Approximate Degree and the Method of Dual Polynomials

Mark Bun[1] Justin Thaler[2]

[1]Harvard University [2]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

- 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

Why Care About Approximate Degree?

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

- $\epsilon \to 1$: PAC learning [KS01]
- ϵ "close to" 1: Attribute-Efficient Learning [KS04, STT12]
- $\epsilon < 1$ a constant: Agnostic Learning [KKMS05]

Why Care About Approximate 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]

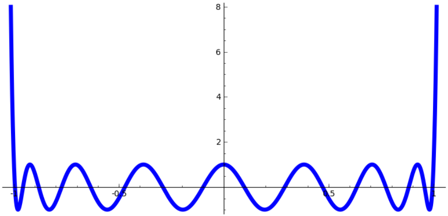
Example: What is the Approximate Degree of AND_n?

$\widetilde{\deg}(\mathrm{AND}_n) = \Theta(\sqrt{n})$.

- Upper bound: Use **Chebyshev Polynomials**.
- Markov's Inequality: Let $G(t)$ be a univariate polynomial s.t. $\deg(G) \leq d$ and $\sup_{t \in [-1,1]} |G(t)| \leq 1$. Then

$$\sup_{t \in [-1,1]} |G'(t)| \leq d^2.$$

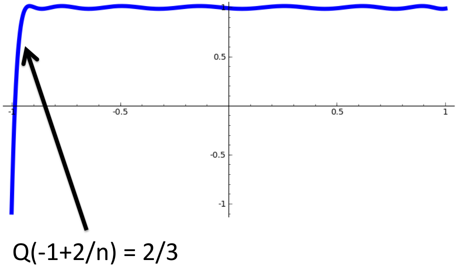
- Chebyshev polynomials are the extremal case.



Example: What is the Approximate Degree of AND_n?

$\widetilde{\deg}(\text{AND}_n) = O(\sqrt{n})$.

- After shifting a scaling, can turn degree $O(\sqrt{n})$ Chebyshev polynomial into a univariate polynomial $Q(t)$ that looks like:

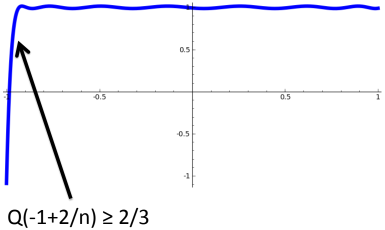


- Define n-variate polynomial p via $p(x) = Q(\sum_{i=1}^{n} x_i/n)$.
- Then $|p(x) - \text{AND}_n(x)| \le 1/3 \quad \forall x \in \{-1, 1\}^n$.

Example: What is the Approximate Degree of AND_n?

[NS92] $\widetilde{\deg}(\mathrm{AND}_n) = \Omega(\sqrt{n})$.

- Lower bound: Use **symmetrization**.
- Suppose $|p(x) - \mathrm{AND}_n(x)| \leq 1/3 \quad \forall x \in \{-1,1\}^n$.
- There is a way to turn p into a univariate polynomial p^{sym} that looks like this:

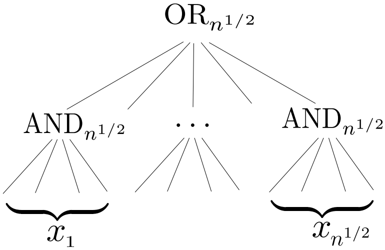


- Claim 1: $\deg(p^{\mathsf{sym}}) \leq \deg(p)$.
- Claim 2: Markov's inequality $\Longrightarrow \deg(p^{\mathsf{sym}}) = \Omega(n^{1/2})$.

Beyond Symmetrization: Analyzing the OR-AND Tree

Beyond Symmetrization

- Symmetrization is "lossy": in turning an n-variate poly p into a univariate poly p^{sym}, we throw away information about p.
- Challenge problem: What is $\widetilde{\deg}(\text{OR-AND}_n)$?



