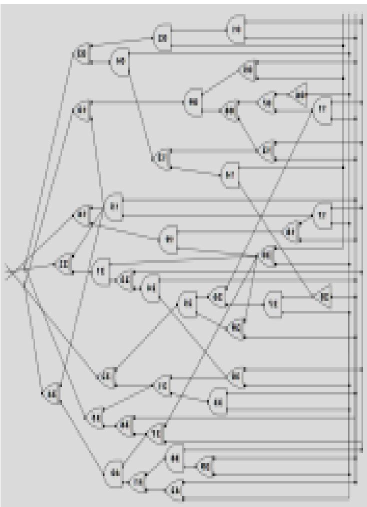
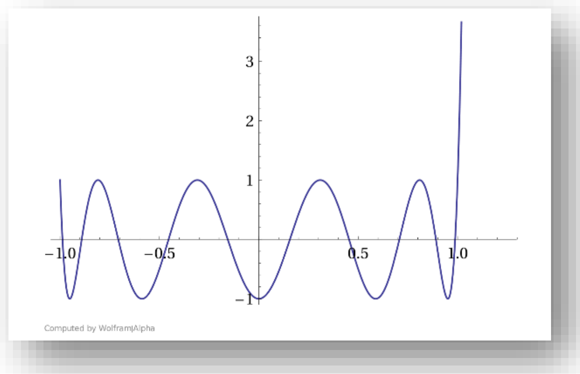
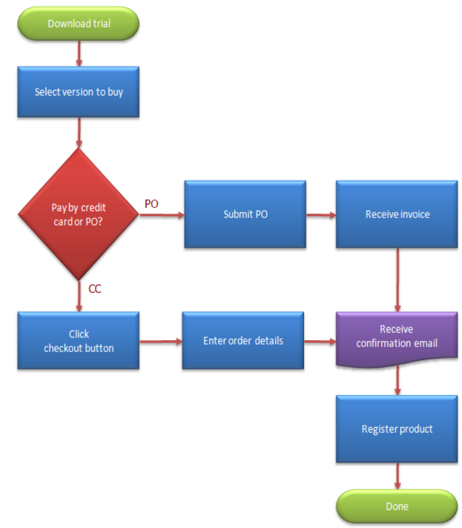
Polynomial Representations of Threshold Functions and Algorithmic Applications







Ryan Williams **Stanford**

Joint with Josh Alman (Stanford) and Timothy M. Chan (Waterloo)

Outline

- The Context: *Polynomial Representations, Polynomials for Algorithms, Nearest Nbrs*
- The New Results
- Some Details
- Conclusion

Exact Polynomial Representations

Let $f : \{0,1\}^n \rightarrow \{0,1\}$; let $D = \mathbb{F}_q$ or $\mathbb{R}$ in the following.

Def. **A polynomial $p : D^n \rightarrow D$ is an (exact) polynomial for f if for all $x \in \{0,1\}^n, p(x) = f(x)$.**

Example: *OR*

$$OR(x_1, \ldots, x_n) = \begin{cases} 0 & \text{if } x_1 = \cdots = x_n = 0 \\ 1 & \text{otherwise} \end{cases}$$

We can write *OR* as a polynomial over D

$$OR(x_1, \ldots, x_n) = 1 - (1 - x_1)(1 - x_2) \cdots (1 - x_n)$$

This has degree *n* and $\Omega(2^n)$ monomials when expanded out.
We need other representations to get smaller polynomials.

Probabilistic Polynomials

Let $f : \{0,1\}^n \to \{0,1\}$; let $D = \mathbb{F}_q$ or $\mathbb{R}$ in the following.

Def. **A distribution P on degree d polynomials $p : D^n \to D$ is a probabilistic polynomial for f with error at most ε if for all $x \in \{0,1\}^n$,**

$$\Pr_{p \sim P}[p(x) = f(x)] > 1 - \varepsilon.$$

Example: ***OR with*** $D = \mathbb{F}_2$

[R'87] Pick a uniformly random $R \subseteq \{1, 2, \ldots, n\}$, then pick the polynomial

$$p(x_1, \ldots, x_n) = \sum_{i \in R} x_i .$$

$p(x) = 0$ **when** $x_1 = \cdots = x_n = 0$ **Odd with probability 1/2 otherwise**

Can amplify to error ϵ with degree $k = O(\log(1/\epsilon))$:

$$p(x_1, \ldots, x_n) = 1 - \left(1 - \sum_{i \in R_1} x_i\right) \cdots \left(1 - \sum_{i \in R_k} x_i\right).$$

Polynomial Threshold Functions (PTF)

Let $f : \{0,1\}^n \rightarrow \{0,1\}$; let D = ~~$\mathbb{F}_q$ or~~ $\mathbb{R}$ in the following.

