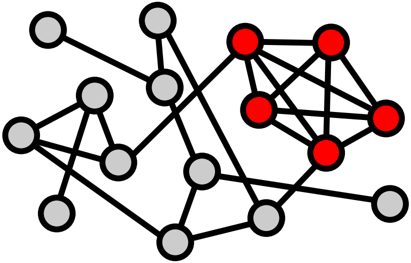
Introduction to the Low-Degree Polynomial Method

Alex Wein
Courant Institute, New York University

Part I: Why Low-Degree Polynomials?

Problems in High-Dimensional Statistics

Example: finding a large clique in a random graph



Problems in High-Dimensional Statistics

Example: finding a large clique in a random graph

- **Detection**: distinguish between a random graph and a graph with a planted clique

Problems in High-Dimensional Statistics

Example: finding a large clique in a random graph

- **Detection**: distinguish between a random graph and a graph with a planted clique
- **Recovery**: given a graph with a planted clique, find the clique

Problems in High-Dimensional Statistics

Example: finding a large clique in a random graph

- **Detection**: distinguish between a random graph and a graph with a planted clique
- **Recovery**: given a graph with a planted clique, find the clique
- **Optimization**: given a random graph (with no planted clique), find as large a clique as possible

Problems in High-Dimensional Statistics

Example: finding a large clique in a random graph

- **Detection**: distinguish between a random graph and a graph with a planted clique
- **Recovery**: given a graph with a planted clique, find the clique
- **Optimization**: given a random graph (with no planted clique), find as large a clique as possible

Common to have information-computation gaps

Problems in High-Dimensional Statistics

Example: finding a large clique in a random graph

- **Detection**: distinguish between a random graph and a graph with a planted clique
- **Recovery**: given a graph with a planted clique, find the clique
- **Optimization**: given a random graph (with no planted clique), find as large a clique as possible

Common to have information-computation gaps

E.g. planted k-clique (either detection or recovery)

Problems in High-Dimensional Statistics

Example: finding a large clique in a random graph

- **Detection**: distinguish between a random graph and a graph with a planted clique
- **Recovery**: given a graph with a planted clique, find the clique
- **Optimization**: given a random graph (with no planted clique), find as large a clique as possible

Common to have information-computation gaps

E.g. planted *k*-clique (either detection or recovery)

What makes problems easy vs hard?

The Low-Degree Polynomial Method

A framework for predicting/explaining average-case computational complexity

The Low-Degree Polynomial Method

A framework for predicting/explaining average-case computational complexity

Originated from sum-of-squares literature (for detection)

[Barak, Hopkins, Kelner, Kothari, Moitra, Potechin '16]

[Hopkins, Steurer '17]

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

[Hopkins '18 (PhD thesis)]

The Low-Degree Polynomial Method

A framework for predicting/explaining average-case computational complexity

Originated from sum-of-squares literature (for detection)

[Barak, Hopkins, Kelner, Kothari, Moitra, Potechin '16]

[Hopkins, Steurer '17]

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

[Hopkins '18 (PhD thesis)]

Today: self-contained motivation (without SoS)

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^N \to \mathbb{R}^M$

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^N \to \mathbb{R}^M$
  - Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^N \to \mathbb{R}^M$
  - Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
  - Output: e.g. $b \in \{0,1\}$

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^N \to \mathbb{R}^M$
  - Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
  - Output: e.g. $b \in \{0,1\}$ or $v \in \mathbb{R}^n$

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^N \to \mathbb{R}^M$
  - Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
  - Output: e.g. $b \in \{0,1\}$ or $v \in \mathbb{R}^n$
- "Low" means $O(\log n)$ where n is dimension

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^N \to \mathbb{R}^M$
  - Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
  - Output: e.g. $b \in \{0,1\}$ or $v \in \mathbb{R}^n$
- "Low" means $O(\log n)$ where n is dimension

Examples of low-degree algorithms:

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^N \to \mathbb{R}^M$
  - Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
  - Output: e.g. $b \in \{0,1\}$ or $v \in \mathbb{R}^n$
- "Low" means $O(\log n)$ where n is dimension

Examples of low-degree algorithms: input $Y \in \mathbb{R}^{n \times n}$

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^N \to \mathbb{R}^M$
  - Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
  - Output: e.g. $b \in \{0,1\}$ or $v \in \mathbb{R}^n$
- "Low" means $O(\log n)$ where n is dimension

Examples of low-degree algorithms: input $Y \in \mathbb{R}^{n \times n}$

- Power iteration: $Y^k \mathbf{1}$ or $\mathrm{Tr}(Y^k)$ $k = O(\log n)$

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^N \to \mathbb{R}^M$
  - Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
  - Output: e.g. $b \in \{0,1\}$ or $v \in \mathbb{R}^n$
- "Low" means $O(\log n)$ where n is dimension

Examples of low-degree algorithms: input $Y \in \mathbb{R}^{n \times n}$

- Power iteration: $Y^k \mathbf{1}$ or $\mathrm{Tr}(Y^k)$ $k = O(\log n)$
- Approximate message passing: $v \leftarrow Y h(v)$ $O(1)$ rounds

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^N \to \mathbb{R}^M$
  - Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
  - Output: e.g. $b \in \{0,1\}$ or $v \in \mathbb{R}^n$
- "Low" means $O(\log n)$ where n is dimension

Examples of low-degree algorithms: input $Y \in \mathbb{R}^{n \times n}$

- Power iteration: $Y^k \mathbf{1}$ or $\mathrm{Tr}(Y^k)$ $k = O(\log n)$
- Approximate message passing: $v \leftarrow Y h(v)$ $O(1)$ rounds
- Local algorithms on sparse graphs radius $O(1)$

The Low-Degree Polynomial Method

Study a restricted class of algorithms: low-degree polynomials

- Multivariate polynomial $f : \mathbb{R}^N \to \mathbb{R}^M$
  - Input: e.g. graph $Y \in \{0,1\}^{\binom{n}{2}}$
  - Output: e.g. $b \in \{0,1\}$ or $v \in \mathbb{R}^n$
- "Low" means $O(\log n)$ where n is dimension

Examples of low-degree algorithms: input $Y \in \mathbb{R}^{n \times n}$

- Power iteration: $Y^k \mathbf{1}$ or $\mathrm{Tr}(Y^k)$ — $k = O(\log n)$
- Approximate message passing: $v \leftarrow Y h(v)$ — $O(1)$ rounds
- Local algorithms on sparse graphs — radius $O(1)$
- Or any of the above applied to $\tilde{Y} = g(Y)$ — $\deg g = O(1)$

The Low-Degree Polynomial Method

Claim: low-degree polynomials provide a unified explanation of information-computation gaps in detection/recovery/optimization

The Low-Degree Polynomial Method

Claim: low-degree polynomials provide a unified explanation of information-computation gaps in detection/recovery/optimization

For all of these problems...

The Low-Degree Polynomial Method

Claim: low-degree polynomials provide a unified explanation of information-computation gaps in detection/recovery/optimization

For all of these problems...

planted clique, sparse PCA, community detection, tensor PCA, planted CSPs, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, p-spin optimization, max independent set

The Low-Degree Polynomial Method

Claim: low-degree polynomials provide a unified explanation of information-computation gaps in detection/recovery/optimization

For all of these problems...

planted clique, sparse PCA, community detection, tensor PCA, planted CSPs, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, p-spin optimization, max independent set

...it is the case that

The Low-Degree Polynomial Method

Claim: low-degree polynomials provide a unified explanation of information-computation gaps in detection/recovery/optimization

For all of these problems...

planted clique, sparse PCA, community detection, tensor PCA, planted CSPs, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, p-spin optimization, max independent set

...it is the case that

- the best known poly-time algorithms are low-degree (spectral/AMP/local)

The Low-Degree Polynomial Method

Claim: low-degree polynomials provide a unified explanation of information-computation gaps in detection/recovery/optimization

For all of these problems...

planted clique, sparse PCA, community detection, tensor PCA, planted CSPs, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, p-spin optimization, max independent set

...it is the case that

- the best known poly-time algorithms are low-degree (spectral/AMP/local)
- low-degree polynomials fail in the "hard" regime

The Low-Degree Polynomial Method

Claim: low-degree polynomials provide a unified explanation of information-computation gaps in detection/recovery/optimization

For all of these problems...

planted clique, sparse PCA, community detection, tensor PCA, planted CSPs, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, p-spin optimization, max independent set

...it is the case that

- the best known poly-time algorithms are low-degree (spectral/AMP/local)
- low-degree polynomials fail in the "hard" regime

"Low-degree conjecture" (informal): low-degree polynomials are as powerful as all poly-time algorithms for "natural" high-dimensional problems [Hopkins '18]

Overview

This talk: techniques to prove that **all** low-degree polynomials fail

Overview

This talk: techniques to prove that **all** low-degree polynomials fail

- Gives evidence for computational hardness

Overview

This talk: techniques to prove that **all** low-degree polynomials fail

- Gives evidence for computational hardness

Settings:

- Detection

  [Hopkins, Steurer '17]

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

  [Hopkins '18] (PhD thesis)

  [Kunisky, W., Bandeira '19] (survey)

- Recovery

  [Schramm, W. '20]