History of the OR-AND Tree

Upper bounds

[HMW03] $\widetilde{\deg}(\text{OR-AND}_n) = O(n^{1/2})$

Lower bounds

[NS92]	$\Omega(n^{1/4})$
[Shi01]	$\Omega(n^{1/4}\sqrt{\log n})$
[Amb03]	$\Omega(n^{1/3})$
[Aar08]	Reposed Question
[She09]	$\Omega(n^{3/8})$
[BT13]	$\Omega(n^{1/2})$
[She13]	$\Omega(n^{1/2})$, independently

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: $\widetilde{\deg}_\epsilon(f) > d$ iff there exists a "dual polynomial" ψ 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) $\sum_{x\in\{-1,1\}^n} \psi(x) q(x) = 0, \deg q \leq d$ "pure high degree d"

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

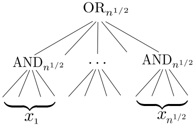
A **lossless** technique. Strong duality implies any approximate degree lower bound can be witnessed by dual polynomial.

Goal: Construct an explicit dual polynomial $\psi_{\textbf{OR-AND}}$ for $\mathrm{OR}\text{-}\mathrm{AND}$

Constructing a Dual Polynomial

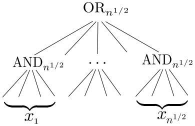
- By [NS92], there are dual polynomials
  $\psi_{\textsf{OUT}}$ for $\widetilde{\deg}\,(\mathrm{OR}_{n^{1/2}}) = \Omega(n^{1/4})$ and
  $\psi_{\textsf{IN}}$ for $\widetilde{\deg}\,(\mathrm{AND}_{n^{1/2}}) = \Omega(n^{1/4})$
- Can we combine $\psi_{\textsf{OUT}}$ and $\psi_{\textsf{IN}}$ to obtain a dual polynomial $\psi_{\textsf{OR-AND}}$ for OR-AND?

A First Attempt



$$\psi_{\textbf{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := \psi_{\textbf{OUT}}(\ldots, \psi_{\textbf{IN}}(x_i), \ldots)$$

A First Attempt

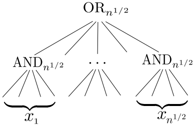


$$\psi_{\textbf{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := \psi_{\textbf{OUT}}(\ldots, \psi_{\textbf{IN}}(x_i), \ldots)$$

- Easy to check: $\psi_{\textbf{OR-AND}}$ has pure high degree at least $n^{1/4} \cdot n^{1/4} = n^{1/2}$.
- E.g. If $\psi_{\textbf{OUT}}(y_1, y_2) = y_1 y_2$ and $\psi_{\textbf{IN}}(z_1, z_2) = z_1 z_2$, then

  $$\psi_{\textbf{OR-AND}}(x_{11}, x_{12}, x_{21}, x_{22}) = (x_{11}x_{12})(x_{21}x_{22}) = x_{11}x_{12}x_{21}x_{22}.$$

A First Attempt



$$\psi_{\textbf{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := \psi_{\textbf{OUT}}(\ldots, \psi_{\textbf{IN}}(x_i), \ldots)$$

- Easy to check: $\psi_{\textbf{OR-AND}}$ has pure high degree at least $n^{1/4} \cdot n^{1/4} = n^{1/2}$.
- E.g. If $\psi_{\textbf{OUT}}(y_1, y_2) = y_1 y_2$ and $\psi_{\textbf{IN}}(z_1, z_2) = z_1 z_2$, then

  $$\psi_{\textbf{OR-AND}}(x_{11}, x_{12}, x_{21}, x_{22}) = (x_{11}x_{12})(x_{21}x_{22}) = x_{11}x_{12}x_{21}x_{22}.$$

- Does $\psi_{\textbf{OR-AND}}$ have high correlation with $\mathrm{OR}\text{-}\mathrm{AND}_n$?
- Problem: Proposed definition of $\psi_{\textbf{OR-AND}}$ may feed non-Boolean values into $\psi_{\textbf{OUT}}$. But we only have control over $\psi_{\textbf{OUT}}$ on **Boolean** inputs.

A Second (and Final) Attempt [She09, Lee09]

$$\psi_{\textsf{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\textsf{OUT}}(\ldots, \operatorname{sgn}(\psi_{\textsf{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\textsf{IN}}(x_i)|$$

(C chosen to ensure $\psi_{\textsf{OR-AND}}$ has L_1-norm 1).

