New Techniques for Low-Degree Lower Bounds

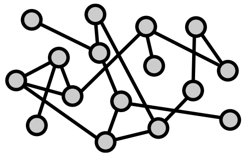
Alex Wein

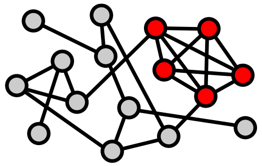
UC Davis

Based on joint works with: Tselil Schramm; Cindy Rush, Fiona Skerman, Dana Yang; Pravesh Kothari, Santosh Vempala, Jeff Xu

Average-Case Algorithmic Tasks

chosen at random

$\mathbb{Q}$: G(n,1/2) 

$\mathbb{P}$: G(n,1/2) + {k-clique} 

- **Detection**: distinguish $\mathbb{P}$ vs $\mathbb{Q}$ w.h.p.
- **Recovery**: given G ~ $\mathbb{P}$, identify the clique vertices
- **Refutation**: given G ~ $\mathbb{Q}$, *prove* there is no k-clique
  - If graph has a k-clique, output is *always* MAYBE
  - If graph is drawn from $\mathbb{Q}$, output is NO w.h.p.

refutation task

- All have poly-time algorithms when $k = \Omega(\sqrt{n})$ [Alon, Krivelevich, Sudakov '98]
- No poly-time algorithms known when $k = o(\sqrt{n})$

3 tasks not equivalent in general!

Low-Degree Polynomial (LDP) Algorithms

[Hopkins, Steurer '17; Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer '17; …]

- **Degree-D algorithm**: multivariate polynomial of degree D

$$f: \{0,1\}^{\binom{n}{2}} \to \mathbb{R}$$

Input: graph

Output: number

proxy for runtime

deg O(1) $\lesssim$ poly-time $\lesssim$ deg O(log n)

n^D terms

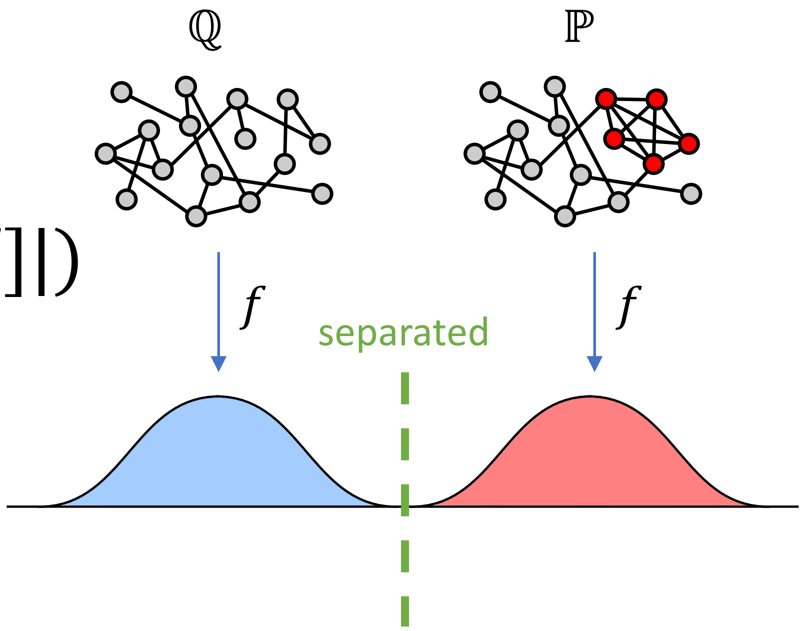
captures spectral, AMP, …

- Examples:
  - Edge count: $f(A) = \sum_{i<j} A_{ij}$
  - Triangle count: $f(A) = \sum_{i<j<k} A_{ij} A_{ik} A_{jk}$
  - Degree of vertex 1: $f(A) = \sum_{1<i} A_{1i}$
  - Count triangles containing vertex 1: $f(A) = \sum_{1<i<j} A_{1i} A_{1j} A_{ij}$
  - Spectral (approx): $f(A) = \mathrm{Tr}(A^{2m}) = \sum_i \lambda_i^{2m} \approx \lambda_{\max}^{2m}$

How to Define "Success" for an LDP Algorithm?

