Hardness of Certification for Random Optimization Problems

Tim Kunisky

(joint work with Afonso Bandeira and Alex Wein)

Courant Institute of Mathematical Sciences

March 21, 2019

Our Question:

How tight can certifiable bounds on random optimization problems be while remaining computationally tractable?

Random Optimization—A Model Problem

$\boxed{\text{Maximizing a gaussian quadratic form over the hypercube}}$

Random Optimization—A Model Problem

$\boxed{\text{Maximizing a gaussian quadratic form over the hypercube}}$

1. Build random data:

$$\boldsymbol{W} \sim \mathrm{GOE}(n)$$

$$(\text{meaning } \boldsymbol{W} \in \mathbb{R}^{n\times n}_{\mathrm{sym}}, W_{ij} \overset{(\perp)}{\sim} \mathcal{N}(0, \tfrac{1+\delta_{ij}}{n}) \text{ for } i \leq j)$$

Random Optimization—A Model Problem

Maximizing a gaussian quadratic form over the hypercube

1. Build random data:

$$\boldsymbol{W} \sim \mathrm{GOE}(n)$$

$$\left(\text{meaning } \boldsymbol{W} \in \mathbb{R}^{n\times n}_{\mathrm{sym}}, W_{ij} \overset{(\perp)}{\sim} \mathcal{N}\left(0, \tfrac{1+\delta_{ij}}{n}\right) \text{ for } i \le j\right)$$

2. Set an optimization task:

$$\mathrm{OPT}(\boldsymbol{W}) = \left\{ \begin{array}{ll} \text{maximize} & f_{\boldsymbol{W}}(\boldsymbol{x}) := \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x} \\ \text{subject to} & \boldsymbol{x} \in \{\pm 1/\sqrt{n}\}^n \end{array} \right\}$$

Random Optimization—A Model Problem

Maximizing a gaussian quadratic form over the hypercube

1. Build random data:

$$\boldsymbol{W} \sim \mathrm{GOE}(n)$$

$$\left(\text{meaning } \boldsymbol{W} \in \mathbb{R}^{n\times n}_{\text{sym}}, W_{ij} \stackrel{(\perp)}{\sim} \mathcal{N}\left(0, \tfrac{1+\delta_{ij}}{n}\right) \text{ for } i \le j\right)$$

2. Set an optimization task:

$$\mathrm{OPT}(\boldsymbol{W}) = \left\{ \begin{array}{ll} \text{maximize} & f_W(\boldsymbol{x}) := \boldsymbol{x}^\top \boldsymbol{W} \boldsymbol{x} \\ \text{subject to} & \boldsymbol{x} \in \{\pm 1/\sqrt{n}\}^n \end{array} \right\}$$

Why this problem? $-f_W$ is the Hamiltonian and $-\mathrm{OPT}(\boldsymbol{W})$ is the ground state energy of the Sherrington-Kirkpatrick spin glass model $\rightsquigarrow$ well-studied in statistical physics.

Random Optimization—The True Value

Random Optimization—The True Value

Physicists in the '70s and '80s developed a deep theory of the structure of the optimization landscape of $f_{\boldsymbol{W}}$. One of the results was:

$$\lim_{n \to \infty} \mathbb{E}_{\boldsymbol{W} \sim \mathrm{GOE}(n)} \mathrm{OPT}(\boldsymbol{W}) =: 2\mathrm{P}_* \approx 1.526.^{(*)}$$

$^{(*)}$ General gaussian process theory $\rightsquigarrow$ strong concentration.

Random Optimization—The True Value

Physicists in the '70s and '80s developed a deep theory of the structure of the optimization landscape of $f_{\boldsymbol{W}}$. One of the results was:

$$\lim_{n \to \infty} \mathbb{E}_{\boldsymbol{W} \sim \mathrm{GOE}(n)} \mathrm{OPT}(\boldsymbol{W}) =: 2\mathrm{P}_* \approx 1.526.^{(*)}$$

P_* is determined as the limit of the optimal values of a sequence of functional optimization problems over probability distributions on $[0, 1]$.

[Parisi '79-80; Guerra, Talagrand, Panchenko, et al. '00s]

$^{(*)}$ General gaussian process theory $\rightsquigarrow$ strong concentration.

Two Fundamental Algorithmic Questions

Two Fundamental Algorithmic Questions

Question 1 (Search): How large can you make $f_{\boldsymbol{W}}(\boldsymbol{x}^{\mathrm{alg}}(\boldsymbol{W}))$ for an efficiently computable $\boldsymbol{x}^{\mathrm{alg}}(\boldsymbol{W}) \in \{\pm 1/\sqrt{n}\}^n$?

Two Fundamental Algorithmic Questions

Question 1 (Search): How large can you make $f_{\boldsymbol{W}}(\boldsymbol{x}^{\mathrm{alg}}(\boldsymbol{W}))$ for an efficiently computable $\boldsymbol{x}^{\mathrm{alg}}(\boldsymbol{W}) \in \{\pm 1/\sqrt{n}\}^n$?

