A Nearly Optimal Lower Bound on The Approximate Degree of AC^0

Justin Thaler

Georgetown University

Joint work with Mark Bun, Princeton University

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
- $\widetilde{\deg}(f) := \deg_{1/3}(f)$ is the approximate degree of f

Why Care About Approximate Degree?

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

- $\epsilon \approx 1/3$: Agnostic Learning [KKMS05]
- $\epsilon \approx 1 - 2^{-n^\delta}$: Attribute-Efficient Learning [KS04, STT12]
- $\epsilon \to 1$ (i.e., threshold degree, $\deg_\pm(f)$): PAC learning [KS01]

Why Care About Approximate Degree?

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

- $\epsilon \approx 1/3$: Agnostic Learning [KKMS05]
- $\epsilon \approx 1 - 2^{-n^\delta}$: Attribute-Efficient Learning [KS04, STT12]
- $\epsilon \to 1$ (i.e., threshold degree, $\deg_\pm(f)$): PAC learning [KS01]

- Upper bounds on $\widetilde{\deg}_{1/3}(f)$ also:
  - Imply fast algorithms for differentially private data release [TUV12, CTUW14].

Why Care About Approximate Degree?

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

- $\epsilon \approx 1/3$: Agnostic Learning [KKMS05]
- $\epsilon \approx 1 - 2^{-n^\delta}$: Attribute-Efficient Learning [KS04, STT12]
- $\epsilon \to 1$ (i.e., threshold degree, $\deg_\pm(f)$): PAC learning [KS01]

- Upper bounds on $\widetilde{\deg}_{1/3}(f)$ also:
  - Imply fast algorithms for differentially private data release [TUV12, CTUW14].
  - Underly the best known <u>lower bounds</u> on formula complexity and graph complexity [Tal2014, 2016a, 2016b]

Why Care About Approximate Degree?

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

- Quantum query complexity [BBCMW98, AS01, Amb03, KSW04]
- Circuit complexity [MP69, Bei93, Bei94, She08]
- Communication complexity [She08, SZ08, CA08, LS08, She12]
  - Lower bounds hold for a communication problem **related** to f.
  - Technique is called the **Pattern Matrix Method** [She08].
  - A lower bound on $\widetilde{\deg}_{1/3}(f)$ implies that the pattern matrix of f has high quantum communication complexity, even with prior entanglement.

Why Care About Approximate Degree?

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

- Quantum query complexity [BBCMW98, AS01, Amb03, KSW04]
- Circuit complexity [MP69, Bei93, Bei94, She08]
- Communication complexity [She08, SZ08, CA08, LS08, She12]
  - Lower bounds hold for a communication problem **related** to f.
  - Technique is called the **Pattern Matrix Method** [She08].
  - A lower bound on $\widetilde{\deg}_{1/3}(f)$ implies that the pattern matrix of f has high <u>quantum</u> communication complexity, even with prior entanglement.
- Lower bounds on $\widetilde{\deg}(f)$ also yield efficient secret-sharing schemes [BIVW16] and oracle separations [Bei94, BCHTV16].

Example 1: The Approximate Degree of AND_n

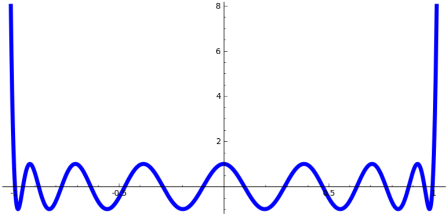
Example: What is the Approximate Degree of AND_n?

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

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

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

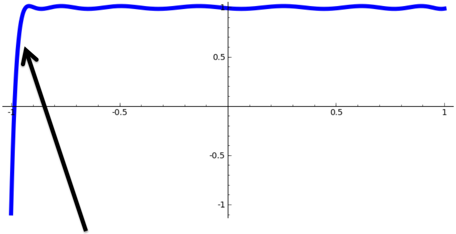
- Chebyshev polynomials are the extremal case.



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:



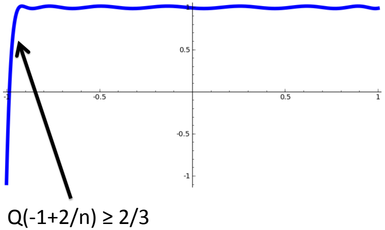
Q(-1+2/n) = 2/3

- 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:



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

Focus of This Talk

- Approximate degree is a key tool for understanding AC^0.
- At the heart of the best known bounds on the complexity of AC^0 under measures such as:
  - Quantum Communication Complexity
  - Approximate Rank
  - Sign-rank $\approx$ UPP^{cc}
  - Discrepancy $\approx$ Margin complexity $\approx$ PP^{cc}
  - Majority-of-Threshold circuit size
  - Threshold-of-Majority circuit size
  - and more.