$n \to \infty$

- **Detection**: f "strongly separates" $\mathbb{P}$ and $\mathbb{Q}$
  $$\sqrt{\max\{\mathrm{Var}_{\mathbb{P}}(f), \mathrm{Var}_{\mathbb{Q}}(f)\}} = o(|\mathrm{E}_{\mathbb{P}}[f] - \mathrm{E}_{\mathbb{Q}}[f]|)$$
- **Recovery**: small mean squared error
  $$\mathrm{E}_{\mathbb{P}}[(f(A) - x)^2] \ll \mathrm{Var}_{\mathbb{P}}(x), \quad x = \mathbb{1}_{1 \in \text{clique}}$$
- **Refutation**: f "strongly separates" $\mathbb{Q}$ and R_k
  - If A has a k-clique then $f(A) \geq 1$
  - $\mathrm{E}_{\mathbb{Q}}[f^2] = o(1)$



property: exists a k-clique

- These are natural *sufficient* conditions
- Won't cover: random optimization problems
  [Gamarnik, Jagannath, **W** '20; **W** '21; Bresler, Huang '21; Huang, Sellke '21; …]

Prototypical Result: Planted Clique

Theorem (lower bound) If $k \leq n^{1/2-\epsilon}$, for some $D = D_n = \omega(\log n)$,

- (Detection) no degree-D polynomial strongly separates $\mathbb{P}$, $\mathbb{Q}$
  [Hopkins '18; Barak, Hopkins, Kelner, Kothari, Moitra, Potechin '16]
- (Recovery) no degree-D polynomial has small MSE
  [Schramm, **W** '22]
- (Refutation) no degree-D polynomial strongly separates $\mathbb{Q}$, R_k
  [Kothari, Vempala, **W**, Xu '23]

Theorem (upper bound) If $k \geq cn^{1/2}$, for some $D = D_n = O(\log n)$,

- (Detection) some degree-D polynomial strongly separates $\mathbb{P}$, $\mathbb{Q}$
- (Recovery) some degree-D polynomial has small MSE
- (Refutation) some degree-D polynomial strongly separates $\mathbb{Q}$, R_k

Focus of This Talk

- Not the focus of this talk:
  - Failure of $O(\log n)$-degree polynomials is “evidence” for inherent hardness
  - Relation to sum-of-squares, statistical query model, …
  - State-of-the-art results for specific problems
- Instead:
  - Proof ideas for lower bounds (failure of all degree-D algorithms)

Reformulation as a Ratio

- **Detection**: to rule out strong separation of $\mathbb{P}$, $\mathbb{Q}$, suffices to show
  $$\chi^2_{\le D}(\mathbb{P}\|\mathbb{Q}) + 1 := \max_{f \deg D} \frac{\mathrm{E}_{\mathbb{P}}[f]}{\sqrt{\mathrm{E}_{\mathbb{Q}}[f^2]}} = O(1) \qquad \|L^{\le D}\|$$
  e.g. [Hopkins '18]
  - or $\chi^2_{\le D}(\mathbb{P}'\|\mathbb{Q}) = O(1)$ for conditional $\mathbb{P}'$ [Bandeira, El Alaoui, Hopkins, Schramm, **W**, Zadik '22; Coja-Oghlan, Gebhard, Hahn-Klimroth, **W**, Zadik '22; Dhawan, Mao, **W** '23]
- **Recovery**: to rule out small MSE, suffices to show
  $$\max_{f \deg D} \frac{\mathrm{E}_{\mathbb{P}}[f \cdot x]}{\sqrt{\mathrm{E}_{\mathbb{P}}[f^2]}} \ll \cdots \qquad x = \mathbb{1}_{1 \in \text{clique}}$$
- **Refutation**: to rule out strong separation of $\mathbb{Q}$, R_k, suffices to construct a distribution $\widetilde{\mathbb{P}}$ supported on R_k, and show $\chi^2_{\le D}(\widetilde{\mathbb{P}}\|\mathbb{Q}) = O(1)$