Answer: For any $\epsilon > 0$, there is $\boldsymbol{x}_{\epsilon}^{\mathrm{alg}}(\boldsymbol{W})$ computable in time $\mathrm{poly}_{\epsilon}(n)$ such that

$$\Pr_{\boldsymbol{W}\sim \mathrm{GOE}(n)}\left[f_{\boldsymbol{W}}(\boldsymbol{x}_{\epsilon}^{\mathrm{alg}}(\boldsymbol{W})) \geq \underbrace{2\mathrm{P}_{*}}_{\mathrm{OPT}(\boldsymbol{W})} - \epsilon \right] \to 1.$$

[Montanari '18; Subag '18; Addario-Berry, Maillard '18]

Two Fundamental Algorithmic Questions

Question 2 (Certification): How small can you make the typical value of $c(\boldsymbol{W})$ for c that is efficiently computable and satisfies $\mathrm{OPT}(\boldsymbol{A}) \leq c(\boldsymbol{A})$ *for all* $\boldsymbol{A} \in \mathbb{R}^{n\times n}_{\mathrm{sym}}$ (a *certificate*)?

Two Fundamental Algorithmic Questions

Question 2 (Certification): How small can you make the typical value of $c(\boldsymbol{W})$ for c that is efficiently computable and satisfies $\mathrm{OPT}(\boldsymbol{A}) \leq c(\boldsymbol{A})$ *for all* $\boldsymbol{A} \in \mathbb{R}^{n \times n}_{\mathrm{sym}}$ (a *certificate*)?

Example: $c(\boldsymbol{A}) := \lambda_{\max}(\boldsymbol{A})$ works, and for any $\epsilon > 0$,

$$\Pr_{\boldsymbol{W} \sim \mathrm{GOE}(n)}\left[2 - \epsilon \leq \lambda_{\max}(\boldsymbol{W}) \leq 2 + \epsilon\right] \to 1.$$

Two Fundamental Algorithmic Questions

Question 2 (Certification): How small can you make the typical value of $c(\boldsymbol{W})$ for c that is efficiently computable and satisfies $\mathrm{OPT}(\boldsymbol{A}) \leq c(\boldsymbol{A})$ *for all* $\boldsymbol{A} \in \mathbb{R}^{n \times n}_{\mathrm{sym}}$ (a *certificate*)?

Example: $c(\boldsymbol{A}) := \lambda_{\max}(\boldsymbol{A})$ works, and for any $\epsilon > 0$,

$$\Pr_{\boldsymbol{W} \sim \mathrm{GOE}(n)}\left[2 - \epsilon \leq \lambda_{\max}(\boldsymbol{W}) \leq 2 + \epsilon\right] \to 1.$$

Answer: Assuming a complexity theory conjecture, for any $\epsilon > 0$, there is *no* certificate $c(\boldsymbol{A})$ that is computable in time poly(n) and that satisfies

$$\Pr_{\boldsymbol{W} \sim \mathrm{GOE}(n)}\left[c(\boldsymbol{W}) \leq 2 - \epsilon\right] \to 1.$$

[Bandeira, K., Wein '19]

Example: Relaxation Algorithms

Example: Relaxation Algorithms

To formulate relaxations, first *linearize*. Recall the *cut polytope*:

$$\mathscr{C}^n = \text{ convex hull of } \{\boldsymbol{x}\boldsymbol{x}^\top : \boldsymbol{x} \in \{\pm 1/\sqrt{n}\}^n\} \subset \mathbb{R}^{n\times n}_{\mathrm{sym}}$$

Example: Relaxation Algorithms

To formulate relaxations, first *linearize*. Recall the *cut polytope*:

$$\mathscr{C}^n = \text{ convex hull of } \{\boldsymbol{x}\boldsymbol{x}^\top : \boldsymbol{x} \in \{\pm 1/\sqrt{n}\}^n\} \subset \mathbb{R}^{n\times n}_{\mathrm{sym}}$$

Computing $\mathrm{OPT}(\boldsymbol{A}) \leftrightarrow$ linear optimization over $\mathscr{C}^n$:

$$\begin{aligned}\mathrm{OPT}(\boldsymbol{A}) &= \left\{\begin{array}{ll} \text{maximize} & \boldsymbol{x}^\top \boldsymbol{A}\boldsymbol{x} \\ \text{subject to} & \boldsymbol{x} \in \{\pm 1/\sqrt{n}\}^n \end{array}\right\} \\ &= \left\{\begin{array}{ll} \text{maximize} & \langle \boldsymbol{A}, \boldsymbol{X}\rangle \\ \text{subject to} & \boldsymbol{X} \in \mathscr{C}^n \end{array}\right\}.\end{aligned}$$

(Though it is convex, the intricate discrete geometry of $\mathscr{C}^n$ makes this problem hard in general.)