- Optimization

  [Gamarnik, Jagannath, W. '20]

Part II: Detection

Detection (e.g. [Hopkins, Steurer '17])

Goal: hypothesis test with error probability $o(1)$ between:

- Null model $Y \sim \mathbb{Q}_n$ e.g. $G(n, 1/2)$
- Planted model $Y \sim \mathbb{P}_n$ e.g. $G(n, 1/2) \cup \{\text{random } k\text{-clique}\}$

Detection (e.g. [Hopkins, Steurer '17])

Goal: hypothesis test with error probability $o(1)$ between:

- Null model $Y \sim \mathbb{Q}_n$ e.g. $G(n, 1/2)$
- Planted model $Y \sim \mathbb{P}_n$ e.g. $G(n, 1/2) \cup \{\text{random } k\text{-clique}\}$

Look for a degree-D polynomial $f : \mathbb{R}^{n \times n} \to \mathbb{R}$ that distinguishes $\mathbb{P}$ from $\mathbb{Q}$

Detection (e.g. [Hopkins, Steurer '17])

Goal: hypothesis test with error probability $o(1)$ between:

- Null model $Y \sim \mathbb{Q}_n$ e.g. $G(n, 1/2)$
- Planted model $Y \sim \mathbb{P}_n$ e.g. $G(n, 1/2) \cup \{\text{random } k\text{-clique}\}$

Look for a degree-D polynomial $f : \mathbb{R}^{n \times n} \to \mathbb{R}$ that distinguishes $\mathbb{P}$ from $\mathbb{Q}$

- $f(Y)$ is "big" when $Y \sim \mathbb{P}$ and "small" when $Y \sim \mathbb{Q}$

Detection (e.g. [Hopkins, Steurer '17])

Goal: hypothesis test with error probability $o(1)$ between:

- Null model $Y \sim \mathbb{Q}_n$ e.g. $G(n, 1/2)$
- Planted model $Y \sim \mathbb{P}_n$ e.g. $G(n, 1/2) \cup \{\text{random } k\text{-clique}\}$

Look for a degree-D polynomial $f : \mathbb{R}^{n \times n} \to \mathbb{R}$ that distinguishes $\mathbb{P}$ from $\mathbb{Q}$

- $f(Y)$ is "big" when $Y \sim \mathbb{P}$ and "small" when $Y \sim \mathbb{Q}$

Compute "advantage":

$$\mathrm{Adv}_{\leq D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}} \qquad \frac{\text{mean in } \mathbb{P}}{\text{fluctuations in } \mathbb{Q}}$$

Detection (e.g. [Hopkins, Steurer '17])

Goal: hypothesis test with error probability $o(1)$ between:

- Null model $Y \sim \mathbb{Q}_n$ e.g. $G(n, 1/2)$
- Planted model $Y \sim \mathbb{P}_n$ e.g. $G(n, 1/2) \cup \{\text{random } k\text{-clique}\}$

Look for a degree-D polynomial $f : \mathbb{R}^{n \times n} \to \mathbb{R}$ that distinguishes $\mathbb{P}$ from $\mathbb{Q}$

- $f(Y)$ is "big" when $Y \sim \mathbb{P}$ and "small" when $Y \sim \mathbb{Q}$

Compute "advantage":

$$\text{Adv}_{\leq D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}} \qquad \frac{\text{mean in } \mathbb{P}}{\text{fluctuations in } \mathbb{Q}}$$

$$= \begin{cases} \omega(1) & \text{"degree-}D\text{ polynomial succeed"} \\ O(1) & \text{"degree-}D\text{ polynomials fail"} \end{cases}$$

Detection (e.g. [Hopkins, Steurer '17])

Prototypical result (planted clique):

Detection (e.g. [Hopkins, Steurer '17])

Prototypical result (planted clique):

Theorem [BHKKMP16,Hop18]: For a planted k-clique in $G(n, 1/2)$,

Detection (e.g. [Hopkins, Steurer '17])

Prototypical result (planted clique):

Theorem [BHKKMP16,Hop18]: For a planted k-clique in $G(n, 1/2)$,

- if $k = \Omega(\sqrt{n})$ then $\text{Adv}_{\leq D} = \omega(1)$ for some $D = O(\log n)$
  low-degree polynomials succeed when $k \gtrsim \sqrt{n}$

Detection (e.g. [Hopkins, Steurer '17])

Prototypical result (planted clique):

Theorem [BHKKMP16,Hop18]: For a planted k-clique in $G(n, 1/2)$,

- if $k = \Omega(\sqrt{n})$ then $\mathrm{Adv}_{\le D} = \omega(1)$ for some $D = O(\log n)$
  low-degree polynomials succeed when $k \gtrsim \sqrt{n}$
- if $k = O(n^{1/2-\epsilon})$ then $\mathrm{Adv}_{\le D} = O(1)$ for any $D = O(\log n)$
  low-degree polynomials fail when $k \ll \sqrt{n}$

Detection (e.g. [Hopkins, Steurer '17])

Prototypical result (planted clique):

Theorem [BHKKMP16,Hop18]: For a planted k-clique in $G(n, 1/2)$,

- if $k = \Omega(\sqrt{n})$ then $\mathrm{Adv}_{\leq D} = \omega(1)$ for some $D = O(\log n)$
  low-degree polynomials succeed when $k \gtrsim \sqrt{n}$
- if $k = O(n^{1/2-\epsilon})$ then $\mathrm{Adv}_{\leq D} = O(1)$ for any $D = O(\log n)$
  low-degree polynomials fail when $k \ll \sqrt{n}$

Sometimes can rule out polynomials of degree $D = n^{\delta}$

Detection (e.g. [Hopkins, Steurer '17])

Prototypical result (planted clique):

Theorem [BHKKMP16,Hop18]: For a planted k-clique in $G(n, 1/2)$,

- if $k = \Omega(\sqrt{n})$ then $\text{Adv}_{\leq D} = \omega(1)$ for some $D = O(\log n)$
  low-degree polynomials succeed when $k \gtrsim \sqrt{n}$
- if $k = O(n^{1/2-\epsilon})$ then $\text{Adv}_{\leq D} = O(1)$ for any $D = O(\log n)$
  low-degree polynomials fail when $k \ll \sqrt{n}$

Sometimes can rule out polynomials of degree $D = n^{\delta}$

Extended low-degree conjecture [Hopkins '18]:

$$\text{degree-}D \text{ polynomials} \Leftrightarrow n^{\tilde{\Theta}(D)}\text{-time algorithms}$$

$$D = n^{\delta} \quad \Leftrightarrow \quad \exp(n^{\delta \pm o(1)}) \text{ time}$$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\leq D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}}$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathsf{Adv}_{\leq D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\le D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S|\le D} \hat{f}_S Y^S$ $\qquad Y^S := \prod_{i\in S} Y_i \qquad S \subseteq [m]$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\leq D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S|\leq D} \hat{f}_S Y^S$ $\quad Y^S := \prod_{i\in S} Y_i \quad S \subseteq [m]$

$\{Y^S\}_{S\subseteq[m]}$ are orthonormal: $\mathbb{E}_{Y\sim\mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\le D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S|\le D} \hat{f}_S Y^S$ $\qquad Y^S := \prod_{i\in S} Y_i \qquad S \subseteq [m]$

$\{Y^S\}_{S\subseteq[m]}$ are orthonormal: $\mathbb{E}_{Y\sim\mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

<u>Numerator</u>: $\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\leq D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S|\leq D} \hat{f}_S Y^S$ $\qquad Y^S := \prod_{i\in S} Y_i \qquad S \subseteq [m]$

$\{Y^S\}_{S\subseteq[m]}$ are orthonormal: $\mathbb{E}_{Y\sim\mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

<u>Numerator</u>: $\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[f(Y)] = \sum_{|S|\leq D} \hat{f}_S \mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[Y^S]$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\leq D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S|\leq D} \hat{f}_S Y^S$ $\quad Y^S := \prod_{i\in S} Y_i \quad S \subseteq [m]$

$\{Y^S\}_{S\subseteq[m]}$ are orthonormal: $\mathbb{E}_{Y\sim\mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

Numerator: $\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[f(Y)] = \sum_{|S|\leq D} \hat{f}_S \mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[Y^S] =: \langle \hat{f}, c\rangle$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\le D} := \max_{f \text{ deg } D} \dfrac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S|\le D} \hat{f}_S Y^S$ $\qquad Y^S := \prod_{i\in S} Y_i \qquad S \subseteq [m]$

$\{Y^S\}_{S\subseteq[m]}$ are orthonormal: $\mathbb{E}_{Y\sim\mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

Numerator: $\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[f(Y)] = \sum_{|S|\le D} \hat{f}_S \mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[Y^S] =: \langle \hat{f}, c\rangle$

Denominator: $\mathop{\mathbb{E}}_{Y\sim\mathbb{Q}}[f(Y)^2]$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\leq D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S| \leq D} \hat{f}_S Y^S$ $\qquad Y^S := \prod_{i \in S} Y_i \qquad S \subseteq [m]$

$\{Y^S\}_{S \subseteq [m]}$ are orthonormal: $\mathbb{E}_{Y \sim \mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

