The Algorithmic Phase Transition for Correlated Spiked Models

Zhangsong Li

School of Mathematical Sciences, Peking University

December 5, 2025

Sino-Russian Student Seminar

Motivation

• A common theme in high-dimensional statistics & machine learning: recovering a low-dimensional structure from high-dimensional noise.

- Recognize certain features in an image;
- Determine which combination of genes cause a certain disease;
- Find "communities" in a social network;
- Predict which users will click which ads;
- etc.

An idealized model: Spiked random matrix

- **Spiked Wigner model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} x x^\top \in \mathbb{R}^{n*n}.$$

An idealized model: Spiked random matrix

- **Spiked Wigner model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} \color{red}{x x^\top} \in \mathbb{R}^{n*n} .$$

Noise $\boldsymbol{W}$: an $n * n$ Wigner matrix (i.e., $\boldsymbol{W}$ is symmetric and $\boldsymbol{W}_{i,j} \sim \mathcal{N}(0, 1 + \mathbf{1}_{i=j})$ independently for $i \leq j$).

An idealized model: Spiked random matrix

- **Spiked Wigner model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} x x^\top \in \mathbb{R}^{n*n}.$$

Noise $\boldsymbol{W}$: an $n * n$ Wigner matrix (i.e., $\boldsymbol{W}$ is symmetric and $\boldsymbol{W}_{i,j} \sim \mathcal{N}(0, 1 + \mathbf{1}_{i=j})$ independently for $i \leq j$).

λ: signal-to-noise ratio.

An idealized model: Spiked random matrix

- **Spiked Wigner model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} x x^\top \in \mathbb{R}^{n*n}.$$

Noise $\boldsymbol{W}$: an $n*n$ Wigner matrix (i.e., $\boldsymbol{W}$ is symmetric and $\boldsymbol{W}_{i,j} \sim \mathcal{N}(0, 1 + \mathbf{1}_{i=j})$ independently for $i \le j$).

λ: signal-to-noise ratio.

Signal vector ("spike") x: sampled from π such that $\|x\| \approx \sqrt{n}$, e.g.,

An idealized model: Spiked random matrix

- **Spiked Wigner model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} x x^\top \in \mathbb{R}^{n*n}.$$

Noise $\boldsymbol{W}$: an $n*n$ Wigner matrix (i.e., $\boldsymbol{W}$ is symmetric and $\boldsymbol{W}_{i,j} \sim \mathcal{N}(0, 1 + \mathbf{1}_{i=j})$ independently for $i \le j$).

λ: signal-to-noise ratio.

Signal vector ("spike") x: sampled from π such that $\|x\| \approx \sqrt{n}$, e.g.,

1. uniform $\{-1, +1\}^n$ hypercube (Rademacher);

An idealized model: Spiked random matrix

- **Spiked Wigner model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} x x^\top \in \mathbb{R}^{n*n}.$$

Noise $\boldsymbol{W}$: an $n * n$ Wigner matrix (i.e., $\boldsymbol{W}$ is symmetric and $\boldsymbol{W}_{i,j} \sim \mathcal{N}(0, 1 + \mathbf{1}_{i=j})$ independently for $i \leq j$).

λ: signal-to-noise ratio.

Signal vector ("spike") x: sampled from π such that $\|x\| \approx \sqrt{n}$, e.g.,

1. uniform $\{-1, +1\}^n$ hypercube (Rademacher);
2. Gaussian distribution $\mathcal{N}(0, \mathbb{I}_n)$.

An idealized model: Spiked random matrix

- **Spiked Wigner model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} x x^\top \in \mathbb{R}^{n*n}.$$

Noise $\boldsymbol{W}$: an $n * n$ Wigner matrix (i.e., $\boldsymbol{W}$ is symmetric and $\boldsymbol{W}_{i,j} \sim \mathcal{N}(0, 1 + \mathbf{1}_{i=j})$ independently for $i \leq j$).

λ: signal-to-noise ratio.

Signal vector ("spike") x: sampled from π such that $\|x\| \approx \sqrt{n}$, e.g.,

1. uniform $\{-1, +1\}^n$ hypercube (Rademacher);
2. Gaussian distribution $\mathcal{N}(0, \mathbb{I}_n)$.
3. normalized "sparse" Rademacher.

