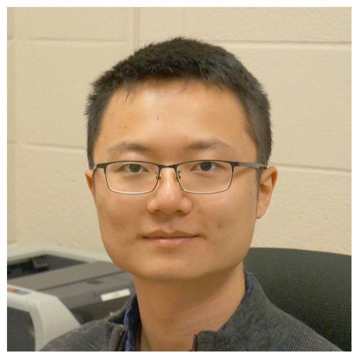
Optimal Spectral Recovery of a Planted Vector in a Subspace

Alex Wein
Simons Institute

Joint work with:



Cheng Mao
Georgia Tech

Planted Vector Problem

Goal: recover a **structured** vector $v \in \mathbb{R}^n$ planted in a **random** d-dimensional subspace of $\mathbb{R}^n$

Planted Vector Problem

$d \ll n$

Goal: recover a **structured** vector $v \in \mathbb{R}^n$ planted in a **random** d-dimensional subspace of $\mathbb{R}^n$

Planted Vector Problem

$d \ll n$

Goal: recover a **structured** vector $v \in \mathbb{R}^n$ planted in a **random** d-dimensional subspace of $\mathbb{R}^n$

- Structure: e.g. sparsity $\|v\|_0 = \rho n$ with nonzero entries $\pm\frac{1}{\sqrt{\rho}}$

Planted Vector Problem

$d \ll n$

Goal: recover a **structured** vector $v \in \mathbb{R}^n$ planted in a **random** d-dimensional subspace of $\mathbb{R}^n$

$\|v\| \approx \sqrt{n}$

- Structure: e.g. sparsity $\|v\|_0 = \rho n$ with nonzero entries $\pm\frac{1}{\sqrt{\rho}}$

Planted Vector Problem

$d \ll n$

Goal: recover a **structured** vector $v \in \mathbb{R}^n$ planted in a **random** d-dimensional subspace of $\mathbb{R}^n$

$\|v\| \approx \sqrt{n}$

- Structure: e.g. sparsity $\|v\|_0 = \rho n$ with nonzero entries $\pm\frac{1}{\sqrt{\rho}}$

Generic task in machine learning: related to dictionary learning, matrix sparsification, sparse PCA, ...

Planted Vector Problem

$d \ll n$

Goal: recover a **structured** vector $v \in \mathbb{R}^n$ planted in a **random** d-dimensional subspace of $\mathbb{R}^n$

$\|v\| \approx \sqrt{n}$

- Structure: e.g. sparsity $\|v\|_0 = \rho n$ with nonzero entries $\pm \frac{1}{\sqrt{\rho}}$

Generic task in machine learning: related to dictionary learning, matrix sparsification, sparse PCA, ...

Model: observe $Y = BR$

basis → B

rotation → R

Planted Vector Problem

$d \ll n$

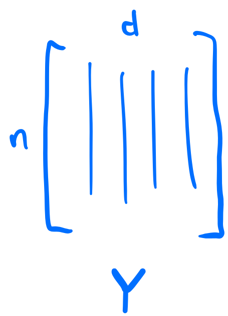
Goal: recover a **structured** vector $v \in \mathbb{R}^n$ planted in a **random** d-dimensional subspace of $\mathbb{R}^n$

- Structure: e.g. sparsity $\|v\|_0 = \rho n$ with nonzero entries $\pm\frac{1}{\sqrt{\rho}}$ ($\|v\| \approx \sqrt{n}$)

Generic task in machine learning: related to dictionary learning, matrix sparsification, sparse PCA, ...

Model: observe $Y = BR$

basis → B, rotation → R



Planted Vector Problem

d ≪ n

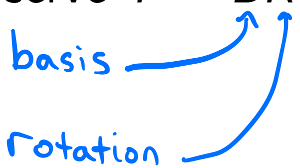
Goal: recover a **structured** vector $v \in \mathbb{R}^n$ planted in a **random** d-dimensional subspace of $\mathbb{R}^n$

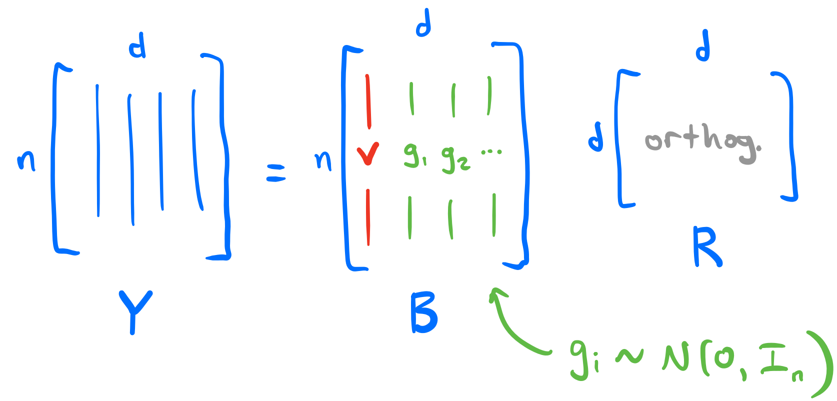
$\|v\| \approx \sqrt{n}$

- Structure: e.g. sparsity $\|v\|_0 = \rho n$ with nonzero entries $\pm \frac{1}{\sqrt{\rho}}$

Generic task in machine learning: related to dictionary learning, matrix sparsification, sparse PCA, ...