Example: Relaxation Algorithms

Typically, certify by choosing $\mathscr{R}^n \supseteq \mathscr{C}^n$ and computing

$$c(\boldsymbol{A}) = \left\{ \begin{array}{ll} \text{maximize} & \langle \boldsymbol{A}, \boldsymbol{X} \rangle \\ \text{subject to} & \boldsymbol{X} \in \mathscr{R}^n \end{array} \right\} \geq \mathrm{OPT}(\boldsymbol{A}).$$

Example: Relaxation Algorithms

Typically, certify by choosing $\mathscr{R}^n \supseteq \mathscr{C}^n$ and computing

$$c(\boldsymbol{A}) = \left\{ \begin{array}{ll} \text{maximize} & \langle \boldsymbol{A}, \boldsymbol{X} \rangle \\ \text{subject to} & \boldsymbol{X} \in \mathscr{R}^n \end{array} \right\} \geq \mathrm{OPT}(\boldsymbol{A}).$$

Semidefinite programming examples:

$c(\boldsymbol{A})$	$\mathscr{R}^n$	$\mathbb{E}_{\boldsymbol{W} \sim \mathrm{GOE}(n)} c(\boldsymbol{W})$
$\lambda_{\max}(\boldsymbol{A})$	$\{\boldsymbol{X} \succeq \boldsymbol{0}, \mathrm{Tr}(\boldsymbol{X}) = 1\}$	$2 + o(1)$

Example: Relaxation Algorithms

Typically, certify by choosing $\mathscr{R}^n \supseteq \mathscr{C}^n$ and computing

$$c(\boldsymbol{A}) = \left\{ \begin{array}{ll} \text{maximize} & \langle \boldsymbol{A}, \boldsymbol{X} \rangle \\ \text{subject to} & \boldsymbol{X} \in \mathscr{R}^n \end{array} \right\} \geq \mathrm{OPT}(\boldsymbol{A}).$$

Semidefinite programming examples:

$c(\boldsymbol{A})$	$\mathscr{R}^n$	$\mathbb{E}_{\boldsymbol{W} \sim \mathrm{GOE}(n)} c(\boldsymbol{W})$
$\lambda_{\max}(\boldsymbol{A})$	$\{\boldsymbol{X} \succeq \boldsymbol{0}, \mathrm{Tr}(\boldsymbol{X}) = 1\}$	$2 + o(1)$
SOS Degree 2	$\{\boldsymbol{X} \succeq \boldsymbol{0}, X_{ii} = 1/n\}$	$2 + o(1)^{(*)}$

$^{(*)}$ *[Montanari, Sen '15]*

Example: Relaxation Algorithms

Typically, certify by choosing $\mathscr{R}^n \supseteq \mathscr{C}^n$ and computing

$$c(\boldsymbol{A}) = \left\{ \begin{array}{ll} \text{maximize} & \langle \boldsymbol{A}, \boldsymbol{X} \rangle \\ \text{subject to} & \boldsymbol{X} \in \mathscr{R}^n \end{array} \right\} \geq \mathrm{OPT}(\boldsymbol{A}).$$

Semidefinite programming examples:

$c(\boldsymbol{A})$	$\mathscr{R}^n$	$\mathbb{E}_{\boldsymbol{W} \sim \mathrm{GOE}(n)} c(\boldsymbol{W})$
$\lambda_{\max}(\boldsymbol{A})$	$\{\boldsymbol{X} \succeq \boldsymbol{0}, \mathrm{Tr}(\boldsymbol{X}) = 1\}$	$2 + o(1)$
SOS Degree 2	$\{\boldsymbol{X} \succeq \boldsymbol{0}, X_{ii} = 1/n\}$	$2 + o(1)^{(*)}$
SOS Degree d	$\{$**complicated!**$\}$	**?**

$^{(*)}$ *[Montanari, Sen '15]*

Our Main Result (Again):

Assuming a complexity theory conjecture, for any $\epsilon > 0$, there is *no* certificate $c(\boldsymbol{A})$ that is computable in time $\mathrm{poly}(n)$ and that satisfies

$$\Pr_{\boldsymbol{W} \sim \mathrm{GOE}(n)}[c(\boldsymbol{W}) \leq 2 - \epsilon] \to 1.$$

Proof Strategy

Argue by contradiction.

Proof Strategy

Argue by contradiction.

$c(\boldsymbol{W})$ efficient certificate:

$\Pr_{\boldsymbol{W}}[c(\boldsymbol{W}) \leq 2 - \epsilon] \to 1.$

Proof Strategy

Argue by contradiction.

