Computational Barriers to Estimation from Low-Degree Polynomials

Alex Wein
Courant Institute, New York University

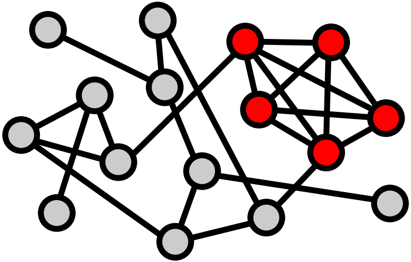
Joint work with:

Tselil Schramm
Stanford

Part I: Why Low-Degree Polynomials?

Problems in High-Dimensional Statistics

Example: planted k-clique in a random graph $G(n, 1/2)$



Problems in High-Dimensional Statistics

Example: planted k-clique in a random graph $G(n, 1/2)$

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

Problems in High-Dimensional Statistics

Example: planted k-clique in a random graph $G(n, 1/2)$

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

Problems in High-Dimensional Statistics

Example: planted k-clique in a random graph $G(n, 1/2)$

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

Both problems have an information-computation gap

Problems in High-Dimensional Statistics

Example: planted k-clique in a random graph $G(n, 1/2)$

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

Both problems have an information-computation gap

Problems in High-Dimensional Statistics

Example: planted k-clique in a random graph $G(n, 1/2)$

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

Both problems have an information-computation gap

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: $b \in \{0,1\}$ (detection)

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: $b \in \{0,1\}$ (detection) or $v \in \mathbb{R}^n$ (recovery)

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: $b \in \{0,1\}$ (detection) or $v \in \mathbb{R}^n$ (recovery)
- "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: $b \in \{0,1\}$ (detection) or $v \in \mathbb{R}^n$ (recovery)
- "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: $b \in \{0,1\}$ (detection) or $v \in \mathbb{R}^n$ (recovery)
- "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: $b \in \{0,1\}$ (detection) or $v \in \mathbb{R}^n$ (recovery)
- "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: $b \in \{0,1\}$ (detection) or $v \in \mathbb{R}^n$ (recovery)
- "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: $b \in \{0,1\}$ (detection) or $v \in \mathbb{R}^n$ (recovery)
- "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: $b \in \{0,1\}$ (detection) or $v \in \mathbb{R}^n$ (recovery)
- "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 $\tilde{Y} = g(Y)$ — $\deg g = O(1)$

Optimality of Low-Degree Polynomials?

Low-degree polynomials seem to be optimal for many problems!

Optimality of Low-Degree Polynomials?

Low-degree polynomials seem to be optimal for many problems!

For all of these problems...

Optimality of Low-Degree Polynomials?

Low-degree polynomials seem to be optimal for many problems!

For all of these problems...

planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, ...

Optimality of Low-Degree Polynomials?

Low-degree polynomials seem to be optimal for many problems!

For all of these problems...

planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, ...

...it is the case that

Optimality of Low-Degree Polynomials?

Low-degree polynomials seem to be optimal for many problems!

For all of these problems...

planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, ...

...it is the case that

- the best known poly-time algorithms are captured by $O(\log n)$-degree polynomials (spectral/AMP)

Optimality of Low-Degree Polynomials?

Low-degree polynomials seem to be optimal for many problems!

For all of these problems...

planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, ...

...it is the case that

- the best known poly-time algorithms are captured by $O(\log n)$-degree polynomials (spectral/AMP)
- low-degree polynomials fail in the "hard" regime

Optimality of Low-Degree Polynomials?

Low-degree polynomials seem to be optimal for many problems!

For all of these problems...
planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, ...
...it is the case that

- the best known poly-time algorithms are captured by $O(\log n)$-degree polynomials (spectral/AMP)
- low-degree polynomials fail in the "hard" regime

"Low-degree conjecture" (informal): for "natural" problems, if low-degree polynomials fail then all poly-time algorithms fail [Hopkins '18]

Optimality of Low-Degree Polynomials?

Low-degree polynomials seem to be optimal for many problems!

For all of these problems...
planted clique, sparse PCA, community detection, tensor PCA, spiked Wigner/Wishart, planted submatrix, planted dense subgraph, ...
...it is the case that

- the best known poly-time algorithms are captured by $O(\log n)$-degree polynomials (spectral/AMP)
- low-degree polynomials fail in the "hard" regime

"Low-degree conjecture" (informal): for "natural" problems, if low-degree polynomials fail then all poly-time algorithms fail [Hopkins '18]

Caveat: Gaussian elimination for planted XOR-SAT

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 (prior work)

  [Hopkins, Steurer '17]

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

  [Hopkins '18] (PhD thesis)

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

Overview

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

- Gives evidence for computational hardness

Settings:

- Detection (prior work)

  [Hopkins, Steurer '17]

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

  [Hopkins '18] (PhD thesis)

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

- Recovery (this work)

  [Schramm, W. '20]

Overview

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

- Gives evidence for computational hardness

Settings:

- Detection (prior work)

  [Hopkins, Steurer '17]

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

  [Hopkins '18] (PhD thesis)

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

- Recovery (this work)

  [Schramm, W. '20]

- Optimization

  [Gamarnik, Jagannath, W. '20]

Relation to Other Frameworks

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]
- Statistical query lower bounds [FGRVX12,...]