Model: observe $Y = BR$





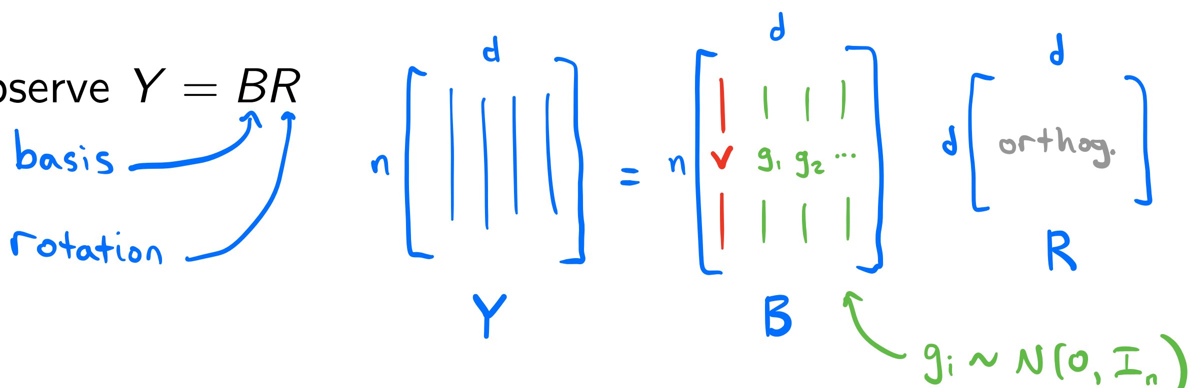
Planted Vector Problem

Goal: recover a **structured** vector $v \in \mathbb{R}^n$ planted in a **random** d-dimensional subspace of $\mathbb{R}^n$

- Structure: e.g. sparsity $\|v\|_0 = \rho n$ with nonzero entries $\pm\frac{1}{\sqrt{\rho}}$

Generic task in machine learning: related to dictionary learning, matrix sparsification, sparse PCA, ...

Model: observe $Y = BR$



(Slightly) harder variant: given any basis for the column span of B

Prior Work

Scaling regime: $n \to \infty$

$\rho = n^{-\alpha}$, $\alpha \in (0, 1)$ ← sparsity

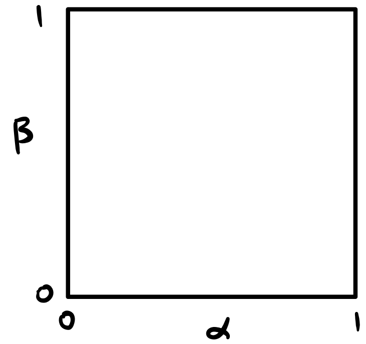
$d = n^{\beta}$, $\beta \in (0, 1)$ ← dimension

Prior Work

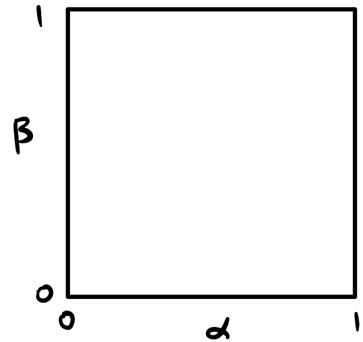
Scaling regime: $n \to \infty$

$\rho = n^{-\alpha}$, $\alpha \in (0, 1)$ ← sparsity

$d = n^{\beta}$, $\beta \in (0, 1)$ ← dimension



Prior Work



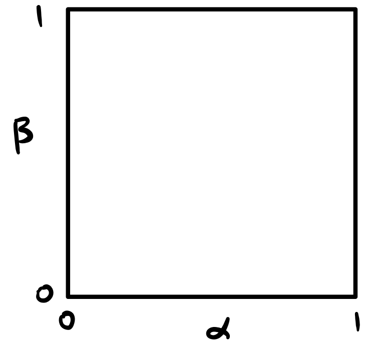
Scaling regime: $n \to \infty$

$\rho = n^{-\alpha}$, $\alpha \in (0, 1)$ ← sparsity

$d = n^{\beta}$, $\beta \in (0, 1)$ ← dimension

Information-theoretically possible (brute-force) [Qu, Sun, Wright '14]

Prior Work



Scaling regime: $n \to \infty$

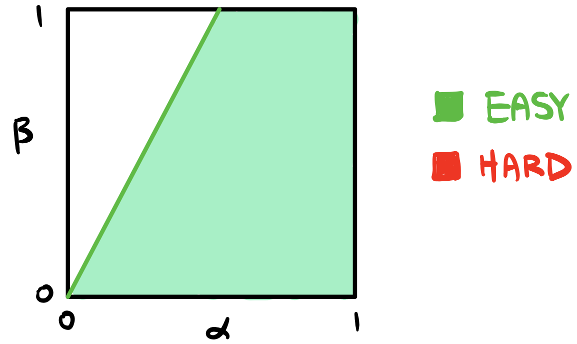
$\rho = n^{-\alpha}$, $\alpha \in (0, 1)$ ← sparsity

$d = n^{\beta}$, $\beta \in (0, 1)$ ← dimension

Information-theoretically possible (brute-force) [Qu, Sun, Wright '14]

Computationally (poly-time) feasible when:

Prior Work



Scaling regime: $n \to \infty$

$\rho = n^{-\alpha}$, $\alpha \in (0, 1)$ ← sparsity

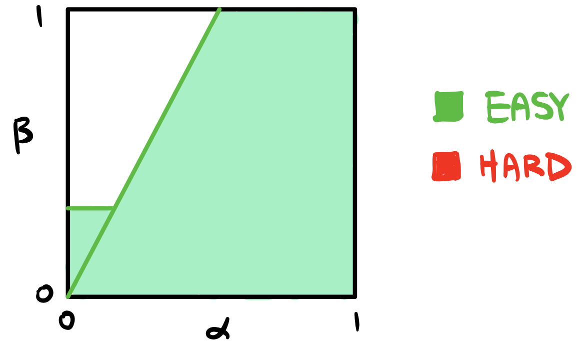
$d = n^{\beta}$, $\beta \in (0, 1)$ ← dimension

Information-theoretically possible (brute-force) [Qu, Sun, Wright '14]

Computationally (poly-time) feasible when:

- $\rho \ll 1/\sqrt{d}$ (linear programming) [Demanet, Hand '13]

