Improved Approximate Degree Bounds For k-distinctness

Shuchen Zhu
Georgetown University

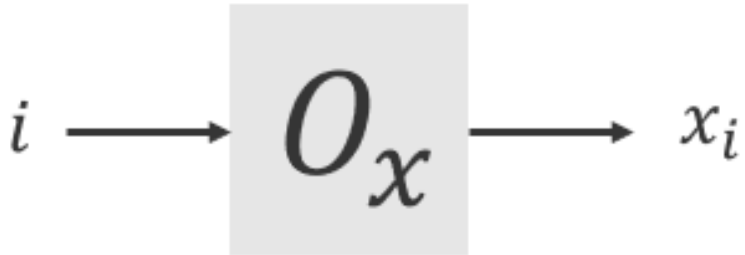
Nikhil Mande
Georgetown

Justin Thaler
Georgetown

arXiv:2002.08389

Query complexity

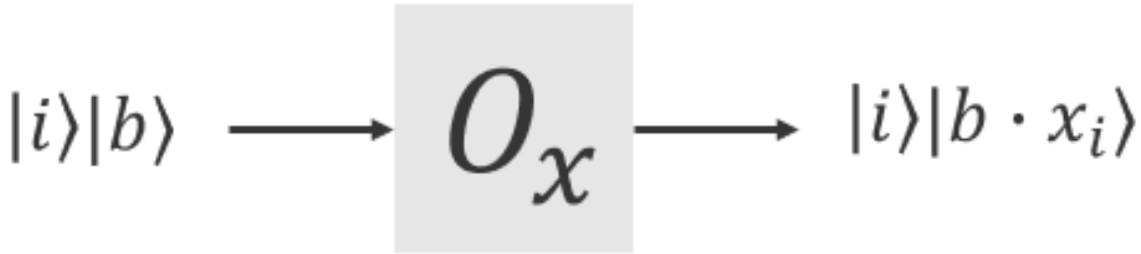
Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function.
Goal: for any given input $x \in \{-1,1\}^n$, compute $f(x)$ by reading as few bits as possible from x.
Equivalently, compute $f(x)$ using an algorithm that invokes the following oracle the least number of times:



- f is known to the algorithm.
- input x is not known to the algorithm.

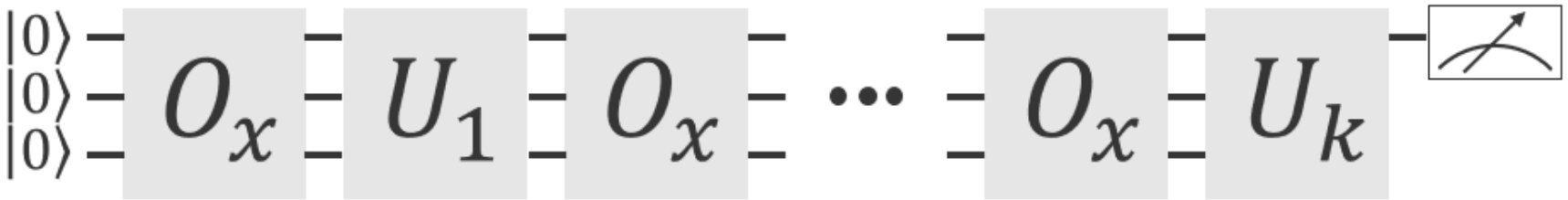
Quantum query complexity

In the quantum setting we have the following quantum oracle:



Quantum query complexity $Q(f)$

Minimum number of quantum oracle O_x in a quantum circuit that for every input x, outputs $f(x)$ with error $\leq 1/3$.



Why query complexity

- Algorithmic Motivation.
  - Most quantum algorithms are naturally phrased as query algorithms. E.g., Shor, Grover, Hidden Subgroup, Linear systems (HHL), etc.
  - Algorithms often transfer to the circuit model, while the query complexity abstraction gets rid of unnecessary details.
- Complexity Motivation.
  - We can prove statements about the power of different computational models!
  - E.g., exponential separation between classical and quantum algorithms.

The k-distinctness problem

$\mathrm{DIST}^k_{N,R}$

Given N numbers in range of size R, does any number appear $\geq k$ times?

- For $k = 2$ it becomes Element Distinctness problem, which is an important function with a long history throughout TCS and is well-understood.
- For $k > 2$, quantum query complexity of k-distinctness remains open.
- It has connections to finding multi-collisions in hash functions, which is highly relevant to cryptography.
- k is constant throughout the talk unless explicitly stated otherwise.