$c(\boldsymbol{W})$ efficient certificate:

$$\Pr_{\boldsymbol{W}}[c(\boldsymbol{W}) \leq 2 - \epsilon] \to 1.$$

$$\Downarrow$$

For hypothesis testing of $\mathbb{P}_n$, $\mathbb{Q}_n$ distributions over $\mathcal{X}_n$, **exists** an efficient test $f : \mathcal{X}_n \to \{\mathrm{p}, \mathrm{q}\}$:

$$\Pr_{\boldsymbol{Y} \sim \mathbb{P}_n}[f(\boldsymbol{Y}) = \mathrm{p}] \to 1,$$
$$\Pr_{\boldsymbol{Y} \sim \mathbb{Q}_n}[f(\boldsymbol{Y}) = \mathrm{q}] \to 1.$$

Proof Strategy

Argue by contradiction.

$c(\boldsymbol{W})$ efficient certificate:

$$\Pr_{\boldsymbol{W}}[c(\boldsymbol{W}) \leq 2 - \epsilon] \to 1.$$

$$\Downarrow$$

For hypothesis testing of $\mathbb{P}_n$, $\mathbb{Q}_n$ distributions over $\mathcal{X}_n$, **exists** an efficient test $f : \mathcal{X}_n \to \{\mathrm{p}, \mathrm{q}\}$:

$$\Pr_{\boldsymbol{Y} \sim \mathbb{P}_n}[f(\boldsymbol{Y}) = \mathrm{p}] \to 1,$$
$$\Pr_{\boldsymbol{Y} \sim \mathbb{Q}_n}[f(\boldsymbol{Y}) = \mathrm{q}] \to 1.$$

"Low-degree polynomials conjecture" on hardness of hypothesis testing

Proof Strategy

Argue by contradiction.

$c(\boldsymbol{W})$ efficient certificate:

$$\Pr_{\boldsymbol{W}}[c(\boldsymbol{W}) \leq 2 - \epsilon] \to 1.$$

$$\Downarrow$$

For hypothesis testing of $\mathbb{P}_n$, $\mathbb{Q}_n$ distributions over $\mathcal{X}_n$, **exists** an efficient test $f : \mathcal{X}_n \to \{\mathrm{p}, \mathrm{q}\}$:

$$\Pr_{\boldsymbol{Y} \sim \mathbb{P}_n}[f(\boldsymbol{Y}) = \mathrm{p}] \to 1,$$
$$\Pr_{\boldsymbol{Y} \sim \mathbb{Q}_n}[f(\boldsymbol{Y}) = \mathrm{q}] \to 1.$$

"Low-degree polynomials conjecture" on hardness of hypothesis testing

$$\Downarrow$$

For hypothesis testing of $\mathbb{P}_n$, $\mathbb{Q}_n$ distributions over $\mathcal{X}_n$, there **does not exist** any efficient test.

Proof Strategy

Argue by contradiction.

$c(\boldsymbol{W})$ efficient certificate:

$$\Pr_{\boldsymbol{W}}[c(\boldsymbol{W}) \leq 2 - \epsilon] \to 1.$$

$$\Downarrow$$

For hypothesis testing of $\mathbb{P}_n$, $\mathbb{Q}_n$ distributions over $\mathcal{X}_n$, **exists** an efficient test $f : \mathcal{X}_n \to \{\mathrm{p}, \mathrm{q}\}$:

$$\Pr_{\boldsymbol{Y} \sim \mathbb{P}_n}[f(\boldsymbol{Y}) = \mathrm{p}] \to 1,$$
$$\Pr_{\boldsymbol{Y} \sim \mathbb{Q}_n}[f(\boldsymbol{Y}) = \mathrm{q}] \to 1.$$

$\Rightarrow\Leftarrow$

"Low-degree polynomials conjecture" on hardness of hypothesis testing

$$\Downarrow$$

For hypothesis testing of $\mathbb{P}_n$, $\mathbb{Q}_n$ distributions over $\mathcal{X}_n$, there **does not exist** any efficient test.

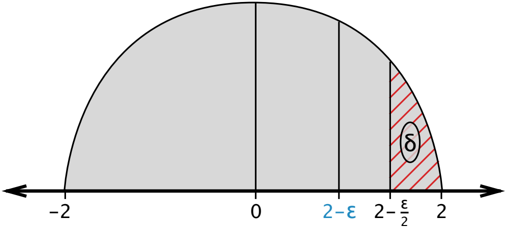
Part I: Certification ⇒ Hypothesis Testing

Key Idea: Certify below $\lambda_{\max}(\boldsymbol{W}) \rightsquigarrow$ distinguish uniform subspace from subspace with a hypercube vector nearby.

Part I: Certification ⇒ Hypothesis Testing

Key Idea: Certify below $\lambda_{\max}(\boldsymbol{W})$ $\rightsquigarrow$ distinguish uniform subspace from subspace with a hypercube vector nearby.