Prior Work



Scaling regime: $n \to \infty$

$\rho = n^{-\alpha}$, $\alpha \in (0,1)$ ← sparsity

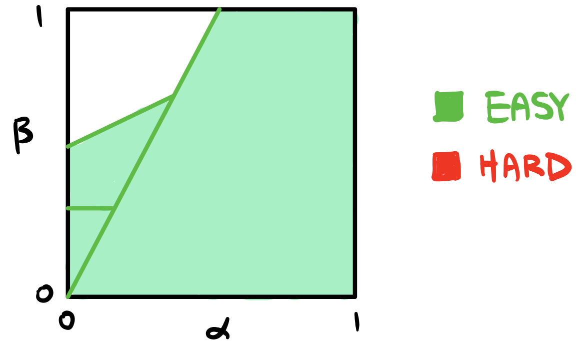
$d = n^{\beta}$, $\beta \in (0,1)$ ← dimension

Information-theoretically possible (brute-force) [Qu, Sun, Wright '14]

Computationally (poly-time) feasible when:

- $\rho \ll 1/\sqrt{d}$ (linear programming) [Demanet, Hand '13]
- $d \ll n^{1/4}$ (non-convex) [Qu, Sun, Wright '14]

Prior Work



Scaling regime: $n \to \infty$

$\rho = n^{-\alpha}$, $\alpha \in (0, 1)$ ← sparsity

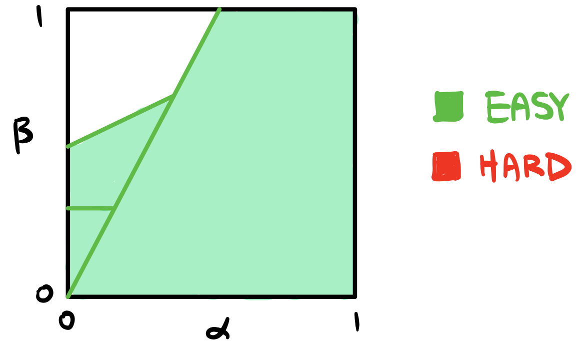
$d = n^{\beta}$, $\beta \in (0, 1)$ ← dimension

Information-theoretically possible (brute-force) [Qu, Sun, Wright '14]

Computationally (poly-time) feasible when:

- $\rho \ll 1/\sqrt{d}$ (linear programming) [Demanet, Hand '13]
- $d \ll n^{1/4}$ (non-convex) [Qu, Sun, Wright '14]
- $\sqrt{\rho} d \ll \sqrt{n}$ (sum-of-squares) [Barak, Kelner, Steurer '13]

Prior Work



Scaling regime: $n \to \infty$

$\rho = n^{-\alpha}$, $\alpha \in (0,1)$ ← sparsity

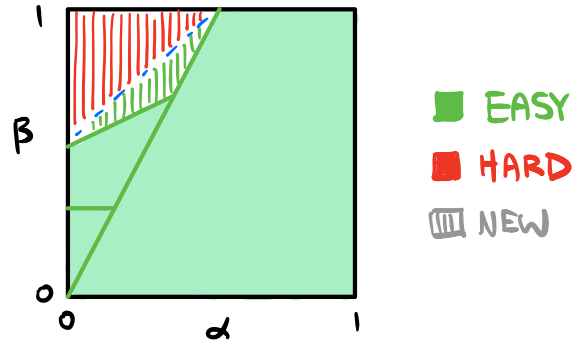
$d = n^{\beta}$, $\beta \in (0,1)$ ← dimension

Information-theoretically possible (brute-force) [Qu, Sun, Wright '14]

Computationally (poly-time) feasible when:

- $\rho \ll 1/\sqrt{d}$ (linear programming) [Demanet, Hand '13]
- $d \ll n^{1/4}$ (non-convex) [Qu, Sun, Wright '14]
- $\sqrt{\rho}d \ll \sqrt{n}$ (sum-of-squares) [Barak, Kelner, Steurer '13]
- $d \ll \sqrt{n}$ (spectral method) [Hopkins, Schramm, Shi, Steurer '15]

Prior Work



Scaling regime: $n \to \infty$

$\rho = n^{-\alpha}$, $\alpha \in (0,1)$ ← sparsity

$d = n^{\beta}$, $\beta \in (0,1)$ ← dimension

Information-theoretically possible (brute-force) [Qu, Sun, Wright '14]

Computationally (poly-time) feasible when:

- $\rho \ll 1/\sqrt{d}$ (linear programming) [Demanet, Hand '13]
- $d \ll n^{1/4}$ (non-convex) [Qu, Sun, Wright '14]
- $\sqrt{\rho} d \ll \sqrt{n}$ (sum-of-squares) [Barak, Kelner, Steurer '13]
- $d \ll \sqrt{n}$ (spectral method) [Hopkins, Schramm, Shi, Steurer '15]

Our contributions:

- Spectral method succeeds when $\rho d \ll \sqrt{n}$
- Evidence for computational hardness when $\rho d \gg \sqrt{n}$

A Helpful Reformulation

(P1) Observe $n \times d$ matrix $Y = BR$

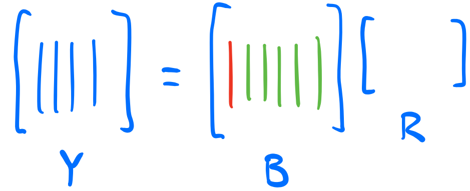
- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix

(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

Claim: P1 and P2 are equivalent (same distribution over Y)

A Helpful Reformulation



(P1) Observe $n \times d$ matrix $Y = BR$

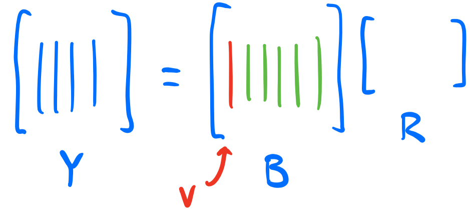
- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix

(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

Claim: P1 and P2 are equivalent (same distribution over Y)