An idealized model: Spiked random matrix

- **Spiked Wishart (covariance) model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top \in \mathbb{R}^{n*N}.$$

$$\Longleftrightarrow \boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N),\ \boldsymbol{Y}_i \sim \mathcal{N}(0, \mathbb{I}_n + \frac{\lambda}{n} x x^\top).$$

An idealized model: Spiked random matrix

- **Spiked Wishart (covariance) model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top \in \mathbb{R}^{n*N} .$$

$$\Longleftrightarrow \boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N), \, \boldsymbol{Y}_i \sim \mathcal{N}(0, \mathbb{I}_n + \frac{\lambda}{n} x x^\top) .$$

Noise $\boldsymbol{W}$: i.i.d. $\mathcal{N}(0, 1)$ entries.

An idealized model: Spiked random matrix

- **Spiked Wishart (covariance) model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top \in \mathbb{R}^{n*N} .$$

$$\Longleftrightarrow \boldsymbol{Y} = (\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_N), \, \boldsymbol{Y}_i \sim \mathcal{N}(0, \mathbb{I}_n + \frac{\lambda}{n} x x^\top) .$$

Noise $\boldsymbol{W}$: i.i.d. $\mathcal{N}(0, 1)$ entries.

λ: signal-to-noise ratio.

An idealized model: Spiked random matrix

- **Spiked Wishart (covariance) model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top \in \mathbb{R}^{n*N}.$$

$$\Longleftrightarrow \boldsymbol{Y} = (\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_N),\ \boldsymbol{Y}_i \sim \mathcal{N}(0, \mathbb{I}_n + \frac{\lambda}{n} x x^\top).$$

Noise $\boldsymbol{W}$: i.i.d. $\mathcal{N}(0,1)$ entries.

λ: signal-to-noise ratio.

Signal vector ("spike") x: sampled from π such that $\|x\| \approx \sqrt{n}$; $\boldsymbol{u}$: sampled from $\mathcal{N}(0, \mathbb{I}_N)$.

An idealized model: Spiked random matrix

- **Spiked Wishart (covariance) model:**

$$\boldsymbol{Y} = \boldsymbol{W} + \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top \in \mathbb{R}^{n*N}.$$

$$\iff \boldsymbol{Y} = (\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N),\, \boldsymbol{Y}_i \sim \mathcal{N}(0, \mathbb{I}_n + \frac{\lambda}{n} x x^\top).$$

Noise $\boldsymbol{W}$: i.i.d. $\mathcal{N}(0,1)$ entries.

λ: signal-to-noise ratio.

Signal vector ("spike") x: sampled from π such that $\|x\| \approx \sqrt{n}$; $\boldsymbol{u}$: sampled from $\mathcal{N}(0, \mathbb{I}_N)$.

We will focus on the "critical" regime where $N = \Theta(n)$ and $\lambda = \Theta(1)$.

Spectral approach

Two inference tasks: detection, recovery.

• Detection: distinguish reliably (error prob $\to 0$) the spiked matrix and the pure noise matrix.

• Recovery: non-trivial correlation with the truth: $\frac{|\langle \widehat{x}, x\rangle|}{\|\widehat{x}\|\|x\|} = \Omega(1)$.

Spectral approach

Two inference tasks: detection, recovery.

• Detection: distinguish reliably (error prob $\to 0$) the spiked matrix and the pure noise matrix.

• Recovery: non-trivial correlation with the truth: $\frac{|\langle \widehat{x}, x\rangle|}{\|\widehat{x}\|\|x\|} = \Omega(1)$.

A canonical spectral approach: principle component analysis (PCA).

Spectral approach

Two inference tasks: detection, recovery.

• Detection: distinguish reliably (error prob $\to 0$) the spiked matrix and the pure noise matrix.

• Recovery: non-trivial correlation with the truth: $\frac{|\langle \widehat{x}, x\rangle|}{\|\widehat{x}\|\|x\|} = \Omega(1)$.

A canonical spectral approach: principle component analysis (PCA).

- Idea: use the top eigenpair $\varsigma_1(\boldsymbol{YY}^\top), v_1(\boldsymbol{YY}^\top)$ for detection and recovery, i.e.,

Spectral approach

Two inference tasks: detection, recovery.

• Detection: distinguish reliably (error prob $\to 0$) the spiked matrix and the pure noise matrix.

• Recovery: non-trivial correlation with the truth: $\frac{|\langle \widehat{x}, x\rangle|}{\|\widehat{x}\|\|x\|} = \Omega(1)$.