$\boldsymbol{W} \sim \mathrm{GOE}(n)$	$\boldsymbol{W}' \sim \mathrm{GOE}'(n)$
Law of top eigenspace:	Replace with:
$V \overset{(\mathrm{d})}{=} \mathrm{span}(\underbrace{\boldsymbol{g}_1, \dots, \boldsymbol{g}_{\delta n}}_{\mathbb{Q}_n})$	$V' = \mathrm{span}(\underbrace{\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\delta n}}_{\mathbb{P}_n})$



Part I: Certification $\Rightarrow$ Hypothesis Testing

Key Idea: Certify below $\lambda_{\max}(\boldsymbol{W}) \rightsquigarrow$ distinguish uniform subspace from subspace with a hypercube vector nearby.

$\boldsymbol{W} \sim \mathrm{GOE}(n)$	$\boldsymbol{W}' \sim \mathrm{GOE}'(n)$
Law of top eigenspace:	Replace with:
$V \overset{(\mathrm{d})}{=} \mathrm{span}(\underbrace{\boldsymbol{g}_1, \dots, \boldsymbol{g}_{\delta n}}_{\mathbb{Q}_n})$	$V' = \mathrm{span}(\underbrace{\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\delta n}}_{\mathbb{P}_n})$

If $(\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\delta n}) \sim \mathbb{P}_n \Rightarrow \exists \boldsymbol{x} \in \{\pm 1/\sqrt{n}\}^n$ "close to" V', then

$$c(\boldsymbol{W}) \leq 2 - \epsilon$$

$$c(\boldsymbol{W}') \geq \left(2 - \frac{\epsilon}{2}\right) \|\boldsymbol{P}_{V'}\boldsymbol{x}\|^2 - 2(1 - \|\boldsymbol{P}_{V'}\boldsymbol{x}\|^2) \geq 2 - \frac{2\epsilon}{3},$$

so thresholding c distinguishes $\mathbb{P}_n$ and $\mathbb{Q}_n = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)^{\otimes \delta n}$.

Part I: Certification $\Rightarrow$ Hypothesis Testing

Remaining: How to define $(\boldsymbol{y}_1, \dots, \boldsymbol{y}_{\delta n}) \sim \mathbb{P}_n$ with:

- Hard to distinguish from $\mathbb{Q}_n = \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)^{\otimes \delta n}$, and
- with high probability there exists $\boldsymbol{x} \in \{\pm 1/\sqrt{n}\}^n$ such that $\|\boldsymbol{P}_{\mathrm{span}(\boldsymbol{y}_i)} \boldsymbol{x}\|^2 \geq 1 - \kappa$?

Part I: Certification $\Rightarrow$ Hypothesis Testing

Remaining: How to define $(\boldsymbol{y}_1, \dots, \boldsymbol{y}_{(1-\delta)n}) \sim \mathbb{P}_n$ with:

- Hard to distinguish from $\mathbb{Q}_n = \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)^{\otimes(1-\delta)n}$, and
- with high probability there exists $\boldsymbol{x} \in \{\pm 1/\sqrt{n}\}^n$ such that $\|\boldsymbol{P}_{\mathrm{span}(\boldsymbol{y}_i)}\boldsymbol{x}\|^2 \leq \kappa$?

Part I: Certification $\Rightarrow$ Hypothesis Testing

Remaining: How to define $(\boldsymbol{y}_1, \dots, \boldsymbol{y}_{(1-\delta)n}) \sim \mathbb{P}_n$ with:

- Hard to distinguish from $\mathbb{Q}_n = \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)^{\otimes(1-\delta)n}$, and
- with high probability there exists $\boldsymbol{x} \in \{\pm 1/\sqrt{n}\}^n$ such that $\|\boldsymbol{P}_{\mathrm{span}(\boldsymbol{y}_i)}\boldsymbol{x}\|^2 \leq \kappa$?

(Negatively-Spiked) Wishart Model: $\beta \in [-1, \infty)$.

- Under $\mathbb{Q}_n$,

$$\boldsymbol{y}_1, \dots, \boldsymbol{y}_{(1-\delta)n} \overset{(\perp)}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n).$$

- Under $\mathbb{P}_n$, choose $\boldsymbol{x} \sim \mathrm{Unif}(\{\pm 1/\sqrt{n}\}^n)$. Then,

$$\boldsymbol{y}_1, \dots, \boldsymbol{y}_{(1-\delta)n} \overset{(\perp)}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n + \beta \boldsymbol{x}\boldsymbol{x}^\top).$$

Part I: Certification ⇒ Hypothesis Testing

Remaining: How to define $(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{(1-\delta)n}) \sim \mathbb{P}_n$ with:

- Hard to distinguish from $\mathbb{Q}_n = \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n)^{\otimes(1-\delta)n}$, and
- with high probability there exists $\boldsymbol{x} \in \{\pm 1/\sqrt{n}\}^n$ such that $\|\boldsymbol{P}_{\mathrm{span}(\boldsymbol{y}_i)}\boldsymbol{x}\|^2 \leq \kappa$?

(Negatively-Spiked) Wishart Model: $\beta \in [-1, \infty)$.