A Helpful Reformulation



(P1) Observe $n \times d$ matrix $Y = BR$

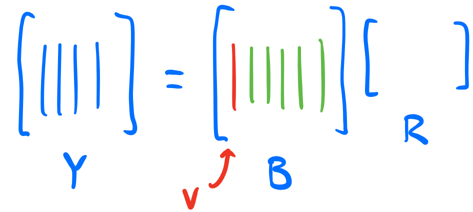
- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix

(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

Claim: P1 and P2 are equivalent (same distribution over Y)

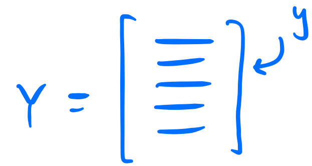
A Helpful Reformulation



(P1) Observe $n \times d$ matrix $Y = BR$

- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix

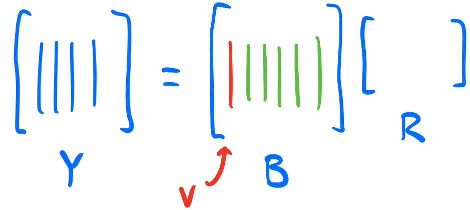
(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$



- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

Claim: P1 and P2 are equivalent (same distribution over Y)

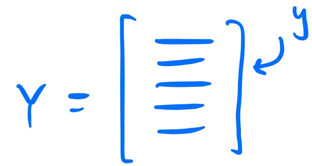
A Helpful Reformulation



(P1) Observe $n \times d$ matrix $Y = BR$

- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix

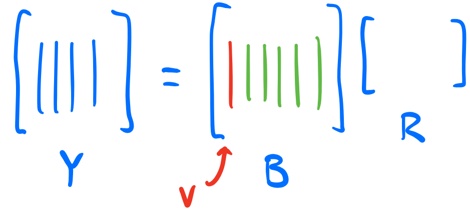
(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$



- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

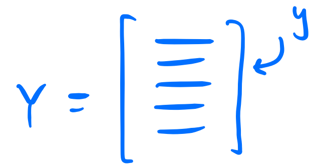
Claim: P1 and P2 are equivalent (same distribution over Y)

A Helpful Reformulation



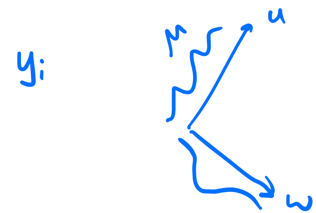
(P1) Observe $n \times d$ matrix $Y = BR$

- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix



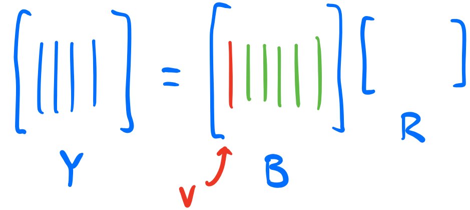
(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0,1)$ for all $w \perp u$, $\|w\| = 1$



Claim: P1 and P2 are equivalent (same distribution over Y)

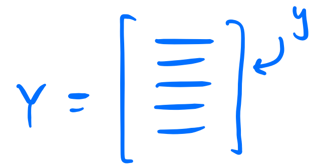
A Helpful Reformulation



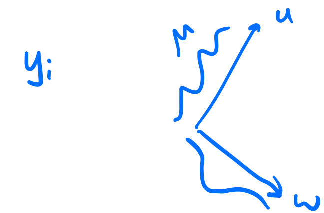
(P1) Observe $n \times d$ matrix $Y = BR$

- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix

(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$



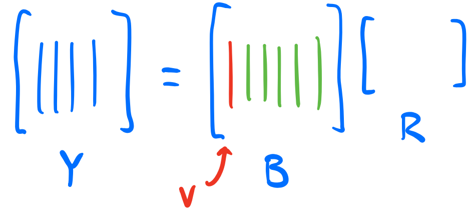
- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$



$y_i \sim N(v_i u, I_d - uu^T) \qquad v_i \sim \mu$

Claim: P1 and P2 are equivalent (same distribution over Y)

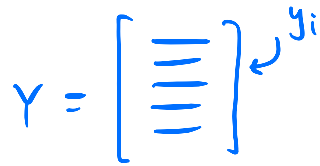
A Helpful Reformulation



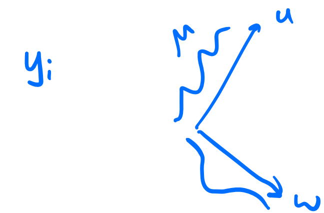
(P1) Observe $n \times d$ matrix $Y = BR$

- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix

(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$



- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

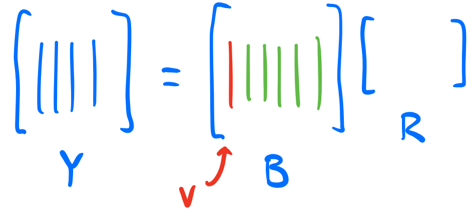


$y_i \sim N(v_i u, I_d - uu^T) \qquad v_i \sim \mu$

Claim: P1 and P2 are equivalent (same distribution over Y)

- Proof: $u^\top$ corresponds to the first row of R
- $Yu = v$

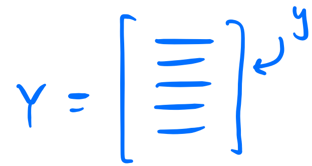
A Helpful Reformulation



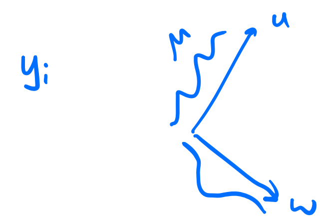
(P1) Observe $n \times d$ matrix $Y = BR$

- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix

(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$



- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

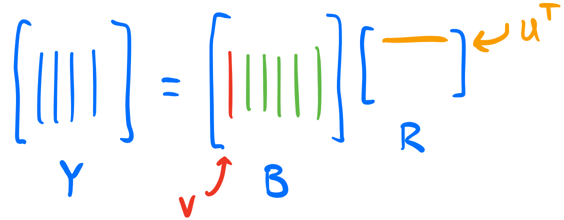


$y_i \sim N(v_i u, I_d - uu^T) \qquad v_i \sim \mu$

Claim: P1 and P2 are equivalent (same distribution over Y)