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]
- Statistical query lower bounds [FGRVX12,...]
  - Need i.i.d. samples

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]
- Statistical query lower bounds [FGRVX12,...]
  - Need i.i.d. samples
  - Equivalent to low-degree [BBHLS20]

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]
- Statistical query lower bounds [FGRVX12,...]
  - Need i.i.d. samples
  - Equivalent to low-degree [BBHLS20]
- Approximate message passing (AMP) [DMM09, LKZ15,...]

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]
- Statistical query lower bounds [FGRVX12,...]
  - Need i.i.d. samples
  - Equivalent to low-degree [BBHLS20]
- Approximate message passing (AMP) [DMM09, LKZ15,...]
  - AMP algorithms are low-degree

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]
- Statistical query lower bounds [FGRVX12,...]
  - Need i.i.d. samples
  - Equivalent to low-degree [BBHLS20]
- Approximate message passing (AMP) [DMM09, LKZ15,...]
  - AMP algorithms are low-degree
  - AMP can be sub-optimal (e.g. tensor PCA) [MR14]

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]
- Statistical query lower bounds [FGRVX12,...]
  - Need i.i.d. samples
  - Equivalent to low-degree [BBHLS20]
- Approximate message passing (AMP) [DMM09, LKZ15,...]
  - AMP algorithms are low-degree
  - AMP can be sub-optimal (e.g. tensor PCA) [MR14]
- Overlap gap property / MCMC lower bounds [GS13, GZ17,...]

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]
- Statistical query lower bounds [FGRVX12,...]
  - Need i.i.d. samples
  - Equivalent to low-degree [BBHLS20]
- Approximate message passing (AMP) [DMM09, LKZ15,...]
  - AMP algorithms are low-degree
  - AMP can be sub-optimal (e.g. tensor PCA) [MR14]
- Overlap gap property / MCMC lower bounds [GS13, GZ17,...]
  - MCMC algorithms are not low-degree (?)

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]
- Statistical query lower bounds [FGRVX12,...]
  - Need i.i.d. samples
  - Equivalent to low-degree [BBHLS20]
- Approximate message passing (AMP) [DMM09, LKZ15,...]
  - AMP algorithms are low-degree
  - AMP can be sub-optimal (e.g. tensor PCA) [MR14]
- Overlap gap property / MCMC lower bounds [GS13, GZ17,...]
  - MCMC algorithms are not low-degree (?)
  - MCMC can be sub-optimal (e.g. tensor PCA) [BGJ18]

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]
- Statistical query lower bounds [FGRVX12,...]
  - Need i.i.d. samples
  - Equivalent to low-degree [BBHLS20]
- Approximate message passing (AMP) [DMM09, LKZ15,...]
  - AMP algorithms are low-degree
  - AMP can be sub-optimal (e.g. tensor PCA) [MR14]
- Overlap gap property / MCMC lower bounds [GS13, GZ17,...]
  - MCMC algorithms are not low-degree (?)
  - MCMC can be sub-optimal (e.g. tensor PCA) [BGJ18]
- Average-case reductions [BR13,...]

Relation to Other Frameworks

- Sum-of-squares lower bounds [BHKKMP16,...]
  - Actually for certification
  - Connected to low-degree [HKPRSS17]
- Statistical query lower bounds [FGRVX12,...]
  - Need i.i.d. samples
  - Equivalent to low-degree [BBHLS20]
- Approximate message passing (AMP) [DMM09, LKZ15,...]
  - AMP algorithms are low-degree
  - AMP can be sub-optimal (e.g. tensor PCA) [MR14]
- Overlap gap property / MCMC lower bounds [GS13, GZ17,...]
  - MCMC algorithms are not low-degree (?)
  - MCMC can be sub-optimal (e.g. tensor PCA) [BGJ18]
- Average-case reductions [BR13,...]
  - Need to argue that starting problem is hard [BB20]

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":

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

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":

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

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

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 $\mathrm{Adv}_{\leq D} = \omega(1)$ 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 $\mathrm{Adv}_{\le D} = \omega(1)$ 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 $k \ll \sqrt{n}$

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 $\mathrm{Adv}_{\leq D} = \omega(1)$ 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 $D = n^\delta$

Extended low-degree conjecture [Hopkins '18]:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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)

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

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. $\langle f, g \rangle := \mathop{\mathbb{E}}_{Y \sim \mathbb{Q}}[f(Y)g(Y)]$

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 $\mathrm{Adv}_{\le D}$) is

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

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

"low-degree likelihood ratio"

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

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 $\mathrm{Adv}_{\le D}$) is

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

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

  "low-degree likelihood ratio"

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

  "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"

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

"norm of low-degree likelihood ratio"

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

Part III: Recovery

Planted Submatrix

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. $\mathcal{N}(0,1)$

Planted Submatrix

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. $\mathcal{N}(0, 1)$

Regime: $1/\sqrt{n} \ll \rho \ll 1$

Planted Submatrix

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. $\mathcal{N}(0, 1)$

Regime: $1/\sqrt{n} \ll \rho \ll 1$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.