- Under $\mathbb{Q}_n$,

$$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{(1-\delta)n} \overset{(\perp)}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n).$$

- Under $\mathbb{P}_n$, choose $\boldsymbol{x} \sim \mathrm{Unif}(\{\pm 1/\sqrt{n}\}^n)$. Then,

$$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{(1-\delta)n} \overset{(\perp)}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{I}_n + \beta \boldsymbol{x}\boldsymbol{x}^\top).$$

Lemma: For all $\kappa > 0$, there exists $\beta \in (-1, 0)$ such that

$$\Pr_{(x;(\boldsymbol{y}_1,\ldots,\boldsymbol{y}_{(1-\delta)n}))\sim\mathbb{P}_n}\left[\|\boldsymbol{P}_{\mathrm{span}(\boldsymbol{y}_i)}\boldsymbol{x}\|^2 \leq \kappa\right] \to 1.$$

Proof Strategy (Reminder)

Argue by contradiction.

$c(\boldsymbol{W})$ efficient certificate:

$$\Pr_{\boldsymbol{W}}[c(\boldsymbol{W}) \leq 2 - \epsilon] \to 1.$$

$$\Downarrow$$

For hypothesis testing of $\mathbb{P}_n$, $\mathbb{Q}_n$ distributions over $\mathbb{R}^{n \times (1-\delta)n}$, **exists** efficient test $f : \mathbb{R}^{n \times (1-\delta)n} \to \{\mathrm{p}, \mathrm{q}\}$:

$$\Pr_{\boldsymbol{Y} \sim \mathbb{P}_n}[f(\boldsymbol{Y}) = \mathrm{p}] \to 1,$$

$$\Pr_{\boldsymbol{Y} \sim \mathbb{Q}_n}[f(\boldsymbol{Y}) = \mathrm{q}] \to 1.$$

$\Rightarrow\Leftarrow$

"Low-degree polynomials conjecture" on hardness of hypothesis testing

$$\Downarrow$$

For hypothesis testing of $\mathbb{P}_n$, $\mathbb{Q}_n$ distributions over $\mathbb{R}^{n \times (1-\delta)n}$, there **does not exist** any efficient test.

Part II: Hardness of Hypothesis Testing

Negatively-Spiked Wishart Model: $\beta \in (-1, 0)$, $\delta \in (0, 1)$.

- Under $\mathbb{Q}_n$,

$$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{(1-\delta)n} \overset{(\perp)}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n).$$

- Under $\mathbb{P}_n$, choose $\boldsymbol{x} \sim \mathrm{Unif}(\{\pm 1/\sqrt{n}\}^n)$. Then,

$$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{(1-\delta)n} \overset{(\perp)}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n + \beta \boldsymbol{x}\boldsymbol{x}^\top).$$

Part II: Hardness of Hypothesis Testing

Negatively-Spiked Wishart Model: $\beta \in (-1, 0)$, $\delta \in (0, 1)$.

- Under $\mathbb{Q}_n$,

$$\boldsymbol{y}_1, \dots, \boldsymbol{y}_{(1-\delta)n} \overset{(\perp)}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n).$$

- Under $\mathbb{P}_n$, choose $\boldsymbol{x} \sim \mathrm{Unif}(\{\pm 1/\sqrt{n}\}^n)$. Then,

$$\boldsymbol{y}_1, \dots, \boldsymbol{y}_{(1-\delta)n} \overset{(\perp)}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n + \beta \boldsymbol{x}\boldsymbol{x}^\top).$$

We will be finished if we can show...

Lemma: Assuming the "low-degree polynomials conjecture" on hardness of hypothesis testing, if $\beta^2(1-\delta) < 1$, then there is no test $f : \mathbb{R}^{n \times (1-\delta)n} \to \{\mathrm{p}, \mathrm{q}\}$ distinguishing $\mathbb{P}_n$ and $\mathbb{Q}_n$ and computable in time poly(n).

The Low-Degree Polynomials Conjecture

The Low-Degree Polynomials Conjecture

Technique developed by *[Hopkins, Steurer '17; Hopkins '18]* for predicting hardness of hypothesis testing, when $\mathbb{P}_n$ is a structured distribution and $\mathbb{Q}_n$ is highly symmetric.

Key Idea: Restrict testing algorithms to those that evaluate low-degree ($\leq D$) polynomials on a sample.

The Low-Degree Polynomials Conjecture

Technique developed by *[Hopkins, Steurer '17; Hopkins '18]* for predicting hardness of hypothesis testing, when $\mathbb{P}_n$ is a structured distribution and $\mathbb{Q}_n$ is highly symmetric.

Key Idea: Restrict testing algorithms to those that evaluate low-degree ($\leq D$) polynomials on a sample.

