriate constant chosen),* $\text{wt-}\widetilde{\deg}\,(\text{AND}_m) \geq \Omega\left(\left(\sum_{i=1}^{m} T(i)^2\right)^{1/2}\right)$. *Then,*

$$\widetilde{\deg}(f) \geq \Omega\left(\left(\sum_{i=1}^{m} n_i\right)^{1/2}\right)$$

Proof. See Appendix. □

5.2.1 Approach to Conjecture 1 using generalized robust polynomials

In a very beautiful result, Sherstov [She09] showed that for any polynomial $p : \{1, -1\}^n \to [-1, 1]$, there exists a *robust* polynomial p_{robust} such that for all $\mathbf{x} \in \{1, -1\}^n$ and all $\mathbf{e} \in [-1/3, 1/3]^n$, $|p(\mathbf{x}) - p_{\text{robust}}(\mathbf{x} + \mathbf{e})| \leq \delta$ and $\deg(p_{\text{robust}}) = O(\deg(p) + \log(1/\delta))$. We propose a generalization of this result, based on the notions of weighted degree. We then show that if this generalization holds, then it implies Conjecture 1.

Conjecture 2. *For any weight function $T : [n] \to \mathbb{N}$, and any polynomial $p : \{1, -1\}^n \to [-1, 1]$, there exists a robust polynomial p_{robust} such that for all $\mathbf{x} \in \{1, -1\}^n$ and all $\mathbf{e} \in [-1/3, 1/3]^n$, $|p(\mathbf{x}) - p_{\text{robust}}(\mathbf{x} + \mathbf{e})| \leq \delta$ and* $\text{wt-}\deg^T(p_{\text{robust}}) = O(\text{wt-}\deg^T(p) + \log(1/\delta))$

We now prove that Conjecture 2 implies Conjecture 1. Combining this with Theorem 7, we conclude that existence of these generalized robust polynomials implies the desired tight lower bound on the approximate weighted degree of depth-2 AND-OR tree.

Theorem 8. *If Conjecture 2 is true, then for any weight function $T : [n] \to \mathbb{N}$,*

$$\text{wt-}\widetilde{\deg}\,(\text{AND}_n) \geq \Omega\left(\left(\sum_{i=1}^{n} T(i)^2\right)^{1/2}\right)$$

6 Future directions

The immediate open questions that need to be answered are,

1. Prove a tight lower bound of $\Omega(N^{1/2})$ on approximate degree of non-regular depth-2 AND-OR tree.
2. Prove a tight lower bound of $\Omega(N^{1/2})$ on approximate degree of regular depth-3 AND-OR tree.

Acknowledgements

We would like to thank Scott for teaching a wonderful course and also giving us constant encouragement and valuable suggestions throughout the timeline of this project. We would also like to thank Robin Kothari for many helpful discussions. In particular, it was in one of these discussions that we figured out the connections to quantum search with variable times.

References

[Aar14] Scott Aaronson. Personal communication. 2014.

[Amb06] Andris Ambainis. Quantum search with variable times. *arXiv:quant-ph/0609168*, 2006.

[BT13a] Mark Bun and Justin Thaler. Dual lower bounds for approximate degree and markov-bernstein inequalities. *ICALP*, 2013.

[BT13b] Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. *ECCC*, (TR13-151), 2013.

[Rei11] Ben Reichardt. Reflections for quantum query complexity: The general adversary bound is tight for every boolean function. *SODA*, 2011.

[She09] Alexander Sherstov. Making polynomials robust to noise. *Theory of Computing*, 2009.

[She13] Alexander Sherstov. Approximating the and-or tree. *Theory of Computing*, 2013.

[Spa08] Robert Spalek. A dual polynomial for or. *Manuscript*, (arXiv 0803.4516), 2008.

A Proofs from Section 5

Proof of Theorem 5. For simplicity, we first prove this theorem for depth $d = 2$. Let $f(\mathbf{x}) = \bigwedge_{i=1}^{m} \bigvee_{j=1}^{n_i} x_{ij}$. Our approach is to find a large enough *regular* tree embedded in f, and subsequently apply Theorem 1.