A Second (and Final) Attempt [She09, Lee09]

$$\psi_{\textsf{OR-AND}}(x_1, \dots, x_{n^{1/2}}) := C \cdot \psi_{\textsf{OUT}}(\dots, \operatorname{sgn}(\psi_{\textsf{IN}}(x_i)), \dots) \prod_{i=1}^{n^{1/2}} |\psi_{\textsf{IN}}(x_i)|$$

(C chosen to ensure $\psi_{\textsf{OR-AND}}$ has L_1-norm 1).

Must verify:

1. $\psi_{\textsf{OR-AND}}$ has pure high degree $\geq n^{1/4} \cdot n^{1/4} = n^{1/2}$.
2. $\psi_{\textsf{OR-AND}}$ has high correlation with $\mathrm{OR\text{-}AND}$.

A Second (and Final) Attempt [She09, Lee09]

$$\psi_{\textbf{OR-AND}}(x_1, \dots, x_{n^{1/2}}) := C \cdot \psi_{\textbf{OUT}}(\dots, \mathrm{sgn}(\psi_{\textbf{IN}}(x_i)), \dots) \prod_{i=1}^{n^{1/2}} |\psi_{\textbf{IN}}(x_i)|$$

(C chosen to ensure $\psi_{\textbf{OR-AND}}$ has L_1-norm 1).

Must verify:

1. $\psi_{\textbf{OR-AND}}$ has pure high degree $\geq n^{1/4} \cdot n^{1/4} = n^{1/2}$.✓[She09]
2. $\psi_{\textbf{OR-AND}}$ has high correlation with $\mathrm{OR\text{-}AND}$. [BT13]

(Sub)Goal: Show $\psi_{\textbf{OR-AND}}$ has pure high degree at least $n^{1/2}$ [She09]

Pure High Degree Analysis [She09]

$$\psi_{\textbf{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\textbf{OUT}}(\ldots, \mathrm{sgn}(\psi_{\textbf{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\textbf{IN}}(x_i)|$$

- Intuition: Consider $\psi_{\textbf{OUT}}(y_1, y_2, y_3) = y_1 y_2$. Then $\psi_{\textbf{OR-AND}}(x_1, x_2, x_3)$ equals:

$$C \cdot \mathrm{sgn}(\psi_{\textbf{IN}}(x_1)) \cdot \mathrm{sgn}(\psi_{\textbf{IN}}(x_2)) \cdot \prod_{i=1}^{3} |\psi_{\textbf{IN}}(x_i)|$$
$$= \psi_{\textbf{IN}}(x_1) \cdot \psi_{\textbf{IN}}(x_2) \cdot |\psi_{\textbf{IN}}(x_3)|$$

Pure High Degree Analysis [She09]

$$\psi_{\textbf{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\textbf{OUT}}(\ldots, \mathrm{sgn}(\psi_{\textbf{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\textbf{IN}}(x_i)|$$

- Intuition: Consider $\psi_{\textbf{OUT}}(y_1, y_2, y_3) = y_1 y_2$. Then $\psi_{\textbf{OR-AND}}(x_1, x_2, x_3)$ equals:

$$C \cdot \mathrm{sgn}(\psi_{\textbf{IN}}(x_1)) \cdot \mathrm{sgn}(\psi_{\textbf{IN}}(x_2)) \cdot \prod_{i=1}^{3} |\psi_{\textbf{IN}}(x_i)|$$
$$= \psi_{\textbf{IN}}(x_1) \cdot \psi_{\textbf{IN}}(x_2) \cdot |\psi_{\textbf{IN}}(x_3)|$$

- Each term of $\psi_{\textbf{OR-AND}}$ is the product of PHD($\psi_{\textbf{OUT}}$) polynomials over disjoint variable sets, each of pure high degree at least PHD($\psi_{\textbf{IN}}$).
- So $\psi_{\textbf{OR-AND}}$ has pure high degree at least PHD($\psi_{\textbf{OUT}}$) $\cdot$ PHD($\psi_{\textbf{IN}}$).