A canonical spectral approach: principle component analysis (PCA).

- Idea: use the top eigenpair $\varsigma_1(\boldsymbol{Y}\boldsymbol{Y}^\top), v_1(\boldsymbol{Y}\boldsymbol{Y}^\top)$ for detection and recovery, i.e.,
  - Detection: $\varsigma_1(\boldsymbol{Y}\boldsymbol{Y}^\top)$ is large $\Longrightarrow$ spiked.

Spectral approach

Two inference tasks: detection, recovery.

• Detection: distinguish reliably (error prob $\to 0$) the spiked matrix and the pure noise matrix.

• Recovery: non-trivial correlation with the truth: $\frac{|\langle \widehat{x}, x\rangle|}{\|\widehat{x}\|\|x\|} = \Omega(1)$.

A canonical spectral approach: principle component analysis (PCA).

- Idea: use the top eigenpair $\varsigma_1(\boldsymbol{YY}^\top), v_1(\boldsymbol{YY}^\top)$ for detection and recovery, i.e.,
  - Detection: $\varsigma_1(\boldsymbol{YY}^\top)$ is large $\Longrightarrow$ spiked.
  - Recovery: estimate x by $v_1(\boldsymbol{YY}^\top)$.

Spectral approach

Two inference tasks: detection, recovery.

• Detection: distinguish reliably (error prob $\to 0$) the spiked matrix and the pure noise matrix.

• Recovery: non-trivial correlation with the truth: $\frac{|\langle \widehat{x}, x\rangle|}{\|\widehat{x}\|\|x\|} = \Omega(1)$.

A canonical spectral approach: principle component analysis (PCA).

- Idea: use the top eigenpair $\varsigma_1(\boldsymbol{YY}^\top), v_1(\boldsymbol{YY}^\top)$ for detection and recovery, i.e.,
  - Detection: $\varsigma_1(\boldsymbol{YY}^\top)$ is large $\Longrightarrow$ spiked.
  - Recovery: estimate x by $v_1(\boldsymbol{YY}^\top)$.
- Intuition: for large λ one expects the rank-one deformation to create an outlier eigenvalue.

Spectral (BBP) transition

Spiked Wishart model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top \in \mathbb{R}^{n*N}$ with $n = \gamma N$.

Spectral (BBP) transition

Spiked Wishart model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top \in \mathbb{R}^{n*N}$ with $n = \gamma N$.

- $\lambda = 0$ (pure Wishart matrix) [Marčenko-Pastur 1967]: $\varsigma_1(\frac{1}{N}\boldsymbol{W}\boldsymbol{W}^\top) \to (1+\sqrt{\gamma})^2$.

Spectral (BBP) transition

Spiked Wishart model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top \in \mathbb{R}^{n*N}$ with $n = \gamma N$.

- $\lambda = 0$ (pure Wishart matrix) [Marčenko-Pastur 1967]: $\varsigma_1(\frac{1}{N}\boldsymbol{W}\boldsymbol{W}^\top) \to (1+\sqrt{\gamma})^2$.
- $\lambda > 0$ (spiked Wishart matrix) [Baik-Ben Arous-Péché 2005]:

Spectral (BBP) transition

Spiked Wishart model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top \in \mathbb{R}^{n*N}$ with $n = \gamma N$.

- $\lambda = 0$ (pure Wishart matrix) [Marčenko-Pastur 1967]: $\varsigma_1(\frac{1}{N}\boldsymbol{W}\boldsymbol{W}^\top) \to (1+\sqrt{\gamma})^2$.
- $\lambda > 0$ (spiked Wishart matrix) [Baik-Ben Arous-Péché 2005]:
  - $\lambda^2 \leq \gamma$ (subcritical): $\varsigma_1(\frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^\top) \to (1+\sqrt{\gamma})^2$ and $\langle v_1(\frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^\top), x\rangle \to 0$.

Spectral (BBP) transition

Spiked Wishart model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top \in \mathbb{R}^{n*N}$ with $n = \gamma N$.