Focus of This Talk

- Approximate degree is a key tool for understanding AC^0.
- At the heart of the best known bounds on the complexity of AC^0 under measures such as:
  - Quantum Communication Complexity
  - Approximate Rank
  - Sign-rank $\approx$ $\mathsf{UPP}^{\mathsf{cc}}$
  - Discrepancy $\approx$ Margin complexity $\approx$ $\mathsf{PP}^{\mathsf{cc}}$
  - Majority-of-Threshold circuit size
  - Threshold-of-Majority circuit size
  - and more.

Problem 1: Is there a function on n variables that is in AC^0, and has approximate degree $\Omega(n)$?

Approximate Degree of AC^0: Details

- Best known result: $\tilde{\Omega}(n^{2/3})$ for the Element Distinctness function (Aaronson and Shi, 2004).

Approximate Degree of AC^0: Details

- Best known result: $\tilde{\Omega}(n^{2/3})$ for the Element Distinctness function (Aaronson and Shi, 2004).
- Our result: For any constant $\delta > 0$, a function in AC^0 with approximate degree $\Omega(n^{1-\delta})$.
  - More precisely, circuit depth is $O(\log(1/\delta))$.

Approximate Degree of AC^0: Details

- Best known result: $\tilde{\Omega}(n^{2/3})$ for the Element Distinctness function (Aaronson and Shi, 2004).
- Our result: For any constant $\delta > 0$, a function in AC^0 with approximate degree $\Omega(n^{1-\delta})$.
  - More precisely, circuit depth is $O(\log(1/\delta))$.
  - Lower bound also applies to DNFs of polylogarithmic width (and quasipolynomial size).

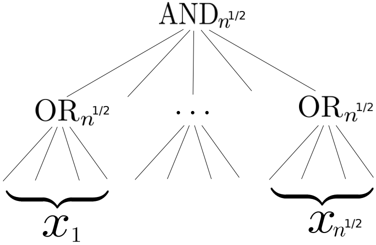
Applications

- Nearly optimal $\Omega(n^{1-\delta})$ lower bounds on quantum communication complexity of AC^0.
- Essentially optimal (quadratic) separation of certificate complexity and approximate degree.
- Better secret sharing schemes with reconstruction in AC^0.

Prior Work: The Method of Dual Polynomials and the AND-OR 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 $\widetilde{\deg}(\text{AND-OR}_n)$?



History of the AND-OR Tree

Theorem

$\widetilde{\deg}(\text{AND-OR}_n) = \Theta(n^{1/2})$.

History of the AND-OR Tree

Theorem

$\widetilde{\deg}(\text{AND-OR}_n) = \Theta(n^{1/2})$.

Tight Upper Bound of $O(n^{1/2})$

[HMW03]	via quantum algorithms
[She12]	different proof (via robustification)

History of the AND-OR Tree

Theorem

$\widetilde{\deg}(\text{AND-OR}_n) = \Theta(n^{1/2})$.

Tight Upper Bound of $O(n^{1/2})$

[HMW03] via quantum algorithms
[She12] different proof (via robustification)

Tight Lower Bound of $\Omega(n^{1/2})$

[BT13] and [She13] via the method of dual polynomials

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):

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

Dual LP:

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

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

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

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

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

Dual Characterization of Approximate Degree

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

Example: $2^{-n} \cdot \mathsf{PARITY}_n$ witnesses the fact that $\lim_{\epsilon \to 1} \widetilde{\deg}_\epsilon(\mathsf{PARITY}_n) = n$.

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

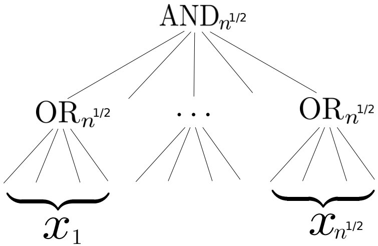
Constructing a Dual Polynomial

- By [NS92], there are dual polynomials
  $\psi_{\textbf{OUT}}$ for $\widetilde{\deg}(\mathrm{AND}_{n^{1/2}}) = \Omega(n^{1/4})$ and
  $\psi_{\textbf{IN}}$ for $\widetilde{\deg}(\mathrm{OR}_{n^{1/2}}) = \Omega(n^{1/4})$
- Both [She13] and [BT13] combine $\psi_{\textbf{OUT}}$ and $\psi_{\textbf{IN}}$ to obtain a dual polynomial $\psi_{\textbf{AND-OR}}$ for $\mathrm{AND\text{-}OR}$.
- The combining method was proposed in earlier work by [SZ09, Lee09, She09].