(Sub)Goal: Show $\psi_{\textsf{\textbf{OR-AND}}}$ has high correlation with OR-AND

Correlation Analysis

$$\psi_{\textsf{\textbf{OR-AND}}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\textsf{\textbf{OUT}}}(\ldots, \operatorname{sgn}(\psi_{\textsf{\textbf{IN}}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\textsf{\textbf{IN}}}(x_i)|$$

- Idea: Show

$$\sum_{x \in \{-1,1\}^n} \psi_{\textsf{\textbf{OR-AND}}}(x) \cdot \mathrm{OR\text{-}AND}_n(x) \approx \sum_{y \in \{-1,1\}^{n^{1/2}}} \psi_{\textsf{\textbf{OUT}}}(y) \cdot \mathrm{OR}_{n^{1/2}}(y).$$

- Intuition: We are feeding $\operatorname{sgn}(\psi_{\textsf{\textbf{IN}}}(x_i))$ into $\psi_{\textsf{\textbf{OUT}}}$.
- $\psi_{\textsf{\textbf{IN}}}$ is **correlated** with $\mathrm{AND}_{n^{1/2}}$, so $\operatorname{sgn}(\psi_{\textsf{\textbf{IN}}}(x_i))$ is a "decent predictor" of $\mathrm{AND}_{n^{1/2}}$.
- But there are errors. Need to show errors don't "build up".

Correlation Analysis

$$\psi_{\textbf{OR-AND}}(x_1, \ldots, x_{n^{1/2}}) := C \cdot \psi_{\textbf{OUT}}(\ldots, \mathrm{sgn}(\psi_{\textbf{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/2}} |\psi_{\textbf{IN}}(x_i)|$$

- Goal: Show

$$\sum_{x \in \{-1,1\}^n} \psi_{\textbf{OR-AND}}(x) \cdot \mathrm{OR\text{-}AND}_n(x) \approx \sum_{y \in \{-1,1\}^{n^{1/2}}} \psi_{\textbf{OUT}}(y) \cdot \mathrm{OR}_{n^{1/2}}(y).$$

- Case 1: Consider any $y = (\mathrm{sgn}\,\psi_{\textbf{IN}}(x_1), \ldots, \mathrm{sgn}\,\psi_{\textbf{IN}}(x_{n^{1/2}})) \neq$ **All-False**.
- There is some coordinate of y that equals TRUE. Only need to "trust" this coordinate to force $\mathrm{OR\text{-}AND}_n$ to evaluate to True on $(x_1, \ldots, x_{n^{1/2}})$. So errors do not build up!

Correlation Analysis

- Case 2: Consider $y =$ **All-False**.
- $\mathrm{OR}_{n^{1/2}}(y) = \mathrm{OR\text{-}AND}_n(x_1, \ldots, x_{n^{1/2}})$ only if <u>all</u> coordinates of y are "error-free".
- Fortunately, ψ_{IN} has a special **one-sided error** property:

  If $\mathrm{sgn}(\psi_{\mathsf{IN}}(x_i)) = 1$, then $\mathrm{AND}_{n^{1/2}}(x_i)$ is **guaranteed** to equal 1.

Summary of Correlation Analysis

- Two Cases.
- In first case (feeding at least one TRUE into $\psi_{\textbf{OUT}}$), errors did not build up, because we only needed to "trust" the TRUE value.
- In second case (all values fed into $\psi_{\textbf{OUT}}$ are FALSE), we needed to trust <u>all</u> values. But we could do this because $\psi_{\textbf{IN}}$ had one-sided error.

(Negative) One-Sided Approximate 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 one-sided ϵ-approximation for f.

