Hardness Amplification and the Approximate Degree of Constant Depth Circuits

Mark Bun[1] and Justin Thaler[2]

[1]Harvard University

[2]Yahoo! Labs

July 10, 2015

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

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

- $\widetilde{\deg}_\epsilon(f) =$ minimum degree needed to ϵ-approximate f
- $\widetilde{\deg}(f) := \widetilde{\deg}_{1/2}(f)$ is the approximate degree of f
- E.g. $\widetilde{\deg}(\mathrm{OR}_n) = \widetilde{\deg}(\mathrm{AND}_n) = \Theta(\sqrt{n})$ [NisanSzegedy92]

Why Care About Approximate Degree?

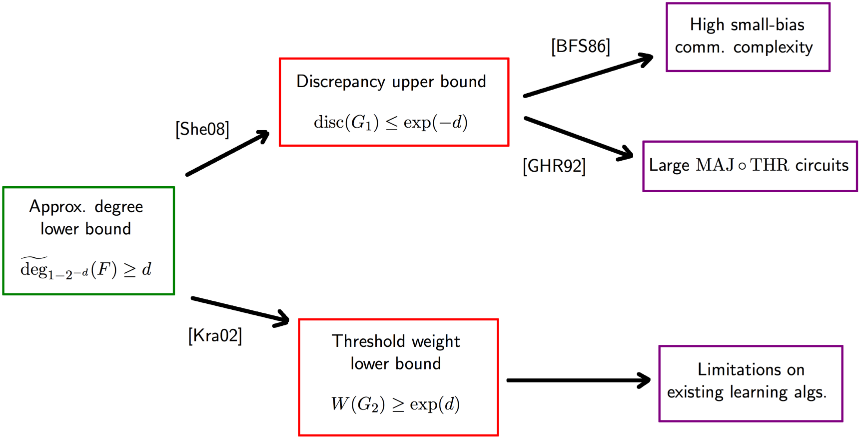
Upper bounds on $\widetilde{\deg}_\epsilon(f)$ give algorithms for

- Efficient learning [KS01, KS04, KKMS05, STT12]
- Approximate inclusion-exclusion [LN93, KLS96, She08]
- Differentially private query release [TUV12, CTUW14]

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

- Quantum query complexity [BBCMW98] [AS01] [Amb03] [KSW04]
- Communication complexity [BVW07] [She08] [SZ07] [CA08] [LS08] [She12]
- Circuit complexity [MP69] [Bei93] [Bei94] [She08]

Complexity of AC^0



Motivating Question

$f \in \mathrm{AC}^0$ hard to approximate by degree d polynomials with constant error

$$\Downarrow ?$$

$F \in \mathrm{AC}^0$ hard to approximate with very high error

Direct Product Theorems for Approximate Degree

- Direct product theorems: Computing $g(f, \ldots, f)$ requires more

  Resources (polynomial degree) and Error (ϵ)

  than computing f alone
- XOR lemma for approximate degree [OS03, Sherstov11]:

$$\widetilde{\deg}_{1-2^{-t}}(\underbrace{f \oplus f \oplus \cdots \oplus f}_{t \text{ copies}}) \gtrsim t \cdot \widetilde{\deg}(f)$$

Problem 1: $\mathrm{PARITY} \notin \mathrm{AC}^0$
Problem 2: $\widetilde{\deg}_{1-\frac{1}{2mt}}(\mathrm{OR}_t(\mathrm{OR}_m, \ldots, \mathrm{OR}_m)) = 1$

Our Contributions

- Identify the relaxed notion of "one-sided" approximate degree.
  - $\widetilde{\mathrm{deg}} \geq \widetilde{\mathrm{odeg}}$
  - Used implicitly in prior work [GS09] [BT13] [She13]
- **Theorem:** $\widetilde{\mathrm{odeg}}$ obeys hardness amplification within AC^0:

$$\widetilde{\mathrm{odeg}}_{1-2^{-t}}(\mathrm{OR}_t(f, \ldots, f)) \geq \widetilde{\mathrm{odeg}}(f)$$

- **Applications:**
  - New discrepancy upper bound and threshold weight lower bound for AC^0
  - Nearly tight approx. degree lower bound for regular $\mathrm{AND\text{-}OR}$ trees
  - Weight-degree tradeoffs for read-once DNF