The Combining Method [SZ09, She09, Lee09]

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

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



The Combining Method [SZ09, She09, Lee09]

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

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

Must verify:

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

The Combining Method [SZ09, She09, Lee09]

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

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

Must verify:

1. $\psi_{\textsf{AND-OR}}$ has pure high degree $\geq n^{1/4} \cdot n^{1/4} = n^{1/2}$.✓[She09]
2. $\psi_{\textsf{AND-OR}}$ has high correlation with $\mathrm{AND\text{-}OR}$. [BT13, She13]

Recent Progress on the Complexity of AC^0: Applying the Method of Dual Polynomials to Block-Composed Functions

(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.
- Examples: $\widetilde{\mathrm{odeg}}_{-,1/3}(\mathrm{AND}_n) = \Theta(\sqrt{n})$; $\widetilde{\mathrm{odeg}}_{-,1/3}(\mathrm{OR}_n) = 1$.

Recent Theorems

Theorem (BT13, She13)

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

Recent Theorems

Theorem (BT13, She13)

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

Theorem (BT14)

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

Recent Theorems

Theorem (BT13, She13)

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

Theorem (BT14)

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

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 $\deg_{\pm}(F) = \Omega(\min\{d, t\})$.

Recent Theorems

Theorem (BT13, She13)

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

Theorem (BT14)

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

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 $\deg_{\pm}(F) = \Omega(\min\{d, t\})$.

Theorem (BCHTV16)

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

Reminder

Problem 1: Is there a function on n variables that is in AC^0, and has approximate degree $\Omega(n)$?

Our Techniques

Approximate Degree of AC^0: Details

- Major technical obstacle to progress on lower bounds: By Robustification [She12]:

$$\widetilde{\deg}(f(g,\ldots,g)) \leq O(\widetilde{\deg}(f)\cdot\widetilde{\deg}(g)).$$

  - i.e., the approximate degree of $f_M \circ g_N$ (as a function of the number of inputs $M \cdot N$) is never larger than that of f or g individually.

Approximate Degree of AC^0: Details

- Major technical obstacle to progress on lower bounds: By Robustification [She12]:

$$\widetilde{\deg}(f(g,\ldots,g)) \leq O(\widetilde{\deg}(f)\cdot\widetilde{\deg}(g)).$$

  - i.e., the approximate degree of $f_M \circ g_N$ (as a function of the number of inputs $M \cdot N$) is never larger than that of f or g individually.
  - So must move **beyond block-composed functions** to make progress on Problem 1.

A General Hardness Amplification Result

Theorem (Main Theorem)

Let $f\colon \{-1,1\}^n \to \{-1,1\}$ *with* $\widetilde{\deg}(f) = d$. *Then* f *can be transformed into a function* g *on* $O(n \log^4 n)$ *variables with*

$$\widetilde{\deg}(g) \geq n^{1/3} \cdot d^{2/3}.$$

A General Hardness Amplification Result

Theorem (Main Theorem)

Let $f\colon \{-1,1\}^n \to \{-1,1\}$ with $\widetilde{\deg}(f) = d$. Then f can be transformed into a function g on $O(n \log^4 n)$ variables with

$$\widetilde{\deg}(g) \geq n^{1/3} \cdot d^{2/3}.$$

- *f computed by circuit of depth $d \Longrightarrow$ g computed by circuit of depth $d+3$.*

A General Hardness Amplification Result

Theorem (Main Theorem)

Let $f\colon \{-1,1\}^n \to \{-1,1\}$ *with* $\widetilde{\deg}(f) = d$. *Then* f *can be transformed into a function* g *on* $O(n \log^4 n)$ *variables with*

$$\widetilde{\deg}(g) \geq n^{1/3} \cdot d^{2/3}.$$

- f *computed by circuit of depth* $d \Longrightarrow$ g *computed by circuit of depth* $d+3$.
- f *computed by monotone circuit of depth* $d \Longrightarrow$ g *computed by monotone circuit of depth* $d+2$.

A General Hardness Amplification Result

Theorem (Main Theorem)

Let $f\colon \{-1,1\}^n \to \{-1,1\}$ *with* $\widetilde{\deg}(f) = d$. *Then* f *can be transformed into a function* g *on* $O(n \log^4 n)$ *variables with*

$$\widetilde{\deg}(g) \geq n^{1/3} \cdot d^{2/3}.$$

- f *computed by circuit of depth* $d \implies$ g *computed by circuit of depth* $d+3$.
- f *computed by monotone circuit of depth* $d \implies$ g *computed by monotone circuit of depth* $d+2$.
- f *computed by monotone DNF of width* $w \implies$ g *computed by monotone DNF of width* $O(w \cdot \log^2 n)$.

A General Hardness Amplification Result

Theorem (Main Theorem)

Let $f\colon \{-1,1\}^n \to \{-1,1\}$ *with* $\widetilde{\deg}(f) = d$. *Then* f *can be transformed into a function* g *on* $O(n \log^4 n)$ *variables with*

$$\widetilde{\deg}(g) \geq n^{1/3} \cdot d^{2/3}.$$

- f *computed by circuit of depth* $d \Longrightarrow$ g *computed by circuit of depth* $d+3$.
- f *computed by monotone circuit of depth* $d \Longrightarrow$ g *computed by monotone circuit of depth* $d+2$.
- f *computed by monotone DNF of width* $w \Longrightarrow$ g *computed by monotone DNF of width* $O(w \cdot \log^2 n)$.

