Hardness Amplification and the Approximate Degree of Constant Depth Circuits

Mark Bun[1] and Justin Thaler[2]

[1]Harvard University

[2]Simons Institute for the Theory of Computing, UC Berkeley

5 December 2013

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 a Boolean function 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 [BVW07] [She07] [SZ07] [CA08] [LS08] [She12]
- Circuit complexity [MP69] [Bei93] [Bei94] [She08]

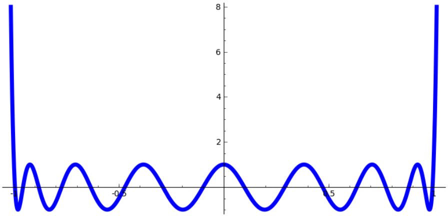
Example: What is the Approximate Degree of AND_n?

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

- 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

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

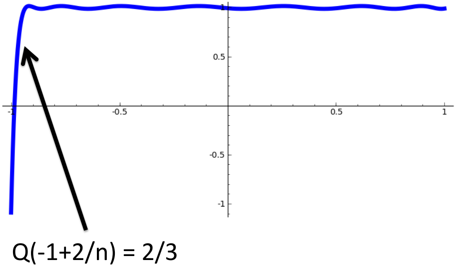
- Chebyshev polynomials are the extremal case.



Example: What is the Approximate Degree of AND_n?

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

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

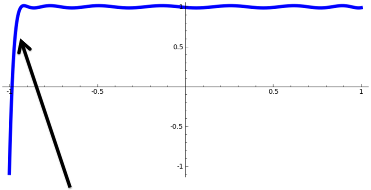


- Define n-variate polynomial p via $p(x) = Q(\sum_{i=1}^{n} x_i/n)$.
- Then $|p(x) - \mathrm{AND}_n(x)| \leq 1/3 \quad \forall x \in \{-1,1\}^n$.

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:



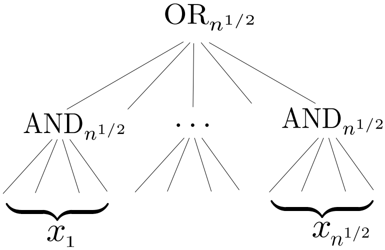
$p^{\mathrm{sym}}(-1+2/n) \geq 2/3$

- 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 $\widetilde{\deg}(\mathrm{OR}\text{-}\mathrm{AND}_n)$?



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] $\Omega(n^{3/8})$
[BT13a] $\Omega(n^{1/2})$
[She13a] $\Omega(n^{1/2})$, 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):

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

Dual LP:

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

Dual Characterization of Approximate Degree

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

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

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

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

(3) equivalent to: $\hat{\psi}(S) = 0$ for all $|S| \leq d$.

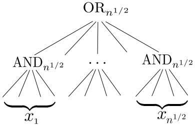
Key technique in, e.g., [She07] [Lee09] [She09]

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

Constructing a Dual Polynomial

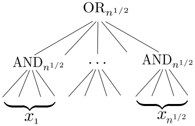
- By [NS92], there are dual polynomials
  $\psi_{\textbf{OUT}}$ for $\widetilde{\deg}\,(\mathrm{OR}_{n^{1/2}}) = \Omega(n^{1/4})$ and
  $\psi_{\textbf{IN}}$ for $\widetilde{\deg}\,(\mathrm{AND}_{n^{1/2}}) = \Omega(n^{1/4})$
- Can we combine $\psi_{\textbf{OUT}}$ and $\psi_{\textbf{IN}}$ to obtain a dual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

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

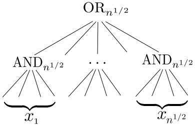


$$\psi_{\textbf{OR-AND}}(x_1, \dots, x_{n^{1/2}}) := \psi_{\textbf{OUT}}(\dots, \psi_{\textbf{IN}}(x_i), \dots)$$

- 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]

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

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

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).