Numerator: $\mathop{\mathbb{E}}_{Y \sim \mathbb{P}}[f(Y)] = \sum_{|S| \leq D} \hat{f}_S \mathop{\mathbb{E}}_{Y \sim \mathbb{P}}[Y^S] =: \langle \hat{f}, c \rangle$

Denominator: $\mathop{\mathbb{E}}_{Y \sim \mathbb{Q}}[f(Y)^2] = \sum_{|S| \leq D} \hat{f}_S^2$ (orthonormality)

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\le D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S|\le D} \hat{f}_S Y^S$ $\qquad Y^S := \prod_{i\in S} Y_i \qquad S \subseteq [m]$

$\{Y^S\}_{S\subseteq[m]}$ are orthonormal: $\mathbb{E}_{Y\sim\mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

Numerator: $\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[f(Y)] = \sum_{|S|\le D} \hat{f}_S \mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[Y^S] =: \langle \hat{f}, c\rangle$

Denominator: $\mathop{\mathbb{E}}_{Y\sim\mathbb{Q}}[f(Y)^2] = \sum_{|S|\le D} \hat{f}_S^2 = \|\hat{f}\|^2$ (orthonormality)

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\le D} := \max_{f \text{ deg } D} \dfrac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S|\le D} \hat{f}_S Y^S$ $\qquad Y^S := \prod_{i\in S} Y_i \qquad S \subseteq [m]$

$\{Y^S\}_{S\subseteq[m]}$ are orthonormal: $\mathbb{E}_{Y\sim\mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

Numerator: $\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)] = \sum_{|S|\le D} \hat{f}_S \mathbb{E}_{Y\sim\mathbb{P}}[Y^S] =: \langle \hat{f}, c\rangle$

Denominator: $\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2] = \sum_{|S|\le D} \hat{f}_S^2 = \|\hat{f}\|^2$ (orthonormality)

$$\mathrm{Adv}_{\le D} = \max_{\hat{f}} \frac{\langle \hat{f}, c\rangle}{\|\hat{f}\|}$$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\leq D} := \max_{f \text{ deg } D} \dfrac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S| \leq D} \hat{f}_S Y^S$ $\qquad Y^S := \prod_{i \in S} Y_i \qquad S \subseteq [m]$

$\{Y^S\}_{S \subseteq [m]}$ are orthonormal: $\mathbb{E}_{Y \sim \mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

Numerator: $\mathop{\mathbb{E}}_{Y \sim \mathbb{P}}[f(Y)] = \sum_{|S| \leq D} \hat{f}_S \mathop{\mathbb{E}}_{Y \sim \mathbb{P}}[Y^S] =: \langle \hat{f}, c \rangle$

Denominator: $\mathop{\mathbb{E}}_{Y \sim \mathbb{Q}}[f(Y)^2] = \sum_{|S| \leq D} \hat{f}_S^2 = \|\hat{f}\|^2$ $\qquad$ (orthonormality)

$$\mathrm{Adv}_{\leq D} = \max_{\hat{f}} \frac{\langle \hat{f}, c \rangle}{\|\hat{f}\|}$$

Optimizer: $\hat{f}^* = c$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\leq D} := \max_{f \text{ deg } D} \dfrac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S|\leq D} \hat{f}_S Y^S$ $\qquad Y^S := \prod_{i\in S} Y_i \qquad S \subseteq [m]$

$\{Y^S\}_{S\subseteq[m]}$ are orthonormal: $\mathbb{E}_{Y\sim\mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

Numerator: $\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[f(Y)] = \sum_{|S|\leq D} \hat{f}_S \mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[Y^S] =: \langle \hat{f}, c\rangle$

Denominator: $\mathop{\mathbb{E}}_{Y\sim\mathbb{Q}}[f(Y)^2] = \sum_{|S|\leq D} \hat{f}_S^2 = \|\hat{f}\|^2$ $\qquad$ (orthonormality)

$$\mathrm{Adv}_{\leq D} = \max_{\hat{f}} \frac{\langle \hat{f}, c\rangle}{\|\hat{f}\|} = \frac{\langle c, c\rangle}{\|c\|}$$

Optimizer: $\hat{f}^* = c$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\le D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S|\le D} \hat{f}_S Y^S$ $\qquad Y^S := \prod_{i\in S} Y_i \qquad S \subseteq [m]$

$\{Y^S\}_{S\subseteq[m]}$ are orthonormal: $\mathbb{E}_{Y\sim\mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

Numerator: $\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)] = \sum_{|S|\le D} \hat{f}_S \mathbb{E}_{Y\sim\mathbb{P}}[Y^S] =: \langle \hat{f}, c\rangle$

Denominator: $\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2] = \sum_{|S|\le D} \hat{f}_S^2 = \|\hat{f}\|^2$ $\qquad$ (orthonormality)

$$\mathrm{Adv}_{\le D} = \max_{\hat{f}} \frac{\langle \hat{f}, c\rangle}{\|\hat{f}\|} = \frac{\langle c, c\rangle}{\|c\|} = \|c\|$$

Optimizer: $\hat{f}^* = c$

Detection (e.g. [Hopkins, Steurer '17])

Goal: compute $\mathrm{Adv}_{\leq D} := \max_{f \text{ deg } D} \dfrac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}}$

Suppose $\mathbb{Q}$ is i.i.d. $\mathrm{Unif}(\pm 1)$

Write $f(Y) = \sum_{|S|\leq D} \hat{f}_S Y^S$ $\qquad Y^S := \prod_{i\in S} Y_i \qquad S \subseteq [m]$

$\{Y^S\}_{S\subseteq[m]}$ are orthonormal: $\mathbb{E}_{Y\sim\mathbb{Q}}[Y^S Y^T] = \mathbb{1}_{S=T}$

Numerator: $\displaystyle \mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[f(Y)] = \sum_{|S|\leq D} \hat{f}_S \mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[Y^S] =: \langle \hat{f}, c\rangle$

Denominator: $\displaystyle \mathop{\mathbb{E}}_{Y\sim\mathbb{Q}}[f(Y)^2] = \sum_{|S|\leq D} \hat{f}_S^2 = \|\hat{f}\|^2$ $\qquad$ (orthonormality)

$$\mathrm{Adv}_{\leq D} = \max_{\hat{f}} \frac{\langle \hat{f}, c\rangle}{\|\hat{f}\|} = \frac{\langle c, c\rangle}{\|c\|} = \|c\| = \sqrt{\sum_{|S|\leq D} \left(\mathop{\mathbb{E}}_{Y\sim\mathbb{P}}[Y^S]\right)^2}$$

Optimizer: $\hat{f}^* = c$

Detection (e.g. [Hopkins, Steurer '17])

Remarks:

Detection (e.g. [Hopkins, Steurer '17])

Remarks:

- Best test is likelihood ratio (Neyman-Pearson lemma)

$$L(Y) = \frac{d\mathbb{P}}{d\mathbb{Q}}(Y)$$

Detection (e.g. [Hopkins, Steurer '17])

Remarks:

- Best test is likelihood ratio (Neyman-Pearson lemma)

$$L(Y) = \frac{d\mathbb{P}}{d\mathbb{Q}}(Y)$$

- Best degree-D test (maximizer of $\mathrm{Adv}_{\le D}$) is

$$f^* = L^{\le D} := \text{projection of } L \text{ onto deg-}D \text{ subspace}$$

Detection (e.g. [Hopkins, Steurer '17])

Remarks:

- Best test is likelihood ratio (Neyman-Pearson lemma)

$$L(Y) = \frac{d\mathbb{P}}{d\mathbb{Q}}(Y)$$

- Best degree-D test (maximizer of $\mathrm{Adv}_{\leq D}$) is

$$f^* = L^{\leq D} := \text{projection of } L \text{ onto deg-}D \text{ subspace}$$

orthogonal projection w.r.t. $\langle f, g \rangle := \mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)g(Y)]$

Detection (e.g. [Hopkins, Steurer '17])

Remarks:

- Best test is likelihood ratio (Neyman-Pearson lemma)

$$L(Y) = \frac{d\mathbb{P}}{d\mathbb{Q}}(Y)$$

- Best degree-D test (maximizer of $\mathrm{Adv}_{\leq D}$) is

$$f^* = L^{\leq D} := \text{projection of } L \text{ onto deg-}D \text{ subspace}$$

orthogonal projection w.r.t. $\langle f, g \rangle := \mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)g(Y)]$

"low-degree likelihood ratio"

Detection (e.g. [Hopkins, Steurer '17])

Remarks:

- Best test is likelihood ratio (Neyman-Pearson lemma)

$$L(Y) = \frac{d\mathbb{P}}{d\mathbb{Q}}(Y)$$

- Best degree-D test (maximizer of $\mathrm{Adv}_{\leq D}$) is

$$f^* = L^{\leq D} := \text{projection of } L \text{ onto deg-}D \text{ subspace}$$

orthogonal projection w.r.t. $\langle f, g \rangle := \mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)g(Y)]$

"low-degree likelihood ratio"

- $\mathrm{Adv}_{\leq D} = \|L^{\leq D}\|$ $\qquad \|f\| := \sqrt{\langle f, f \rangle} = \mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]$

