Lower Bounds for the Approximate Degree of Block-Composed Functions

Justin Thaler[1]

[1]Yahoo Labs

Boolean Functions

- Boolean function $f : \{-1,1\}^n \to \{-1,1\}$
- $$\mathrm{AND}_n(x) = \begin{cases} -1 \quad \text{(TRUE)} & \text{if } x = (-1)^n \\ 1 \quad \text{(FALSE)} & \text{otherwise} \end{cases}$$

Approximate Degree and Threshold Degree

- A real polynomial p ϵ-approximates f if

$$|p(x) - f(x)| < \epsilon \quad \forall x \in \{-1, 1\}^n$$

- $\widetilde{\deg}_\epsilon(f) =$ minimum degree needed to ϵ-approximate f
- $\widetilde{\deg}(f) := \widetilde{\deg}_{1/3}(f)$ is the approximate degree of f

Approximate Degree and Threshold Degree

- A real polynomial p ϵ-approximates f if

$$|p(x) - f(x)| < \epsilon \quad \forall x \in \{-1, 1\}^n$$

- $\widetilde{\deg}_\epsilon(f) =$ minimum degree needed to ϵ-approximate f
- $\widetilde{\deg}(f) := \deg_{1/3}(f)$ is the approximate degree of f
- $\deg_\pm(f) := \lim_{\epsilon \to 1} \widetilde{\deg}_\epsilon(f)$ is the threshold degree of f
- Equivalent to the least degree of a polynomial p such that $p(x) \cdot f(x) > 0$ for all $x \in \{-1, 1\}^n$.

Approximate Degree and Threshold Degree: Example

- OR_n has threshold degree 1, since $p(x) = \sum_i (1 - x_i)/2 - 1$ sign-represents OR_n.
- OR_n has approximate degree $\Theta(\sqrt{n})$ [NS94].

Why Care About Approximate and Threshold Degree?

Upper bounds on $\widetilde{\deg}_\epsilon(f)$ yield efficient learning algorithms

- $\epsilon \to 1$ (i.e., threshold degree): PAC learning [KS01]
- ϵ "close to" 1: Attribute-Efficient Learning [KS04, STT12]
- $\epsilon < 1$ a constant: Agnostic Learning [KKMS05]

Why Care About Approximate and Threshold Degree?

Lower bounds on $\widetilde{\deg}_\epsilon(f)$ yield lower bounds on:

- Quantum query complexity [BBCMW98] [AS01] [Amb03] [KSW04]
- Communication complexity [BVdW08] [She07] [SZ07] [CA08] [LS08] [She12]
- Circuit complexity [MP69] [Bei93] [Bei94] [She08]

Hardness-Amplification for Approximate Degree

- Approximate degree remains poorly understood.
- However, several recent works have established various forms of "hardness amplification" for approximate degree.

Hardness-Amplification for Approximate Degree

- Approximate degree remains poorly understood.
- However, several recent works have established various forms of "hardness amplification" for approximate degree.
- The goal of these results is:
  - Given: A "simple" Boolean function f that is "hard to approximate to low error" by degree d polynomials.
  - Turn f into a "still-simple" F that is hard to approximate even to very high error.

Prior Results on Hardness Amplification for Approximate Degree

(Negative) One-Sided Approximate Degree

- Negative one-sided approximate degree is an intermediate notion between approximate degree and threshold degree.
- A real polynomial p is a negative one-sided ϵ-approximation for f if

$$|p(x) - 1| < \epsilon \quad \forall x \in f^{-1}(1)$$

$$p(x) \leq -1 \quad \forall x \in f^{-1}(-1)$$

- $\widetilde{\mathrm{odeg}}_{-,\epsilon}(f)$ = min degree of a negative one-sided ϵ-approximation for f.

(Negative) One-Sided Approximate Degree

- Negative one-sided approximate degree is an intermediate notion between approximate degree and threshold degree.
- A real polynomial p is a negative one-sided ϵ-approximation for f if

$$|p(x) - 1| < \epsilon \quad \forall x \in f^{-1}(1)$$

$$p(x) \leq -1 \quad \forall x \in f^{-1}(-1)$$

- $\widetilde{\text{odeg}}_{-,\epsilon}(f) =$ min degree of a negative one-sided ϵ-approximation for f.
- Examples: $\widetilde{\text{odeg}}_{-,1/3}(\mathrm{AND}_n) = \Theta(\sqrt{n})$; $\widetilde{\text{odeg}}_{-,1/3}(\mathrm{OR}_n) = 1$.