- $\lambda = 0$ (pure Wishart matrix) [Marčenko-Pastur 1967]: $\varsigma_1(\frac{1}{N}\boldsymbol{W}\boldsymbol{W}^\top) \to (1+\sqrt{\gamma})^2$.
- $\lambda > 0$ (spiked Wishart matrix) [Baik-Ben Arous-Péché 2005]:
  - $\lambda^2 \leq \gamma$ (subcritical): $\varsigma_1(\frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^\top) \to (1+\sqrt{\gamma})^2$ and $\langle v_1(\frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^\top), x\rangle \to 0$.
  - $\lambda^2 > \gamma$ (supercritical): $\varsigma_1(\frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^\top) \to (1+\lambda)(1+\frac{\gamma}{\lambda}) > (1+\sqrt{\gamma})^2$ and $\langle v_1(\frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^\top), x\rangle = \Omega(1)$.

Spectral (BBP) transition

Spiked Wishart model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top \in \mathbb{R}^{n*N}$ with $n = \gamma N$.

- $\lambda = 0$ (pure Wishart matrix) [Marčenko-Pastur 1967]: $\varsigma_1(\frac{1}{N}\boldsymbol{W}\boldsymbol{W}^\top) \to (1+\sqrt{\gamma})^2$.
- $\lambda > 0$ (spiked Wishart matrix) [Baik-Ben Arous-Péché 2005]:
  - $\lambda^2 \leq \gamma$ (subcritical): $\varsigma_1(\frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^\top) \to (1+\sqrt{\gamma})^2$ and $\langle v_1(\frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^\top), x\rangle \to 0$.
  - $\lambda^2 > \gamma$ (supercritical): $\varsigma_1(\frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^\top) \to (1+\lambda)(1+\frac{\gamma}{\lambda}) > (1+\sqrt{\gamma})^2$ and $\langle v_1(\frac{1}{N}\boldsymbol{Y}\boldsymbol{Y}^\top), x\rangle = \Omega(1)$.

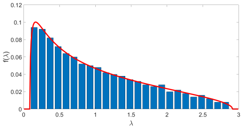


Figure: $\gamma = 0.5, \lambda = 0$

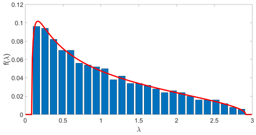


Figure: $\gamma = 0.5, \lambda = 0.5$

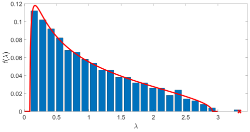


Figure: $\gamma = 0.5, \lambda = 1.5$

Spectral (BBP) transition

Spiked Wigner model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} x x^\top \in \mathbb{R}^{n*n}$.

Spectral (BBP) transition

Spiked Wigner model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} x x^\top \in \mathbb{R}^{n*n}$.

- $\lambda = 0$ (pure Wigner matrix) [Wigner 1958]: $\varsigma_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}) \to 2$.

Spectral (BBP) transition

Spiked Wigner model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} xx^\top \in \mathbb{R}^{n*n}$.

- $\lambda = 0$ (pure Wigner matrix) [Wigner 1958]: $\varsigma_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}) \to 2$.
- $\lambda > 0$ (spiked Wigner matrix) [Féral-Péché 2007]:

Spectral (BBP) transition

Spiked Wigner model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} x x^\top \in \mathbb{R}^{n*n}$.

- $\lambda = 0$ (pure Wigner matrix) [Wigner 1958]: $\varsigma_1(\frac{1}{\sqrt{n}} \boldsymbol{Y}) \to 2$.
- $\lambda > 0$ (spiked Wigner matrix) [Féral-Péché 2007]:
  - $\lambda \leq 1$ (subcritical): $\varsigma_1(\frac{1}{\sqrt{n}} \boldsymbol{Y}) \to 2$ and $\langle v_1(\frac{1}{\sqrt{n}} \boldsymbol{Y}), x \rangle \to 0$.

Spectral (BBP) transition

Spiked Wigner model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} x x^\top \in \mathbb{R}^{n*n}$.

- $\lambda = 0$ (pure Wigner matrix) [Wigner 1958]: $\varsigma_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}) \to 2$.
- $\lambda > 0$ (spiked Wigner matrix) [Féral-Péché 2007]:
    - $\lambda \leq 1$ (subcritical): $\varsigma_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}) \to 2$ and $\langle v_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}), x\rangle \to 0$.
    - $\lambda > 1$ (supercritical): $\varsigma_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}) \to \lambda + \frac{1}{\lambda} > 2$ and $\langle v_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}), x\rangle = \Omega(1)$.