Historical results of k-distinctness

- For $k = 2$, Element Distinctness (ED) had been shown to satisfy $Q(\text{ED}) = \Theta(N^{\frac{2}{3}})$ [AS04, Amb07].
- For $k > 2$
  - Upper bound:
    - $Q(\text{DIST}^k_{N,R}) = O(N^{\frac{k}{k+1}})$, quantum walks [Amb07].
    - $Q(\text{DIST}^k_{N,R}) = O(N^{\frac{3}{4}-\frac{1}{2^{k+2}-4}})$, learning graphs [Bel12].
  - Lower bound:
    - $Q(\text{DIST}^k_{N,R}) = \Omega(Q(\text{ED})) = \Omega(N^{\frac{2}{3}})$ [AS04].
    - $Q(\text{DIST}^k_{N,R}) = \widetilde{\Omega}(N^{\frac{3}{4}-\frac{1}{2k}})$, polynomial method [BKT18].
    - $Q(\text{DIST}^k_{N,R}) = \widetilde{\Omega}(N^{\frac{3}{4}-\frac{1}{4k}})$, our result, polynomial method.
- Our lower bound result shows for the first time that for 4-distinctness is strictly harder than Element Distinctness.
- Our lower bound result also applies to more general approximate degree.

Approximate degree

ϵ-approximation

A polynomial $p : \mathbb{R}^n \to \mathbb{R}$ ϵ-approximates a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ 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.

The connection between approximate degree and quantum query complexity is due to the seminal result [BBC+01]:

$$Q(f) \geq \frac{1}{2}\widetilde{\deg}(f).$$

- We show that $\widetilde{\deg}(\mathrm{DIST}^k_{N,R}) \geq \widetilde{\Omega}(N^{\frac{3}{4}-\frac{1}{4k}})$, for constant k.

Summary of results

Lower bound result

For any constant $k \geq 2$, the approximate degree and quantum query complexity of the k-distinctness function with domain size N and range size $R \geq N$ is $\widetilde{\Omega}(N^{\frac{3}{4}-\frac{1}{4k}})$.

Upper bound result

For any $k \leq \text{polylog}(N)$, the approximate degree of k-distinctness is $\widetilde{O}(N^{\frac{3}{4}})$.

Approximate degree upper bound

Upper bound result

For any $k \leq \text{polylog}(N)$, the approximate degree of k-distinctness is $\widetilde{O}(N^{\frac{3}{4}})$.

- The previous best result [Bel12]

$$Q(\text{DIST}_{N,R}^{k}) = \exp(O(k)) \cdot O(N^{\frac{3}{4} - \frac{1}{2^{k+2}-4}}).$$

- This becomes linear for $k \geq \Omega(\log(N))$.
- The approximate degree upper bound result does not imply a quantum query complexity upper bound, but it implies that polynomial method cannot yield a better than $N^{\frac{3}{4}}$ lower bound for $Q(\text{DIST}_{N,R}^{k})$.
- An upper bound on the quantum query complexity of $(\log n)$-distinctness would imply an upper bound for min-entropy estimation [LW19].

Lower bound techniques

Approximate degree lower bound technique

What is best error achievable by **any** degree d approximation of f?

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

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

Dual LP:

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

Dual Characterization of Approximate Degree

Fact: $\widetilde{\deg}_\epsilon(f) > d$ iff there exists a function $\psi\colon \{-1,1\}^n \to \mathbb{R}$ 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) $\sum_{x\in\{-1,1\}^n} \psi(x)q(x) = 0$, when $\deg q \leq d$ "$\mathrm{phd}(\psi) > d$"

Such a ψ is called a dual polynomial.

Connection between $\mathrm{DIST}^k_{N,R}$ and composed functions

Theorem [BKT18]

Let $N, R \in \mathbb{N}$ and $2 \leq k \leq N$ be any integer. Then for any $\epsilon > 0$,

$$\widetilde{\deg}_\epsilon(\mathrm{DIST}^k_{N,R+N}) = \Omega\left(\frac{1}{\log R} \cdot \widetilde{\deg}_\epsilon(\mathrm{OR}_R \circ \mathrm{THR}^k_N)^{\leq N}\right).$$

- $\leq N$ denotes the the domain is restricted to inputs of Hamming weight less than N.
- $\mathrm{OR}_N : \{-1,1\}^N \to \{-1,1\}$ equals 1 if $x = \mathbf{1}^N$, and -1 otherwise.
- *Threshold* function $\mathrm{THR}^k_N : \{-1,1\}^N \to \{-1,1\}$ equals 1 for inputs of Hamming weight less than k, and -1 otherwise.
- Hamming weight is the number of -1 in a given input string.

