Fine-Grained Extensions of the Low-Degree Testing Framework

Alex Wein
UC Davis

Based on joint works with: {Jay Mardia, Kabir Aladin Verchand}, {Ankur Moitra}
Planted Clique Problem

• Find a planted \( k \)-clique in an \( n \)-vertex random graph
  • \( G(n,1/2) + \{\text{random } k\text{-clique}\} \)
  • Believed to have a statistical-computational gap

Any estimator fails

Statistical threshold

[Alon, Krivelevich, Sudakov ’98]

Statistically possible but no poly-time algorithm known!

Polynomial-time algorithm succeeds (w.h.p.)

[Alon, Krivelevich, Sudakov ’98]
**“Hard” Regime**

- How to show computational hardness?
- Average-case complexity is difficult...
- Instead:
  - Average-case reductions
  - Failure of restricted classes of algorithms
    - Statistical query (SQ) algorithms
    - Sum-of-squares (SoS) hierarchy
    - “Local” algorithms
    - Approximate message passing
    - ...
- This talk: **low-degree polynomial algorithms for hypothesis testing**
  - As opposed to recovery/estimation, optimization, refutation, ...
Low-Degree Testing
[Hopkins, Steurer ’17; Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer ’17; Hopkins ‘18, ...]

• **Degree-D test:** multivariate polynomial of degree $D = D_n$

\[ f : \{0,1\}^{(n)} \rightarrow \mathbb{R} \]

Input: graph \hspace{1cm} Output: number

• E.g. count edges, triangles, subgraphs, ...
  
  • $f(A) = \sum_{i<j<k} A_{ij}A_{ik}A_{jk}$ (triangle count)

• “Success”: $f = f_n$ strongly separates $\mathbb{P}$ and $\mathbb{Q}$ if

\[ \sqrt{\max\{\text{Var}_\mathbb{P}(f), \text{Var}_\mathbb{Q}(f)\}} = o(|E_\mathbb{P}[f] - E_\mathbb{Q}[f]|) \]

as $n \rightarrow \infty$
Consideration #1: Runtime

• Heuristic:
  • $\deg D \approx \text{time } n^{\Theta(D)}$
  • $\deg O(1) < \text{poly time } < \deg O(\log n)$
  • $\deg n^c \approx \text{time } \exp(\tilde{\Theta}(n^c))$

• “Low-degree conjecture” [Hopkins ‘18]
  • If low-degree polynomials fail, so do algorithms of the corresponding runtime
  • Not true for all distributions $\mathbb{P}, \mathbb{Q}…$
  • Can’t hope to prove it…
  • Or think of low-degree lower bounds as ruling out restricted algorithms

• Question: can we differentiate fine-grained time complexities such as $O(n)$ versus $O(n^2)$?
Consideration #2: Testing Error

• Strong separation ⇒ strong detection
  \[ \sqrt{\max\{\text{Var}_P(f), \text{Var}_Q(f)\}} = o \left( |E_P[f] - E_Q[f]| \right) \Rightarrow \text{type I + type II} = o(1) \]

• Weak separation ⇒ weak detection
  \[ \sqrt{\max\{\text{Var}_P(f), \text{Var}_Q(f)\}} = O \left( |E_P[f] - E_Q[f]| \right) \Rightarrow \text{type I + type II} \leq 1 - \epsilon \]

• Heuristic: if low-degree polynomials fail at strong (or weak) separation then efficient algorithms fail at strong (or weak) detection

• Question: can we identify the optimal tradeoff between type I and type II errors in a regime where weak (but not strong) detection is tractable?
Part 1: Fine-Grained Runtime

Based on joint work with Jay Mardia and Kabir Aladin Verchand

*arXiv, “Low-degree phase transitions for detecting a planted clique in sublinear time,” 2024*
Planted Clique in Sub-Linear Time

• Distinguish $\mathbb{Q}$: $G(n,1/2)$ versus $\mathbb{P}$: $G(n,1/2) + \{\text{random k-clique}\}$

• What runtime is required in the “easy” regime $k = \Theta(n^{1/2+\delta})$?