Def. A polynomial $p : \mathbb{R}^n \rightarrow \mathbb{R}$ is a polynomial threshold function (PTF) for f if for all $x \in \{0,1\}^n$,

- $p(x) \geq 0$ when $f(x) = 1$
- $p(x) < 0$ when $f(x) = 0$

Example: *OR*

Just take a sum!

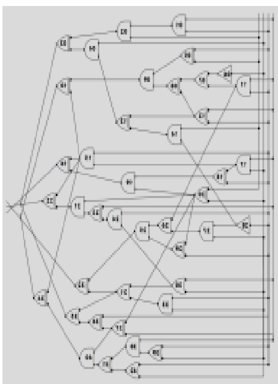
$$p(x_1, \ldots, x_n) = x_1 + x_2 + \cdots + x_n - 1/2.$$

Def. A distribution P on degree d polynomials $p : \mathbb{R}^n \rightarrow \mathbb{R}$ is a probabilistic PTF for f with error at most ε if for all $x \in \{0,1\}^n$,

- $\Pr_{p \sim P}[p(x) \geq 0] > 1 - \varepsilon$ when $f(x) = 1$
- $\Pr_{p \sim P}[p(x) < 0] > 1 - \varepsilon$ when $f(x) = 0$

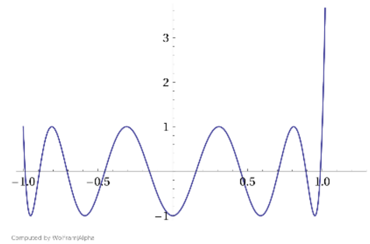
Polynomials For Algorithms

low complexity

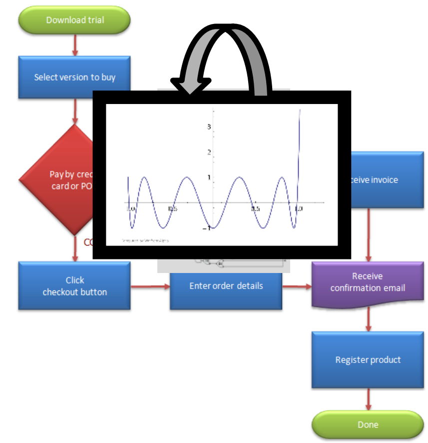


reduction

"nice" polynomials



"classical" algorithm



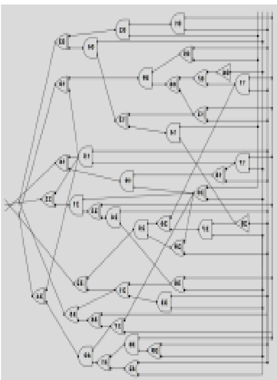
Multipoint evaluation of polynomials (MM/FFT) ➔ faster algorithm!

Has led to faster algorithms for:

- **All-pairs shortest paths [W'14]**
- **All-points nearest neighbors in Hamming metric [AW'15]**
- **#k-SAT [CW'16]**
- **Circuit-SAT in many regimes (and thus, circuit lower bounds)**
- **Succinct stable matching [MPS'16]**
- **Partial match queries [AWY'15] …**

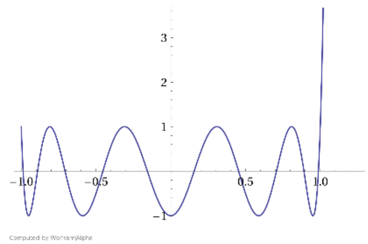
Polynomials For Algorithms

low complexity

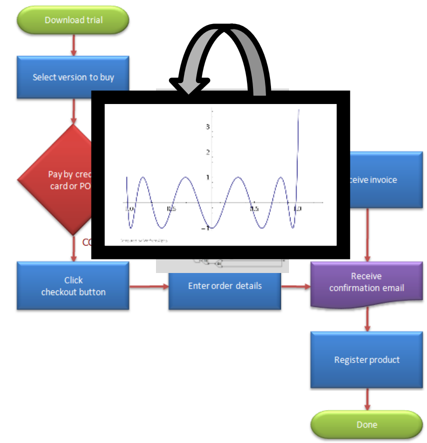


randomized reduction

"nice" polynomials



"classical" algorithm



Multipoint evaluation of polynomials (MM/FFT) ➔ faster algorithm!