Detection (e.g. [Hopkins, Steurer '17])

Remarks:

- Best test is likelihood ratio (Neyman-Pearson lemma)

$$L(Y) = \frac{d\mathbb{P}}{d\mathbb{Q}}(Y)$$

- Best degree-D test (maximizer of $\mathrm{Adv}_{\leq D}$) is

$$f^* = L^{\leq D} := \text{projection of } L \text{ onto deg-} D \text{ subspace}$$

orthogonal projection w.r.t. $\langle f, g \rangle := \mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)g(Y)]$

"low-degree likelihood ratio"

- $\mathrm{Adv}_{\leq D} = \|L^{\leq D}\|$ $\qquad \|f\| := \sqrt{\langle f, f \rangle} = \mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]$

"norm of low-degree likelihood ratio"

Detection (e.g. [Hopkins, Steurer '17])

Remarks:

- Best test is likelihood ratio (Neyman-Pearson lemma)

$$L(Y) = \frac{d\mathbb{P}}{d\mathbb{Q}}(Y)$$

- Best degree-D test (maximizer of $\mathrm{Adv}_{\leq D}$) is

$$f^* = L^{\leq D} := \text{projection of } L \text{ onto deg-}D \text{ subspace}$$

orthogonal projection w.r.t. $\langle f, g \rangle := \mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)g(Y)]$

"low-degree likelihood ratio"

- $\mathrm{Adv}_{\leq D} = \|L^{\leq D}\|$ $\quad \|f\| := \sqrt{\langle f, f \rangle} = \mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]$

"norm of low-degree likelihood ratio"

Proof: $\hat{L}_S = \mathbb{E}_{Y \sim \mathbb{Q}}[L(Y)Y^S] = \mathbb{E}_{Y \sim \mathbb{P}}[Y^S]$ $\qquad \hat{f}^*_S = \mathbb{E}_{Y \sim \mathbb{P}}[Y^S]\,\mathbb{1}_{|S| \leq D}$

Detection (e.g. [Hopkins, Steurer '17])

User-friendly results:

Detection (e.g. [Hopkins, Steurer '17])

User-friendly results:

- Additive Gaussian model:
  $\mathbb{P} : Y = X + Z \quad \text{vs} \quad \mathbb{Q} : Y = Z$

Detection (e.g. [Hopkins, Steurer '17])

User-friendly results:

- Additive Gaussian model:
  $\mathbb{P} : Y = X + Z \quad \text{vs} \quad \mathbb{Q} : Y = Z$

$$\mathrm{Adv}^2_{\leq D} = \sum_{d=0}^{D} \frac{1}{d!} \mathop{\mathbb{E}}_{X,X'} \langle X, X' \rangle^d$$

Detection (e.g. [Hopkins, Steurer '17])

User-friendly results:

- Additive Gaussian model:
  $\mathbb{P} : Y = X + Z \quad \text{vs} \quad \mathbb{Q} : Y = Z$

$$\mathrm{Adv}^2_{\leq D} = \sum_{d=0}^{D} \frac{1}{d!} \mathop{\mathbb{E}}_{X,X'} \langle X, X' \rangle^d$$

- Rademacher model $Y \in \{\pm 1\}^m$:
  $\mathbb{P} : \mathbb{E}[Y|X] = X \quad \text{vs} \quad \mathbb{Q} : \mathbb{E}[Y] = 0$

Detection (e.g. [Hopkins, Steurer '17])

User-friendly results:

- Additive Gaussian model:
  $\mathbb{P} : Y = X + Z \quad \text{vs} \quad \mathbb{Q} : Y = Z$

$$\mathrm{Adv}^2_{\leq D} = \sum_{d=0}^{D} \frac{1}{d!} \mathop{\mathbb{E}}_{X,X'} \langle X, X' \rangle^d$$

- Rademacher model $Y \in \{\pm 1\}^m$:
  $\mathbb{P} : \mathbb{E}[Y|X] = X \quad \text{vs} \quad \mathbb{Q} : \mathbb{E}[Y] = 0$

$$\mathrm{Adv}^2_{\leq D} \leq \sum_{d=0}^{D} \frac{1}{d!} \mathop{\mathbb{E}}_{X,X'} \langle X, X' \rangle^d$$

Detection (e.g. [Hopkins, Steurer '17])

Recap (detection):

Detection (e.g. [Hopkins, Steurer '17])

Recap (detection):

- Given $\mathbb{P}, \mathbb{Q}$, can compute (via linear algebra)

$$\mathsf{Adv}_{\leq D} = \|L^{\leq D}\| = \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}}$$

Detection (e.g. [Hopkins, Steurer '17])

Recap (detection):

- Given $\mathbb{P}, \mathbb{Q}$, can compute (via linear algebra)

$$\text{Adv}_{\leq D} = \|L^{\leq D}\| = \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}}$$

- Need to know orthogonal polynomials w.r.t. $\mathbb{Q}$
  - Possible when $\mathbb{Q}$ has independent coordinates

Detection (e.g. [Hopkins, Steurer '17])

Recap (detection):

- Given $\mathbb{P}, \mathbb{Q}$, can compute (via linear algebra)

$$\mathrm{Adv}_{\leq D} = \|L^{\leq D}\| = \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}}$$

- Need to know orthogonal polynomials w.r.t. $\mathbb{Q}$
  - Possible when $\mathbb{Q}$ has independent coordinates
- To predict computational complexity: for $D \approx \log n$,

$$\mathrm{Adv}_{\leq D} = \begin{cases} \omega(1) & \Rightarrow \text{"easy"} \\ O(1) & \Rightarrow \text{"hard"} \end{cases}$$

Detection (e.g. [Hopkins, Steurer '17])

Recap (detection):

- Given $\mathbb{P}, \mathbb{Q}$, can compute (via linear algebra)

$$\mathrm{Adv}_{\leq D} = \|L^{\leq D}\| = \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}}$$

- Need to know orthogonal polynomials w.r.t. $\mathbb{Q}$
  - Possible when $\mathbb{Q}$ has independent coordinates

- To predict computational complexity: for $D \approx \log n$,

$$\mathrm{Adv}_{\leq D} = \begin{cases} \omega(1) & \Rightarrow \text{ "easy"} \\ O(1) & \Rightarrow \text{ "hard"} \end{cases}$$

- These predictions are "correct" for: planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, ... [BHKKMP16,HS17,HKPRSS17,Hop18,BK**W**19,K**W**B19,DK**W**B19]

Part III: Recovery

Recovery [Schramm, W. '20]

Example (planted submatrix): observe $n \times n$ matrix $Y = X + Z$

- Signal: $X = \lambda v v^\top \quad \lambda > 0 \quad v_i \sim \mathrm{Bernoulli}(\rho)$
- Noise: Z i.i.d. $\mathcal{N}(0, 1)$

Recovery [Schramm, W. '20]

Example (planted submatrix): observe $n \times n$ matrix $Y = X + Z$

- Signal: $X = \lambda v v^\top \quad \lambda > 0 \quad v_i \sim \mathrm{Bernoulli}(\rho)$
- Noise: Z i.i.d. $\mathcal{N}(0,1)$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.
Recovery: given $Y \sim \mathbb{P}$, recover v

Recovery [Schramm, W. '20]

Example (planted submatrix): observe $n \times n$ matrix $Y = X + Z$

- Signal: $X = \lambda v v^\top$ $\quad \lambda > 0 \quad$ $v_i \sim \mathrm{Bernoulli}(\rho)$
- Noise: Z i.i.d. $\mathcal{N}(0,1)$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.
Recovery: given $Y \sim \mathbb{P}$, recover v

If you can recover then you can detect (poly-time reduction)

Recovery [Schramm, W. '20]

Example (planted submatrix): observe $n \times n$ matrix $Y = X + Z$

- Signal: $X = \lambda v v^\top \quad \lambda > 0 \quad v_i \sim \mathrm{Bernoulli}(\rho)$
- Noise: Z i.i.d. $\mathcal{N}(0, 1)$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.
Recovery: given $Y \sim \mathbb{P}$, recover v

If you can recover then you can detect (poly-time reduction)

- How: run recovery algorithm to get $\hat{v} \in \{0, 1\}^n$; check $\hat{v}^\top Y \hat{v}$

Recovery [Schramm, W. '20]

Example (planted submatrix): observe $n \times n$ matrix $Y = X + Z$

- Signal: $X = \lambda v v^\top \quad \lambda > 0 \quad v_i \sim \mathrm{Bernoulli}(\rho)$
- Noise: Z i.i.d. $\mathcal{N}(0,1)$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.
Recovery: given $Y \sim \mathbb{P}$, recover v

If you can recover then you can detect (poly-time reduction)

- How: run recovery algorithm to get $\hat{v} \in \{0,1\}^n$; check $\hat{v}^\top Y \hat{v}$

So if $\mathrm{Adv}_{\leq D} = O(1)$, this suggests recovery is hard

Recovery [Schramm, W. '20]

Example (planted submatrix): observe $n \times n$ matrix $Y = X + Z$

- Signal: $X = \lambda v v^\top \quad \lambda > 0 \quad v_i \sim \mathrm{Bernoulli}(\rho)$
- Noise: Z i.i.d. $\mathcal{N}(0,1)$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.
Recovery: given $Y \sim \mathbb{P}$, recover v

If you can recover then you can detect (poly-time reduction)

- How: run recovery algorithm to get $\hat{v} \in \{0,1\}^n$; check $\hat{v}^\top Y \hat{v}$

So if $\mathrm{Adv}_{\leq D} = O(1)$, this suggests recovery is hard

But planted submatrix has a detection-recovery gap

Recovery [Schramm, W. '20]

Example (planted submatrix): observe $n \times n$ matrix $Y = X + Z$

- Signal: $X = \lambda vv^\top \quad \lambda > 0 \quad v_i \sim \mathrm{Bernoulli}(\rho)$
- Noise: Z i.i.d. $\mathcal{N}(0,1)$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.
Recovery: given $Y \sim \mathbb{P}$, recover v

If you can recover then you can detect (poly-time reduction)

- How: run recovery algorithm to get $\hat{v} \in \{0,1\}^n$; check $\hat{v}^\top Y \hat{v}$

So if $\mathrm{Adv}_{\leq D} = O(1)$, this suggests recovery is hard

But planted submatrix has a detection-recovery gap

How to show hardness of recovery when detection is easy?