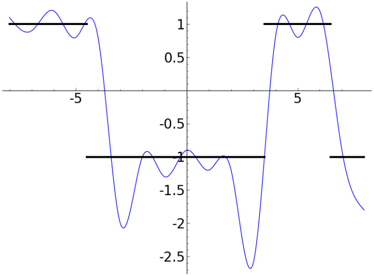
One-Sided Approximate Degree

- A real polynomial p is a one-sided ϵ-approximation for f if

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

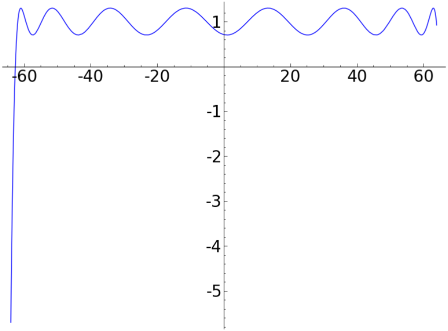
$$p(x) \leq -1 + \epsilon \quad \forall x \in f^{-1}(-1)$$

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



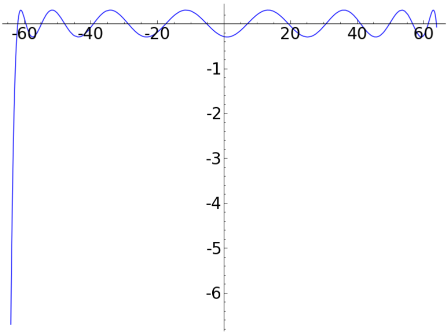
Some Observations about $\widetilde{\mathrm{odeg}}$

- $\widetilde{\mathrm{odeg}}(\mathrm{AND}_n) = \widetilde{\deg}(\mathrm{AND}_n) = \Omega(\sqrt{n})$
  I.e. a one-sided approximation to AND_n can be turned into an ordinary approximation with the same degree:



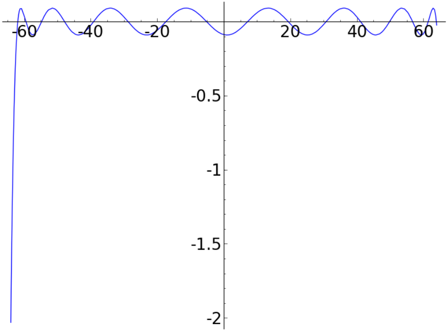
Some Observations about $\widetilde{\mathrm{odeg}}$

- $\widetilde{\mathrm{odeg}}(\mathrm{AND}_n) = \widetilde{\mathrm{deg}}(\mathrm{AND}_n) = \Omega(\sqrt{n})$
  I.e. a one-sided approximation to AND_n can be turned into an ordinary approximation with the same degree:



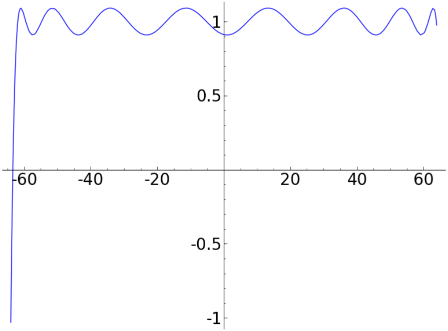
Some Observations about $\widetilde{\mathrm{odeg}}$

- $\widetilde{\mathrm{odeg}}(\mathrm{AND}_n) = \widetilde{\mathrm{deg}}(\mathrm{AND}_n) = \Omega(\sqrt{n})$
  I.e. a one-sided approximation to AND_n can be turned into an ordinary approximation with the same degree:



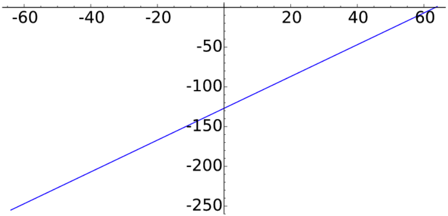
Some Observations about $\widetilde{\mathrm{odeg}}$

- $\widetilde{\mathrm{odeg}}(\mathrm{AND}_n) = \widetilde{\deg}(\mathrm{AND}_n) = \Omega(\sqrt{n})$
  I.e. a one-sided approximation to AND_n can be turned into an ordinary approximation with the same degree:



Some Observations about $\widetilde{\mathrm{odeg}}$

- $\widetilde{\mathrm{odeg}}(\mathrm{AND}_n) = \widetilde{\deg}(\mathrm{AND}_n) = \Omega(\sqrt{n})$
- $\widetilde{\mathrm{odeg}}(\mathrm{OR}_n) = \Theta(1)$



No amplification for OR!