Has led to faster algorithms for:

- All-pairs shortest paths [W'14]
- All-points nearest neighbors in Hamming metric [AW'15]
- #k-SAT [CW'16]
- Circuit-SAT in many regimes (and thus, circuit lower bounds)
- Succinct stable matching [MPS'16]
- Partial match queries [AWY'15] …

Multipoint Evaluation

Reduce multipoint polynomial evaluation to matrix multiplication

Suppose we want to evaluate polynomial $p(x_1, \dots, x_m, y_1, \dots, y_m)$ on all pairs of $x \in A$ and $y \in B$.

First expand p out in terms of monomials, for instance:

$$p(x, y) = 2x_1y_1 + 3x_4x_7y_8 - 12y_9y_{13} + \cdots$$

Then p can also be written as an inner product:

$$p(x, y) = (2x_1, 3x_4x_7, -12, \dots) \cdot (y_1, y_8, y_9y_{13}, \dots)$$

Hence we can evaluate p on a **combinatorial rectangle** using fast (rectangular) matrix multiplication [Coppersmith'82]

Evaluation Lemma [C'82,W'14] Given sets $A, B \subseteq \{0,1\}^m$, $|A| = |B| = n$, and a polynomial $p(x_1, \dots, x_m, y_1, \dots, y_m)$, with $|p| \le (n)^{0.1}$, can evaluate p on **all** $(x, y) \in A \times B$ in $(n^2 + n^{1.1} \cdot m) poly(\log n)$ time.

All-Points (Hamming) Nearest Neighbors

<u>Given:</u> Sets A, B of n points in $\{0,1\}^d$, $d = c \cdot \log n$
<u>Task:</u> For all $x \in A$, output $y \in B$ that *minimizes* $h(x, y)$

Can easily be computed in $O(n \cdot 2^d)$ time and $O(n^2 \cdot d)$ time

Thm [AW'15] All-Points Hamming Nearest Neighbors can be solved on n points in $c \log n$ dimensions in randomized $n^{2 - \frac{1}{O(c \log^2 c)}}$ time.

"Truly subquadratic time" for O(log n) dimensions

Idea 1: Set p to be a PTF for testing whether two vectors of length d have Hamming distance $\leq d - k$, for each $k = d, d-1, \ldots, 2, 1$.

$$p(x_1, \ldots, x_d, y_1, \ldots, y_d) = k - \frac{1}{2} - \sum_{i=1}^{d} \underbrace{x_i y_i + (1 - x_i)(1 - y_i)}_{EQ(x_i, y_i)}$$

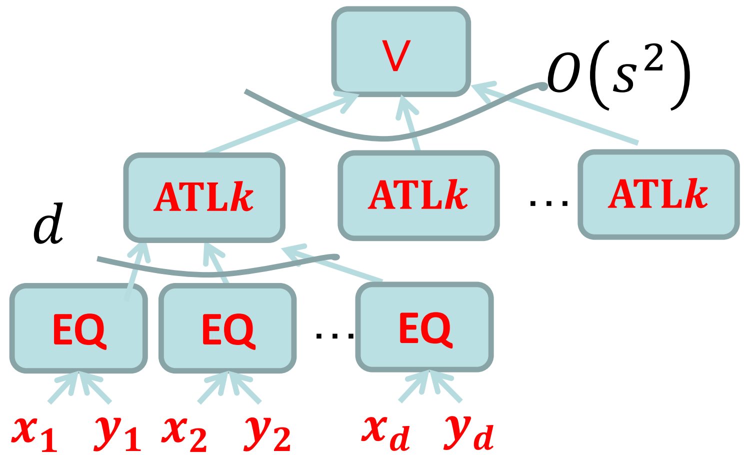
When you eval p on all pairs of vectors, only gives $\Omega(n^2)$ time...

All-Points (Hamming) Nearest Neighbors

Given: Sets A, B of n points in $\{0,1\}^d$, $d = c \cdot \log n$
Task: For all $x \in A$, output $y \in B$ that *minimizes* $h(x,y)$

Idea 2: Group A, B into $\sim n/s$ groups of size s, and let p determine whether there is a close pair among a group from A and a group from B.