- Proof: $u^\top$ corresponds to the first row of R
- $Yu = v$

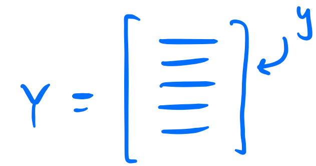
A Helpful Reformulation



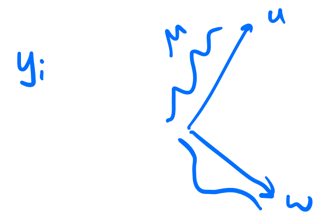
(P1) Observe $n \times d$ matrix $Y = BR$

- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix

(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$



- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

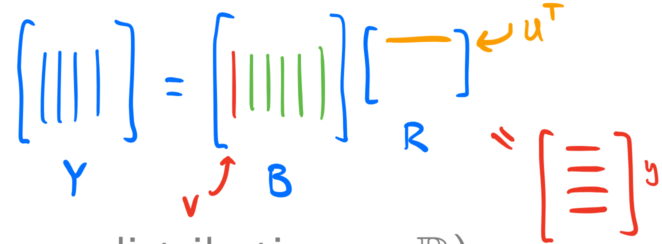


$y_i \sim N(v_i u, I_d - uu^T)$ $\quad v_i \sim \mu$

Claim: P1 and P2 are equivalent (same distribution over Y)

- Proof: $u^\top$ corresponds to the first row of R
- $Yu = v$

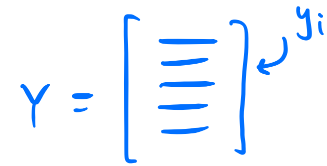
A Helpful Reformulation



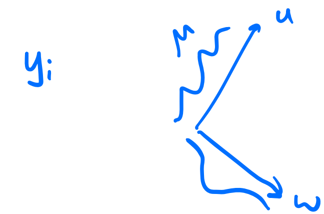
(P1) Observe $n \times d$ matrix $Y = BR$

- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix

(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$



- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

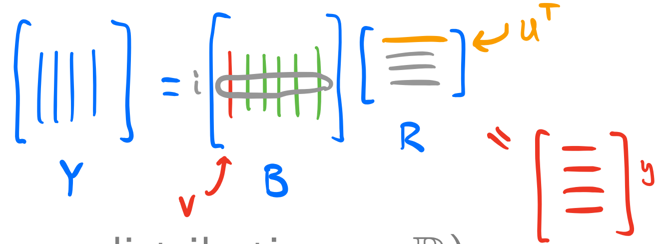


★ $y_i \sim N(v_i u, I_d - uu^T)$ $v_i \sim \mu$

Claim: P1 and P2 are equivalent (same distribution over Y)

- Proof: $u^\top$ corresponds to the first row of R
- $Yu = v$

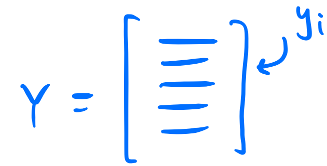
A Helpful Reformulation



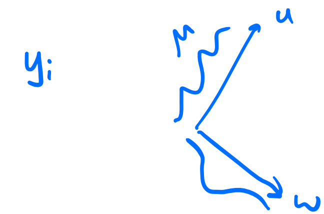
(P1) Observe $n \times d$ matrix $Y = BR$

- v has i.i.d. entries drawn from μ (some distribution on $\mathbb{R}$)
- R is a random orthogonal matrix

(P2) Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$



- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ is independent (given u) with distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle = v_i \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$



★ $y_i \sim N(v_i u, I_d - uu^T)$ $v_i \sim \mu$

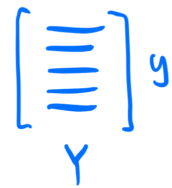
Claim: P1 and P2 are equivalent (same distribution over Y)

- Proof: $u^\top$ corresponds to the first row of R
- $Yu = v$

Upper Bound (Algorithm)

Algorithm [Hopkins, Schramm, Shi, Steurer '15]: leading eigenvector of $d \times d$ matrix

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d \right) y_i y_i^\top - 3n\, I_d$$

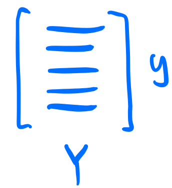


Hope: leading eigenvector $\approx u$

Upper Bound (Algorithm)

Algorithm [Hopkins, Schramm, Shi, Steurer '15]: leading eigenvector of $d \times d$ matrix

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d \right) y_i y_i^\top - 3n \, I_d$$

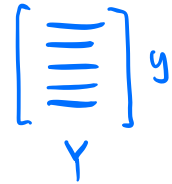


Hope: leading eigenvector $\approx u$

Upper Bound (Algorithm)

Algorithm [Hopkins, Schramm, Shi, Steurer '15]: leading eigenvector of $d \times d$ matrix

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d\right) y_i y_i^\top - 3n\, I_d$$

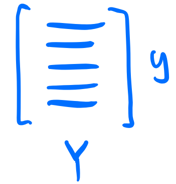


Hope: leading eigenvector $\approx u$ $v = Yu$

Upper Bound (Algorithm)

Algorithm [Hopkins, Schramm, Shi, Steurer '15]: leading eigenvector of $d \times d$ matrix

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d\right) y_i y_i^\top - 3n\, I_d$$



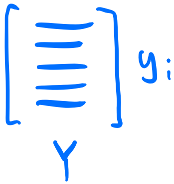
Hope: leading eigenvector $\approx u$ $\quad v = Yu$

Our improvements to analysis:

Upper Bound (Algorithm)

Algorithm [Hopkins, Schramm, Shi, Steurer '15]: leading eigenvector of $d \times d$ matrix