Recovery [Schramm, W. '20]

Example (planted submatrix): observe $n \times n$ matrix $Y = X + Z$

- Signal: $X = \lambda v v^\top \quad \lambda > 0 \quad v_i \sim \mathrm{Bernoulli}(\rho)$
- Noise: Z i.i.d. $\mathcal{N}(0,1)$

Recovery [Schramm, W. '20]

Example (planted submatrix): observe $n \times n$ matrix $Y = X + Z$

- Signal: $X = \lambda vv^\top \quad \lambda > 0 \quad v_i \sim \mathrm{Bernoulli}(\rho)$
- Noise: Z i.i.d. $\mathcal{N}(0,1)$

Goal: given Y, estimate v_1 via polynomial $f : \mathbb{R}^{n \times n} \to \mathbb{R}$

Recovery [Schramm, W. '20]

Example (planted submatrix): observe $n \times n$ matrix $Y = X + Z$

- Signal: $X = \lambda v v^\top \quad \lambda > 0 \quad v_i \sim \mathrm{Bernoulli}(\rho)$
- Noise: Z i.i.d. $\mathcal{N}(0,1)$

Goal: given Y, estimate v_1 via polynomial $f : \mathbb{R}^{n \times n} \to \mathbb{R}$

Low-degree minimum mean squared error:

$$\mathrm{MMSE}_{\leq D} = \min_{f \text{ deg } D} \mathbb{E}(f(Y) - v_1)^2$$

Recovery [Schramm, W. '20]

Example (planted submatrix): observe $n \times n$ matrix $Y = X + Z$

- Signal: $X = \lambda v v^\top \quad \lambda > 0 \quad v_i \sim \mathrm{Bernoulli}(\rho)$
- Noise: Z i.i.d. $\mathcal{N}(0,1)$

Goal: given Y, estimate v_1 via polynomial $f : \mathbb{R}^{n \times n} \to \mathbb{R}$

Low-degree minimum mean squared error:

$$\mathrm{MMSE}_{\le D} = \min_{f \text{ deg } D} \mathbb{E}(f(Y) - v_1)^2$$

Equivalent to low-degree maximum correlation:

$$\mathrm{Corr}_{\le D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$$

Fact: $\mathrm{MMSE}_{\le D} = \mathbb{E}[v_1^2] - \mathrm{Corr}_{\le D}^2$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Same proof as detection?

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\text{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Same proof as detection?

$$f = \sum_{|S| \leq D} \hat{f}_S Y^S$$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Same proof as detection?

$$f = \sum_{|S| \leq D} \hat{f}_S Y^S$$

<u>Numerator</u>: $\mathbb{E}[f(Y) \cdot v_1]$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Same proof as detection?

$$f = \sum_{|S| \leq D} \hat{f}_S Y^S$$

<u>Numerator</u>: $\mathbb{E}[f(Y) \cdot v_1] = \sum_{|S| \leq D} \hat{f}_S \, \mathbb{E}[Y^S \cdot v_1]$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\le D} = \max\limits_{f \text{ deg } D} \dfrac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Same proof as detection?

$$f = \sum_{|S| \le D} \hat{f}_S Y^S$$

<u>Numerator</u>: $\mathbb{E}[f(Y) \cdot v_1] = \sum_{|S| \le D} \hat{f}_S \, \mathbb{E}[Y^S \cdot v_1] =: \langle \hat{f}, c \rangle$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Same proof as detection?

$$f = \sum_{|S| \leq D} \hat{f}_S Y^S$$

<u>Numerator</u>: $\mathbb{E}[f(Y) \cdot v_1] = \sum_{|S| \leq D} \hat{f}_S \, \mathbb{E}[Y^S \cdot v_1] =: \langle \hat{f}, c \rangle$

<u>Denominator</u>: $\mathbb{E}[f(Y)^2]$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Same proof as detection?

$$f = \sum_{|S| \leq D} \hat{f}_S Y^S$$

<u>Numerator</u>: $\mathbb{E}[f(Y) \cdot v_1] = \sum_{|S| \leq D} \hat{f}_S \, \mathbb{E}[Y^S \cdot v_1] =: \langle \hat{f}, c \rangle$

<u>Denominator</u>: $\mathbb{E}[f(Y)^2] = \sum_{S,T} \hat{f}_S \hat{f}_T \, \mathbb{E}[Y^S \cdot Y^T]$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Same proof as detection?

$$f = \sum_{|S| \leq D} \hat{f}_S Y^S$$

<u>Numerator</u>: $\mathbb{E}[f(Y) \cdot v_1] = \sum_{|S| \leq D} \hat{f}_S \, \mathbb{E}[Y^S \cdot v_1] =: \langle \hat{f}, c \rangle$

<u>Denominator</u>: $\mathbb{E}[f(Y)^2] = \sum_{S,T} \hat{f}_S \hat{f}_T \, \mathbb{E}[Y^S \cdot Y^T] = \hat{f}^\top M \hat{f}$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Same proof as detection?

$$f = \sum_{|S| \leq D} \hat{f}_S Y^S$$

<u>Numerator</u>: $\mathbb{E}[f(Y) \cdot v_1] = \sum_{|S| \leq D} \hat{f}_S \, \mathbb{E}[Y^S \cdot v_1] =: \langle \hat{f}, c \rangle$

<u>Denominator</u>: $\mathbb{E}[f(Y)^2] = \sum_{S,T} \hat{f}_S \hat{f}_T \, \mathbb{E}[Y^S \cdot Y^T] = \hat{f}^\top M \hat{f}$

$$\mathrm{Corr}_{\leq D} = \max_{\hat{f}} \frac{\langle \hat{f}, c \rangle}{\sqrt{\hat{f}^\top M \hat{f}}}$$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Same proof as detection?

$$f = \sum_{|S| \leq D} \hat{f}_S Y^S$$

<u>Numerator</u>: $\mathbb{E}[f(Y) \cdot v_1] = \sum_{|S| \leq D} \hat{f}_S \, \mathbb{E}[Y^S \cdot v_1] =: \langle \hat{f}, c \rangle$

<u>Denominator</u>: $\mathbb{E}[f(Y)^2] = \sum_{S,T} \hat{f}_S \hat{f}_T \, \mathbb{E}[Y^S \cdot Y^T] = \hat{f}^\top M \hat{f}$

$$\mathrm{Corr}_{\leq D} = \max_{\hat{f}} \frac{\langle \hat{f}, c \rangle}{\sqrt{\hat{f}^\top M \hat{f}}} = \sqrt{c^\top M^{-1} c}$$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Trick: bound denominator via Jensen's inequality on "signal" X

$$\mathbb{E}[f(Y)^2] = \mathop{\mathbb{E}}_{Z} \mathop{\mathbb{E}}_{X}[f(X+Z)^2] \geq \mathop{\mathbb{E}}_{Z} \left(\mathop{\mathbb{E}}_{X} f(X+Z) \right)^2$$

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Trick: bound denominator via Jensen's inequality on "signal" X

$$\mathbb{E}[f(Y)^2] = \mathop{\mathbb{E}}_{Z} \mathop{\mathbb{E}}_{X}[f(X+Z)^2] \geq \mathop{\mathbb{E}}_{Z} \left(\mathop{\mathbb{E}}_{X} f(X+Z) \right)^2$$

Why is this tight?