Let $ATLk(x)$ output 1 iff sum of the bits in x is at least k.
Consider the circuit C_k defined on s vectors from A and B:



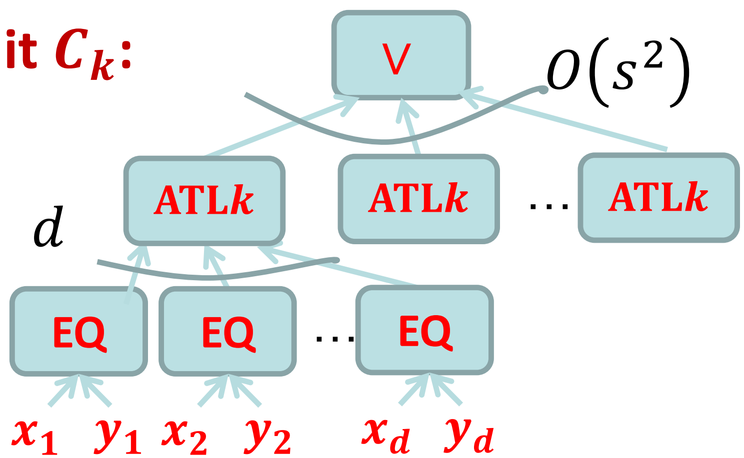
C_k outputs 1 iff *some* pair of points in the input has Hamming distance at most $d - k$

Goal: Construct polynomial representing C_k with few monomials

All-Points (Hamming) Nearest Neighbors

Thm [AW'15] All-Points Hamming Nearest Neighbors can be solved on n points in $c \log n$ dimensions in randomized $n^{2-\frac{1}{O(c\log^2 c)}}$ time.

Consider the circuit C_k:



C_k outputs 1 iff some pair of points has Hamming distance at most $d-k$

Goal: Construct a "nice" probabilistic polynomial simulating C_k

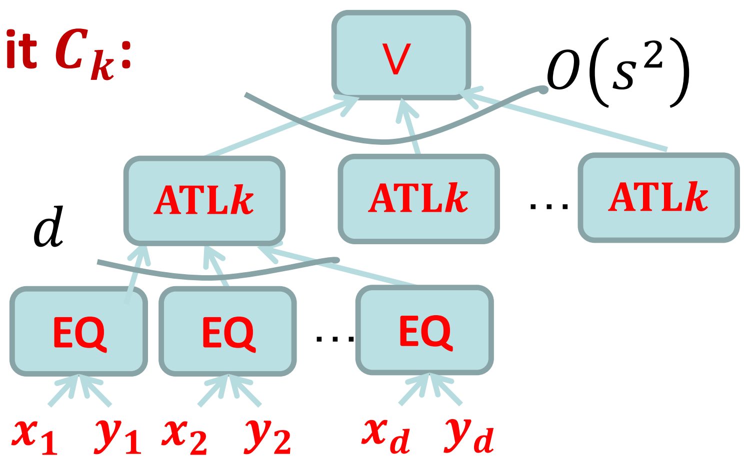
Thm [AW'15] Every symmetric function on d inputs has a probabilistic polynomial of degree $O\left(\sqrt{d\log\left(\frac{1}{\epsilon}\right)}\right)$ with error at most ϵ

Replace each $ATLk$ in C_k with a prob. polynomial having error $\leq \frac{1}{s^3}$: obtain a "somewhat sparse" probabilistic polynomial for C_k

All-Points (Hamming) Nearest Neighbors

Thm [AW'15] All-Points Hamming Nearest Neighbors can be solved on n points in $c \log n$ dimensions in randomized $n^{2 - \frac{1}{O(c \log^2 c)}}$ time.

Consider the circuit C_k:



C_k outputs 1 iff some pair of points has Hamming distance at most $d - k$

Goal: Construct a "nice" probabilistic polynomial simulating C_k

Nearest Neighbors on n points reduces to Evaluating C_k on $O\left(\frac{n^2}{s^2}\right)$ points, $\forall k$

<u>Evaluation Lemma</u> [C'82,W'14] Given sets $A, B \subseteq \{0,1\}^m$, $|A| = |B| = n/s$, and a polynomial $q(x_1, \dots, x_m, y_1, \dots, y_m)$ over $\mathbb{F}$, with $|q| \leq (n/s)^{0.1}$, can evaluate q on all $(x, y) \in A \times B$ in $\left(\left(\frac{n}{s}\right)^2 + \left(\frac{n}{s}\right) \cdot m\right) poly(log\, n)$ time.

New Results

All allow you to compute an OR of s At-Least-K functions!

Prob Polys with less randomness:

Thm: Every symmetric function on d inputs has a prob. poly. of degree $O\left(\sqrt{d\log(s)}\right)$ and error $1/s$, using $O(\log d \cdot \log(d \cdot s))$ random bits

- degree $O\left(\sqrt{d\log(s)}\right)$ using $\Omega(d)$ random bits was known [AW'15]
- degree lower bound $\Omega\left(\sqrt{d\log(s)}\right)$ [R'87,S'87]

New Results

All allow you to compute an OR of s At-Least-K functions!