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]

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

(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 $\mathrm{OR\text{-}AND}$. [BT13a]

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

Pure High Degree Analysis [She09]

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

- 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:

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

$$= \psi_{\textbf{IN}}(x_1) \cdot \psi_{\textbf{IN}}(x_2) \cdot |\psi_{\textbf{IN}}(x_3)|$$

Pure High Degree Analysis [She09]

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

- 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:

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

- 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 PHD($\psi_{\textbf{OR-AND}}$) $\geq$ PHD($\psi_{\textbf{OUT}}$)$\cdot$PHD($\psi_{\textbf{IN}}$).

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

Correlation Analysis

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

- Idea: Show

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

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

Correlation Analysis

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

- Goal: Show

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

- Case 1: Consider any $y = (\operatorname{sgn} \psi_{\textsf{\textbf{IN}}}(x_1), \ldots, \operatorname{sgn} \psi_{\textsf{\textbf{IN}}}(x_{n^{1/2}})) \neq$ **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**.
- $\mathrm{OR}_{n^{1/2}}(y) = \mathrm{OR}\text{-}\mathrm{AND}_n(x_1, \ldots, x_{n^{1/2}})$ 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.

One-Sided Approximate Degree

- A real polynomial p is a 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.
- $\widetilde{\mathrm{odeg}}(f) := \widetilde{\mathrm{odeg}}_{1/3}(f)$ is the one-sided approximate degree of f.

Dual Formulation of $\widetilde{\mathrm{odeg}}$

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

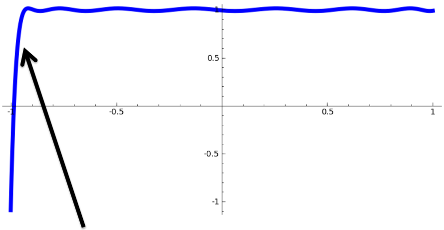
$$\begin{aligned}
\text{min}_{p,\epsilon} \quad & \epsilon \\
\text{s.t.} \quad & |p(x) - 1| \leq \epsilon && \text{for all } x \in f^{-1}(1) \\
& p(x) \leq -1 && \text{for all } x \in f^{-1}(-1) \\
& \deg p \leq d
\end{aligned}$$

Dual LP:

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

Proof that $\widetilde{\mathrm{odeg}}(\mathrm{AND}_n) = \Omega(\sqrt{n})$

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



$p^{sym}(-1+2/n) \geq 2/3$

Hardness Amplification for Constant-Depth Circuits [BT13b]

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

Proof of Main Theorem

- Define $\psi_{\textbf{IN}}$ to be any dual witness to the fact that $\widetilde{\mathrm{odeg}}(f) \geq d$.
- 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 $\psi_{\textsf{\textbf{OUT}}}$ 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

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

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

Proof of Main Theorem: Correlation Analysis

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

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

A New $\widetilde{\mathrm{odeg}}$ 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 $\mathrm{ED} : \{-1,1\}^n \to \{-1,1\}$ denote the $\textsc{Element Distinctness}$ function.

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

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* $\widetilde{\mathrm{odeg}}_\epsilon(F) = \tilde{\Omega}(n^{2/5})$.

Proof: Combine lower bound on $\widetilde{\mathrm{odeg}}(\mathrm{ED})$ 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, BVW07].

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 $\exp\left(\Omega(n^{2/5})\right)$.

Previous bests were both $\exp(\Omega(n^{1/3}))$ [Sher08, BVW07, KP97].

Back to OR-AND Trees

- Let $\mathrm{OR\text{-}AND}_{d,n}$ denote the balanced $\mathrm{OR\text{-}AND}$ tree of depth d (with an OR gate at the top).
- Earlier, we proved $\widetilde{\deg}(\mathrm{OR\text{-}AND}_{2,n}) = \Theta(n^{1/2})$.
- But proving equivalent lower bound for depth 3 or greater remained open.

Back to OR-AND Trees

- Let $\mathrm{OR\text{-}AND}_{d,n}$ denote the balanced $\mathrm{OR\text{-}AND}$ tree of depth d (with an OR gate at the top).
- Earlier, we proved $\widetilde{\deg}(\mathrm{OR\text{-}AND}_{2,n}) = \Theta(n^{1/2})$.
- But proving equivalent lower bound for depth 3 or greater remained open.