Spectral (BBP) transition

Spiked Wigner model: $\boldsymbol{Y} = \boldsymbol{W} + \frac{\lambda}{\sqrt{n}} x x^\top \in \mathbb{R}^{n*n}$.

- $\lambda = 0$ (pure Wigner matrix) [Wigner 1958]: $\varsigma_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}) \to 2$.
- $\lambda > 0$ (spiked Wigner matrix) [Féral-Péché 2007]:
  - $\lambda \le 1$ (subcritical): $\varsigma_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}) \to 2$ and $\langle v_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}), x\rangle \to 0$.
  - $\lambda > 1$ (supercritical): $\varsigma_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}) \to \lambda + \frac{1}{\lambda} > 2$ and $\langle v_1(\frac{1}{\sqrt{n}}\boldsymbol{Y}), x\rangle = \Omega(1)$.

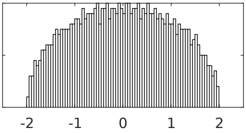


Figure: $\lambda = 0$

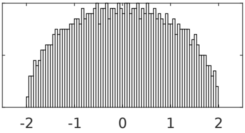


Figure: $\lambda = 0.5$

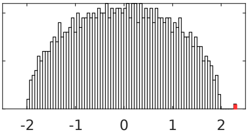


Figure: $\lambda = 1.5$

Optimality of PCA

Question: is this spectral approach optimal

- Statistically (i.e., among all possible statistics)?
- Computationally (i.e., if we restrict to polynomial-time computable statistics)?

Optimality of PCA

Question: is this spectral approach optimal

- Statistically (i.e., among all possible statistics)?
- Computationally (i.e., if we restrict to polynomial-time computable statistics)?

Statistically: the answer depends on the prior π of x.

- [Perry-Wein-Bandeira-Moitra 2018]: for "simple" and "dense" priors (e.g., $\pi = \text{unif}(\{-1, +1\}^n)$ or $\pi = \mathcal{N}(0, \mathbb{I}_n)$), PCA is optimal.
- [Lesieur-Krzakala-Zdeborová 2015], [Lelarge-Miolane 2019], [El Alaoui-Krzakala-Jordan 2020], etc.: for "sufficiently sparse" prior π, PCA is not optimal.

Optimality of PCA

Question: is this spectral approach optimal

- Statistically (i.e., among all possible statistics)?
- Computationally (i.e., if we restrict to polynomial-time computable statistics)?

Statistically: the answer depends on the prior π of x.

- [Perry-Wein-Bandeira-Moitra 2018]: for "simple" and "dense" priors (e.g., $\pi = \mathrm{unif}(\{-1, +1\}^n)$ or $\pi = \mathcal{N}(0, \mathbb{I}_n)$), PCA is optimal.
- [Lesieur-Krzakala-Zdeborová 2015], [Lelarge-Miolane 2019], [El Alaoui-Krzakala-Jordan 2020], etc.: for "sufficiently sparse" prior π, PCA is not optimal.

Computationally: [Kunisky-Wein-Bandeira 2022]: evidence suggests that PCA is always optimal.

Why correlated spiked models?

Why correlated spiked models?

- Modern data analysis increasingly involves multiple, related datasets.

Why correlated spiked models?

- Modern data analysis increasingly involves multiple, related datasets.
- Belief in multi-modal learning: jointly analyzing related datasets can yield more powerful inferences than processing each one in isolation.

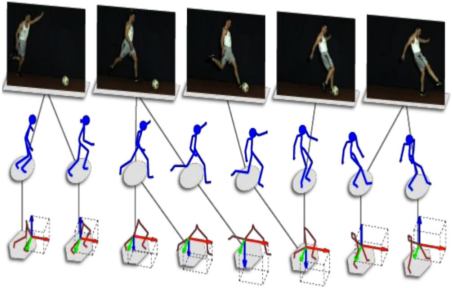


Figure: Multiple related datasets

Correlated spiked models

[Krzakala-Zdeborová 2025]: a natural toy model for multi-modal inference:

$$\boldsymbol{X} = \sum_{k=1}^{r} \frac{\lambda_k}{\sqrt{n}} x_k u_k^\top + \boldsymbol{W} \in \mathbb{R}^{n*N}, \quad \boldsymbol{Y} = \sum_{k=1}^{r} \frac{\mu_k}{\sqrt{n}} y_k v_k^\top + \boldsymbol{Z} \in \mathbb{R}^{n*N}.$$

Here (x_k, y_k) are correlated spikes such that $\|x_k\|, \|y_k\| \approx \sqrt{n}$ and $\langle x_k, y_k \rangle \approx \rho_k \|x_k\| \|y_k\|$.