Prob Polys with less randomness:

Thm: Every symmetric function on d inputs has a prob. poly. of degree $O\left(\sqrt{d\log(s)}\right)$ and error $1/s$, using $O(\log d\cdot\log(d\cdot s))$ random bits

A "Nice" PTF for the *At-Least-K* (threshold) function:

Thm: For all $s>0$, there is a PTF $P(x_1,\dots,x_d)$ of degree $O\left(\sqrt{d\log(s)}\right)$ for $ATLk$ on d variables such that for all x,

- $ATLk(x)=0\Rightarrow -1<P(x)<0$
- $ATLk(x)=1\Rightarrow P(x)>s$

An "Even Nicer" Probabilistic PTF for *At-Least-K*:

Thm: For all $s>0$, there is a ***probabilistic*** PTF $P(x_1,\dots,x_d,y)$ of degree $O\left(d^{\frac{1}{3}}\cdot\log^{\frac{2}{3}}(ds)\right)$ for $ATLk$ on d variables such that for all x,

- $ATLk(x)=0\Rightarrow \Pr_y[-1<P(x,y)<0]\geq 1-\frac{1}{s}$
- $ATLk(x)=1\Rightarrow \Pr_y[P(x,y)>s]\geq 1-\frac{1}{s}$

Main Applications

All-Points Nearest Neighbors in Hamming, ℓ_1, and ℓ_2

Given n red and n blue points in $D^{c \log n}$, we can:

- Find a Hamming nearest blue nbr, for all red points,
  - In randomized $dn + n^{2-1/\tilde{O}(c^{0.5})}$ time
  - In deterministic $dn + n^{2-1/\tilde{O}(c)}$ time *("derandomizes"* [AW'15]*)*
- Find $(1+\varepsilon)$-*approximate* nearest blue nbr for all red points in randomized $dn + n^{2-\tilde{\Omega}(\varepsilon^{1/3})}$ time, in Hamming, ℓ_1, and ℓ_2 metrics *(improving over Locality Sensitive Hashing* [IM'98]*, G. Valiant* [V'13]*)*

Circuit Complexity

- ACC∘THR∘THR circuits can be evaluated on all 2^n inputs in $2^n \cdot poly(n)$ (deterministic) time*

  *circuits with $n^{2-\varepsilon}$ bottom THR gates, and $2^{n^{\varepsilon}}$ gates elsewhere

  Implies analogous lower bounds for this circuit class
- Satisfiability of subexp-size MAJORITY∘AC0∘THR∘AC0∘THR circuits can be decided in randomized << 2^n time*

  *where MAJORITY and THR gates have fan-in $n^{6/5-\varepsilon}$

Prob. Polys With Less Randomness

Thm: ATL_k on d inputs has a prob. poly. of degree $O\left(\sqrt{d\log(s)}\right)$ and error $1/s$, using $O(\log d \cdot \log(d \cdot s))$ random bits

Sketch: Let's outline the construction from [AW'15] for ATL_k.

Let $\delta = \Theta\left(\frac{\log^{1/2}(s)}{d^{1/2}}\right)$.

Recursive construction. For an input $x \in \{0,1\}^d$, we have two cases:

1. If $|x| \notin [k - \delta d, k + \delta d]$: Construct a shorter input $\widetilde{x}$ by random sampling a $\frac{1}{10}$-fraction of x. Let $\widetilde{k} = k/10$. By Chernoff-Hoeffding and our choice of δ, it's likely that $ATL_{\widetilde{k}}(\widetilde{x}) = ATL_k(x)$, so use the polynomial $ATL_{\widetilde{k}}(\widetilde{x})$.
2. If $|x| \in [k - \delta d, k + \delta d]$: Use an exact polynomial A of degree $O(\delta d)$ that's guaranteed to give the correct answer (polynomial interpolation).

To determine which of the cases we're in, use $ATL_{(\widetilde{k}+\delta d)}(\widetilde{x})$ and $ATL_{(\widetilde{k}-\delta d)}(\widetilde{x})$.

$$ATL_k(x) \approx$$
$$\left(1 - ATL_{(\widetilde{k}+\delta d)}(\widetilde{x})\right) ATL_{(\widetilde{k}-\delta d)}(\widetilde{x}) \cdot A(x) + \left(1 - \left(1 - ATL_{(\widetilde{k}+\delta d)}(\widetilde{x})\right) ATL_{(\widetilde{k}-\delta d)}(\widetilde{x})\right) \cdot ATL_{\widetilde{k}}(\widetilde{x})$$

Observation: In the analysis, we only need $O(\log d)$-wise independence to generate a good random sample $\widetilde{x}$ of x

“Nice” PTFs for Threshold Functions

A “Nice” PTF for the *At-Least-K* function:

Thm: For all $s > 0$, there is a PTF $P(x_1, \dots, x_d)$ of degree $O\left(\sqrt{d \log(s)}\right)$ for $ATLk$ on d variables such that for all x,

- $ATLk(x) = 0 \Rightarrow -1 < P(x) < 0$
- $ATLk(x) = 1 \Rightarrow P(x) > s$

Idea: Use Chebyshev polynomials!

$$T_q(x) = \sum_{i=0}^{\lfloor q/2 \rfloor} \binom{q}{2i} \left(x^2 - 1\right)^i x^{q-2i}$$

$$T_q(x) = 2xT_{q-1}(x) - T_{q-2}(x),$$
$$T_0(x) = 1,$$
$$T_1(x) = x.$$

$$T_q(x) = \begin{cases} cos\left(q \cdot arccos(x)\right), |x| \leq 1 \\ \frac{1}{2}\left(x - \sqrt{x^2 - 1}\right)^q + \frac{1}{2}\left(x + \sqrt{x^2 - 1}\right)^q, |x| \geq 1 \end{cases}$$

"Nice" PTFs for Threshold Functions

A "Nice" PTF for the *At-Least-K* function:

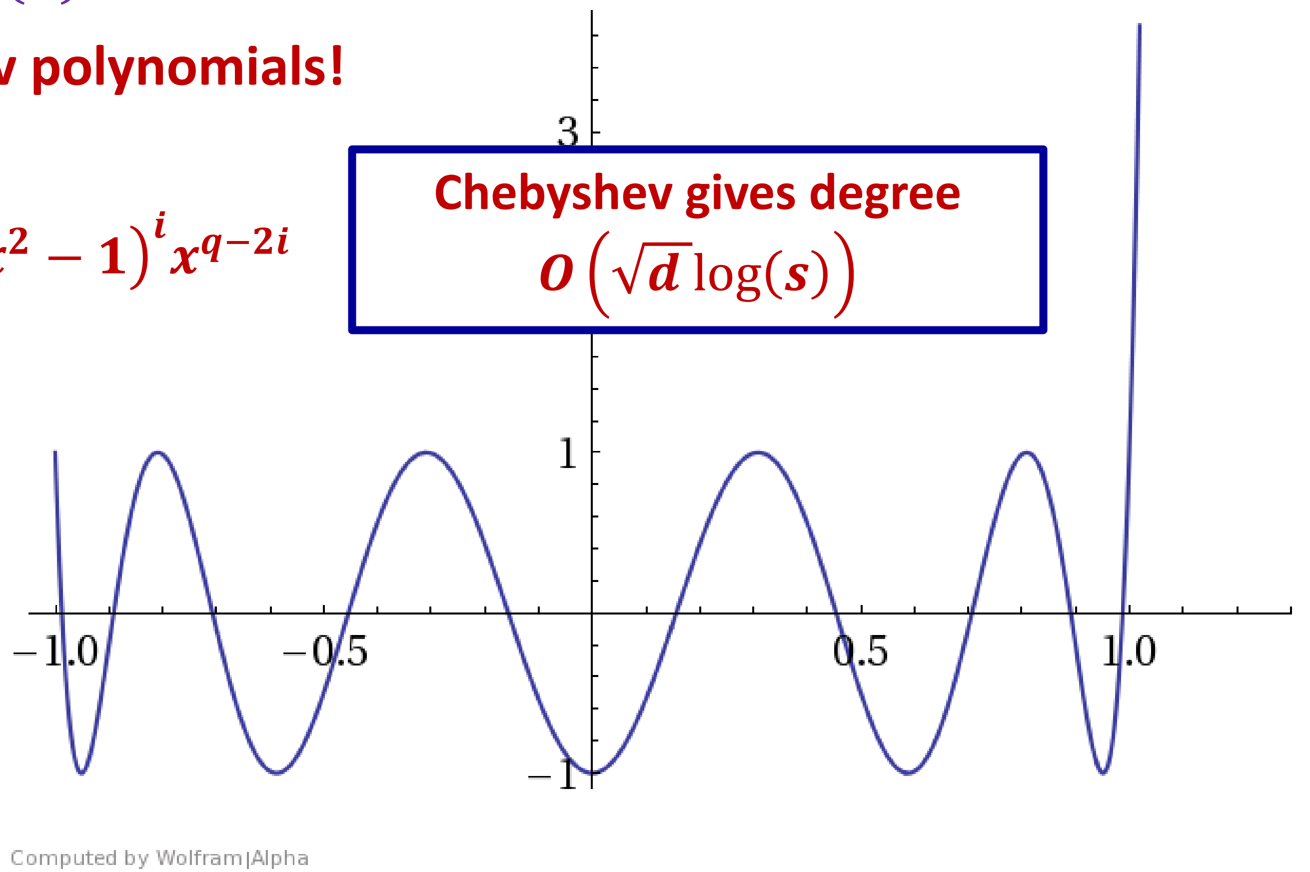
Thm: For all $s > 0$, there is a PTF $P(x_1, \ldots, x_d)$ of degree $O\left(\sqrt{d}\log(s)\right)$ for $ATLk$ on d variables such that for all x,