Suppose the fan-ins of the OR gates are $n_1 \le n_2 \le \cdots \le n_m$, where $\sum_{i=1}^{m} n_i = N$. We define buckets $B_0, B_1, \cdots \subseteq [m]$, as $B_k = \{i \in [m] : 2^{k-1} < n_i \le 2^k\}$. Since each $n_i < N$, there can at most be $\log N$ buckets. Define $S_k = \sum_{i \in B_k} n_i$. Clearly, since $\sum_{k=1}^{\log N} S_k = N$, there exists a k such that $S_k \ge N/\log N$. We have that for such a k, the size of the bucket $|B_k| \ge \frac{S_k}{2^k} \ge \frac{N}{2^k \log N}$. Also, the fan-in of each OR gate is at least 2^{k-1}.

Thus, we have that the function

$$g(\mathbf{y}) = \bigwedge_{p=1}^{|B_k|} \bigvee_{q=1}^{2^{k-1}} y_{pq}$$

can be *embedded* into the function f (that is, it is possible to replace the x_{ij}'s by either y_{pq}'s or ± 1 such that f computes g exactly. Since g is a regular AND-OR tree, by Theorem 1, we get that

$$\widetilde{\deg}(f) \ge \widetilde{\deg}(g) \ge \Omega\left(\sqrt{|B_k| \cdot 2^{k-1}}\right) \ge \Omega\left(\sqrt{\frac{N}{\log N}}\right)$$

We now turn to prove the theorem for all $d \ge 2$. Our approach is again to find a large enough regular tree embedded in f, and subsequently apply Theorem 2.

For any variable, define it's signature as $(m_1, \cdots, m_d)$, which is the tuple of fan-ins of all the gates on the path from the root to that variable.[‡] We partition these variables into buckets according to their signature values. That is, define buckets $B(k_2, k_3, \cdots, k_d)$ as a partition of the N variables with signature $(m_1, m_2, \cdots, m_d)$ such that for each $i \in \{2, \cdots, d\}$, it holds that $2^{k_i - 1} \le m_i < 2^{k_i}$.[§]

Since there are only N variables, each k_i ranges from 1 to at most $\log N$. Thus, we have a total of at most $\log^{d-1} N$ buckets. Hence, there exists a bucket $B(k_2^*, \cdots, k_d^*)$ with at least $\frac{N}{\log^{d-1} N}$ variables. We set all variables outside this bucket to appropriate values so that the function becomes equal to an AND-OR tree over this bucket. Furthermore, for every $i \in \{2, \cdots, d\}$, we pull down the fan-ins of the i-th layer to $2^{k_i - 1}$, by setting more variables appropriately. This process reduces the number of variables in bucket $B(k_2^*, \cdots, k_d^*)$ by a factor of at most 2^{d-1}.

Thus, we end up with a function g which is a regular AND-OR tree of depth d, with at least $N' = \frac{N}{(2 \log N)^{d-1}}$ leaves in all. From Theorem 2, we conclude that

$$\widetilde{\deg}(f) \ge \widetilde{\deg}(g) \ge \Omega\left(\sqrt{\frac{N'}{\log^{(d-2)} N'}}\right) = \Omega\left(\sqrt{\frac{N}{2^d \log^{(2d-3)} N}}\right)$$

□

Proof of Lemma 1. This proof is identical to that of its unweighted version : $\widetilde{\deg}_{2\varepsilon}(f) \le 2 \cdot Q_\varepsilon(f)$. Suppose for some subset $S \subseteq [n]$, the algorithm makes a query of the form,

$$\sum_{i \in S} \alpha_{i,b,z} |i\rangle |b\rangle |z\rangle \mapsto \sum_{i \in S} \alpha_{i,b,z} |i\rangle |b \oplus x_i\rangle |z_i\rangle = \sum_{i \in S} \alpha_{i,b,z} \left[\left(\frac{1 + x_i}{2}\right) |i\rangle |b\rangle |z_i\rangle + \left(\frac{1 - x_i}{2}\right) |i\rangle |\neg b\rangle |z_i\rangle\right]$$

To make this query, the algorithm consumes time equal to $\max_{i \in S} T(i)$. At the same time, the weighted degree of any entry in the amplitude vector goes up by at most $\max_{i \in S} T(i)$. Thus, after any number of

[‡] We can assume without loss of generality that all the variables in the formula are at depth d. Because, if a variable is at depth less than d, we can simply add gates with fan-in 1 and push it to depth d.

[§] Since m_1 is the same for all variables, we don't need to consider m_1 as a parameter for the buckets.

queries, the time consumed by the algorithm is an upper bound on the weighted degree of any entry of the amplitude vector.