Explicit Solution

- In any case, our goal is to upper-bound something of the form

$$\mathrm{Adv}_{\leq D} := \max_{f \deg D} \frac{\mathrm{E}_{\mathbb{P}}[f \cdot y]}{\sqrt{\mathrm{E}_{\mathbb{H}}[f^2]}}$$

  - For detection: $y = 1$, $\mathbb{H} = \mathbb{Q}$
  - For recovery: $y = x$, $\mathbb{H} = \mathbb{P}$
- Choose a basis $\{h_\alpha\}$ for degree-D polynomials, expand $f(A) = \sum_\alpha \hat{f}_\alpha h_\alpha(A)$
- Define $c_\alpha = \mathrm{E}_{\mathbb{P}}[h_\alpha \cdot y]$ and $P_{\alpha\beta} = \mathrm{E}_{\mathbb{H}}[h_\alpha \cdot h_\beta]$
- Conclude:

$$\mathrm{Adv}_{\leq D} = \max_{\hat{f}} \frac{c^\top \hat{f}}{\sqrt{\hat{f}^\top P \hat{f}}} = \sqrt{c^\top P^{-1} c}$$

When $\mathbb{H}$ is a Product Measure…

- Recall: $\mathrm{Adv}_{\leq D} \coloneqq \max_{f \deg D} \frac{\mathrm{E}_{\mathbb{P}}[f \cdot y]}{\sqrt{\mathrm{E}_{\mathbb{H}}[f^2]}} = \sqrt{c^\top P^{-1} c}$
  - $c_\alpha = \mathrm{E}_{\mathbb{P}}[h_\alpha \cdot y],\ P_{\alpha\beta} = \mathrm{E}_{\mathbb{H}}[h_\alpha \cdot h_\beta]$
- If $\mathbb{H}$ has independent coordinates (product measure), choose $\{h_\alpha\}$ to be an orthonormal basis of polynomials: $\mathrm{E}_{\mathbb{H}}\left[h_\alpha \cdot h_\beta\right] = \mathbb{1}_{\alpha=\beta}$
  - $P = I,\ \mathrm{Adv}_{\leq D} = \|c\|$
  - Gives low-degree lower bounds for detection: $\mathbb{P}$ vs product measure $\mathbb{Q}$ [Hopkins, Steurer ’17; Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer ’17; …]
- This talk: what to do when $\mathbb{H}$ is not a product measure
  - Recovery: $\mathbb{H} = \mathbb{P}$ (mixture of product measures) [Schramm, **W** ’22]
  - Planted-vs-planted testing, e.g. distinguish 1 planted clique vs 2 planted cliques [Rush, Skerman, **W**, Yang ’23; Kothari, Vempala, **W**, Xu ‘23]

Overview

- I'll cover two approaches
  - Jensen trick [Schramm, **W** '22; …]
  - Tensor decomposition [**W** '23]
- I'll present these two in a unified way (credit: Jon Niles-Weed)
- Setup
  - Goal: lower bound on $\mathrm{E}_{\mathbb{H}}[f^2] = \|f\|^2$
  - Inner product / norm for functions: $\langle f, g\rangle := \mathrm{E}_{\mathbb{H}}[f \cdot g]$, $\|f\| := \sqrt{\langle f, f\rangle}$
  - For orthonormal basis $\{t_\gamma\}$, $\|f\|^2 = \sum_\gamma \langle t_\gamma, f\rangle^2$
  - For orthonormal set $\{t_\gamma\}$, $\|f\|^2 \geq \sum_\gamma \langle t_\gamma, f\rangle^2$