- $ATLk(x) = 0 \Rightarrow -1 < P(x) < 0$
- $ATLk(x) = 1 \Rightarrow P(x) > s$

Idea: Use Chebyshev polynomials!

$$T_q(x) = \sum_{i=0}^{\lfloor q/2 \rfloor} \binom{q}{2i} \left(x^2 - 1\right)^i x^{q-2i}$$

Chebyshev gives degree $O\left(\sqrt{d}\log(s)\right)$



If $x \in [-1, 1]$ then $T_q(x) \in [-1, 1]$

If $x \geq 1 + \varepsilon$ then $T_q(x) \geq \frac{1}{2} e^{q\sqrt{\varepsilon}}$

"Nice" PTFs for Threshold Functions

A "Nice" PTF for the *At-Least-K* function:

Thm: For all $s > 0$, there is a PTF $P(x_1, \ldots, x_d)$ of degree $O\left(\sqrt{d \log(s)}\right)$ for $ATLk$ on d variables such that for all x,

- $ATLk(x) = 0 \Rightarrow -1 < P(x) < 0$
- $ATLk(x) = 1 \Rightarrow P(x) > s$

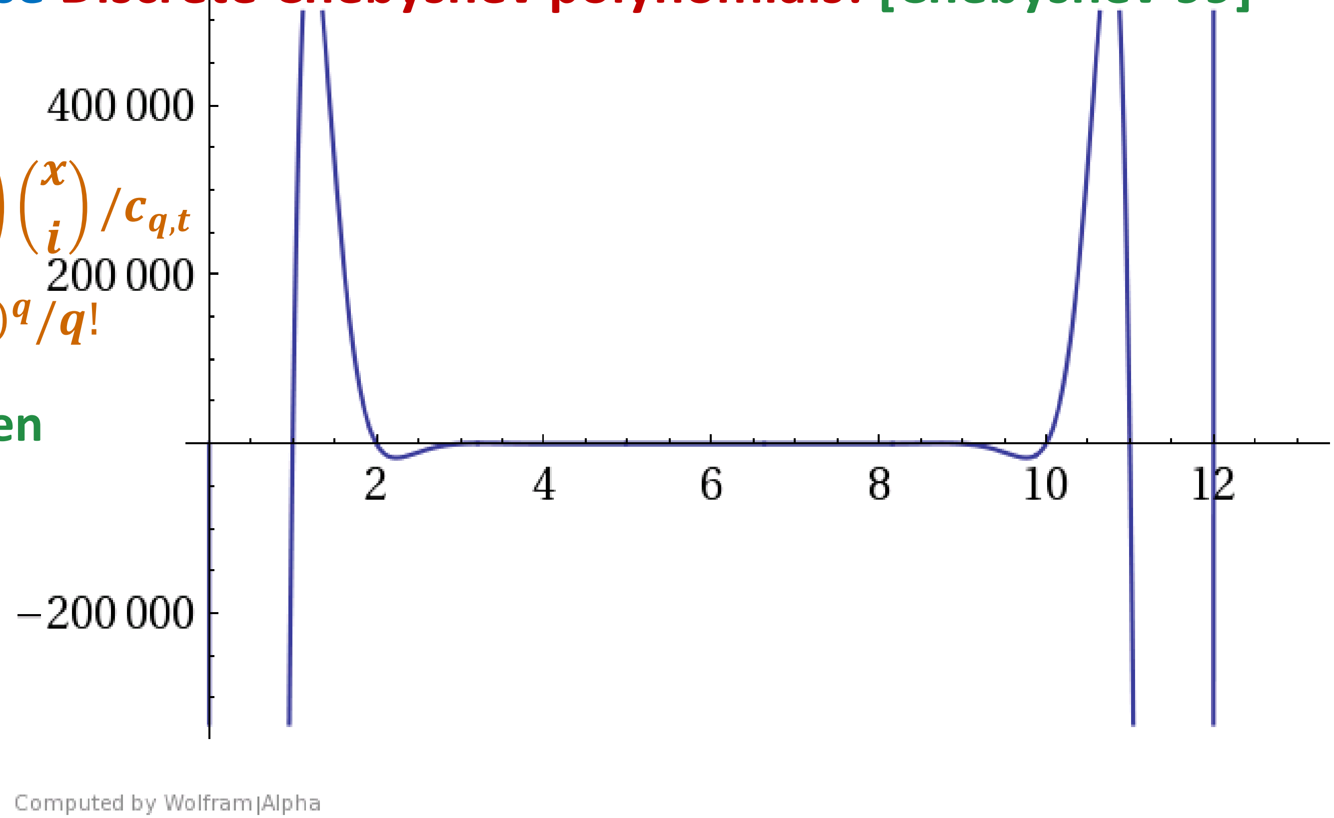
Better degree: use Discrete Chebyshev polynomials! [Chebyshev'99]