(Negative) One-Sided Approximate Degree

- Negative one-sided approximate degree is an intermediate notion between approximate degree and threshold degree.
- A real polynomial p is a <u>negative one-sided ϵ-approximation</u> for f if

$$|p(x) - 1| < \epsilon \quad \forall x \in f^{-1}(1)$$

$$p(x) \leq -1 \quad \forall x \in f^{-1}(-1)$$

- $\widetilde{\mathrm{odeg}}_{-,\epsilon}(f)$ = min degree of a negative one-sided ϵ-approximation for f.
- Examples: $\widetilde{\mathrm{odeg}}_{-,1/3}(\mathrm{AND}_n) = \Theta(\sqrt{n})$; $\widetilde{\mathrm{odeg}}_{-,1/3}(\mathrm{OR}_n) = 1$.
- Positive one-sided approximate degree is defined similarly, with the rule of +1 and -1 reversed.
- Examples: $\widetilde{\mathrm{odeg}}_{+,1/3}(\mathrm{AND}_n) = 1$; $\widetilde{\mathrm{odeg}}_{-,1/3}(\mathrm{OR}_n) = \Theta(\sqrt{n})$.

Prior Hardness Amplification Results

Theorem (Bun and Thaler)

Let f be a Boolean function with $\widetilde{odeg}_{-,1/2}(f) \geq d$. Let $F = \mathrm{OR}_t(f, \ldots, f)$. Then $\widetilde{odeg}_{-,1-2^{-t}}(F) \geq d$.

Prior Hardness Amplification Results

Theorem (Bun and Thaler)

Let f be a Boolean function with $\widetilde{odeg}_{-,1/2}(f) \geq d$. Let $F = \mathrm{OR}_t(f, \ldots, f)$. Then $\widetilde{odeg}_{-,1-2^{-t}}(F) \geq d$.

Theorem (Sherstov)

Let f be a Boolean function with $\widetilde{odeg}_{-,1/2}(f) \geq d$. Let $F = \mathrm{OR}_t(f, \ldots, f)$. Then $\deg_{\pm}(F) = \Omega(\min\{d, t\})$.

Our Hardness Amplification Result

- For some applications in complexity theory, one needs a hardness amplification theorem that yields lower bounds even for functions with low threshold degree.
- This is what we achieve.

Our Hardness Amplification Result

- For some applications in complexity theory, one needs a hardness amplification theorem that yields lower bounds even for functions with low threshold degree.
- This is what we achieve.
- Define $\mathrm{OMB}_t \colon \{-1,1\}^t \to \{-1,1\}$ via:

$$\mathrm{OMB}_t(x_1, \ldots, x_t) = (-1)^{i^*-1},$$

where i^* is the largest index such that $x_{i^*} = -1$.

Theorem

Let f be a Boolean function with $\widetilde{odeg}_{+,1/2}(f) \geq d$. Let $F = \mathrm{OMB}_t(f, \ldots, f)$. Then $\widetilde{odeg}_{+,1-2^{-t}}(F) \geq d$.

Our Hardness Amplification Result

- For some applications in complexity theory, one needs a hardness amplification theorem that yields lower bounds even for functions with low threshold degree.
- This is what we achieve.
- Define $\mathrm{OMB}_t \colon \{-1,1\}^t \to \{-1,1\}$ via:

$$\mathrm{OMB}_t(x_1, \ldots, x_t) = (-1)^{i^*-1},$$

where i^* is the largest index such that $x_{i^*} = -1$.

Theorem

Let f be a Boolean function with $\widetilde{\mathrm{odeg}}_{+,1/2}(f) \geq d$. Let $F = \mathrm{OMB}_t(f, \ldots, f)$. Then $\widetilde{\mathrm{odeg}}_{+,1-2^{-t}}(F) \geq d$.