Dual Formulation of $\widetilde{\text{odeg}}_{-,\epsilon}$

Primal LP (Linear in ϵ and coefficients of p):

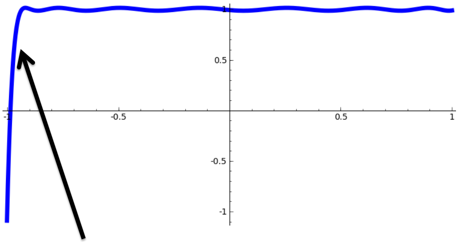
$$\begin{aligned} \min_{p,\epsilon} \quad & \epsilon \\ \text{s.t.} \quad & |p(x) - 1| \le \epsilon && \text{for all } x \in f^{-1}(1) \\ & p(x) \le -1 && \text{for all } x \in f^{-1}(-1) \\ & \deg p \le d \end{aligned}$$

Dual LP:

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

Proof that $\widetilde{\mathrm{odeg}}_{-,1/3}(\mathrm{AND}_n) = \Omega(\sqrt{n})$

We argued that the symmetrization of any $1/3$-approximation to AND_n had to look like this:



Q(-1+2/n) ≥ 2/3

Hardness Amplification for Constant-Depth Circuits [BT14]

Main Theorem

- Given: A “simple” Boolean function f that is “hard to approximate to low error” by degree d polynomials.
- Can we turn f into a “still-simple” F that is hard to approximate even to very high error?

Main Theorem

- Given: A "simple" Boolean function f that is "hard to approximate to low error" by degree d polynomials.
- Can we turn f into a "still-simple" F that is hard to approximate even to very high error?

A: Yes.

Theorem

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

Proof of Main Theorem

- Define $\psi_{\textbf{IN}}$ to be any dual witness to the fact that $\widetilde{\text{odeg}}_{-,1/3}(f) \geq d$.
- Define $\psi_{\textbf{OUT}} : \{-1,1\}^t \to \mathbb{R}$ via:

$$\psi_{\textbf{OUT}}(y) = \begin{cases} 1/2 & \text{if } y = \textbf{ ALL-FALSE} \\ -1/2 & \text{if } y = \textbf{ ALL-TRUE} \\ 0 & \text{otherwise} \end{cases}$$

- Combine $\psi_{\textbf{OUT}}$ and $\psi_{\textbf{IN}}$ exactly as before to obtain a dual witness ψ_F for F.

Must verify:

1. ψ_F has pure high degree d.
2. ψ_F has correlation at least $1 - 2^{-t}$ with F.

Proof of Main Theorem: Pure High Degree

- Notice $\psi_{\textbf{OUT}}$ is balanced (i.e., it has pure high degree 1).
- So previous analysis shows ψ_F has pure high degree at least $1 \cdot d = d$.

Proof of Main Theorem: Correlation Analysis

$$\psi_F(x_1, \ldots, x_t) := C \cdot \psi_{\textbf{OUT}}(\ldots, \operatorname{sgn}(\psi_{\textbf{IN}}(x_i)), \ldots) \prod_{i=1}^{t} |\psi_{\textbf{IN}}(x_i)|$$

- Idea: Show

$$\sum_{x \in \{-1,1\}^n} \psi_F(x) \cdot F(x) \geq \sum_{y \in \{-1,1\}^t} \psi_{\textbf{OUT}}(y) \cdot \mathrm{OR}_t(y) - 2^{-t} = 1 - 2^{-t}.$$

- Case 1: Consider $y = (\operatorname{sgn} \psi_{\textbf{IN}}(x_1), \ldots, \operatorname{sgn} \psi_{\textbf{IN}}(x_t)) =$ **All-True**.
- If even a single coordinate y_i of y is "error-free", then $F(x) = \mathrm{OR}_t(f(x_1), \ldots, f(x_t)) = -1$. :-D
- Any individual coordinate of y is in error with probability at most $1/2$, since $\psi_{\textbf{IN}}$ is well-correlated with f.
- So **all** coordinates of y are in error with probability only 2^{-t}.