At the end of the algorithm, the probability that the algorithm declares the function value to be -1 is the sum of squares of entries in the amplitude vector. Thus, we get that for any $\mathbf{x} \in \{1,-1\}^n$, $|\Pr[A(x) = -1] - \widetilde{f}(\mathbf{x})| \leq \varepsilon$, where $|\Pr[A(\mathbf{x}) = -1]|$ is a polynomial in $\mathbf{x}$ of weighted-degree at most $Q^T_\varepsilon(f)$, and $\widetilde{f}(\mathbf{x}) = (1-f(\mathbf{x}))/2$ is the 0-1 version of f. Thus, we get an ε-approximating polynomial for $\widetilde{f}$ and hence a 2ε-approximating polynomial for f, with weighted degree at most $Q^T_\varepsilon(f)$. □

Proof of Theorem 7. We begin by noting that the notion of dual distributions generalises directly to handle weighted-degree instead of degree. This may be done by replacing every occurrence of the condition $|S| < d$ with $\text{wt-}\deg(\chi_S) < d$ in the statements of Section 4 to give the following lemma.

Lemma 2. *For any $f : \{-1,1\}^n \to \{-1,1\}$ and any weight function $T : [n] \to \mathbb{N}$, $\widetilde{\text{wt-}\deg}^T_\varepsilon(f) > d$ iff there is a pair of disjoint distributions (μ_1, μ_{-1}) on $\{-1,1\}^n$ such that:*

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_1}[f(\mathbf{x})] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_{-1}}[f(\mathbf{x})] &> 2\varepsilon \\ \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_1}[\chi_S(\mathbf{x})] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_{-1}}[\chi_S(\mathbf{x})] &= 0 \quad \textit{for all } S : \text{wt-}\deg^T(\chi_S) \leq d \end{aligned}$$

By the hypothesis of the theorem, there exists a constant c_1, such that $\widetilde{\text{wt-}\deg}_\varepsilon(\text{AND}_m) \geq D \geq c_1 \left(\left(\sum_{i=1}^m T(i)^2\right)^{1/2} \right)$. From the above lemma, there exists a pair of dual distributions $(\mu_1^{\text{AND}}, \mu_{-1}^{\text{AND}})$ over $\{-1,1\}^m$ such that,

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_1^{\text{AND}}}[\text{AND}(\mathbf{x})] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_{-1}^{\text{AND}}}[\text{AND}(\mathbf{x})] &> 2\varepsilon \\ \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_1^{\text{AND}}}[\chi_S(\mathbf{x})] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_{-1}^{\text{AND}}}[\chi_S(\mathbf{x})] &= 0 \quad \forall S : \text{wt-}\deg^T(\chi_S) \leq D \end{aligned}$$

We know that there exists a constant c_2 for any n, $\widetilde{\deg}_{1-\delta}(\text{OR}_n) \geq c_2\Omega(\sqrt{n})$, which implies that there exist dual distributions $(\mu_1^{\text{OR}_n}, \mu_{-1}^{\text{OR}_n})$ over $\{-1,1\}^n$ satisfying the following:

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_1^{\text{OR}_n}}[\text{OR}_n(\mathbf{x})] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_{-1}^{\text{OR}_n}}[\text{OR}_n(\mathbf{x})] &> 2(1-\delta) \\ \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_1^{\text{OR}_n}}[\chi_S(\mathbf{x})] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_{-1}^{\text{OR}_n}}[\chi_S(\mathbf{x})] &= 0 \quad \forall S : |S| \leq c_2\sqrt{n} \end{aligned}$$

Now consider the pair of distributions (μ_1^f, μ_{-1}^f), where μ_b^f is sampled as follows:

1. Sample $(z_1, \ldots, z_m) \leftarrow \mu_b^{\text{AND}}$.
2. For each $i \in [m]$, sample $(x_{i1}, \ldots, x_{in_i}) \leftarrow \mu_{z_i}^{\text{OR}_{n_i}}$.

$(\mathbb{E}_{\mathbf{x}\sim\mu_1^f}[f(\mathbf{x})] - \mathbb{E}_{\mathbf{x}\sim\mu_{-1}^f}[f(\mathbf{x})]) \geq 2(\varepsilon - \delta)$. This follows in the same as in Section 4 [For precise details about how this may be done, see [She13]]