Planted Submatrix

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. $\mathcal{N}(0,1)$

Regime: $1/\sqrt{n} \ll \rho \ll 1$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.

- Sum of all entries succeeds when $\lambda \gg (\rho\sqrt{n})^{-2}$

Planted Submatrix

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. $\mathcal{N}(0,1)$

Regime: $1/\sqrt{n} \ll \rho \ll 1$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.

- Sum of all entries succeeds when $\lambda \gg (\rho\sqrt{n})^{-2}$

Recovery: given $Y \sim \mathbb{P}$, recover v

Planted Submatrix

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. $\mathcal{N}(0,1)$

Regime: $1/\sqrt{n} \ll \rho \ll 1$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.

- Sum of all entries succeeds when $\lambda \gg (\rho\sqrt{n})^{-2}$

Recovery: given $Y \sim \mathbb{P}$, recover v

- Leading eigenvector succeeds when $\lambda \gg (\rho\sqrt{n})^{-1}$

Planted Submatrix

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. $\mathcal{N}(0,1)$

Regime: $1/\sqrt{n} \ll \rho \ll 1$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.

- Sum of all entries succeeds when $\lambda \gg (\rho\sqrt{n})^{-2}$

Recovery: given $Y \sim \mathbb{P}$, recover v

- Leading eigenvector succeeds when $\lambda \gg (\rho\sqrt{n})^{-1}$
- Exhaustive search succeeds when $\lambda \gg (\rho n)^{-1/2}$

Planted Submatrix

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. $\mathcal{N}(0, 1)$

Regime: $1/\sqrt{n} \ll \rho \ll 1$

Detection: distinguish $\mathbb{P} : Y = X + Z$ vs $\mathbb{Q} : Y = Z$ w.h.p.

- Sum of all entries succeeds when $\lambda \gg (\rho\sqrt{n})^{-2}$

Recovery: given $Y \sim \mathbb{P}$, recover v

- Leading eigenvector succeeds when $\lambda \gg (\rho\sqrt{n})^{-1}$
- Exhaustive search succeeds when $\lambda \gg (\rho n)^{-1/2}$

Detection-recovery gap

Recovery Hardness from Detection Hardness?

Recovery Hardness from Detection Hardness?

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

Recovery Hardness from Detection Hardness?

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 Hardness from Detection Hardness?

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

Recovery Hardness from Detection Hardness?

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 how to show hardness of recovery when detection is easy?

Recovery Hardness from Detection Hardness?

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

But how to show hardness of recovery when detection is easy?

Attempt: choose a better null distribution?

Recovery Hardness from Detection Hardness?

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

But how to show hardness of recovery when detection is easy?

Attempt: choose a better null distribution?

- Match mean of planted distribution?

Recovery Hardness from Detection Hardness?

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 how to show hardness of recovery when detection is easy?

Attempt: choose a better null distribution?

- Match mean of planted distribution?
- Gaussian matching first 2 moments of planted distribution?

Recovery Hardness from Detection Hardness?

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

But how to show hardness of recovery when detection is easy?

Attempt: choose a better null distribution?

- Match mean of planted distribution?
- Gaussian matching first 2 moments of planted distribution?

This closes detection-recovery gap partially but not all the way

Low-Degree Recovery

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. $\mathcal{N}(0,1)$

Low-Degree Recovery

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 $f : \mathbb{R}^{n \times n} \to \mathbb{R}$

Low-Degree Recovery

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:

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

Low-Degree Recovery

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

Hardness of Recovery

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

Hardness of Recovery

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?

Hardness of Recovery

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?

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

Hardness of Recovery

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?

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

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

Hardness of Recovery

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?

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

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

Hardness of Recovery

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

Same proof as detection?

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

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

Hardness of Recovery

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

Hardness of Recovery

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

Hardness of Recovery

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?

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

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

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

Hardness of Recovery

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?

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

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

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

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

Hardness of Recovery

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?

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

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

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

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

Hardness of Recovery

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

Hardness of Recovery

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

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

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

Hardness of Recovery

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?

Hardness of Recovery

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

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

Hardness of Recovery

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

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:

Hardness of Recovery

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:

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

where M is upper triangular

Hardness of Recovery

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

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:

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

where M is upper triangular (can invert)

Main Result

Main Result

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

Main 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 $\{x\} \cup \{Y_i \,:\, i \in S\}$

Main 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)

Main 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

Main 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

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) sharp threshold for planted submatrix
  - AMP succeeds when $\lambda > (\rho\sqrt{en})^{-1}$ [Hajek, Wu, Xu '15]

Future Directions?

- (Detection) bound $\mathrm{Adv}_{\le D}$ when $\mathbb{Q}$ is not a product measure
  - E.g. random regular graphs
- (Recovery) bound $\mathrm{MMSE}_{\le D}$ when not "signal + noise"
  - E.g. sparse regression, phase retrieval
- (Recovery) sharp threshold for planted submatrix
  - AMP succeeds when $\lambda > (\rho\sqrt{en})^{-1}$ [Hajek, Wu, Xu '15]
- 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)