- AC^0 results obtained by recursively applying Main Theorem, starting with f equal to OR_n.

Idea of the Hardness Amplification Construction

Idea of the Hardness-Amplifying Construction

- Consider the function SURJECTIVITY: $\{-1,1\}^n \to \{-1,1\}$.
  - Let $n = N \log R$. SURJ interprets its input x as a list of N numbers $(x_1, \ldots, x_N)$ from a range $[R]$.
  - SURJ$(x) = -1$ if and only if every element of the range $[R]$ appears at least once in the list.
- When we apply Main Theorem to $f = \mathrm{AND}_R$, the "harder" function g is precisely SURJ.

Getting to Know SURJECTIVITY

- It is known that $\widetilde{\deg}(\text{SURJ}) = \tilde{\Omega}(n^{2/3})$ for $R = N/2$ [AS04].
- Best known upper bound on $\widetilde{\deg}(\text{SURJ})$ is trivial $O(n)$.

Getting to Know SURJECTIVITY

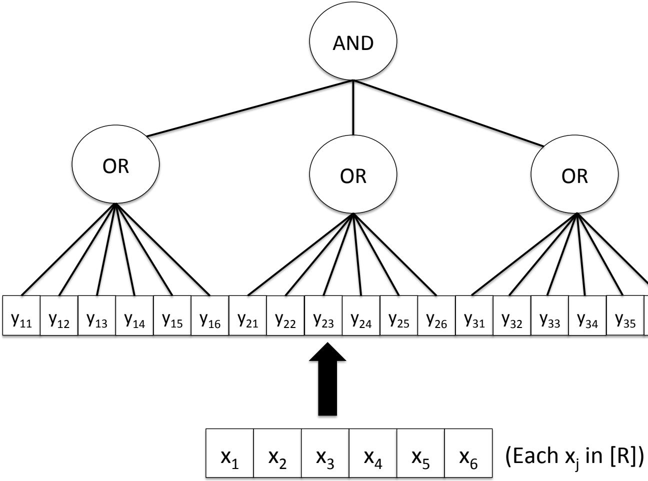
- It is known that $\widetilde{\deg}(\text{SURJ}) = \tilde{\Omega}(n^{2/3})$ for $R = N/2$ [AS04].
- Best known upper bound on $\widetilde{\deg}(\text{SURJ})$ is trivial $O(n)$.
- An instructive way to achieve this trivial upper bound:
  - Let
    $$y_{ij} = \begin{cases} -1 \text{ if } x_j = i \\ +1 \text{ otherwise} \end{cases}$$
  - Then

$$\text{SURJ}(x) = \text{AND}_R(\text{OR}_N(y_{1,1}, \ldots, y_{1,N}), \ldots, \text{OR}_N(y_{R,1} \ldots, y_{R,N})).$$

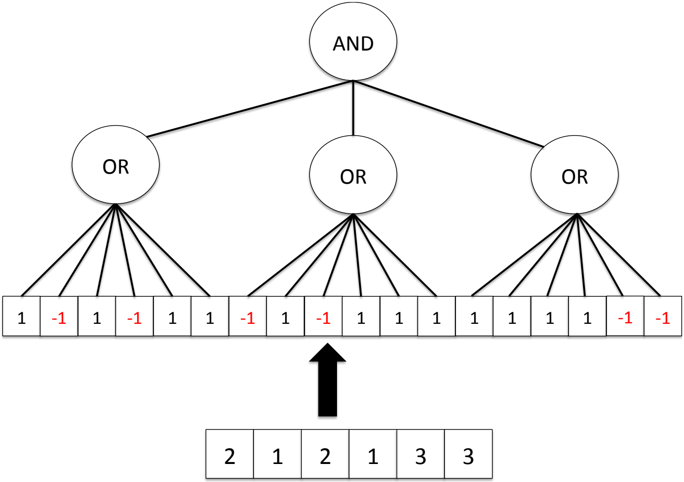
Getting to Know SURJECTIVITY

- It is known that $\widetilde{\deg}(\text{SURJ}) = \tilde{\Omega}(n^{2/3})$ for $R = N/2$ [AS04].
- Best known upper bound on $\widetilde{\deg}(\text{SURJ})$ is trivial $O(n)$.
- An instructive way to achieve this trivial upper bound:
  - Let
$$y_{ij} = \begin{cases} -1 \text{ if } x_j = i \\ +1 \text{ otherwise} \end{cases}$$
  - Then
$$\text{SURJ}(x) = \text{AND}_R(\text{OR}_N(y_{1,1}, \ldots, y_{1,N}), \ldots, \text{OR}_N(y_{R,1} \ldots, y_{R,N})).$$
  - Let p be a degree $O(\sqrt{R \cdot N}) = O(N)$ polynomial approximating $\text{AND}_R(\text{OR}_N, \ldots, \text{OR}_N)$.
    - Can construct p via robustification.
  - Then $p(y_{1,1}, \ldots, y_{1,N}, \ldots, y_{R,1}, \ldots, y_{R,N})$ approximates SURJ, and has degree $O(\deg(p) \cdot \log R) = O(n)$.