• Naïve methods (max degree / total edges) have “linear” runtime $\Theta(n^2)$

• “Subsampled” max degree has runtime $\Theta(n^{3(1/2-\delta)})$ [MAC’20]
  • Is this optimal?

• How to approach this?
  • Polynomial degree doesn’t capture fine-grained runtime: even naïve method (count total edges) is a degree-1 polynomial
  • The bottleneck seems to be reading the input...
Non-Adaptive Edge Query Model

[Feige, Gamarnik, Neeman, Rácz, Tetali ‘20; Rácz, Schiffer ‘20]

• Restricted class of algorithms
  • First choose a subset of edges ("mask") $M \subseteq \binom{[n]}{2}$ to observe (hard-coded)
  • Then perform a computation to decide $\mathbb{P}_M$ vs $\mathbb{Q}_M$
    • Runtime may be bounded or unbounded
    • Or require a low-degree test, as a proxy for bounded runtime

• Main result: low-degree tests require a mask of size $|M| \approx n^{3(1/2-\delta)}$

• Theorem: Let $k = \Theta(n^{1/2+\delta})$ for a constant $\delta \in (0,1)$
  • (Easy) If $\gamma > 3(1/2 - \delta)$ there exists $|M| = O(n^{\gamma})$ and a degree-$O(\log n)$ polynomial that strongly separates $\mathbb{P}_M$ and $\mathbb{Q}_M$
  • (Hard) If $\gamma < 3(1/2 - \delta)$ then for every $|M| = O(n^{\gamma})$, every degree-$o(\log^2 n)$ polynomial fails to weakly separate $\mathbb{P}_M$ and $\mathbb{Q}_M$
Non-Adaptive Edge Query Model: Phase Diagram

- **IT threshold:** query a complete subgraph, brute-force search for clique; adaptivity doesn’t help [Rácz, Schiffer ‘20]
- **Our result:** non-adaptive low-degree algorithms cannot improve the best known sub-linear runtime $O(n^{3(1/2-\delta)})$
  - **Open:** does adaptivity help?

![Phase Diagram](image)

- $k = n^{1/2+\delta}$
- $\delta = \frac{1}{2}$
- $\delta = 0$
- $\delta = -\frac{1}{2}$
- $\gamma = 0$, $\gamma = 1$, $\gamma = \frac{3}{2}$, $\gamma = 2$
- $|M| = n^\gamma$

[Easy] [Mardia, Asi, Chandrasekher ‘20]
[Low-degree hard] (our result)
[Impossible] [Rácz, Schiffer ‘20]
Proof Overview

• Standard tool: low-degree likelihood ratio

\[ \|L_M^{\leq D}\| := \sup_{f \text{ deg } D} \frac{\mathbb{E}_{Y \sim \mathbb{P}_M}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}_M}[f(Y)^2]}} \]

• Goal: show \( \|L_M^{\leq D}\| = 1 + o(1) \), implying that no deg-D polynomial weakly separates \( \mathbb{P}_M \) and \( \mathbb{Q}_M \)

• Convenient upper bound

\[ \|L_M^{\leq D}\|^2 \leq 1 + \sum_{d=1}^{D} \frac{1}{d!} \mathbb{E}_{X,X'}[\langle X, X' \rangle^d] \]

where \( X \in \{0,1\}^M \) is the indicator for mask edges with both endpoints in the clique, and \( X' \) is an independent copy (for a different clique)

• Difficulty: need to bound this for every possible \( M \) of a given size
Conditioning

• Issue: \( \| L^{\leq D}_M \| = \omega(1) \) for some choices of \( M \), e.g. the “star”
  \( M = \{(1,2), (1,3), (1,4), \ldots, (1,n)\} \)
  • In this case, \( \| L^{\leq D}_M \| \) is dominated by the unlikely event (under \( \mathbb{P}_M \)) that vertex 1 is in the clique