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Trick: bound denominator via Jensen's inequality on "signal" X

$$\mathbb{E}[f(Y)^2] = \mathop{\mathbb{E}}_{Z} \mathop{\mathbb{E}}_{X}[f(X+Z)^2] \geq \mathop{\mathbb{E}}_{Z} \left(\mathop{\mathbb{E}}_{X} f(X+Z) \right)^2$$

Why is this tight? In hard regime, f depends mostly on Z

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\leq D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Trick: bound denominator via Jensen's inequality on "signal" X

$$\mathbb{E}[f(Y)^2] = \mathop{\mathbb{E}}_{Z} \mathop{\mathbb{E}}_{X}[f(X+Z)^2] \geq \mathop{\mathbb{E}}_{Z} \left(\mathop{\mathbb{E}}_{X} f(X+Z) \right)^2$$

Why is this tight? In hard regime, f depends mostly on Z

This simplifies expression enough to find a closed form:

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\le D} = \max\limits_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Trick: bound denominator via Jensen's inequality on "signal" X

$$\mathbb{E}[f(Y)^2] = \mathop{\mathbb{E}}_Z \mathop{\mathbb{E}}_X [f(X+Z)^2] \ge \mathop{\mathbb{E}}_Z \left(\mathop{\mathbb{E}}_X f(X+Z) \right)^2$$

Why is this tight? In hard regime, f depends mostly on Z

This simplifies expression enough to find a closed form:

$$\mathrm{Corr}_{\le D} \le \max_{\hat{f}} \frac{\langle \hat{f}, c \rangle}{\|M\hat{f}\|}$$

where M is upper triangular

Recovery [Schramm, W. '20]

For hardness, want upper bound on $\mathrm{Corr}_{\le D} = \max_{f \text{ deg } D} \frac{\mathbb{E}[f(Y) \cdot v_1]}{\sqrt{\mathbb{E}[f(Y)^2]}}$

Trick: bound denominator via Jensen's inequality on "signal" X

$$\mathbb{E}[f(Y)^2] = \mathop{\mathbb{E}}_{Z} \mathop{\mathbb{E}}_{X}[f(X+Z)^2] \ge \mathop{\mathbb{E}}_{Z} \left(\mathop{\mathbb{E}}_{X} f(X+Z) \right)^2$$

Why is this tight? In hard regime, f depends mostly on Z

This simplifies expression enough to find a closed form:

$$\mathrm{Corr}_{\le D} \le \max_{\hat{f}} \frac{\langle \hat{f}, c \rangle}{\|M\hat{f}\|} = \|c^\top M^{-1}\|$$

where M is upper triangular (can invert)

Recovery [Schramm, W. '20]

End result:

Recovery [Schramm, W. '20]

End result:

Theorem [Schramm, W. '20]
Additive Gaussian model $Y = X + Z$
Scalar value to recover: x

Recovery [Schramm, W. '20]

End result:

Theorem [Schramm, W. '20]
Additive Gaussian model $Y = X + Z$
Scalar value to recover: x

$$\mathrm{Corr}^2_{\leq D} \leq \sum_{|S| \leq D} \kappa_S^2$$

where κ_S is the joint cumulant of $\{x\} \cup \{Y_i \,:\, i \in S\}$

Recovery [Schramm, W. '20]

End result:

Theorem [Schramm, W. '20]
Additive Gaussian model $Y = X + Z$
Scalar value to recover: x

$$\text{Corr}^2_{\leq D} \leq \sum_{|S| \leq D} \kappa_S^2$$

where κ_S is the joint cumulant of $\{x\} \cup \{Y_i \,:\, i \in S\}$

Corollary (tight bounds for planted submatrix recovery)

Recovery [Schramm, W. '20]

End result:

Theorem [Schramm, W. '20]
Additive Gaussian model $Y = X + Z$
Scalar value to recover: x

$$\mathrm{Corr}_{\le D}^2 \le \sum_{|S| \le D} \kappa_S^2$$

where κ_S is the joint cumulant of $\{x\} \cup \{Y_i \,:\, i \in S\}$

Corollary (tight bounds for planted submatrix recovery)

- if $\lambda \ll \min\{1, \frac{1}{\rho\sqrt{n}}\}$ then $\mathrm{MMSE}_{\le n^{\Omega(1)}} \approx \rho(1-\rho)$
  low-degree polynomials have trivial MSE in the "hard" regime

Recovery [Schramm, W. '20]

End result:

Theorem [Schramm, W. '20]
Additive Gaussian model $Y = X + Z$
Scalar value to recover: x

$$\mathrm{Corr}^2_{\leq D} \leq \sum_{|S| \leq D} \kappa_S^2$$

where κ_S is the joint cumulant of $\{x\} \cup \{Y_i \,:\, i \in S\}$

Corollary (tight bounds for planted submatrix recovery)

- if $\lambda \ll \min\{1, \frac{1}{\rho\sqrt{n}}\}$ then $\mathrm{MMSE}_{\leq n^{\Omega(1)}} \approx \rho(1-\rho)$
  low-degree polynomials have trivial MSE in the "hard" regime
- if $\lambda \gg \min\{1, \frac{1}{\rho\sqrt{n}}\}$ then $\mathrm{MMSE}_{\leq O(\log n)} = o(\rho)$
  low-degree polynomials succeed in the "easy" regime

Part IV: Optimization

Optimization [Gamarnik, Jagannath, W. '20]

Example (spherical spin glass): for $Y \in \mathbb{R}^{n \times n \times n \times n}$ i.i.d. $\mathcal{N}(0,1)$,

$$\max_{\|v\|=1} \frac{1}{\sqrt{n}} \langle Y, v^{\otimes 4} \rangle$$

Optimization [Gamarnik, Jagannath, W. '20]

Example (spherical spin glass): for $Y \in \mathbb{R}^{n \times n \times n \times n}$ i.i.d. $\mathcal{N}(0,1)$,

$$\max_{\|v\|=1} \frac{1}{\sqrt{n}} \langle Y, v^{\otimes 4} \rangle$$

Optimum value: $\mathrm{OPT} = \max_{\|v\|=1} H(v) = \Theta(1)$ [ABC'13]

Optimization [Gamarnik, Jagannath, W. '20]

Example (spherical spin glass): for $Y \in \mathbb{R}^{n \times n \times n \times n}$ i.i.d. $\mathcal{N}(0,1)$,

$$\max_{\|v\|=1} \frac{1}{\sqrt{n}} \langle Y, v^{\otimes 4} \rangle$$

Optimum value: $\mathrm{OPT} = \max_{\|v\|=1} H(v) = \Theta(1)$ [ABC'13]

Best known algorithms achieve value $\mathrm{ALG} < \mathrm{OPT}$ [Subag'18, EMS'20]

Optimization [Gamarnik, Jagannath, W. '20]

Example (spherical spin glass): for $Y \in \mathbb{R}^{n \times n \times n \times n}$ i.i.d. $\mathcal{N}(0,1)$,

$$\max_{\|v\|=1} \frac{1}{\sqrt{n}} \langle Y, v^{\otimes 4} \rangle$$

Optimum value: $\mathrm{OPT} = \max_{\|v\|=1} H(v) = \Theta(1)$ [ABC'13]

Best known algorithms achieve value $\mathrm{ALG} < \mathrm{OPT}$ [Subag'18, EMS'20]

Result: no low-degree polynomial can achieve value $\mathrm{OPT} - \epsilon$

Optimization [Gamarnik, Jagannath, W. '20]

Example (spherical spin glass): for $Y \in \mathbb{R}^{n\times n\times n\times n}$ i.i.d. $\mathcal{N}(0,1)$,

$$\max_{\|v\|=1} \frac{1}{\sqrt{n}} \langle Y, v^{\otimes 4} \rangle$$

Optimum value: $\mathrm{OPT} = \max_{\|v\|=1} H(v) = \Theta(1)$ [ABC'13]

Best known algorithms achieve value $\mathrm{ALG} < \mathrm{OPT}$ [Subag'18, EMS'20]

Result: no low-degree polynomial can achieve value $\mathrm{OPT} - \epsilon$

Theorem [Gamarnik, Jagannath, W. '20]
For some $\epsilon > 0$, no $f : \mathbb{R}^{n\times n\times n\times n} \to \mathbb{R}^n$ of degree $\mathrm{polylog}(n)$ achieves both of the following with probability $1 - \exp(-n^{\Omega(1)})$:

Optimization [Gamarnik, Jagannath, W. '20]

Example (spherical spin glass): for $Y \in \mathbb{R}^{n \times n \times n \times n}$ i.i.d. $\mathcal{N}(0,1)$,

$$\max_{\|v\|=1} \frac{1}{\sqrt{n}} \langle Y, v^{\otimes 4} \rangle$$

Optimum value: $\text{OPT} = \max_{\|v\|=1} H(v) = \Theta(1)$ [ABC'13]

Best known algorithms achieve value $\text{ALG} < \text{OPT}$ [Subag'18, EMS'20]

Result: no low-degree polynomial can achieve value $\text{OPT} - \epsilon$

Theorem [Gamarnik, Jagannath, W. '20]
For some $\epsilon > 0$, no $f : \mathbb{R}^{n \times n \times n \times n} \to \mathbb{R}^n$ of degree $\mathrm{polylog}(n)$ achieves both of the following with probability $1 - \exp(-n^{\Omega(1)})$:

- Objective: $H(f(Y)) \geq \text{OPT} - \epsilon$

Optimization [Gamarnik, Jagannath, W. '20]

Example (spherical spin glass): for $Y \in \mathbb{R}^{n \times n \times n \times n}$ i.i.d. $\mathcal{N}(0,1)$,

$$\max_{\|v\|=1} \frac{1}{\sqrt{n}} \langle Y, v^{\otimes 4} \rangle$$

Optimum value: $\mathrm{OPT} = \max_{\|v\|=1} H(v) = \Theta(1)$ [ABC'13]