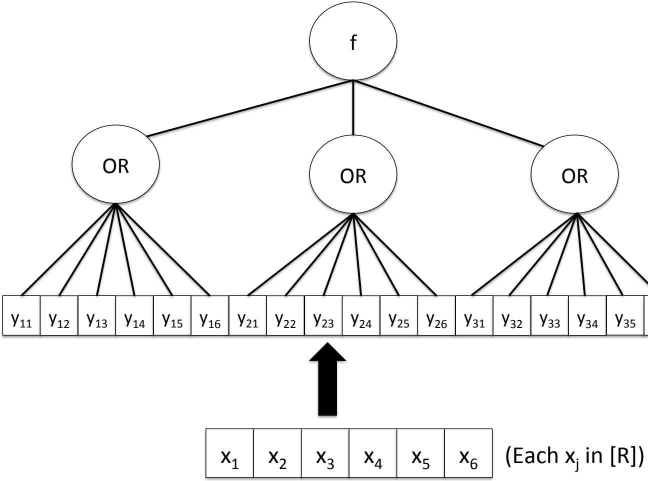
SURJ Illustrated $(R = 3, N = 6)$



SURJ Illustrated $(R = 3, N = 6)$



First Attempt: Amplifying Hardness of $f\colon\{-1,1\}^R \to \{-1,1\}$ ($R=3, N=6$)



Hardness-Amplifying Construction: Second Attempt

- First attempt at handling general f fails when $f = \mathrm{OR}$.
  - $g(x) = \mathrm{OR}_R(\mathrm{OR}_N(y_{1,1}, \ldots, y_{1,N}), \ldots, \mathrm{OR}_N(y_{R,1} \ldots, y_{R,N}))$ has (exact) degree 1.

Hardness-Amplifying Construction: Second Attempt

- First attempt at handling general f fails when $f = \mathrm{OR}$.
  - $g(x) = \mathrm{OR}_R(\mathrm{OR}_N(y_{1,1}, \ldots, y_{1,N}), \ldots, \mathrm{OR}_N(y_{R,1} \ldots, y_{R,N}))$ has (exact) degree 1.
- Let $R' = R \log R$. For $f\colon \{-1,1\}^R \to \{-1,1\}$, the real* definition of g is:

$$g(x)=(f \circ \mathrm{AND}_{\log R})(\mathrm{OR}_N(y_{1,1},\ldots,y_{1,N}),\ldots,\mathrm{OR}_N(y_{R',1},\ldots,y_{R',N}))$$

*This is still a slight simplification.

Idea of the Analysis for SURJECTIVITY

Idea of the Analysis for SURJECTIVITY

- Let $n = N \log R$.
- Recall: to approximate SURJ: $\{-1,1\}^n \to \{-1,1\}$, it is **sufficient** to approximate the block-composed function $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ on $N \cdot R$ bits.
- Goal is to show this approximation method is close to optimal.

Idea of the Analysis for SURJECTIVITY

- Let $n = N \log R$.
- Recall: to approximate SURJ: $\{-1,1\}^n \to \{-1,1\}$, it is **sufficient** to approximate the block-composed function $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ on $N \cdot R$ bits.
- Goal is to show this approximation method is close to optimal.
- Step 1: Show that to approximate SURJ(x), it is **necessary** to approximate $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$, under the promise that the input has Hamming weight **at most** N.
    - Follows from a symmetrization argument (Ambainis 2003).

Idea of the Analysis for SURJECTIVITY

- Let $n = N \log R$.
- Recall: to approximate SURJ: $\{-1,1\}^n \to \{-1,1\}$, it is **sufficient** to approximate the block-composed function $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ on $N \cdot R$ bits.
- Goal is to show this approximation method is close to optimal.
- Step 1: Show that to approximate SURJ(x), it is **necessary** to approximate $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$, under the promise that the input has Hamming weight **at most** N.
  - Follows from a symmetrization argument (Ambainis 2003).
- Step 2: Prove that for some $N = \tilde{O}(R)$, this promise problem requires degree $\gtrsim \Omega(R^{2/3})$.