Dual formulation

Find a dual witness Γ for $(\mathrm{OR}_R \circ \mathrm{THR}_N^k)^{\leq N}$. Γ must satisfy the following properties:

- **Normalization:** $\|\Gamma\|_1 = 1$.
- **Pure high degree:** There exists a $D = \widetilde{\Omega}\left(N^{\frac{3}{4}-\frac{1}{4k}}\right)$ such that for every polynomial $p : \{-1,1\}^{RN} \to \mathbb{R}$ of degree less than D, we have $\sum_x p(x)\Gamma(x) = 0$.
- **Correlation:** $\sum_x \Gamma(x)(\mathrm{OR}_R \circ \mathrm{THR}_N^k)(x) > 1/3$.
- **Exponentially little mass on inputs of large Hamming weight:** $\sum_{x \notin (\{-1,1\}^{RN})^{\leq N}} |\Gamma(x)| \leq (2NR)^{-\widetilde{\Omega}\left(R^{\frac{3}{4}-\frac{1}{4k}}\right)}$ for all $x \notin (\{-1,1\}^{RN})^{\leq N}$ (**strong dual decay**).

- We alter dual polynomial Λ in [BKT18].

Dual constructions in [BKT18]

Construct three individual dual polynomials θ, ϕ and ψ.

$$\underbrace{\mathrm{OR}_R \circ \mathrm{THR}_N^k}_{\Lambda} = \underbrace{\mathrm{OR}_{R/4^k}}_{\theta} \underbrace{\circ}_{\star} \underbrace{\mathrm{OR}_{4^k}}_{\phi} \underbrace{\circ}_{\star} \underbrace{\mathrm{THR}_N^k}_{\psi}$$

Dual block composition $\star$

Let $\theta : \{-1,1\}^n \to \mathbb{R}, \phi : \{-1,1\}^m \to \mathbb{R}$ be any functions. Let $x = (x_1, \ldots, x_n)$ where each $x_i \in \{-1,1\}^m$. Define the *dual block composition* $\theta \star \phi$ to be

$$\theta \star \phi(x) = 2^n \theta(\mathrm{sgn}(\phi(x_1)), \ldots, \mathrm{sgn}(\phi(x_n))) \prod_{i=1}^{n} |\phi(x_i)|.$$

We need to make sure four conditions of Λ are satisfied: **normalization**, **pure high degree**, **correlation** and **strong dual decay**.

Four conditions of dual polynomial Λ

- Dual block composition generically preserves necessary conditions for normalization, pure high degree, and dual decay.
- But for correlation it needs novel analysis:
  - Usually correlation does not hold automatically after dual composition.
  - Heavily rely on ψ correlating very well with THR_N^k in [BKT18].
  - Requiring such high correlation between ψ and THR_N^k hurts the final degree lower bound

Our modification to Λ

Our solution to improve correlation: inspired by [She12], alter Λ again by attaching a polynomial p to it:

$$\Gamma(x) = (\theta \star \phi \star \psi')(x) \cdot p(x).$$

This is a variant of dual composition that improves correlation.

- We modify p to account for refined error notions that arise in the analysis of k-distinctness.

Open questions

- Can we do better than our $\widetilde{\Omega}(N^{\frac{3}{4}-\frac{1}{4k}})$ lower bound for k-distinctness?
  - Recall the best upper bound is $O(N^{\frac{3}{4}-\frac{1}{2^{k+2}-4}})$ [Bel12].
  - Liu and Zhandry [LZ19] showed that the quantum query complexity of a certain *search* version of k-distinctness is $\Theta(N^{\frac{1}{2}-\frac{1}{2^k-1}})$. This may suggest $\frac{3}{4}-\frac{1}{\exp(O(k))}$ is the right exponent for k-distinctness.
  - We suspect that techniques based on dual-block-composition have reached their limit.
- Intermediate Goal: improve over the long-standing $\Omega(N^{\frac{2}{3}})$ lower bound for 3-distinctness.
- A quantum query complexity upper bound for $(\log n)$-distinctness?