For any χ_S, we have the following (where $S_i = S \cap \{x_{i1}, x_{i2}, \ldots, x_{in_i}\}$):

$$\begin{aligned} \mathop{\mathbb{E}}_{\mathbf{x}\sim\mu_b^f}[\chi_S(\mathbf{x})] &= \mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_b^{\text{AND}}}\left[\prod_{i=1}^m \mathop{\mathbb{E}}_{\mathbf{x}_i\sim\mu_{z_i}^{\text{OR}_{n_i}}} \chi_{S_i}(\mathbf{x}_i)\right] \\ &= \mathop{\mathbb{E}}_{\mathbf{z}\sim\mu_b^{\text{AND}}}\left[\prod_{i=1}^m \left(\frac{(1+z_i)}{2} \mathop{\mathbb{E}}_{\mathbf{x}_i\sim\mu_1^{\text{OR}_{n_i}}} \chi_{S_i}(\mathbf{x}_i) + \frac{(1-z_i)}{2} \mathop{\mathbb{E}}_{\mathbf{x}_i\sim\mu_{-1}^{\text{OR}_{n_i}}} \chi_{S_i}(\mathbf{x}_i)\right)\right] \end{aligned}$$

The term corresponding to any i in the product above is a constant independent of z_i if $|S_i| \leq c_2\sqrt{n_i}$, by the property of the dual distributions of OR_{n_i}, in which case it doesn't contribute to the weighted degree of the product, and a linear term in z_i otherwise, in which case it contributes $c\sqrt{n_1}$ to the weighted degree. Hence the weighted degree of the product is at most $\frac{c}{c_2}|S|$. We know that if this weighted degree is smaller than $cc_1(\sum_{i=1}^m n_i)^{1/2}$, then the expectation of the product is the same under both μ_1^f and μ_{-1}^f.

This implies that $\mathbb{E}_{\mathbf{x}\sim\mu_1^f}[\chi_S(\mathbf{x})] - \mathbb{E}_{\mathbf{x}\sim\mu_{-1}^f}[\chi_S(\mathbf{x})] = 0$ if $|S| \leq c_1c_2(\sum_{i=1}^m n_i)^{1/2}$.

Hence, as they have the above two properties, (μ_1^f, μ_{-1}^f) are dual distributions certifying that $\widetilde{\deg}(f) \geq \Omega\left(\left(\sum_{i=1}^m n_i\right)^{1/2}\right)$.

$\square$

Proof of Theorem 8. Suppose there exists a polynomial p such that $\text{wt-}\deg^T(p) = D$ and for all $\mathbf{x} \in \{1,-1\}^n$, $|p(\mathbf{x}) - \text{AND}(\mathbf{x})| \leq \varepsilon$. Assuming Conjecture 2, there exists a robust polynomial, say q, such that for all $\mathbf{x} \in \{1,-1\}^n$ and all $\mathbf{e} \in [-1/3, 1/3]^n$, $|p(\mathbf{x}) - q(\mathbf{x}+\mathbf{e})| \leq \delta$ and $\text{wt-}\deg^T(q) = O(\text{wt-}\deg^T(p) + \log(1/\delta)) = O(D + \log(1/\delta))$.

We know that $\widetilde{\deg}_{1/3}(\text{AND}_{T(i)^2}) < C \cdot T(i)$ for some constant C. Thus, we have polynomials $p_1, p_2, \cdots, p_n$ such that for each i, $p_i(\mathbf{y}_i)$ $(1/3)$-approximates $\text{AND}_{T(i)^2}(\mathbf{y}_i)$. It is easy to see that $q(p_1(\mathbf{y}_1), \cdots, p_n(\mathbf{y}_n))$ now $(\varepsilon+\delta)$-approximates $\text{AND}(\text{AND}(\mathbf{y}_1), \cdots, \text{AND}(\mathbf{y}_n))$. Also, it is easy to see that the ordinary degree of $q(p_1, \cdots, p_n)$ is at most $C \cdot \text{wt-}\deg^T(q)$ (since $\deg(p_i) < C \cdot T(i)$).

We know that $\widetilde{\deg}(AND_N) = \Omega\left(\sqrt{N}\right)$, for $N = \sum_{i=1}^n T(i)^2$. Thus, choosing ε and δ such that $\varepsilon + \delta < 1/3$, we get that,

$$C \cdot \text{wt-}\deg^T(q) \geq \Omega\left(\left(\sum_{i=1}^n T(i)^2\right)^{1/2}\right)$$

Thus we get that $\text{wt-}\deg^T(p) + O(1) = \Omega(\text{wt-}\deg^T(q)) \geq \Omega\left(\left(\sum_{i=1}^n T(i)^2\right)^{1/2}\right)$, which proves the desired lower bound on the weighted approximate degree of AND_n. $\square$