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d \right) y_i y_i^\top - 3n\, I_d$$



Hope: leading eigenvector $\approx u$ $\quad v = Yu$

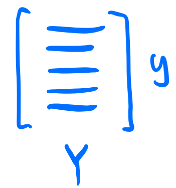
Our improvements to analysis:

- Improved condition $d \ll \sqrt{n} \quad \to \quad \rho d \ll \sqrt{n}$

Upper Bound (Algorithm)

Algorithm [Hopkins, Schramm, Shi, Steurer '15]: leading eigenvector of $d \times d$ matrix

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d \right) y_i y_i^\top - 3n \, I_d$$



Hope: leading eigenvector $\approx u$ $v = Yu$

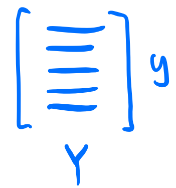
Our improvements to analysis:

- Improved condition $d \ll \sqrt{n} \quad \to \quad \rho d \ll \sqrt{n}$
- Guarantees in ℓ_∞ norm instead of ℓ_2
  - Implies exact recovery of v
  - Proof: "leave-one-out" analysis

Upper Bound (Algorithm)

Algorithm [Hopkins, Schramm, Shi, Steurer '15]: leading eigenvector of $d \times d$ matrix

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d \right) y_i y_i^\top - 3n\, I_d$$



Hope: leading eigenvector $\approx u$ $\quad$ $v = Yu$

Our improvements to analysis:

- Improved condition $d \ll \sqrt{n} \quad \to \quad \rho d \ll \sqrt{n}$
- Guarantees in ℓ_∞ norm instead of ℓ_2
  - Implies exact recovery of v
  - Proof: "leave-one-out" analysis
- Covers dense case $\rho = 1$ (planted ± 1 vector)
  - Spectral method (bottom eigenvector) succeeds when $d \ll \sqrt{n}$
  - (and hard when $d \gg \sqrt{n}$)

Upper Bound: Proof Idea

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d\right) y_i y_i^\top - 3n\, I_d$$

Upper Bound: Proof Idea

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d\right) y_i y_i^\top - 3n\, I_d$$

"Signal part": $\mathbb{E}[M] \approx \left(\|v\|_4^4 - 3n\right) uu^\top$

- Note: $g \sim \mathcal{N}(0, I_n)$ has $\|g\|_4^4 \approx 3n$

Upper Bound: Proof Idea

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d\right) y_i y_i^\top - 3n\, I_d$$

"Signal part": $\mathbb{E}[M] \approx \left(\|v\|_4^4 - 3n\right) uu^\top$

- Note: $g \sim \mathcal{N}(0, I_n)$ has $\|g\|_4^4 \approx 3n$

"Noise part": $\|M - \mathbb{E}[M]\| \lesssim \cdots$

- "Centering" $\|y_i\|^2 - d$ is important here

Upper Bound: Proof Idea

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d\right) y_i y_i^\top - 3n\, I_d$$

"Signal part": $\mathbb{E}[M] \approx \left(\|v\|_4^4 - 3n\right) uu^\top$

- Note: $g \sim \mathcal{N}(0, I_n)$ has $\|g\|_4^4 \approx 3n$

"Noise part": $\|M - \mathbb{E}[M]\| \lesssim \cdots$

- "Centering" $\|y_i\|^2 - d$ is important here

Result: spectral method succeeds when

$$\left|\|v\|_4^4 - 3n\right| \geq \epsilon n \quad \text{and} \quad \frac{d\sqrt{n}}{\|v\|_4^4} + \frac{\sqrt{d}\|v\|_\infty}{\|v\|_4^2} \ll 1$$

Upper Bound: Proof Idea

$$M = \sum_{i=1}^{n} \left(\|y_i\|^2 - d\right) y_i y_i^\top - 3n\, I_d$$

"Signal part": $\mathbb{E}[M] \approx \left(\|v\|_4^4 - 3n\right) uu^\top$

- Note: $g \sim \mathcal{N}(0, I_n)$ has $\|g\|_4^4 \approx 3n$

"Noise part": $\|M - \mathbb{E}[M]\| \lesssim \cdots$

- "Centering" $\|y_i\|^2 - d$ is important here

Result: spectral method succeeds when

$$\left|\|v\|_4^4 - 3n\right| \geq \epsilon n \quad \text{and} \quad \frac{d\sqrt{n}}{\|v\|_4^4} + \frac{\sqrt{d}\|v\|_\infty}{\|v\|_4^2} \ll 1$$

Note: for ρ-sparse Rademacher v: $\|v\|_4^4 \approx \frac{n}{\rho}$

$$\left|\frac{1}{\rho} - 3\right| \geq \epsilon \quad \text{and} \quad \rho d \ll \sqrt{n}$$

Lower Bound (Hardness)

Sufficient to show hardness of detection/testing problem:

Lower Bound (Hardness)

Sufficient to show hardness of detection/testing problem:

$\mathbb{P}$ (planted): Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ has distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

Lower Bound (Hardness)

Sufficient to show hardness of detection/testing problem:

$\mathbb{P}$ (planted): Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ has distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

$\mathbb{Q}$ (null): Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw each $y_i \sim \mathcal{N}(0, I_d)$

Lower Bound (Hardness)

Sufficient to show hardness of detection/testing problem:

$\mathbb{P}$ (planted): Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ has distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

$\mathbb{Q}$ (null): Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw each $y_i \sim \mathcal{N}(0, I_d)$

Goal: distinguish between $\mathbb{P}$ and $\mathbb{Q}$ with high probability, given a single Y

Lower Bound (Hardness)

Sufficient to show hardness of detection/testing problem:

$\mathbb{P}$ (planted): Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw random unit vector $u \in \mathbb{R}^d$
- Each $y_i \in \mathbb{R}^d$ has distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

$\mathbb{Q}$ (null): Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw each $y_i \sim \mathcal{N}(0, I_d)$