- Example Application: Let $F = \mathrm{OMB}_t(\mathrm{OR}_{n/t}, \ldots, \mathrm{OR}_{n/t})$. Then $\deg_{\pm}(F) = 1$, yet $\widetilde{\mathrm{odeg}}_{+,1-2^{-t}}(F) = \Omega(\sqrt{n/t})$.

Intuition: A Matching Upper Bound

Theorem

Let f be a Boolean function with $\widetilde{odeg}_{+,1/2}(f) \geq d$. Let $F = \mathrm{OMB}_t(f, \ldots, f)$. Then $\widetilde{odeg}_{+,1-2^{-t}}(F) \geq d$.

- OMB_t itself can be sign-represented by the degree-1 polynomial $p(x) = 1 + \sum_{i=1}^{t} (-3)^i \cdot (1 - x_i)/2$.

Intuition: A Matching Upper Bound

Theorem

Let f be a Boolean function with $\widetilde{odeg}_{+,1/2}(f) \geq d$. Let $F = \mathrm{OMB}_t(f, \ldots, f)$. Then $\widetilde{odeg}_{+,1-2^{-t}}(F) \geq d$.

- OMB_t itself can be sign-represented by the degree-1 polynomial $p(x) = 1 + \sum_{i=1}^{t} (-3)^i \cdot (1 - x_i)/2$.
- In fact, OMB_t is approximated to error $\approx 1 - 3^{-t}$ by $3^{-t-1} \cdot p(x)$.

Intuition: A Matching Upper Bound

Theorem

Let f be a Boolean function with $\widetilde{odeg}_{+,1/2}(f) \geq d$. Let $F = \mathrm{OMB}_t(f, \ldots, f)$. Then $\widetilde{odeg}_{+,1-2^{-t}}(F) \geq d$.

- OMB_t itself can be sign-represented by the degree-1 polynomial $p(x) = 1 + \sum_{i=1}^{t} (-3)^i \cdot (1 - x_i)/2$.
- In fact, OMB_t is approximated to error $\approx 1 - 3^{-t}$ by $3^{-t-1} \cdot p(x)$.
- Suppose there is a degree d polynomial q such that
  1. $q(x) = 1$ for all $x \in f^{-1}(1)$.
  2. $1 \leq q(x) \leq 4/3$ for all $x \in f^{-1}(-1)$.

  Then $\mathrm{OMB}_t(f, \ldots, f)$ is sign-represented by $p(q, \ldots, q)$.

Intuition: A Matching Upper Bound

Theorem

Let f be a Boolean function with $\widetilde{odeg}_{+,1/2}(f) \geq d$. Let $F = \mathrm{OMB}_t(f, \ldots, f)$. Then $\widetilde{odeg}_{+,1-2^{-t}}(F) \geq d$.

- OMB_t itself can be sign-represented by the degree-1 polynomial $p(x) = 1 + \sum_{i=1}^{t}(-3)^i \cdot (1 - x_i)/2$.
- In fact, OMB_t is approximated to error $\approx 1 - 3^{-t}$ by $3^{-t-1} \cdot p(x)$.
- Suppose there is a degree d polynomial q such that
  1. $q(x) = 1$ for all $x \in f^{-1}(1)$.
  2. $1 \leq q(x) \leq 4/3$ for all $x \in f^{-1}(-1)$.

  Then $\mathrm{OMB}_t(f, \ldots, f)$ is sign-represented by $p(q, \ldots, q)$.
- In fact, it is approximated to error $\approx 1 - 3^{-t}$ by $3^{-t-1} \cdot p(q, \ldots, q)$.

Overview of the Proof

Symmetrization

- Historically, approximate degree lower bounds were proven via a technique called symmetrization.
- Symmetrization argues any approximating polynomial $p\colon \{-1,1\}^n \to \mathbb{R}$ for f must have large degree via a two-step process.
    1. Turn p into a certain univariate polynomial q such that $\deg(q) \leq \deg(p)$.
    2. Argue that q has to have large degree, and hence p does as well.