Idea of the Analysis for SURJECTIVITY

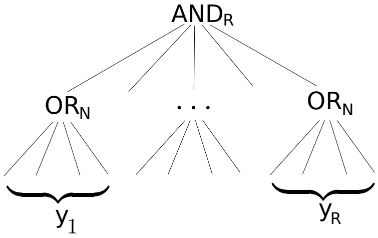
- Let $n = N \log R$.
- Recall: to approximate SURJ: $\{-1,1\}^n \to \{-1,1\}$, it is **sufficient** to approximate the block-composed function $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ on $N \cdot R$ bits.
- Goal is to show this approximation method is close to optimal.
- Step 1: Show that to approximate SURJ(x), it is **necessary** to approximate $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$, under the promise that the input has Hamming weight **at most** N.
  - Follows from a symmetrization argument (Ambainis 2003).
- Step 2: Prove that for some $N = \tilde{O}(R)$, this promise problem requires degree $\gtrsim \Omega(R^{2/3})$.
  - Builds on the "dual combining technique" used earlier to analyze $\mathrm{AND\text{-}OR}_n$ (with no promise).

Overview of Step 2

Prove That For Some $N = \tilde{O}(R)$, Approximating $\mathrm{AND}_R \circ \mathrm{OR}_N$ Under the Promise That The Input Has Hamming Weight **At Most** N Requires Degree $\gtrsim R^{2/3}$.

Attempt 1

- For some $N = \tilde{O}(R)$, want a dual witness for $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ that **only places mass on inputs of Hamming weight at most** N.



Attempt 1

- For some $N = \tilde{O}(R)$, want a dual witness for $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ that **only places mass on inputs of Hamming weight at most** N.
- Attempt 1: Use the dual witness for $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ from prior work [She09, Lee09, BT13, She13].

$$\psi_{\textsf{AND-OR}}(y_1, \ldots, y_R) := C \cdot \psi_{\textsf{AND}}(\ldots, \mathrm{sgn}(\psi_{\textsf{OR}}(y_j)), \ldots) \prod_{j=1}^{R} |\psi_{\textsf{OR}}(y_j)|$$

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

Attempt 1

- For some $N = \tilde{O}(R)$, want a dual witness for $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ that **only places mass on inputs of Hamming weight at most** N.
- Attempt 1: Use the dual witness for $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ from prior work [She09, Lee09, BT13, She13].

$$\psi_{\textsf{AND-OR}}(y_1, \ldots, y_R) := C \cdot \psi_{\textsf{AND}}(\ldots, \mathrm{sgn}(\psi_{\textsf{OR}}(y_j)), \ldots) \prod_{j=1}^{R} |\psi_{\textsf{OR}}(y_j)|$$

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

Must verify:

1. $\psi_{\textsf{AND-OR}}$ has pure high degree $\geq R^{1/2} \cdot N^{1/2} = \Omega(N)$.
2. $\psi_{\textsf{AND-OR}}$ well-correlated with AND-OR.
3. $\psi_{\textsf{AND-OR}}$ places mass only on inputs of Hamming weight $\leq N$.

Attempt 1

- For some $N = \tilde{O}(R)$, want a dual witness for $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ that **only places mass on inputs of Hamming weight at most** N.
- Attempt 1: Use the dual witness for $\mathrm{AND}_R(\mathrm{OR}_N, \ldots, \mathrm{OR}_N)$ from prior work [She09, Lee09, BT13, She13].

$$\psi_{\textsf{AND-OR}}(y_1, \ldots, y_R) := C \cdot \psi_{\textsf{AND}}(\ldots, \mathrm{sgn}(\psi_{\textsf{OR}}(y_j)), \ldots) \prod_{j=1}^{R} |\psi_{\textsf{OR}}(y_j)|$$

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

Must verify:

1. $\psi_{\textsf{AND-OR}}$ has pure high degree $\geq R^{1/2} \cdot N^{1/2} = \Omega(N)$. ✓ [She09]
2. $\psi_{\textsf{AND-OR}}$ well-correlated with $\mathrm{AND\text{-}OR}$. ✓ [BT13, She13]
3. $\psi_{\textsf{AND-OR}}$ places mass only on inputs of Hamming weight $\leq N$. ✗

Patching Attempt 1

- Goal: Fix Property 3 without destroying Properties 1 or 2.

Patching Attempt 1

- Goal: Fix Property 3 without destroying Properties 1 or 2.
- Fact (cf. Razborov and Sherstov 2008): Suppose

$$\sum_{|y|>N} |\psi_{\textbf{AND-OR}}(y)| \ll R^{-D}.$$

  - Then we can "post-process" $\psi_{\textbf{AND-OR}}$ to "zero out" any mass it places it inputs of Hamming weight larger than N.
  - While ensuring that the resulting dual witness still has pure high degree $\min\{D, \mathsf{PHD}(\psi_{\textbf{AND-OR}})\}$.