Goal: distinguish between $\mathbb{P}$ and $\mathbb{Q}$ with high probability, given a single Y

Hardness of detection implies hardness of recovery
(poly-time reduction)

Low-Degree Spectral Methods

Recall spectral method:

$$M = M(Y) = \sum_{i=1}^{n} \left(\|y_i\|^2 - d \right) y_i y_i^\top - 3n\, I_d$$

Low-Degree Spectral Methods

Recall spectral method:

$$M = M(Y) = \sum_{i=1}^{n} \left(\|y_i\|^2 - d \right) y_i y_i^\top - 3n\, I_d$$

- $d \times d$ matrix where each entry is a degree-4 polynomial in the entries of Y

Low-Degree Spectral Methods

Recall spectral method:

$$M = M(Y) = \sum_{i=1}^{n} \left(\|y_i\|^2 - d\right) y_i y_i^\top - 3n\, I_d$$

- $d \times d$ matrix where each entry is a degree-4 polynomial in the entries of Y
- Our analysis shows that when $\rho d \ll \sqrt{n}$, can distinguish $\mathbb{P}$ from $\mathbb{Q}$ by thresholding $\|M(Y)\|$

"easy"

Low-Degree Spectral Methods

Recall spectral method:

$$M = M(Y) = \sum_{i=1}^{n} \left(\|y_i\|^2 - d \right) y_i y_i^\top - 3n\, I_d$$

- $d \times d$ matrix where each entry is a degree-4 polynomial in the entries of Y
- Our analysis shows that when $\rho d \ll \sqrt{n}$, can distinguish $\mathbb{P}$ from $\mathbb{Q}$ by thresholding $\|M(Y)\|$ ← "easy"

Might there be a better spectral method?

Low-Degree Spectral Methods

Recall spectral method:

$$M = M(Y) = \sum_{i=1}^{n} \left(\|y_i\|^2 - d\right) y_i y_i^\top - 3n\, I_d$$

- $d \times d$ matrix where each entry is a degree-4 polynomial in the entries of Y
- Our analysis shows that when $\rho d \ll \sqrt{n}$, can distinguish $\mathbb{P}$ from $\mathbb{Q}$ by thresholding $\|M(Y)\|$ ← "easy"

Might there be a better spectral method?

- Could use different polynomials, maybe of higher degree...

Low-Degree Spectral Methods

Recall spectral method:

$$M = M(Y) = \sum_{i=1}^{n} \left(\|y_i\|^2 - d\right) y_i y_i^\top - 3n\, I_d$$

- $d \times d$ matrix where each entry is a degree-4 polynomial in the entries of Y
- Our analysis shows that when $\rho d \ll \sqrt{n}$, can distinguish $\mathbb{P}$ from $\mathbb{Q}$ by thresholding $\|M(Y)\|$ ↖ "easy"

Might there be a better spectral method?

- Could use different polynomials, maybe of higher degree...

We answer this in the negative: any poly-size spectral method with constant-degree entries cannot distinguish $\mathbb{P}, \mathbb{Q}$ when $\rho d \gg \sqrt{n}$

Limits of Spectral Methods

Theorem: **(i) (Easy regime)** If $\rho d \ll \sqrt{n}$ there exists a $d \times d$ degree-4 matrix M and threshold $\tau > 0$ such that

$$\mathbb{P}(\|M\| \geq 2\tau) \geq 1 - n^{-\omega(1)},$$

$$\mathbb{Q}(\|M\| \leq \tau) \geq 1 - n^{-\omega(1)}.$$

(ii) (Hard regime) If $\rho d \gg \sqrt{n}$ then for any constants $\ell, k, \epsilon > 0$, there is no $n^{\ell} \times n^{\ell}$ degree-k symmetric matrix M and threshold $\tau > 0$ such that

$$\mathbb{P}(\|M\| \geq (1+\epsilon)\tau) \geq 1 - \frac{\epsilon}{4},$$

$$\mathbb{Q}(\|M\| \leq \tau) \geq 1 - n^{-C},$$

for a constant $C = C(\ell, k, \epsilon)$.

So $\rho d \approx \sqrt{n}$ is the precise threshold for spectral methods; suggests a fundamental barrier

Proof: Failure of Spectral Methods

Two steps to show that all spectral methods fail when $\rho d \gg \sqrt{n}$:

Proof: Failure of Spectral Methods

Two steps to show that all spectral methods fail when $\rho d \gg \sqrt{n}$:

(I) Any spectral method $M(Y)$ can be approximated by a low-degree polynomial $f : \mathbb{R}^{nd} \to \mathbb{R}$:

$$f(Y) = \mathrm{Tr}(M^{2q}) = \sum_i \lambda_i^{2q} \approx \lambda_{\max}^{2q} \qquad \text{for } q \approx \log n$$

Degree of f is $2qk \approx \log n$ (where k is degree of each entry of M)

Proof: Failure of Spectral Methods

Two steps to show that all spectral methods fail when $\rho d \gg \sqrt{n}$:

(I) Any spectral method $M(Y)$ can be approximated by a low-degree polynomial $f : \mathbb{R}^{nd} \to \mathbb{R}$:

$$f(Y) = \mathrm{Tr}(M^{2q}) = \sum_i \lambda_i^{2q} \approx \lambda_{\max}^{2q} \qquad \text{for } q \approx \log n$$

Degree of f is $2qk \approx \log n$ (where k is degree of each entry of M)

(II) If $\rho d \gg \sqrt{n}$, any polynomial $f : \mathbb{R}^{nd} \to \mathbb{R}$ of degree $D = O(\log n)$ fails at detection:

$$\mathrm{Adv}_{\leq D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y\sim\mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y\sim\mathbb{Q}}[f(Y)^2]}} = O(1)$$

Low-Degree Polynomial Lower Bounds

$$\mathrm{Adv}_{\le D} := \max_{f \text{ deg } D} \frac{\mathbb{E}_{Y \sim \mathbb{P}}[f(Y)]}{\sqrt{\mathbb{E}_{Y \sim \mathbb{Q}}[f(Y)^2]}} = O(1)$$

(Also called $\|L^{\le D}\|$ or $\sqrt{\chi^2_{\le D}(\mathbb{P}\|\mathbb{Q}) + 1}$)

Follows a long line of work on low-degree polynomial lower bounds:
[Barak, Hopkins, Kelner, Kothari, Moitra, Potechin '16] [Hopkins, Steurer '17]
[Hopkins, Kothari, Potechin, Raghavendra, Schramm, Steurer '17]
[Hopkins '18 (PhD thesis)] [Kunisky, W., Bandeira '19 (survey)] ...