• Need to condition on high-probability “good” event: no vertex of “high” \( M \)-degree is in the clique
  • Why high prob: due to bound on \( |M| \), few vertices have “high” \( M \)-degree

• Formally: conditional low-degree calculation with modified \( \widetilde{\mathbb{P}}_M \)
  • Effectively lets us assume \( M \) has no “high” degree vertices
Key Idea

- Recall goal: bound $\text{LDUB}(M) = 1 + \sum_{d=1}^{D} \frac{1}{d!} E_{X,X'}[\langle X, X' \rangle^d]$ for all $M$
  - Assuming $M$ has no “high” degree vertices
- “Donation” operation simplifies $M$ and only increases $\text{LDUB}(M)$
  - By repeated application, can reduce the total number of vertices in the mask; now straightforward to bound $\text{LDUB}(M)$
Summary (Part 1: Fine-Grained Runtime)

• Main result: non-adaptive $O(\log n)$-degree tests require $\approx n^{3(1/2-\delta)}$ edge queries to detect a clique of size $k = \Theta(n^{1/2+\delta})$

• 2 ways to motivate this model:
  • Barrier to improving the best known sub-linear runtime $O(n^{3(1/2-\delta)})$ for planted clique in the “easy” regime
  • Non-adaptive queries model a scenario where we must decide in advance what data to collect

• Open: can non-adaptive algorithms achieve a better runtime?
  • How to formulate this as a low-degree question?
Part 2: Fine-Grained Error Probability

Based on joint work with Ankur Moitra

arXiv, “Precise Error Rates for Computationally Efficient Testing,” 2023
Spiked Wigner Model

- Testing problem over $n \times n$ symmetric matrices
  - $\mathbb{Q}$: $Y = W$ \quad $W_{ij} = W_{ji} \sim N(0, 1/n)$, $W_{ii} \sim N(0, 2/n)$
  - $\mathbb{P}$: $Y = \frac{\lambda}{n} xx^\top + W$ \quad $x \in \mathbb{R}^n$ i.i.d. from $\pi$, mean 0, variance 1
  - $n \to \infty$ with $\pi$ and $\lambda > 0$ fixed

- **Weak detection** (beat random guess) is always easy: $\text{Tr}(Y)$

- Phase diagram for **strong detection** (type I + type II $\to$ 0)
  [Baik, Ben Arous, Péché ‘05; El Alaoui, Krzakala, Jordan ‘20; Kunisky, W, Bandeira ‘19]

\[
\begin{array}{c}
0 & \lambda^*(\pi) & 1 \\
\text{impossible} & \text{hard} & \text{easy}
\end{array}
\]

- “BBP” eigenvalue transition

- Goal: optimal poly-time weak detection in “hard” regime?
Linear Spectral Statistics (LSS)

- Best known poly-time algorithm for weak detection when $\lambda < 1$
- Threshold $\sum_i f_\lambda(\mu_i)$ where $\mu_i$ are the eigenvalues of $Y$, for some $f_\lambda$
- Achieves a particular ROC (receiver operating characteristic) curve $\phi_\lambda$
  [Chung, Lee ‘22]

$$\phi_\lambda(\alpha) = 1 - \Phi[\Phi^{-1}(1 - \alpha) - \sqrt{-\log(1 - \lambda^2)/2}]$$
$\Phi$ = standard normal CDF

- IT optimal when $\lambda < \lambda^* (\pi)$
  [El Alaoui, Krzakala, Jordan ‘20]
- Poly-time optimal when $\lambda^* (\pi) < \lambda < 1$??