Proof of Main Theorem: Correlation Analysis

$$\psi_F(x_1, \ldots, x_t) := C \cdot \psi_{\textsf{OUT}}(\ldots, \operatorname{sgn}(\psi_{\textsf{IN}}(x_i)), \ldots) \prod_{i=1}^{t} |\psi_{\textsf{IN}}(x_i)|$$

- Idea: Show

$$\sum_{x \in \{-1,1\}^n} \psi_F(x) \cdot F(x) \geq \sum_{y \in \{-1,1\}^t} \psi_{\textsf{OUT}}(y) \cdot \mathrm{OR}_t(y) - 2^{-t} = 1 - 2^{-t}.$$

- Case 2: Consider $y = (\operatorname{sgn}\psi_{\textsf{IN}}(x_1), \ldots, \operatorname{sgn}\psi_{\textsf{IN}}(x_t)) =$ **All-False**. Then $\operatorname{sgn}(\psi_F(x)) = \operatorname{sgn}(\psi_{\textsf{OUT}}(y)) = 1$.
- Then $F(y) = \mathrm{OR}_t(f(x_1), \ldots, f(x_t) = 1$ only if all coordinates of y are "error-free".
- Fortunately, $\psi_{\textsf{IN}}$ has one-sided error: If $\operatorname{sgn}(\psi_{\textsf{IN}}(x_i)) = 1$, then $f(x_i)$ is **guaranteed** to equal 1.

A New $\widetilde{\text{odeg}}_{-,1/3}$ Bound for AC^0

- We want to apply amplification to functions in AC^0, getting out very "hard" functions that are still in AC^0.
- Let $\text{ED} : \{-1, 1\}^n \to \{-1, 1\}$ denote the ELEMENT DISTINCTNESS function.

- [AS04] showed $\widetilde{\deg}(\text{ED}) = \Omega(N^{2/3})$.
- This is the best known lower bound on the approximate degree of an AC^0 function.
- We show that in fact $\widetilde{\text{odeg}}_{-,1/3}(\text{ED}) = \Omega(N^{2/3})$.

New Lower Bounds for AC^0

Theorem

Let $F = \mathrm{OR}_{n^{2/5}}(\mathrm{ED}_{n^{3/5}}, \ldots, \mathrm{ED}_{n^{3/5}})$ *and* $\epsilon = 1 - 2^{-n^{2/5}}$. *Then* $\widetilde{odeg}_{-,\epsilon}(F) = \tilde{\Omega}(n^{2/5})$.

Proof: Combine lower bound on $\widetilde{\mathrm{odeg}}_{-,1/3}(\mathrm{ED})$ with Main Theorem.

New Lower Bounds for AC^0

Definition

Let $f : X \times Y \to \{-1, 1\}$ be a function, and μ a probability distribution on $X \times Y$. The discrepancy of f under μ is

$$\mathrm{disc}_\mu(f) := \max_{S \subseteq X, T \subseteq Y} \left| \sum_{x \in S} \sum_{y \in T} \mu(x, y) f(x, y) \right|.$$

The discrepancy of f is: $\mathrm{disc}(f) := \min_\mu \mathrm{disc}_\mu(f)$.

- Low discrepancy implies high communication complexity in nearly every communication model.
- Also a central quantity in learning theory and circuit complexity.

New Lower Bounds for AC^0

Theorem (She08, "Pattern Matrix Method")

Let $F : \{-1,1\}^n$ *be any function satisfying* $\widetilde{\deg}_{1-1/W}(F) \geq d$. *Let* $F' : \{-1,1\}^{4n} \times \{-1,1\}^{4n} \to \{-1,1\}$ *by*

$$F'(x,y) = F(\ldots, \vee_{j=1}^{4}(x_{i,j} \wedge y_{i,j}), \ldots).$$

Then $\mathrm{disc}(F') \lesssim \max\{1/W, 2^{-d}\}$.

Corollary

There is an AC^0 *function* f *(computed by a depth four circuit) with discrepancy* $\exp\left(-\Omega(n^{2/5})\right)$.

Proof: Apply Pattern Mat. Meth. to $\mathrm{OR}_{n^{2/5}}(\mathrm{ED}_{n^{3/5}}, \ldots, \mathrm{ED}_{n^{3/5}})$.

Previous best bound: $\exp\left(-\Omega(n^{1/3})\right)$ [She08, BVdW08].

More applications

Corollary

There is an AC^0 function f that cannot be computed by $MAJ \circ THR$ circuits of size $\exp\left(\Omega(n^{2/5})\right)$.

Corollary

There is an AC^0 function f with threshold weight $\exp\left(\Omega(n^{2/5})\right)$.