Blueprint

clique edges → X ; iid Bernoulli(1/2) → Z

- Example: $\mathbb{H}$ is planted clique distribution $A = X \vee Z$
- Write $f(A) = g(X, Z)$; every f induces some g
- Choose some orthonormal set of functions $\{t_\gamma(X, Z)\}$
- $\|f\|^2 = \|g\|^2 \geq \sum_\gamma \langle t_\gamma, g \rangle^2 =: \|w\|^2$ $\quad w_\gamma := \langle t_\gamma, g \rangle = \mathrm{E}_{X,Z}[t_\gamma \cdot g]$
- How does w depend on $\hat{f}$? $\quad$ Recall $f(A) = \sum_\alpha \hat{f}_\alpha h_\alpha(A)$
  - $w = M\hat{f}$ where $M_{\gamma\alpha} = \langle t_\gamma, h_\alpha \rangle$
- Will need explicit left inverse M^+ for M, i.e., $M^+ M = I$
- $\mathrm{Adv}_{\leq D} := \max_{f \deg D} \frac{\mathrm{E}_{\mathbb{P}}[f \cdot y]}{\sqrt{\mathrm{E}_{\mathbb{H}}[f^2]}} \leq \max_{\hat{f}} \frac{c^\top \hat{f}}{\|w\|} = \max_{\hat{f}} \frac{c^\top M^+ M \hat{f}}{\|M\hat{f}\|} \leq \|c^\top M^+\|$

More Details: Planted Clique

clique edges → X; iid Bernoulli(1/2) → Z

- Example: $\mathbb{H}$ is planted clique distribution $A = X \vee Z$
- Fourier characters $\alpha \subseteq \binom{[n]}{2}$, $\chi_\alpha(A) = \prod_{(i,j)\in\alpha}(-1)^{A_{ij}}$
- $\{\chi_\alpha(Z)\}$ are orthonormal, $\{\chi_\alpha(A)\}$ are not
- Choose $h_\alpha(A) = \chi_\alpha(A)$, $|\alpha| \le D$ -- basis for f
- Choose $t_\gamma(X, Z) = \chi_\gamma(Z)$, $|\gamma| \le D$ -- orthonormal set of functions
- Fortunately, M is upper-triangular: $M_{\gamma\alpha} := \langle t_\gamma, h_\alpha \rangle = 0 \text{ unless } \gamma \subseteq \alpha$
  - Can find explicit inverse $M^+ = M^{-1}$
- $\mathrm{Adv}_{\le D} \le \|c^\top M^{-1}\|$

Tensor Decomposition

- Given $n \times n \times n$ tensor $T = (1+\delta)a_1^{\otimes 3} + \sum_{j=2}^{r} a_j^{\otimes 3} \quad (\mathbb{P})$
  - $a_j \in \{\pm 1\}^n$ iid Rademacher
- Goal: recover a_{11}
- Poly-time when $r \ll n^{3/2}$ [Ma, Shi, Steurer '16; Ding, d'Orsi, Liu, Tiegel, Steurer '22]
- **Theorem (informal)** [W '23]: low-degree MMSE is small when $r \ll n^{3/2}$, trivial when $r \gg n^{3/2}$
- Recall: suffices to upper-bound

$$\max_{f \deg D} \frac{\mathrm{E}_{\mathbb{P}}[f \cdot a_{11}]}{\sqrt{\mathrm{E}_{\mathbb{P}}[f^2]}}$$

More Details: Tensor Decomposition

- Recall: $T = (1+\delta)a_1^{\otimes 3} + \sum_{j=2}^{r} a_j^{\otimes 3}$, $a_j \in \{\pm 1\}^n$ iid Rademacher
- Write $f(T) = g(a)$; every f induces some g
- Choose $\{h_\alpha(T)\}$ monomial basis -- basis for f
- Choose $t_\gamma(a) = \chi_\gamma(a)$ Fourier characters -- orthonormal set (basis)
- Some freedom to choose left inverse M^+
  - Left inverse: procedure for finding $\{h_\alpha(T)\}$-coefficients given $\{t_\gamma(a)\}$-coefficients
  - Fortunately a simple recursive construction for M^+ works
- $\mathrm{Adv}_{\le D} \le \|c^\top M^+\|$

Comments

- Other methods not mentioned in this talk:
  - Exact constant-degree MMSE for spiked Wigner via AMP [Montanari, **W** ‘22]
  - Annealed Franz-Parisi potential / low-overlap chi-squared
    [Bandeira, El Alaoui, Hopkins, Schramm, **W**, Zadik ’22]
- Open question: random regular graphs?

Thanks!