$\alpha$ = type I error
type I error = false positive rate

$\beta$ = power
type I error = true positive rate

$\phi_{0.9}$, $\phi_{0.7}$

perfect

concave

trivial

$\Phi = $ standard normal CDF
Strengthening of Low-Degree Conjecture

• For any $\lambda < 1$, any* $\pi$, and any $D = D_n = o(n/\log n)$,

\[
\|L^{\leq D}\| := \sup_{f \text{ deg } D} \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]} \to (1 - \lambda^2)^{-1/4} \text{ as } n \to \infty
\]

• This is $O(1)$, implying no strong separation by degree-D polynomials

• “Standard” LD conjecture: strong detection requires exponential runtime $\exp(n^{1-o(1)})$

• **Conjecture** (strong LD conjecture): for spiked Wigner, any $f = f_n$ with

\[
\limsup_{n \to \infty} \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}} > (1 - \lambda^2)^{-1/4}
\]

requires runtime $\exp(n^{1-o(1)})$
Main Result

• Assuming strong LD conjecture, LSS has optimal ROC curve among efficient algorithms

• Theorem:
  • Fix $\lambda \in (0, 1)$
  • Fix any* spike prior $\pi$
  • Assume the strong LD conjecture
  • Suppose $\beta > \phi_\lambda(\alpha)$, i.e., $(\alpha, \beta)$ lies above the ROC curve of LSS
  • Then any test with (type I, power) $\to (\alpha, \beta)$ requires runtime $\exp(n^{1-o(1)})$
Proof Idea

- Given achievable (concave) ROC curve $\phi$, can construct $f$ with
  \[
  \text{ratio}(f) = \text{val}(\phi)
  \]

- \[
  \text{ratio}(f) := \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}}
  \]

- \[
  \text{val}(\phi) := \sqrt{\int_0^1 (\phi'(\alpha))^2 d\alpha}
  \]

- Better curve has better val($\phi$)

\[
\text{val}(\phi') > \text{val}(\phi)
\]
Proof Idea

- Recall: \( \text{ratio}(f) \leq (1 - \lambda^2)^{-1/4} \)
  - For low-degree \( f \), and conjecturally for all efficiently-computable \( f \)
- Given this, what ROC curves are possible?
  - Must have \( \text{val}(\phi) \leq (1 - \lambda^2)^{-1/4} \)
  - Many possibilities...

\[
\text{ratio}(f) := \frac{E_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{E_{Y \sim \mathbb{Q}}[f(Y)^2]}}
\]
\[
\text{val}(\phi) := \sqrt{\int_0^1 (\phi'(\alpha))^2 \, d\alpha}
\]
Proof

- We know $\phi_\lambda$ is achievable in poly time [Chung, Lee ‘22], yielding ratio
  \[ \text{val}(\phi_\lambda) = (1 - \lambda^2)^{-1/4} \]

- Assume for contradiction: some $(\alpha^*, \beta^*)$ above $\phi_\lambda$ is achievable

- Can then achieve an even better ROC curve $\psi$

- Thus achieving ratio
  \[ \text{val}(\psi) > \text{val}(\phi_\lambda) = (1 - \lambda^2)^{-1/4} \]

- Contradicts strong LD conjecture

- Conclude: $(\alpha^*, \beta^*)$ not achievable (in sub-exponential time)
Summary (Part 2: Fine-Grained Error Probability)

• Spiked Wigner model with $\lambda^*(\pi) < \lambda < 1$: strong detection possible—but-hard
• Weak detection is always easy, but what is the optimal ROC curve?
• Assuming “strong low-degree conjecture,” linear spectral statistics (LSS) has the best ROC curve among all poly-time (even sub-exponential time) algorithms
• Consequence (“computational universality“): while IT threshold $\lambda^*(\pi)$ depends on prior $\pi$, the best computationally-efficient test only uses the spectrum and is thus oblivious to the prior
• Akin to optimal low-degree estimation error when $\lambda > 1$ [Montanari, W ’22]
• Open: more “direct” analysis of low-degree tests?

Thanks!