Patching Attempt 1

- New Goal: Show that, for $D \approx R^{2/3}$,

$$\sum_{|y|>N} |\psi_{\textsf{AND-OR}}(y)| \ll R^{-D}. \tag{1}$$

- Recall:

$$\psi_{\textsf{AND-OR}}(y_1, \ldots, y_R) := C \cdot \psi_{\textsf{AND}}(\ldots, \operatorname{sgn}(\psi_{\textsf{OR}}(y_j)), \ldots) \prod_{j=1}^{R} |\psi_{\textsf{OR}}(y_j)|$$

Patching Attempt 1

- New Goal: Show that, for $D \approx R^{2/3}$,

$$\sum_{|y|>N} |\psi_{\textsf{AND-OR}}(y)| \ll R^{-D}. \qquad (1)$$

- Recall:

$$\psi_{\textsf{AND-OR}}(y_1, \ldots, y_R) := C \cdot \psi_{\textsf{AND}}(\ldots, \operatorname{sgn}(\psi_{\textsf{OR}}(y_j)), \ldots) \prod_{j=1}^{R} |\psi_{\textsf{OR}}(y_j)|$$

- Intuition:
  - A dual witness $\psi_{\textsf{OR}}$ for OR can be made "weakly" biased toward low Hamming weight inputs.
    - Specifically: $\sum_{|y_i|=t} |\psi_{\textsf{OR}}(y_i)| \leq t^{-2}$.

Patching Attempt 1

- New Goal: Show that, for $D \approx R^{2/3}$,

$$\sum_{|y|>N} |\psi_{\textbf{AND-OR}}(y)| \ll R^{-D}. \tag{1}$$

- Recall:

$$\psi_{\textbf{AND-OR}}(y_1, \ldots, y_R) := C \cdot \psi_{\textbf{AND}}(\ldots, \operatorname{sgn}(\psi_{\textbf{OR}}(y_j)), \ldots) \prod_{j=1}^{R} |\psi_{\textbf{OR}}(y_j)|$$

- Intuition:
  - A dual witness $\psi_{\textbf{OR}}$ for OR can be made "weakly" biased toward low Hamming weight inputs.
    - Specifically: $\sum_{|y_i|=t} |\psi_{\textbf{OR}}(y_i)| \leq t^{-2}$.
  - $|\psi_{\textbf{AND-OR}}(y_1, \ldots, y_R)|$ "resembles" the product distribution $\prod_{j=1}^{R} |\psi_{\textbf{OR}}(y_j)|$.
  - So it is exponentially more biased toward low Hamming weight inputs than $\psi_{\textbf{OR}}$ itself.

Patching Attempt 1 (Slightly Loose Analysis)

- New Goal: Show that, for $D \approx R^{2/3}$,

$$\sum_{|y| > 2R^{1.01}} |\psi_{\textsf{AND-OR}}(y)| \ll R^{-D}. \tag{2}$$

Patching Attempt 1 (Slightly Loose Analysis)

- New Goal: Show that, for $D \approx R^{2/3}$,

$$\sum_{|y|>2R^{1.01}} |\psi_{\textsf{AND-OR}}(y)| \ll R^{-D}. \tag{2}$$

- Recall:

$$\psi_{\textsf{AND-OR}}(y_1, \dots, y_R) := C \cdot \psi_{\textsf{AND}}(\dots, \operatorname{sgn}(\psi_{\textsf{OR}}(y_j)), \dots) \prod_{j=1}^{R} |\psi_{\textsf{OR}}(y_j)|$$

Patching Attempt 1 (Slightly Loose Analysis)

- New Goal: Show that, for $D \approx R^{2/3}$,

$$\sum_{|y|>2R^{1.01}} |\psi_{\textsf{AND-OR}}(y)| \ll R^{-D}. \tag{2}$$

- Recall:

$$\psi_{\textsf{AND-OR}}(y_1, \ldots, y_R) := C \cdot \psi_{\textsf{AND}}(\ldots, \operatorname{sgn}(\psi_{\textsf{OR}}(y_j)), \ldots) \prod_{j=1}^{R} |\psi_{\textsf{OR}}(y_j)|$$

- We need to modify $\psi_{\textsf{OR}}$ to ensure that Equation (2) holds.

Patching Attempt 1 (Slightly Loose Analysis)

- New Goal: Show that, for $D \approx R^{2/3}$,

$$\sum_{|y| > 2R^{1.01}} |\psi_{\textsf{AND-OR}}(y)| \ll R^{-D}. \qquad (2)$$

- Recall:

$$\psi_{\textsf{AND-OR}}(y_1, \ldots, y_R) := C \cdot \psi_{\textsf{AND}}(\ldots, \operatorname{sgn}(\psi_{\textsf{OR}}(y_j)), \ldots) \prod_{j=1}^{R} |\psi_{\textsf{OR}}(y_j)|$$

- We need to modify $\psi_{\textsf{OR}}$ to ensure that Equation (2) holds.
  1. Modify $\psi_{\textsf{OR}}$ to place no mass whatsoever on inputs of Hamming weight more than $R^{1/3}$.