Best known algorithms achieve value $\mathrm{ALG} < \mathrm{OPT}$ [Subag'18, EMS'20]

Result: no low-degree polynomial can achieve value $\mathrm{OPT} - \epsilon$

Theorem [Gamarnik, Jagannath, W. '20]
For some $\epsilon > 0$, no $f : \mathbb{R}^{n \times n \times n \times n} \to \mathbb{R}^n$ of degree $\mathrm{polylog}(n)$ achieves both of the following with probability $1 - \exp(-n^{\Omega(1)})$:

- Objective: $H(f(Y)) \geq \mathrm{OPT} - \epsilon$
- Normalization: $\|f(Y)\| \approx 1$

Optimization [Gamarnik, Jagannath, W. '20]

Example (max independent set): given sparse graph $G(n, d/n)$,

$$\max_{S \subseteq [n]} |S| \;\; s.t.\; S \text{ independent}$$

Optimization [Gamarnik, Jagannath, W. '20]

Example (max independent set): given sparse graph $G(n, d/n)$,

$$\max_{S \subseteq [n]} |S| \;\; s.t.\; S \text{ independent}$$

$$\mathrm{OPT} = 2\frac{\log d}{d} n$$

Optimization [Gamarnik, Jagannath, W. '20]

Example (max independent set): given sparse graph $G(n, d/n)$,

$$\max_{S \subseteq [n]} |S| \;\; s.t.\; S \text{ independent}$$

$$\mathrm{OPT} = 2\frac{\log d}{d}n \qquad \mathrm{ALG} = \frac{\log d}{d}n$$

Optimization [Gamarnik, Jagannath, W. '20]

Example (max independent set): given sparse graph $G(n, d/n)$,

$$\max_{S \subseteq [n]} |S| \;\; s.t. \; S \text{ independent}$$

$$\mathrm{OPT} = 2\frac{\log d}{d} n \qquad \mathrm{ALG} = \frac{\log d}{d} n$$

Result: no low-degree polynomial can achieve $(1 + \frac{1}{\sqrt{2}})\frac{\log d}{d} n$

Optimization [Gamarnik, Jagannath, W. '20]

Example (max independent set): given sparse graph $G(n, d/n)$,

$$\max_{S \subseteq [n]} |S| \;\; s.t.\; S \text{ independent}$$

$$\mathrm{OPT} = 2\frac{\log d}{d} n \qquad \mathrm{ALG} = \frac{\log d}{d} n$$

Result: no low-degree polynomial can achieve $(1 + \frac{1}{\sqrt{2}})\frac{\log d}{d} n$

Theorem [Gamarnik, Jagannath, W. '20]

No polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}^n$ of degree $\mathrm{polylog}(n)$ achieves both of the following with probability $1 - \exp(-n^{\Omega(1)})$:

Optimization [Gamarnik, Jagannath, W. '20]

Example (max independent set): given sparse graph $G(n, d/n)$,

$$\max_{S \subseteq [n]} |S| \;\; s.t.\ S \text{ independent}$$

$$\mathrm{OPT} = 2\frac{\log d}{d} n \qquad \mathrm{ALG} = \frac{\log d}{d} n$$

Result: no low-degree polynomial can achieve $(1 + \frac{1}{\sqrt{2}})\frac{\log d}{d} n$

Theorem [Gamarnik, Jagannath, W. '20]

No polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}^n$ of degree $\mathrm{polylog}(n)$ achieves both of the following with probability $1 - \exp(-n^{\Omega(1)})$:

- $f_i(Y) \in [0, 1/3] \cup [2/3, 1]$ for most i

Optimization [Gamarnik, Jagannath, W. '20]

Example (max independent set): given sparse graph $G(n, d/n)$,

$$\max_{S \subseteq [n]} |S| \;\; s.t.\; S \text{ independent}$$

$$\mathrm{OPT} = 2\frac{\log d}{d} n \qquad \mathrm{ALG} = \frac{\log d}{d} n$$

Result: no low-degree polynomial can achieve $(1 + \frac{1}{\sqrt{2}})\frac{\log d}{d} n$

Theorem [Gamarnik, Jagannath, W. '20]

No polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}^n$ of degree $\mathrm{polylog}(n)$ achieves both of the following with probability $1 - \exp(-n^{\Omega(1)})$:

- $f_i(Y) \in [0, 1/3] \cup [2/3, 1]$ for most i
- $\{i : f_i(Y) \in [2/3, 1]\}$ is a near-indep set of size $(1 + \frac{1}{\sqrt{2}})\frac{\log d}{d} n$

Optimization [Gamarnik, Jagannath, W. '20]

Example (max independent set): given sparse graph $G(n, d/n)$,

$$\max_{S \subseteq [n]} |S| \;\; s.t.\; S \text{ independent}$$

$$\mathrm{OPT} = 2\frac{\log d}{d} n \qquad \mathrm{ALG} = \frac{\log d}{d} n$$

Result: no low-degree polynomial can achieve $(1 + \frac{1}{\sqrt{2}})\frac{\log d}{d} n$

Theorem [Gamarnik, Jagannath, W. '20]

No polynomial $f : \{0,1\}^{\binom{n}{2}} \to \mathbb{R}^n$ of degree $\mathrm{polylog}(n)$ achieves both of the following with probability $1 - \exp(-n^{\Omega(1)})$:

- $f_i(Y) \in [0, 1/3] \cup [2/3, 1]$ for most i
- $\{i : f_i(Y) \in [2/3, 1]\}$ is a near-indep set of size $(1 + \frac{1}{\sqrt{2}})\frac{\log d}{d} n$

Forthcoming: improve $1 + \frac{1}{\sqrt{2}} \to 1 + \epsilon$ (optimal)

Optimization [Gamarnik, Jagannath, W. '20]

How to prove failure of low-degree polynomials for optimization?

Optimization [Gamarnik, Jagannath, W. '20]

How to prove failure of low-degree polynomials for optimization?

Same proof as before?

Optimization [Gamarnik, Jagannath, W. '20]

How to prove failure of low-degree polynomials for optimization?

Same proof as before?

$$\max_{f \deg D} \mathbb{E}\, H(f(Y)) = \max_{f \deg D} \mathbb{E} \frac{1}{\sqrt{n}} \langle Y, f(Y)^{\otimes 4} \rangle$$

Optimization [Gamarnik, Jagannath, W. '20]

How to prove failure of low-degree polynomials for optimization?

Same proof as before?

$$\max_{f \text{ deg } D} \mathbb{E}\, H(f(Y)) = \max_{f \text{ deg } D} \mathbb{E} \frac{1}{\sqrt{n}} \langle Y, f(Y)^{\otimes 4} \rangle$$

No! High-degree in $\hat{f}$

Optimization [Gamarnik, Jagannath, W. '20]

How to prove failure of low-degree polynomials for optimization?

Same proof as before?

$$\max_{f \deg D} \mathbb{E}\, H(f(Y)) = \max_{f \deg D} \mathbb{E} \frac{1}{\sqrt{n}} \langle Y, f(Y)^{\otimes 4} \rangle$$

No! High-degree in $\hat{f}$

Instead, use 2 ingredients:

Optimization [Gamarnik, Jagannath, W. '20]

How to prove failure of low-degree polynomials for optimization?

Same proof as before?

$$\max_{f \text{ deg } D} \mathbb{E}\, H(f(Y)) = \max_{f \text{ deg } D} \mathbb{E} \frac{1}{\sqrt{n}} \langle Y, f(Y)^{\otimes 4} \rangle$$

No! High-degree in $\hat{f}$

Instead, use 2 ingredients:

- Stability of low-degree polynomials

Optimization [Gamarnik, Jagannath, W. '20]

How to prove failure of low-degree polynomials for optimization?