Proof of Hardness Amplification: The Method of Dual Polynomials

LP Formulation of Approximate Degree [IT68, She08]

What is the 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" $\psi : \{-1,1\}^n \to \mathbb{R}$ 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) $$\sum_{x \in \{-1,1\}^n} \psi(x) q(x) = 0 \text{ if } \deg q \leq d$$ "pure high degree d"

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

Key technique in, e.g., [She08] [Lee09] [She09] [BT13] [She13]

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

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

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

Dual LP:

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

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

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

(4) $\psi(x) \leq 0$ for all $x \in f^{-1}(-1)$ "one-sided error"

Goal: Construct an explicit dual polynomial ψ_F for $\widetilde{\mathrm{odeg}}_{1-2^{-t}}(F) \geq d$

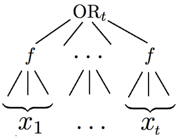
Constructing a Dual Polynomial

- Start with dual polynomials:
  - $\psi_{\textsf{IN}}$ for $\widetilde{\mathrm{odeg}}(f) = d$
  - Define $\psi_{\textsf{OUT}} : \{-1, 1\}^t \to \mathbb{R}$ by:

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

- Combine $\psi_{\textsf{OUT}}$ and $\psi_{\textsf{IN}}$ to obtain a dual polynomial ψ_F for F
- Follows construction used in [Lee09], [Sherstov09], [BunThaler13] with refined analysis

A First Attempt



$$\psi_F(x_1, \dots, x_n) := \psi_{\textsf{OUT}}(\dots, \psi_{\textsf{IN}}(x_i), \dots)$$

- ψ_F has pure high degree at least d because $\psi_{\textsf{OUT}}$ is balanced. E.g. If $\psi_{\textsf{OUT}}(y_1, y_2) = \frac{1}{4}(y_1 + y_2)$ and $\psi_{\textsf{IN}}(z_1, z_2) = z_1 z_2$, then

$$\psi_F(x_{11}, x_{12}, x_{21}, x_{22}) = \frac{1}{4}(x_{11}x_{12} + x_{21}x_{22}).$$

- Does ψ_F have high correlation with F?
- Problem: $\psi_{\textsf{IN}}$ might feed non-Boolean values into $\psi_{\textsf{OUT}}$. But we only have control over $\psi_{\textsf{OUT}}$ on **Boolean** inputs.

The Actual Construction [She09, Lee09]

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

(C chosen to ensure ψ_F has L_1-norm 1).

Must verify:

1. ψ_F has pure high degree d ✓[Lee09, Sherstov09]
2. ψ_F has one-sided error ✓By inspection
3. ψ_F has correlation at least $1 - 2^{-t}$ with F This work
   Builds on [B.Thaler13]

(Sub)Goal: Show ψ_F has high correlation with F

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

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

- Intuition: We are feeding $\operatorname{sgn}(\psi_{\textsf{IN}}(x_i))$ into $\psi_{\textsf{OUT}}$.
- $\psi_{\textsf{IN}}$ is **correlated** with f, so $\operatorname{sgn}(\psi_{\textsf{IN}}(x_i))$ is a "decent predictor" of f.
- But there are errors. Need to show errors decay exponentially.

Correlation Analysis

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

- Goal: 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-FALSE**.
- If even a single coordinate y_i of y is "truthful", then $F(x) = \mathrm{OR}_t(f(x_1), \ldots, f(x_t)) = -1$.
- 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}.

Correlation Analysis

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

- Goal: 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-TRUE**.
- Then $F(y) = \mathrm{OR}_t(f(x_1), \ldots, f(x_t)) = 1$ only if <u>all</u> coordinates of y are "truthful".
- 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.

Summary of Correlation Analysis

- Case 1 (feeding **ALL-TRUE** into $\psi_{\textbf{OUT}}$): Error decays like 2^{-t} because we only need to trust one coordinate.
- Case 2 (feeding **ALL-FALSE** into $\psi_{\textbf{OUT}}$): We need to trust <u>all</u> values. But we can because $\psi_{\textbf{IN}}$ has one-sided error.