Beyond Symmetrization

- Symmetrization is "lossy": in turning an n-variate poly p into a univariate poly p^{sym}, we throw away information about p.
- Recent breakthroughs have exploited a "lossless" approach to proving approximate degree lower bounds.

Linear Programming Formulation of Approximate Degree

What is best error achievable by **any** degree d approximation of f?
Primal LP (Linear in ϵ and coefficients of p):

$$\begin{aligned} \text{min}_{p,\epsilon} \quad & \epsilon \\ \text{s.t.} \quad & |p(x) - f(x)| \leq \epsilon \qquad \text{for all } x \in \{-1,1\}^n \\ & \deg p \leq d \end{aligned}$$

Dual LP:

$$\begin{aligned} \text{max}_{\psi} \quad & \sum_{x \in \{-1,1\}^n} \psi(x) f(x) \\ \text{s.t.} \quad & \sum_{x \in \{-1,1\}^n} |\psi(x)| = 1 \\ & \sum_{x \in \{-1,1\}^n} \psi(x) q(x) = 0 \qquad \text{whenever } \deg q \leq d \end{aligned}$$

Dual Characterization of Approximate Degree

Theorem: $\deg_\epsilon(f) > d$ iff there exists a "dual polynomial" $\psi\colon \{-1,1\}^n \to \mathbb{R}$ with

(1) $\sum_{x \in \{-1,1\}^n} \psi(x) f(x) > \epsilon$ "high correlation with f"

(2) $\sum_{x \in \{-1,1\}^n} |\psi(x)| = 1$ "L_1-norm 1"

(3) $\sum_{x \in \{-1,1\}^n} \psi(x) q(x) = 0$, when $\deg q \leq d$ "pure high degree d"

Key technique in, e.g., [She07] [Lee09] [She09]

A **lossless** technique. Strong duality implies any approximate degree lower bound can be witnessed by dual polynomial.

Our Proof

- Recall our main result:

Theorem

Let f be a Boolean function with $\widetilde{odeg}_{+,1/2}(f) \geq d$. Let $F = \mathrm{OMB}_t(f, \ldots, f)$. Then $\widetilde{odeg}_{+,1-2^{-t}}(F) \geq d$.

- Proved by showing how to take any dual witness to the fact that $\widetilde{\mathrm{odeg}}_{+,1/2}(f) \geq d$ and turn it into a dual witness for the statement in the theorem.

Our Proof

- Recall our main result:

Theorem

Let f be a Boolean function with $\widetilde{\mathrm{odeg}}_{+,1/2}(f) \geq d$. Let $F = \mathrm{OMB}_t(f, \ldots, f)$. Then $\widetilde{\mathrm{odeg}}_{+,1-2^{-t}}(F) \geq d$.

- Proved by showing how to take any dual witness to the fact that $\widetilde{\mathrm{odeg}}_{+,1/2}(f) \geq d$ and turn it into a dual witness for the statement in the theorem.
- Our construction differs substantially from the dual witnesses of prior work (Bun and Thaler, Sherstov).

Our Proof

- Recall our main result:

Theorem

Let f be a Boolean function with $\widetilde{odeg}_{+,1/2}(f) \geq d$. Let $F = \mathrm{OMB}_t(f, \ldots, f)$. Then $\widetilde{odeg}_{+,1-2^{-t}}(F) \geq d$.

- Proved by showing how to take any dual witness to the fact that $\widetilde{\mathrm{odeg}}_{+,1/2}(f) \geq d$ and turn it into a dual witness for the statement in the theorem.
- Our construction differs substantially from the dual witnesses of prior work (Bun and Thaler, Sherstov).
- Such new techniques are essential, as the "primal optimal" (approximating polynomials) for $\mathrm{OMB}_t(f, \ldots, f)$ are very different from the optimal approximating polynomials for $\mathrm{OR}_t(f, \ldots, f)$.