Subsequent Work by Sherstov [She14]

Threshold Degree

Definition

Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be a Boolean function. A polynomial p sign-represents f if $\mathrm{sgn}(p(x)) = f(x)$ for all $x \in \{-1, 1\}^n$.

Definition

The threshold degree of f is $\min \deg(p)$, where the minimum is over all sign-representations of f.

Threshold Degree of AC^0

- Minsky and Papert [MP69] proved an $\Omega(n^{1/3})$ lower bound on the threshold degree of a specific DNF.
- It had been open ever since to prove a lower bound of $\Omega(n^{1/3+\delta})$ for any function in AC^0.
- Only progress: $\Omega(n^{1/3}\log^k n)$ for any constant k [OS03].
- We conjectured in [BT14] that $\mathrm{OR}_{n^{2/5}}(\mathrm{ED}_{n^{3/5}}, \ldots, \mathrm{ED}_{n^{3/5}})$ has threshold degree $\Omega(n^{2/5})$.

Subsequent Work

- Sherstov [She14] proved our conjecture.

Theorem (She14)

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

- More generally, he exhibits a depth k circuit of polynomial size with threshold degree $\Omega(n^{(k-1)/(2k-1)})$.

Additional Intuition and Results [Tha14]

Robustification: A Generic Approximation Technique

- Sherstov [She13] showed that for any polynomial $p : \{-1,1\}^t \to [-1,1]$, and every $\delta > 0$, there is a polynomial $p_{\text{robust}} : \mathbb{R}^t \to \mathbb{R}$ of degree $O(\deg(p) + \log(1/\delta))$ such that

$$|p(y) - p_{\text{robust}}(y + \mathbf{e})| < \delta$$

for all $y \in \{-1,1\}^t$ and $\mathbf{e} \in [1/3, 1/3]^t$.

Robustification: A Generic Approximation Technique

- Sherstov [She13] showed that for any polynomial $p : \{-1,1\}^t \to [-1,1]$, and every $\delta > 0$, there is a polynomial $p_{\text{robust}} : \mathbb{R}^t \to \mathbb{R}$ of degree $O(\deg(p) + \log(1/\delta))$ such that

$$|p(y) - p_{\text{robust}}(y + \mathbf{e})| < \delta$$

  for all $y \in \{-1,1\}^t$ and $\mathbf{e} \in [1/3, 1/3]^t$.
- Given functions g, f, can construct an $(\epsilon + \delta)$-approximating polynomial for $g(f, \ldots, f)$ via:

$$p^* := p_{\text{robust}}(q, \ldots, q),$$

  where p is an ϵ-approximation for g, and q is a $(1/3)$-approximation for f.

Robustification: A Generic Approximation Technique

- Sherstov [She13] showed that for any polynomial $p : \{-1,1\}^t \to [-1,1]$, and every $\delta > 0$, there is a polynomial $p_{\text{robust}} : \mathbb{R}^t \to \mathbb{R}$ of degree $O(\deg(p) + \log(1/\delta))$ such that

$$|p(y) - p_{\text{robust}}(y + \mathbf{e})| < \delta$$

  for all $y \in \{-1,1\}^t$ and $\mathbf{e} \in [1/3, 1/3]^t$.

- Given functions g, f, can construct an $(\epsilon + \delta)$-approximating polynomial for $g(f, \ldots, f)$ via:

$$p^* := p_{\text{robust}}(q, \ldots, q),$$

  where p is an ϵ-approximation for g, and q is a $(1/3)$-approximation for f.

- This technique is optimal for all block-composed functions discussed so far.

Robustification and Complementary Slackness

- Throughout, we used a single "combining" technique to construct dual polynomials for block-composed functions.
- Namely, the dual polynomial for $F = g(f, \ldots, f)$ was:

$$\psi_F(x_1, \ldots, x_t) := C \cdot \psi_{\mathsf{OUT}}(\ldots, \mathrm{sgn}(\psi_{\mathsf{IN}}(x_i)), \ldots) \prod_{i=1}^{t} |\psi_{\mathsf{IN}}(x_i)|.$$

- Handwavy Claim: ψ_F is designed to witness the optimality of Robustification.
- Specifically, ψ_F "almost" obeys complementary slackness with respect to $p^* = p_{\mathsf{robust}}(q, \ldots, q)$.