Recap of the Proof

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

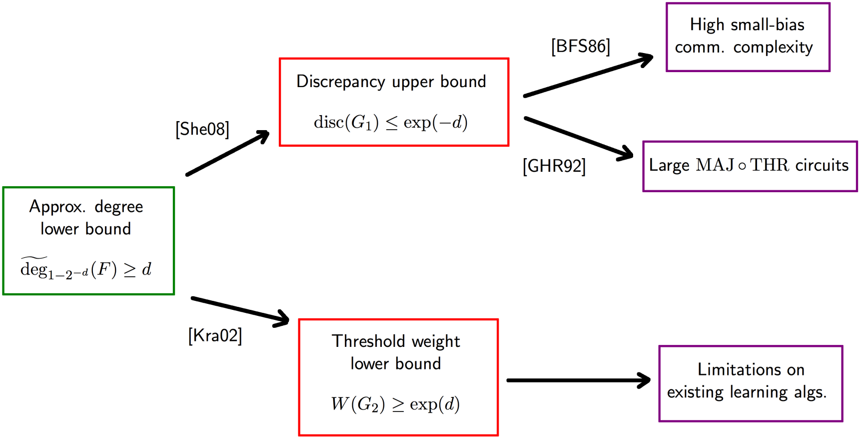
(C chosen to ensure ψ_F has L_1-norm 1).

Properties of ψ_F:

1. ψ_F has pure high degree d ✓[Lee09, Sherstov09]
2. ψ_F has one-sided error ✓By inspection
3. ψ_F has correlation at least $1 - 2^{-t}$ with F ✓

Applications to the Complexity of AC^0

Complexity of AC^0



A New $\widetilde{\text{odeg}}$ Lower 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.
- AC^0 function of interest: Let $\text{ED} : \{-1,1\}^n \to \{-1,1\}$ denote the ELEMENT DISTINCTNESS function.

- [AaronsonShi01] showed $\widetilde{\deg}(\text{ED}) = \Omega(n^{2/3})^*$.
- Best known lower bound on the approximate degree of an AC^0 function.
- This work: $\widetilde{\text{odeg}}(\text{ED}) = \Omega(n^{2/3})$.

*Hiding a logarithmic factor

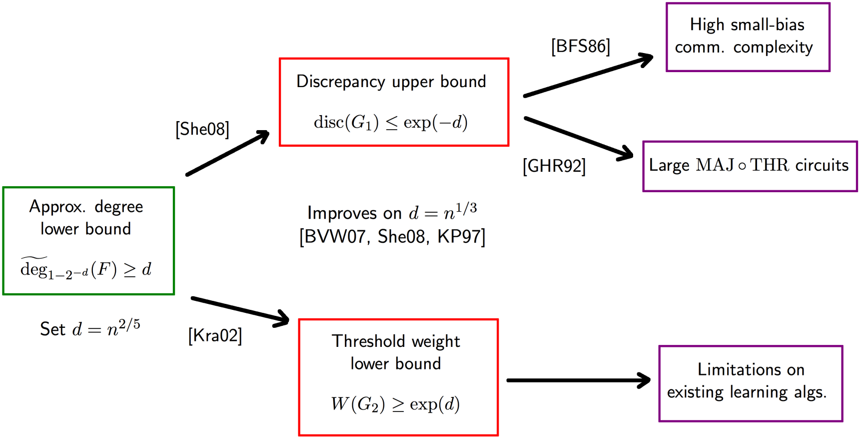
New Lower Bounds for AC^0

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

Proof: Combine lower bound on $\widetilde{\mathrm{odeg}}(\mathrm{ED})$ with Main Theorem.

New Lower Bounds for AC^0



Subsequent work

Further applications of one-sided approximate degree

- Amplification from $\widetilde{\mathrm{odeg}}$ to **threshold degree** [Sherstov14]
- Algorithms for **reliable** agnostic learning [KanadeThaler14]
- Further hardness amplification results [Thaler14]

Thank you!

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. (Equivalent to $\lim_{\epsilon \to 1} \widetilde{\deg}_\epsilon(f)$).

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