Same proof as before?

$$\max_{f \deg D} \mathbb{E}\, H(f(Y)) = \max_{f \deg D} \mathbb{E} \frac{1}{\sqrt{n}} \langle Y, f(Y)^{\otimes 4} \rangle$$

No! High-degree in $\hat{f}$

Instead, use 2 ingredients:

- Stability of low-degree polynomials
- Overlap gap property (OGP)
  [Gamarnik, Sudan '13]
  [Chen, Gamarnik, Panchenko, Rahman '17]
  [Gamarnik, Jagannath '19]

Optimization [Gamarnik, Jagannath, W. '20]

"Low-degree polynomials are stable"

Optimization [Gamarnik, Jagannath, W. '20]

"Low-degree polynomials are stable"

$Y \sim$ i.i.d. Bernoulli(p)

Optimization [Gamarnik, Jagannath, W. '20]

"Low-degree polynomials are stable"

$Y \sim$ i.i.d. Bernoulli(p)

Interpolation path: $Y^{(0)} \quad Y^{(1)} \quad Y^{(2)} \quad \ldots \quad Y^{(m-1)} \quad Y^{(m)}$

Optimization [Gamarnik, Jagannath, W. '20]

"Low-degree polynomials are stable"

$Y \sim$ i.i.d. Bernoulli(p)

Interpolation path: $Y^{(0)} \quad Y^{(1)} \quad Y^{(2)} \quad \ldots \quad Y^{(m-1)} \quad Y^{(m)}$

$f : \{0,1\}^m \to \mathbb{R}^n$ degree D

Optimization [Gamarnik, Jagannath, W. '20]

"Low-degree polynomials are stable"

$Y \sim$ i.i.d. Bernoulli(p)

Interpolation path: $Y^{(0)} \quad Y^{(1)} \quad Y^{(2)} \quad \ldots \quad Y^{(m-1)} \quad Y^{(m)}$

$f : \{0,1\}^m \to \mathbb{R}^n$ degree D

Definition: Index i is "c-bad" if

$$\|f(Y^{(i)}) - f(Y^{(i-1)})\|^2 > c \mathop{\mathbb{E}}_{Y} \|f(Y)\|^2$$

Optimization [Gamarnik, Jagannath, W. '20]

"Low-degree polynomials are stable"

$Y \sim$ i.i.d. Bernoulli(p)

Interpolation path: $Y^{(0)} \quad Y^{(1)} \quad Y^{(2)} \quad \ldots \quad Y^{(m-1)} \quad Y^{(m)}$

$f : \{0,1\}^m \to \mathbb{R}^n$ degree D

Definition: Index i is "c-bad" if

$$\|f(Y^{(i)}) - f(Y^{(i-1)})\|^2 > c \mathop{\mathbb{E}}_{Y} \|f(Y)\|^2$$

Theorem

$$\Pr_{Y^{(0)},\ldots,Y^{(m)}} [\nexists\, c\text{-bad } i] \geq p^{4D/c}$$

Optimization [Gamarnik, Jagannath, W. '20]

"Low-degree polynomials are stable"

$Y \sim$ i.i.d. Bernoulli(p)

Interpolation path: $Y^{(0)} \quad Y^{(1)} \quad Y^{(2)} \quad \ldots \quad Y^{(m-1)} \quad Y^{(m)}$

$f : \{0,1\}^m \to \mathbb{R}^n$ degree D

Definition: Index i is "c-bad" if

$$\|f(Y^{(i)}) - f(Y^{(i-1)})\|^2 > c \mathop{\mathbb{E}}_{Y} \|f(Y)\|^2$$

Theorem

$$\Pr_{Y^{(0)},\ldots,Y^{(m)}} [\nexists\, c\text{-bad } i] \geq p^{4D/c}$$

With non-trivial probability (over path), f's output is "smooth"

Optimization [Gamarnik, Jagannath, W. '20]

Overlap gap property (OGP): with high probability,
$Y \sim G(n, d/n)$ has no occurrence of

Optimization [Gamarnik, Jagannath, W. '20]

Overlap gap property (OGP): with high probability,
$Y \sim G(n, d/n)$ has no occurrence of

- S, T independent sets

Optimization [Gamarnik, Jagannath, W. '20]

Overlap gap property (OGP): with high probability, $Y \sim G(n, d/n)$ has no occurrence of

- S, T independent sets
- $|S|, |T| \approx (1 + \frac{1}{\sqrt{2}})\Phi$

Optimization [Gamarnik, Jagannath, W. '20]

Overlap gap property (OGP): with high probability, $Y \sim G(n, d/n)$ has no occurrence of

- S, T independent sets
- $|S|, |T| \approx (1 + \frac{1}{\sqrt{2}})\Phi$
- $|S \cap T| \approx \Phi$

Optimization [Gamarnik, Jagannath, W. '20]

Overlap gap property (OGP): with high probability,
$Y \sim G(n, d/n)$ has no occurrence of

- S, T independent sets
- $|S|, |T| \approx (1 + \frac{1}{\sqrt{2}})\Phi$
- $|S \cap T| \approx \Phi$

Proof: first moment method [Gamarnik, Sudan '13]

Optimization [Gamarnik, Jagannath, W. '20]

Ensemble OGP: with high probability, $\forall i, j$ on the interpolation path

$$Y^{(0)} \quad Y^{(1)} \quad Y^{(2)} \quad \dots \quad Y^{(m-1)} \quad Y^{(m)}$$

there is no occurrence of

- S independent set in $Y^{(i)}$
- T independent set in $Y^{(j)}$
- $|S|, |T| \approx (1 + \frac{1}{\sqrt{2}})\Phi$
- $|S \cap T| \approx \Phi$

Optimization [Gamarnik, Jagannath, W. '20]

Proof that low-degree polynomials fail:

Optimization [Gamarnik, Jagannath, W. '20]

Proof that low-degree polynomials fail:

Suppose $f(Y)$ outputs independent sets of size $(1 + \frac{1}{\sqrt{2}})\Phi$

Optimization [Gamarnik, Jagannath, W. '20]

Proof that low-degree polynomials fail:

Suppose $f(Y)$ outputs independent sets of size $(1 + \frac{1}{\sqrt{2}})\Phi$

$$Y^{(0)} \quad Y^{(1)} \quad Y^{(2)} \quad \ldots \quad Y^{(m-1)} \quad Y^{(m)}$$

Optimization [Gamarnik, Jagannath, W. '20]

Proof that low-degree polynomials fail:

Suppose $f(Y)$ outputs independent sets of size $(1 + \frac{1}{\sqrt{2}})\Phi$

$$Y^{(0)} \quad Y^{(1)} \quad Y^{(2)} \quad \ldots \quad Y^{(m-1)} \quad Y^{(m)}$$

Separation: $f(Y^{(0)})$ and $f(Y^{(m)})$ are "far apart"

Optimization [Gamarnik, Jagannath, W. '20]

Proof that low-degree polynomials fail:

Suppose $f(Y)$ outputs independent sets of size $(1 + \frac{1}{\sqrt{2}})\Phi$

$$Y^{(0)} \quad Y^{(1)} \quad Y^{(2)} \quad \ldots \quad Y^{(m-1)} \quad Y^{(m)}$$

Separation: $f(Y^{(0)})$ and $f(Y^{(m)})$ are "far apart"

Stability: with probability $\gtrsim n^{-D}$, there are no big "jumps"

$$f(Y^{(i)}) \to f(Y^{(i+1)})$$

Optimization [Gamarnik, Jagannath, W. '20]

Proof that low-degree polynomials fail:

Suppose $f(Y)$ outputs independent sets of size $(1+\frac{1}{\sqrt{2}})\Phi$

$$Y^{(0)} \quad Y^{(1)} \quad Y^{(2)} \quad \ldots \quad Y^{(m-1)} \quad Y^{(m)}$$

Separation: $f(Y^{(0)})$ and $f(Y^{(m)})$ are "far apart"

Stability: with probability $\gtrsim n^{-D}$, there are no big "jumps"

$$f(Y^{(i)}) \to f(Y^{(i+1)})$$

Contradicts OGP

Future Directions?

Future Directions?

- (Detection) bound $\mathrm{Adv}_{\leq D}$ when $\mathbb{Q}$ is not a product measure
  - E.g. random regular graphs

Future Directions?

- (Detection) bound $\mathrm{Adv}_{\leq D}$ when $\mathbb{Q}$ is not a product measure
  - E.g. random regular graphs
- (Recovery) bound $\mathrm{MMSE}_{\leq D}$ when not "signal + noise"
  - E.g. sparse regression, phase retrieval

Future Directions?

- (Detection) bound $\mathrm{Adv}_{\leq D}$ when $\mathbb{Q}$ is not a product measure
  - E.g. random regular graphs
- (Recovery) bound $\mathrm{MMSE}_{\leq D}$ when not "signal + noise"
  - E.g. sparse regression, phase retrieval
- (Recovery) precise value of $\mathrm{MMSE}_{\leq D}$
  - Matching AMP?

Future Directions?

- (Detection) bound $\mathrm{Adv}_{\leq D}$ when $\mathbb{Q}$ is not a product measure
  - E.g. random regular graphs
- (Recovery) bound $\mathrm{MMSE}_{\leq D}$ when not "signal + noise"
  - E.g. sparse regression, phase retrieval
- (Recovery) precise value of $\mathrm{MMSE}_{\leq D}$
  - Matching AMP?
- (Optimization) prove tight results for new settings
  - E.g. p-spin optimization

Future Directions?

- (Detection) bound $\text{Adv}_{\leq D}$ when $\mathbb{Q}$ is not a product measure
  - E.g. random regular graphs
- (Recovery) bound $\text{MMSE}_{\leq D}$ when not "signal + noise"
  - E.g. sparse regression, phase retrieval
- (Recovery) precise value of $\text{MMSE}_{\leq D}$
  - Matching AMP?
- (Optimization) prove tight results for new settings
  - E.g. p-spin optimization
- Implications for other algorithms?
  - E.g. convex programming, MCMC

References

- Detection (survey article)
  Notes on Computational Hardness of Hypothesis Testing: Predictions using the Low-Degree Likelihood Ratio
  Kunisky, W., Bandeira
  arXiv:1907.11636

- Recovery
  Computational Barriers to Estimation from Low-Degree Polynomials
  Schramm, W.
  arXiv:2008.02269

- Optimization
  Low-Degree Hardness of Random Optimization Problems
  Gamarnik, Jagannath, W.
  arXiv:2004.12063

(extra scratch paper)