The Dual Witness

- Let ψ_{IN} be a dual witness for the fact that $\widetilde{\mathrm{odeg}}_{+,1/2}(f) \geq d$.
- Let $x_1, \ldots, x_t$ be inputs to f.
- The dual witness we construct for $F = \mathrm{OMB}_t(f, \ldots, f)$ is:

$$\psi_F(x_1, \ldots, x_t) := \sum_{i=1}^{t} \psi^{(i)}, \text{ where}$$

$$\psi^{(i)} = (-1)^{i-1} \cdot \left(\prod_{j<i} \mathbb{I}_E(x_j) \cdot |\psi_{\mathsf{IN}}(x_j)| \right) \cdot \psi_{\mathsf{IN}}(x_i) \cdot \left(\prod_{j>i} \mathbb{I}_{f^{-1}(1)}(x_j) \cdot |\psi_{\mathsf{IN}}(x_j)| \right),$$

where E is set of inputs on which ψ_{IN} "makes an error" (i.e., disagrees in sign with f).

Applications to Query and Communication Complexity

A Motivating Goal for This Work

- An important question in complexity theory is to determine the relative power of alternation (as captured by the polynomial-hierarchy PH), and counting (as captured by $\#\mathrm{P}$ and its decisional variant PP).
- Both PH and PP generalize NP in natural ways.
- Toda famously showed that their power is related: $\mathrm{PH} \subseteq \mathrm{P}^{\mathrm{PP}}$.
- But it is open how much of PH is contained in PP itself.

A Motivating Goal for This Work

- An important question in complexity theory is to determine the relative power of alternation (as captured by the polynomial-hierarchy PH), and counting (as captured by $\#\mathrm{P}$ and its decisional variant PP).
- Both PH and PP generalize NP in natural ways.
- Toda famously showed that their power is related: $\mathrm{PH} \subseteq \mathrm{P}^{\mathrm{PP}}$.
- But it is open how much of PH is contained in PP itself.
- It is interesting to study the analogous question in the settings of query and communication complexity.

A Motivating Goal for This Work

- An important question in complexity theory is to determine the relative power of alternation (as captured by the polynomial-hierarchy PH), and counting (as captured by $\#\mathrm{P}$ and its decisional variant PP).
- Both PH and PP generalize NP in natural ways.
- Toda famously showed that their power is related: $\mathrm{PH} \subseteq \mathrm{P}^{\mathrm{PP}}$.
- But it is open how much of PH is contained in PP itself.
- It is interesting to study the analogous question in the settings of query and communication complexity.
- Beigel (1992) used OMB to give an oracle (i.e., a query problem) relative to which $\mathrm{P}^{\mathrm{NP}} \not\subseteq \mathrm{PP}$.

A Motivating Goal for This Work

- An important question in complexity theory is to determine the relative power of alternation (as captured by the polynomial-hierarchy PH), and counting (as captured by $\#\mathrm{P}$ and its decisional variant PP).
- Both PH and PP generalize NP in natural ways.
- Toda famously showed that their power is related: $\mathrm{PH} \subseteq \mathrm{P}^{\mathrm{PP}}$.
- But it is open how much of PH is contained in PP itself.
- It is interesting to study the analogous question in the settings of query and communication complexity.
- Beigel (1992) used OMB to give an oracle (i.e., a query problem) relative to which $\mathrm{P}^{\mathrm{NP}} \not\subseteq \mathrm{PP}$.
- Buhrman, Vershchagin, and de Wolf (2008) "lifted" the result to communication complexity.
  - They gave a problem that is in the communication analogue of P^{NP}, but not in the communication analogue of PP.

Our Improvements

- Quantitatively, Beigel and Buhrman et al. gave functions in the query and communication analogues of P^{NP}, but any PP algorithm for the problem has cost $\Omega(n^{1/3})$.
- Our results improve the PP cost to $\Omega(n^{2/5})$.
- Our proof also yields the first explicit distributions under which the functions are "hard" for PP.

Our Improvements

- Quantitatively, Beigel and Buhrman et al. gave functions in the query and communication analogues of P^{NP}, but any PP algorithm for the problem has cost $\Omega(n^{1/3})$.
- Our results improve the PP cost to $\Omega(n^{2/5})$.
- Our proof also yields the first explicit distributions under which the functions are "hard" for PP.
- Upcoming work with Bun: improved the PP cost further to nearly $\Omega(n^{2/3})$, with additional applications to learning theory, communication complexity, and circuit complexity.
- Requires a hardness amplification method that goes beyond block-composed functions!