Correlated spiked models

[Krzakala-Zdeborová 2025]: a natural toy model for multi-modal inference:

$$\boldsymbol{X} = \sum_{k=1}^{r} \frac{\lambda_k}{\sqrt{n}} x_k u_k^\top + \boldsymbol{W} \in \mathbb{R}^{n*N}, \quad \boldsymbol{Y} = \sum_{k=1}^{r} \frac{\mu_k}{\sqrt{n}} y_k v_k^\top + \boldsymbol{Z} \in \mathbb{R}^{n*N}.$$

Here (x_k, y_k) are correlated spikes such that $\|x_k\|, \|y_k\| \approx \sqrt{n}$ and $\langle x_k, y_k \rangle \approx \rho_k \|x_k\| \|y_k\|$. Two special cases:

Correlated spiked Wigner model:

$$\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} xx^\top + \boldsymbol{W} \in \mathbb{R}^{n*n}, \quad \boldsymbol{Y} = \frac{\mu}{\sqrt{n}} yy^\top + \boldsymbol{Z} \in \mathbb{R}^{n*n}.$$

Correlated spiked Wishart model:

$$\boldsymbol{X} = \frac{\sqrt{\lambda}}{\sqrt{n}} x\boldsymbol{u}^\top + \boldsymbol{W} \in \mathbb{R}^{n*N}, \quad \boldsymbol{Y} = \frac{\sqrt{\mu}}{\sqrt{n}} y\boldsymbol{v}^\top + \boldsymbol{Z} \in \mathbb{R}^{n*N}.$$

Special case 1

Correlated spiked Wigner model:

$$\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} x x^\top + \boldsymbol{W} \in \mathbb{R}^{n*n}, \quad \boldsymbol{Y} = \frac{\mu}{\sqrt{n}} y y^\top + \boldsymbol{Z} \in \mathbb{R}^{n*n}.$$

Here $(x, y) \sim \pi$ such that $\|x\|, \|y\| \approx \sqrt{n}$ and $\langle x, y \rangle \approx \rho \|x\| \|y\|$, e.g.,

Special case 1

Correlated spiked Wigner model:

$$\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} x x^\top + \boldsymbol{W} \in \mathbb{R}^{n*n}, \quad \boldsymbol{Y} = \frac{\mu}{\sqrt{n}} y y^\top + \boldsymbol{Z} \in \mathbb{R}^{n*n}.$$

Here $(x, y) \sim \pi$ such that $\|x\|, \|y\| \approx \sqrt{n}$ and $\langle x, y \rangle \approx \rho \|x\| \|y\|$, e.g.,

1. correlated Gaussian: $(x_i, y_i) \overset{iid}{\sim} \mathcal{N}(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$;

Special case 1

Correlated spiked Wigner model:

$$\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} xx^\top + \boldsymbol{W} \in \mathbb{R}^{n*n}, \quad \boldsymbol{Y} = \frac{\mu}{\sqrt{n}} yy^\top + \boldsymbol{Z} \in \mathbb{R}^{n*n}.$$

Here $(x, y) \sim \pi$ such that $\|x\|, \|y\| \approx \sqrt{n}$ and $\langle x, y\rangle \approx \rho \|x\| \|y\|$, e.g.,

1. correlated Gaussian: $(x_i, y_i) \overset{iid}{\sim} \mathcal{N}(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$;
2. correlated Rademacher: $(x_i, y_i) \overset{iid}{\sim}$ such that $x_i, y_i \sim \mathrm{unif}(\{-1, +1\})$ and $\mathbb{E}[x_i y_i] = \rho$;

Special case 1

Correlated spiked Wigner model:

$$\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} xx^\top + \boldsymbol{W} \in \mathbb{R}^{n*n}, \quad \boldsymbol{Y} = \frac{\mu}{\sqrt{n}} yy^\top + \boldsymbol{Z} \in \mathbb{R}^{n*n}.$$

Here $(x, y) \sim \pi$ such that $\|x\|, \|y\| \approx \sqrt{n}$ and $\langle x, y\rangle \approx \rho\|x\|\|y\|$, e.g.,

1. correlated Gaussian: $(x_i, y_i) \overset{iid}{\sim} \mathcal{N}(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$;
2. correlated Rademacher: $(x_i, y_i) \overset{iid}{\sim}$ such that $x_i, y_i \sim \text{unif}(\{-1, +1\})$ and $\mathbb{E}[x_i y_i] = \rho$;
3. correlated sparse Rademacher: $(x_i, y_i) \overset{iid}{\sim} (\frac{BX}{\sqrt{p}}, \frac{BY}{\sqrt{p}})$ such that $B \sim \text{Ber}(p)$ and X, Y are correlated Rademacher.