Patching Attempt 1 (Slightly Loose Analysis)

- New Goal: Show that, for $D \approx R^{2/3}$,

$$\sum_{|y|>2R^{1.01}} |\psi_{\textsf{AND-OR}}(y)| \ll R^{-D}. \qquad (2)$$

- Recall:

$$\psi_{\textsf{AND-OR}}(y_1, \ldots, y_R) := C \cdot \psi_{\textsf{AND}}(\ldots, \operatorname{sgn}(\psi_{\textsf{OR}}(y_j)), \ldots) \prod_{j=1}^{R} |\psi_{\textsf{OR}}(y_j)|$$

- We need to modify $\psi_{\textsf{OR}}$ to ensure that Equation (2) holds.
  1. Modify $\psi_{\textsf{OR}}$ to place no mass whatsoever on inputs of Hamming weight more than $R^{1/3}$.
  2. Suppose $\psi_{\textsf{OR}}$ also satisfies the following "low Hamming weight bias" condition.
     - $\sum_{|y_i|>R^{0.01}} |\psi_{\textsf{OR}}(y_i)| \leq R^{-40}$.

Patching Attempt 1 (Slightly Loose Analysis)

- New Goal: Show that, for $D \approx R^{2/3}$,

$$\sum_{|y|>2R^{1.01}} |\psi_{\textsf{AND-OR}}(y)| \ll R^{-D}. \qquad (2)$$

- Recall:

$$\psi_{\textsf{AND-OR}}(y_1, \ldots, y_R) := C \cdot \psi_{\textsf{AND}}(\ldots, \mathrm{sgn}(\psi_{\textsf{OR}}(y_j)), \ldots) \prod_{j=1}^{R} |\psi_{\textsf{OR}}(y_j)|$$

- We need to modify $\psi_{\textsf{OR}}$ to ensure that Equation (2) holds.
  1. Modify $\psi_{\textsf{OR}}$ to place no mass whatsoever on inputs of Hamming weight more than $R^{1/3}$.
  2. Suppose $\psi_{\textsf{OR}}$ also satisfies the following "low Hamming weight bias" condition.
     - $\sum_{|y_i|>R^{0.01}} |\psi_{\textsf{OR}}(y_i)| \leq R^{-40}$.
- Condition (1) $\implies (|y| > 2R^{1.01} \implies \left|\{i \colon |y_i| > R^{0.01}\}\right| > R^{2/3})$

Patching Attempt 1 (Slightly Loose Analysis)

- New Goal: Show that, for $D \approx R^{2/3}$,

$$\sum_{|y| > 2R^{1.01}} |\psi_{\textsf{AND-OR}}(y)| \ll R^{-D}. \tag{2}$$

- Recall:

$$\psi_{\textsf{AND-OR}}(y_1, \dots, y_R) := C \cdot \psi_{\textsf{AND}}(\dots, \operatorname{sgn}(\psi_{\textsf{OR}}(y_j)), \dots) \prod_{j=1}^{R} |\psi_{\textsf{OR}}(y_j)|$$

- We need to modify $\psi_{\textsf{OR}}$ to ensure that Equation (2) holds.
  1. Modify $\psi_{\textsf{OR}}$ to place <u>no mass whatsoever</u> on inputs of Hamming weight more than $R^{1/3}$.
  2. Suppose $\psi_{\textsf{OR}}$ also satisfies the following "low Hamming weight bias" condition.
     - $\sum_{|y_i| > R^{0.01}} |\psi_{\textsf{OR}}(y_i)| \leq R^{-40}$.
- Condition (1) $\implies (|y| > 2R^{1.01} \implies \left|\{i \colon |y_i| > R^{0.01}\}\right| > R^{2/3})$
- Condition (2) + product-like nature of $\psi_{\textsf{AND-OR}} \implies$ total mass $\psi_{\textsf{AND-OR}}$ places on such inputs is $\ll R^{-R^{2/3}}$.

Completing The Analysis

- Fact: Both properties from previous slide are satisfied by a dual witness $\psi_{\textsf{OR}}$ for OR of pure high degree $\approx R^{1/6}$.
- This ensures $\psi_{\textsf{AND-OR}}$ has pure high degree $\gtrsim R^{1/2} \cdot R^{1/6} = R^{2/3}$. $\square$

Future Directions

- An $\Omega(n)$ lower bound on the approximate degree of AC^0?
- Extend our $\Omega(n^{1-\delta})$ degree lower bound from polylogarithmic width DNFs to polynomial size DNFs?
- Extend our bounds on $\widetilde{\deg}_\epsilon(f)$ from $\epsilon = 1/3$ to ϵ much closer to 1?
  - We believe our techniques can extend to give:
    - A function f in AC^0 with $\widetilde{\deg}_\epsilon(f) \geq n^{1-\delta}$, for $\epsilon = 1 - 2^{-n^{1-\delta}}$.
    - New threshold degree lower bounds for AC^0.

Thank you!