$$D_{q,t}(x) := \sum_{i=0}^{q} (-1)^i \binom{q}{i} \binom{t-x}{q-i} \binom{x}{i} / c_{q,t}$$

where $c_{q,t} = (t+1)^q / q!$

If $x \in \{0, 1, \ldots, t\}$ then $D_{q,t}(x) \in [-1, 1]$

If $x \geq 1 + t$ then $D_{q,t}(x) \geq e^{q^2/8t}$



“Nice” PTFs for Threshold Functions

A “Nice” PTF for the *At-Least-K* function:

Thm: For all $s > 0$, there is a PTF $P(x_1, \dots, x_d)$ of degree $O\left(\sqrt{d \log(s)}\right)$ for $ATLk$ on d variables such that for all x,

- $ATLk(x) = 0 \Rightarrow -1 < P(x) < 0$
- $ATLk(x) = 1 \Rightarrow P(x) > s$

Fact [Chebyshev’99] For all $q \in \left[\Omega\left((t \log t)^{0.5}\right), t\right]$,

- If $x \in \{0, 1, \dots, t\}$ then $D_{q,t}(x) \in [-1, 1]$
- If $x \le -1$ then $D_{q,t}(x) \ge exp\left(\frac{q^2}{8(t+1)}\right)$.

Define $P(x_1, \dots, x_d) \coloneqq D_{q,d}\left((k-1) - \sum_i x_i\right)$,
where $q := \Theta((d \log s)^{0.5})$

Then:

$\sum_i x_i < k \Rightarrow P(x) \in [-1, 1]$

$\sum_i x_i \ge k \Rightarrow P(x) \ge e\hat{}\Theta(d \log(s)/d) = e\hat{}(\Theta(\log(s)) = poly(s)$

Shift P **a little, to get the desired properties in the theorem.**

"Even Nicer" Probabilistic PTF

An "Even Nicer" Probabilistic PTF for *At-Least-K*:

Thm: For all $s > 0$, there is a ***probabilistic*** PTF $P(x_1, \ldots, x_d, y)$ of degree $O\left(d^{\frac{1}{3}} \cdot \log^{\frac{2}{3}}(ns)\right)$ for $ATLk$ on d variables such that for all x,

- $ATLk(x) = 0 \Rightarrow \Pr_y[-1 < P(x,y) < 0] \geq 1 - \frac{1}{s}$
- $ATLk(x) = 1 \Rightarrow \Pr_y[P(x,y) > s] \geq 1 - \frac{1}{s}$

Idea: Combine the previous two constructions!

Really Really Sketchy Sketch: Let $\delta > 0$.

Construct $\widetilde{x}$ by random sampling δd bits of x.

Let $Q(\widetilde{x})$ be a probabilistic polynomial for $ATL_{k'}$ with error at most $\frac{1}{2s}$.

(here the parameter k' is slightly smaller than $\widetilde{k}$)

Take a modified discrete Chebyshev polynomial $D_{q',t'}(x)$ which

- "blows up" to $> s$ when $\sum_i x_i > k - 1$
- And otherwise stays in the interval $[-1, 1]$.

Set $P(x) = D_{q',t'}(x) \cdot Q(\widetilde{x})$ for $t' \approx d/\sqrt{\delta d}$

Careful case analysis and new setting of δ (and q') yields the result!

Conclusion

- What is the power/limit of *probabilistic* PTFs representing Boolean functions?
  - How easy/difficult is it to prove degree lower bounds for such representations?
- Can our SAT algorithm for
  MAJ ∘ AC0 ∘ THR ∘ AC0 ∘ THR
  be derandomized?
  - Would imply stronger circuit lower bounds
- How much can the $\boldsymbol{n}^{2-\widetilde{\Omega}(\varepsilon^{1/3})}$ runtime for $(\mathbf{1}+\boldsymbol{\varepsilon})$-approximate batch nearest neighbor be improved?

Thank you!