Special case 2

Correlated spiked Wishart model:

$$\boldsymbol{X} = \frac{\sqrt{\lambda}}{\sqrt{n}} x \boldsymbol{u}^\top + \boldsymbol{W} \in \mathbb{R}^{n*N}, \quad \boldsymbol{Y} = \frac{\sqrt{\mu}}{\sqrt{n}} y \boldsymbol{v}^\top + \boldsymbol{Z} \in \mathbb{R}^{n*N}.$$

Here $(x, y) \sim \pi$ such that $\|x\|, \|y\| \approx \sqrt{n}$ and $\langle x, y \rangle \approx \rho \|x\| \|y\|$, and $\boldsymbol{u}, \boldsymbol{v} \sim \mathcal{N}(0, \mathbb{I}_N)$. Equivalently,

$$(\boldsymbol{X}, \boldsymbol{Y}) = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_N; \boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N)$$

such that given x, y

$$(\boldsymbol{X}_i, \boldsymbol{Y}_i) \overset{iid}{\sim} \mathcal{N}(0, \mathbb{I}_n + \frac{\lambda}{n} x x^\top) \otimes \mathcal{N}(0, \mathbb{I}_n + \frac{\mu}{n} y y^\top)$$

Approach 1: Canonical correlation analysis

- A widely-used spectral approach [Thompson 2000], [Guo-Wu 2019]: canonical correlation analysis (CCA).

Approach 1: Canonical correlation analysis

- A widely-used spectral approach [Thompson 2000], [Guo-Wu 2019]: canonical correlation analysis (CCA).
- Idea: extract information of the spike from the MANOVA matrix

$$(\boldsymbol{X}\boldsymbol{X}^\top)^{-\frac{1}{2}}(\boldsymbol{X}\boldsymbol{Y}^\top)(\boldsymbol{Y}\boldsymbol{Y}^\top)^{-\frac{1}{2}}.$$

Approach 1: Canonical correlation analysis

- A widely-used spectral approach [Thompson 2000], [Guo-Wu 2019]: canonical correlation analysis (CCA).
- Idea: extract information of the spike from the MANOVA matrix

$$(\boldsymbol{X}\boldsymbol{X}^\top)^{-\frac{1}{2}}(\boldsymbol{X}\boldsymbol{Y}^\top)(\boldsymbol{Y}\boldsymbol{Y}^\top)^{-\frac{1}{2}}\,.$$

- Advantages: widely-used for general inference of hidden directions; seems to be robust in practice.

Approach 1: Canonical correlation analysis

- A widely-used spectral approach [Thompson 2000], [Guo-Wu 2019]: canonical correlation analysis (CCA).
- Idea: extract information of the spike from the MANOVA matrix

$$(\boldsymbol{X}\boldsymbol{X}^\top)^{-\frac{1}{2}}(\boldsymbol{X}\boldsymbol{Y}^\top)(\boldsymbol{Y}\boldsymbol{Y}^\top)^{-\frac{1}{2}}\,.$$

- Advantages: widely-used for general inference of hidden directions; seems to be robust in practice.
- Limitations I: requires $(\boldsymbol{X}\boldsymbol{X}^\top)$ to be invertible $\Longrightarrow$ restricts to the case $N \geq n$ (for Wishart model).

Approach 1: Canonical correlation analysis

- A widely-used spectral approach [Thompson 2000], [Guo-Wu 2019]: canonical correlation analysis (CCA).
- Idea: extract information of the spike from the MANOVA matrix

$$(\boldsymbol{X}\boldsymbol{X}^\top)^{-\frac{1}{2}}(\boldsymbol{X}\boldsymbol{Y}^\top)(\boldsymbol{Y}\boldsymbol{Y}^\top)^{-\frac{1}{2}}.$$

- Advantages: widely-used for general inference of hidden directions; seems to be robust in practice.
- Limitations I: requires $(\boldsymbol{X}\boldsymbol{X}^\top)$ to be invertible $\Longrightarrow$ restricts to the case $N \geq n$ (for Wishart model).
- Limitations II: has unsatisfactory performance for the correlated spiked model [Bao-Hu-Pan-Zhou 2019], [Ma-Yang 2023], [Bykhovskaya-Gorin 2023].