Robustification Is Optimal (Except When It's Not)

- Suppose p_{robust} achieved exactly optimal error ϵ among all degree d approximations to the outer function g.
- Then p_{robust} yields an optimal solution to the linear program capturing $\widetilde{\deg}_\epsilon(g)$.
- Complementary slackness guarantees a dual polynomial $\psi_{\textbf{OUT}}$ that places non-zero weight only on primal constraints that are made tight by p_{robust}.
- i.e. $\psi_{\textbf{OUT}}(y) \neq 0$ only for "maximal error points" $y \in \{-1, 1\}^t$ satisfying $|p_{\text{robust}}(y) - g(y)| = \epsilon$.

Robustification Is Optimal (Except When It's Not)

Recall: $\psi_F(x_1, \ldots, x_t) := C \cdot \psi_{\textsf{OUT}}(\ldots, \mathrm{sgn}(\psi_{\textsf{IN}}(x_i)), \ldots) \prod_{i=1}^{t} |\psi_{\textsf{IN}}(x_i)|$

- If $\psi_{\textsf{IN}}$ were <u>perfectly</u> correlated with f, then $\psi_F(x_1, \ldots, x_t) \neq 0$ only for $(x_1, \ldots, x_t)$ such that

$$\left|p_{\textsf{robust}}\left(q\left(x_1\right), \ldots, q\left(x_t\right)\right) - g\left(f\left(x_1\right), \ldots, f\left(x_t\right)\right)\right| = \epsilon \pm \delta.$$

- That is, ψ_F places non-zero weight only on points on which $p_{\textsf{robust}}(q, \ldots, q)$ has "nearly maximal error" of at least $\epsilon - \delta$.

Q: Is Robustification Always Optimal?

- A: No.
- 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 $x_{i^*} = -1$.

Q: Is Robustification Always Optimal?

- A: No.
- 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 $x_{i^*} = -1$.
- Consider the function $F = \mathrm{OMB}_t(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$.

Q: Is Robustification Always Optimal?

- A: No.
- 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 $x_{i^*} = -1$.
- Consider the function $F = \mathrm{OMB}_t(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$.
- Optimal ϵ-approximating polynomial for F is of the form $p^* = p(q, \ldots, q)$, where p is a non-robust ϵ-approximating polynomial for OMB_t and q is an approximation for OR_N (cf. [KS04, Beigel94]).

A Final 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 $\widetilde{odeg}_{+,1-2^{-t}}(F) \geq d$.

A Final 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 $\widetilde{odeg}_{+,1-2^{-t}}(F) \geq d$.

Compare to our earlier theorem:

Theorem

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

- Our new result applies to functions F with low threshold degree (e.g., $\mathrm{OMB}_t(\mathrm{OR}, \ldots, \mathrm{OR})$ has threshold degree 1).
- The earlier theorem only applied to functions with threshold degree at least t [She14].

Application

- Corollary: we exhibit an AC^0 function F' with dimension complexity $n^{O(\log n)}$ and discrepancy $\exp(-\tilde{\Omega}(n^{2/5}))$.
- Previous best separation for AC^0 was due to [BVdW08], who gave an AC^0 function with dimension-complexity poly(n) and discrepancy $\exp(-\Omega(n^{1/3}))$.
- Key technical ingredient for our result:

Lemma

Let $F = \mathrm{OMB}_{n^{2/5}}(\overline{\mathrm{ED}}_{n^{3/5}}, \ldots, \overline{\mathrm{ED}}_{n^{3/5}})$ *and* $\epsilon = 1 - 2^{-\tilde{\Omega}(n^{2/5})}$. *Then* $\deg_{\pm}(F) = O(\log n)$, *while* $\widetilde{\deg}_{\epsilon}(F) = \tilde{\Omega}(n^{2/5})$.

The Dual Witness

The dual witness we construct for $F = \mathrm{OMB}(f, \ldots, f)$ is:

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

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

where E is set of inputs on which $\psi_{\textbf{IN}}$ "makes an error".

Thank you!