Important Adjustment: To include *spectral algorithms*, need to allow evaluation of $\lambda_{\max}(\boldsymbol{M})$ for $\boldsymbol{M}$ having constant-degree polynomials in the sample $\rightsquigarrow$ via power method enough to take $D(n) = \omega(\log n)$.

The Low-Degree Polynomials Conjecture

Define likelihood ratio

$$L_n(\boldsymbol{Y}) := \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}(\boldsymbol{Y}).$$

The Low-Degree Polynomials Conjecture

Define likelihood ratio

$$L_n(\boldsymbol{Y}) := \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}(\boldsymbol{Y}).$$

Heuristic for best low-degree polynomial:

$$\left\{\begin{array}{ll} \text{maximize} & \mathbb{E}_{\boldsymbol{Y}\sim\mathbb{P}_n} f(\boldsymbol{Y}) \\ \text{subject to} & f \in \mathbb{R}[\boldsymbol{Y}]_{\leq D} \\ & \mathbb{E}_{\boldsymbol{Y}\sim\mathbb{Q}_n} f(\boldsymbol{Y})^2 = 1 \end{array}\right\} = \|L_n^{\leq D}\|_{L^2(\mathbb{Q}_n)},$$

where $L_n^{\leq D}$ is projection of L_n in $L^2(\mathbb{Q}_n)$ to $\mathbb{R}[\boldsymbol{Y}]_{\leq D}$.

The Low-Degree Polynomials Conjecture

Define likelihood ratio

$$L_n(\boldsymbol{Y}) := \frac{d\mathbb{P}_n}{d\mathbb{Q}_n}(\boldsymbol{Y}).$$

Heuristic for best low-degree polynomial:

$$\left\{ \begin{array}{ll} \text{maximize} & \mathbb{E}_{\boldsymbol{Y}\sim\mathbb{P}_n} f(\boldsymbol{Y}) \\ \text{subject to} & f \in \mathbb{R}[\boldsymbol{Y}]_{\leq D} \\ & \mathbb{E}_{\boldsymbol{Y}\sim\mathbb{Q}_n} f(\boldsymbol{Y})^2 = 1 \end{array} \right\} = \|L_n^{\leq D}\|_{L^2(\mathbb{Q}_n)},$$

where $L_n^{\leq D}$ is projection of L_n in $L^2(\mathbb{Q}_n)$ to $\mathbb{R}[\boldsymbol{Y}]_{\leq D}$.

Conjecture: For "nice" $\mathbb{P}_n$, $\mathbb{Q}_n$, and some $D(n) = \omega(\log n)$, if $\|L_n^{\leq D(n)}\|_{L^2(\mathbb{Q}_n)} = O_{n\to\infty}(1)$, then there is no test that distinguishes $\mathbb{P}_n$ and $\mathbb{Q}_n$ and runs in time poly(n).

Part II: Hardness of Hypothesis Testing

Negatively-Spiked Wishart Model: $\beta \in (-1, 0)$, $\delta \in (0, 1)$.

- Under $\mathbb{Q}_n$,

$$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{(1-\delta)n} \stackrel{(\perp)}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n).$$

- Under $\mathbb{P}_n$, choose $\boldsymbol{x} \sim \mathrm{Unif}(\{\pm 1/\sqrt{n}\}^n)$. Then,

$$\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{(1-\delta)n} \stackrel{(\perp)}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n + \beta \boldsymbol{x}\boldsymbol{x}^\top).$$

Part II: Hardness of Hypothesis Testing

Negatively-Spiked Wishart Model: $\beta \in (-1, 0)$, $\delta \in (0, 1)$.

- Under $\mathbb{Q}_n$,

$$\boldsymbol{y}_1, \dots, \boldsymbol{y}_{(1-\delta)n} \overset{(\perp)}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n).$$

- Under $\mathbb{P}_n$, choose $\boldsymbol{x} \sim \mathrm{Unif}(\{\pm 1/\sqrt{n}\}^n)$. Then,

$$\boldsymbol{y}_1, \dots, \boldsymbol{y}_{(1-\delta)n} \overset{(\perp)}{\sim} \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n + \beta \boldsymbol{x}\boldsymbol{x}^\top).$$

Since $\mathbb{Q}_n$ is i.i.d. gaussian, use Hermite polynomials, getting expression in $\boldsymbol{x}^i \sim \mathrm{Unif}(\{\pm 1/\sqrt{n}\}^n)$ independent copies:

$$\|L_n^{\leq D}\|^2 = \mathbb{E}_{\boldsymbol{x}^1, \boldsymbol{x}^2}\left[\phi_n^{\leq D/2}(\beta^2 \langle \boldsymbol{x}^1, \boldsymbol{x}^2 \rangle^2)\right],$$

$$\phi_n^{\leq k} = \text{order } k \text{ Taylor poly. of } \phi_n(t) = (1-t)^{-(1-\delta)n/2}$$