Theorem

For any constant $d > 1$, $\widetilde{\deg}(\mathrm{OR\text{-}AND}_{d,n}) = \Omega(n^{1/2}/\log^{d-2}(n))$.

(Upper bound of $O(n^{1/2})$ for any constant d follows from [She12]).

First Proof Attempt (for the case $d = 3$)

- Goal: construct a dual polynomial for $\mathrm{OR}\text{-}\mathrm{AND}_{3,n}$.
- Let ψ_{IN} denote the dual polynomial for $\mathrm{AND}\text{-}\mathrm{OR}_{2,n^{2/3}}$ constructed earlier.
- Let ψ_{OUT} denote a dual polynomial witnessing $\widetilde{\deg}(\mathrm{OR}_{n^{1/3}}) = \Omega(n^{1/6})$
- Combine ψ_{IN} and ψ_{OUT} exactly as before:

$$\psi_{\mathsf{COMB}}(x_1, \ldots, x_{n^{1/3}}) := C \cdot \psi_{\mathsf{OUT}}(\ldots, \mathrm{sgn}(\psi_{\mathsf{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/3}} |\psi_{\mathsf{IN}}(x_i)|$$

First Proof Attempt (for the case $d = 3$)

- Goal: construct a dual polynomial for $\mathrm{OR}\text{-}\mathrm{AND}_{3,n}$.
- Let $\psi_{\textsf{IN}}$ denote the dual polynomial for $\mathrm{AND}\text{-}\mathrm{OR}_{2,n^{2/3}}$ constructed earlier.
- Let $\psi_{\textsf{OUT}}$ denote a dual polynomial witnessing $\widetilde{\deg}(\mathrm{OR}_{n^{1/3}}) = \Omega(n^{1/6})$
- Combine $\psi_{\textsf{IN}}$ and $\psi_{\textsf{OUT}}$ exactly as before:

$$\psi_{\textsf{COMB}}(x_1, \ldots, x_{n^{1/3}}) := C \cdot \psi_{\textsf{OUT}}(\ldots, \mathrm{sgn}(\psi_{\textsf{IN}}(x_i)), \ldots) \prod_{i=1}^{n^{1/3}} |\psi_{\textsf{IN}}(x_i)|$$

- $\psi_{\textsf{COMB}}$ has p.h.d. $\Omega(n^{1/6} \cdot n^{1/3}) = \Omega(n^{1/2})$. ✓
- But $\psi_{\textsf{COMB}}$ may have poor correlation with $\mathrm{OR}\text{-}\mathrm{AND}_{3,n}$. Problem: $\psi_{\textsf{IN}}$ does not have one-sided error.

Actual Proof (for the case $d = 3$)

- Instead, use a <u>different</u> dual polynomial $\psi_{\mathbf{IN}}$ for $\mathrm{OR\text{-}AND}_{2,n^{2/3}}$.
- Construction of $\psi_{\mathbf{IN}}$ uses hardness amplification to achieve the following:
- $\psi_{\mathbf{IN}}$ has error "on both sides", but the error from the "wrong side" will be very small.
- Hardness amplification step causes $\psi_{\mathbf{IN}}$ to have p.h.d. $\Omega(n^{1/3}/\sqrt{\log n})$, rather than $\Omega(n^{1/3})$.

Subsequent Work by Sherstov [She13b]

Threshold Degree

Definition

Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. A polynomial p sign-represents f if $\operatorname{sgn}(p(x)) = f(x)$ 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. (Equivalent to $\lim_{\epsilon \to 1} \widetilde{\deg}_\epsilon(f)$).

Threshold Degree of AC^0

- Minsky and Papert [MP68] proved an $\Omega(n^{1/3})$ lower bound on the threshold degree of a specific DNF.
- It has 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 [BT13b] 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 [She13b] has recently proved our conjecture.
- More generally, he exhibits a depth k circuit of polynomial size with threshold degree $\Omega(n^{(k-1)/(2k-1)})$.