Similar low-degree lower bounds for many problems:
planted clique (and variants), sparse PCA, community detection, tensor PCA, planted CSPs, spiked Wigner/Wishart matrix, sparse clustering, planted submatrix, planted dense subgraph, p-spin optimization, max independent set, ...

Low-degree polynomials provide a unified explanation for why all these problems are hard in the (conjectured) "hard" regime

A Key Lemma

General analysis of “planted non-gaussian direction” problems

$\mathbb{P}$: Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw random unit vector $u \sim \mathcal{U}$ (some distribution)
- Each $y_i \in \mathbb{R}^d$ has distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

$\mathbb{Q}$: Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw each $y_i \sim \mathcal{N}(0, I_d)$

A Key Lemma

General analysis of “planted non-gaussian direction” problems

$\mathbb{P}$: Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw random unit vector $u \sim \mathcal{U}$ (some distribution)
- Each $y_i \in \mathbb{R}^d$ has distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

$\mathbb{Q}$: Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw each $y_i \sim \mathcal{N}(0, I_d)$

A Key Lemma

General analysis of "planted non-gaussian direction" problems

$\mathbb{P}$: Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw random unit vector $u \sim \mathcal{U}$ (some distribution)
- Each $y_i \in \mathbb{R}^d$ has distribution μ in direction u and otherwise gaussian:
  - $\langle y_i, u \rangle \sim \mu$
  - $\langle y_i, w \rangle \sim \mathcal{N}(0, 1)$ for all $w \perp u$, $\|w\| = 1$

$\mathbb{Q}$: Observe $n \times d$ matrix Y with rows $y_1, \ldots, y_n$

- Draw each $y_i \sim \mathcal{N}(0, I_d)$

Lemma: for any μ (with finite moments) and $\mathcal{U}$,

$$\mathrm{Adv}_{\leq D}^2 = \sum_{k=0}^{D} \mathop{\mathbb{E}}_{u, u' \sim \mathcal{U}} [\langle u, u' \rangle^k] \sum_{\substack{\alpha \in \mathbb{N}^n \\ |\alpha| = k}} \prod_{i=1}^{n} \left(\mathop{\mathbb{E}}_{x \sim \mu} [h_{\alpha_i}(x)] \right)^2$$

Summary

Summary

- Planted ρ-sparse vector in a d-dimensional subspace of $\mathbb{R}^n$

Summary

- Planted ρ-sparse vector in a d-dimensional subspace of $\mathbb{R}^n$
- Reformulation: planted distribution μ in direction u

Summary

- Planted ρ-sparse vector in a d-dimensional subspace of $\mathbb{R}^n$
- Reformulation: planted distribution μ in direction u
- Improved analysis of [HSSS'15] spectral algorithm: succeeds when $\rho d \ll \sqrt{n}$

Summary

- Planted ρ-sparse vector in a d-dimensional subspace of $\mathbb{R}^n$
- Reformulation: planted distribution μ in direction u
- Improved analysis of [HSSS'15] spectral algorithm: succeeds when $\rho d \ll \sqrt{n}$
  - More generally, depends on $\|v\|_4^4 - 3n$

Summary

- Planted ρ-sparse vector in a d-dimensional subspace of $\mathbb{R}^n$
- Reformulation: planted distribution μ in direction u
- Improved analysis of [HSSS'15] spectral algorithm: succeeds when $\rho d \ll \sqrt{n}$
  - More generally, depends on $\|v\|_4^4 - 3n$
- Matching lower bound: no spectral method (or low-degree polynomial) can succeed when $\rho d \gg \sqrt{n}$

Summary

- Planted ρ-sparse vector in a d-dimensional subspace of $\mathbb{R}^n$
- Reformulation: planted distribution μ in direction u
- Improved analysis of [HSSS'15] spectral algorithm: succeeds when $\rho d \ll \sqrt{n}$
  - More generally, depends on $\|v\|_4^4 - 3n$
- Matching lower bound: no spectral method (or low-degree polynomial) can succeed when $\rho d \gg \sqrt{n}$
  - "Low-degree barrier" similar to planted clique, sparse PCA, etc.

Summary

- Planted ρ-sparse vector in a d-dimensional subspace of $\mathbb{R}^n$
- Reformulation: planted distribution μ in direction u
- Improved analysis of [HSSS'15] spectral algorithm: succeeds when $\rho d \ll \sqrt{n}$
  - More generally, depends on $\|v\|_4^4 - 3n$
- Matching lower bound: no spectral method (or low-degree polynomial) can succeed when $\rho d \gg \sqrt{n}$
  - "Low-degree barrier" similar to planted clique, sparse PCA, etc.
- Open: hardness via reduction from planted clique?

Summary

- Planted ρ-sparse vector in a d-dimensional subspace of $\mathbb{R}^n$
- Reformulation: planted distribution μ in direction u
- Improved analysis of [HSSS'15] spectral algorithm: succeeds when $\rho d \ll \sqrt{n}$
  - More generally, depends on $\|v\|_4^4 - 3n$
- Matching lower bound: no spectral method (or low-degree polynomial) can succeed when $\rho d \gg \sqrt{n}$
  - "Low-degree barrier" similar to planted clique, sparse PCA, etc.
- Open: hardness via reduction from planted clique?

Thanks!