Part II: Hardness of Hypothesis Testing

Want: When $\beta^2(1-\delta) < 1$ and $D(n) \sim (\log n)^{1+\alpha}$ for some small $\alpha > 0$, then

$$\|L_n^{\leq D}\|^2 = \mathbb{E}_{\boldsymbol{x}^1,\boldsymbol{x}^2}\left[\phi_n^{\leq D/2}(\beta^2\langle \boldsymbol{x}^1, \boldsymbol{x}^2\rangle^2)\right] = O_{n\to\infty}(1).$$

Part II: Hardness of Hypothesis Testing

Want: When $\beta^2(1-\delta) < 1$ and $D(n) \sim (\log n)^{1+\alpha}$ for some small $\alpha > 0$, then

$$\|L_n^{\leq D}\|^2 = \mathbb{E}_{\boldsymbol{x}^1,\boldsymbol{x}^2}\left[\phi_n^{\leq D/2}(\beta^2\langle \boldsymbol{x}^1, \boldsymbol{x}^2\rangle^2)\right] = O_{n\to\infty}(1).$$

Heuristic Argument: (1) $\langle \boldsymbol{x}^1, \boldsymbol{x}^2\rangle \rightsquigarrow \mathcal{N}(0, \frac{1}{n})$ fast by CLT, (2) $\phi_n^{\leq k} \leq \phi_n$, (3) $n \to \infty$.

Part II: Hardness of Hypothesis Testing

Want: When $\beta^2(1-\delta) < 1$ and $D(n) \sim (\log n)^{1+\alpha}$ for some small $\alpha > 0$, then

$$\| L_n^{\leq D} \|^2 = \mathbb{E}_{\boldsymbol{x}^1, \boldsymbol{x}^2} \left[\phi_n^{\leq D/2}(\beta^2 \langle \boldsymbol{x}^1, \boldsymbol{x}^2 \rangle^2) \right] = O_{n \to \infty}(1).$$

Heuristic Argument: (1) $\langle \boldsymbol{x}^1, \boldsymbol{x}^2 \rangle \rightsquigarrow \mathcal{N}(0, \frac{1}{n})$ fast by CLT, (2) $\phi_n^{\leq k} \leq \phi_n$, (3) $n \to \infty$. Then,

$$\lim_{n\to\infty} \| L_n^{\leq D(n)} \|^2 \lesssim \lim_{n \to \infty} \mathbb{E}_{g \sim \mathcal{N}(0,1)} \Bigg[\underbrace{\left(1 - \frac{\beta^2 g^2}{n}\right)^{-(1-\delta)n/2}}_{\phi_n(\beta^2 g^2/n)} \Bigg]$$

$$= \mathbb{E}_{g \sim \mathcal{N}(0,1)} \left[\exp\left(\beta^2 (1-\delta) g^2 / 2 \right) \right],$$

the moment-generating function of a χ^2 random variable; finite exactly when $\beta^2(1-\delta) < 1$. □

Proof Strategy (One Last Reminder)

Argue by contradiction.

$c(\boldsymbol{W})$ efficient certificate:

$$\Pr_{\boldsymbol{W}}[c(\boldsymbol{W}) \leq 2 - \epsilon] \to 1.$$

$$\Downarrow$$

For hypothesis testing of $\mathbb{P}_n$, $\mathbb{Q}_n$ distributions over $\mathbb{R}^{n \times (1-\delta)n}$, **exists** efficient test $f : \mathbb{R}^{n \times (1-\delta)n} \to \{\mathrm{p}, \mathrm{q}\}$:

$$\Pr_{\boldsymbol{Y} \sim \mathbb{P}_n}[f(\boldsymbol{Y}) = \mathrm{p}] \to 1,$$

$$\Pr_{\boldsymbol{Y} \sim \mathbb{Q}_n}[f(\boldsymbol{Y}) = \mathrm{q}] \to 1.$$

$$\Rightarrow\Leftarrow$$

"Low-degree polynomials conjecture" on hardness of hypothesis testing

$$\Downarrow$$

For hypothesis testing of $\mathbb{P}_n$, $\mathbb{Q}_n$ distributions over $\mathbb{R}^{n \times (1-\delta)n}$, there **does not exist** any efficient test.

Takeaways

On this problem:

- There is a gap between search and certification! As in k-SAT, cuts in hypergraphs, cliques in random graphs, and others. **Q:** What is responsible?

Takeaways

On this problem:

- There is a gap between search and certification! As in k-SAT, cuts in hypergraphs, cliques in random graphs, and others. **Q:** What is responsible?

On general methodology:

- We can prove hardness of certification in random problems using "planted" distributions. (We knew this.)
- But sometimes, the correct planted distribution is not obvious. (We knew this, too—but the "quietness" concept is hard to pin down.)
- The low-degree method can help us predict thresholds for certification. **Q:** How to do it more systematically?

Thank you!

(This talk is based on the paper "Computational Hardness of Certifying Bounds on Constrained PCA Problems" *[arXiv:1902.07324]*.)