Approach 2: Partial least squares

- Another widely-used spectral approach [Wold-Sjőstrőm-Eriksson 2001]: partial least squares (PLS).

Approach 2: Partial least squares

- Another widely-used spectral approach [Wold-Sjőstrőm-Eriksson 2001]: partial least squares (PLS).
- Idea: extract information of the spike from the sample-cross-covariance matrix

$$\boldsymbol{X}\boldsymbol{Y}^\top.$$

Approach 2: Partial least squares

- Another widely-used spectral approach [Wold-Sjőstrőm-Eriksson 2001]: partial least squares (PLS).
- Idea: extract information of the spike from the sample-cross-covariance matrix

$$\boldsymbol{X}\boldsymbol{Y}^\top .$$

- [Mergny-Zdeborová 2025]: outperforms the CCA approach (especially when ρ is close to 1).

Approach 2: Partial least squares

- Another widely-used spectral approach [Wold-Sjőstrőm-Eriksson 2001]: partial least squares (PLS).
- Idea: extract information of the spike from the sample-cross-covariance matrix

$$\boldsymbol{X}\boldsymbol{Y}^\top .$$

- [Mergny-Zdeborová 2025]: outperforms the CCA approach (especially when ρ is close to 1).

Question 1: better algorithms?

Approach 2: Partial least squares

- Another widely-used spectral approach [Wold-Sjőstrőm-Eriksson 2001]: partial least squares (PLS).
- Idea: extract information of the spike from the sample-cross-covariance matrix

$$\boldsymbol{X}\boldsymbol{Y}^\top.$$

- [Mergny-Zdeborová 2025]: outperforms the CCA approach (especially when ρ is close to 1).

Question 1: better algorithms?

Question 2: computational threshold for this model?

Our algorithmic results

Define

$$F(\lambda,\mu,\rho,\gamma)=\max\left\{\frac{\lambda^2}{\gamma},\frac{\mu^2}{\gamma},\frac{\lambda^2\rho^2}{\gamma-\lambda^2+\lambda^2\rho^2}+\frac{\mu^2\rho^2}{\gamma-\mu^2+\mu^2\rho^2}\right\}.$$

Theorem (L.25+)

(1) For the correlated spiked Wigner model, suppose that $F(\lambda,\mu,\rho,1)>1$. Then (under some mild assumptions on the prior π) we can solve the detection/recovery problem efficiently.
(2) For the correlated spiked Wishart model, suppose that $\frac{n}{N}=\gamma$ for some $\gamma=\Theta(1)$ and $F(\lambda,\mu,\rho,\gamma)>1$. Then (under some mild assumptions on the prior π) we can solve the detection/recovery problem efficiently.

Discussions

Discussions

- Shows that an algorithm can leverage the correlation between the spikes to detect and estimate the signals even in regimes where efficiently recovering either x from $\boldsymbol{X}$ alone or y from $\boldsymbol{Y}$ alone is believed to be computationally infeasible.

Discussions

- Shows that an algorithm can leverage the correlation between the spikes to detect and estimate the signals even in regimes where efficiently recovering either x from $\boldsymbol{X}$ alone or y from $\boldsymbol{Y}$ alone is believed to be computationally infeasible.
- Outperforms the PLS/CCA method (see the figure below for a comparison when $\gamma = 0.25$ and $\rho = 0.99$).

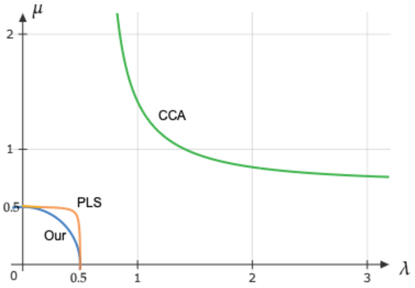


Figure: Phase diagram in the (λ, μ) plane.

Algorithmic heuristics

Let us focus on the spiked Wigner model $\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} x x^\top + \boldsymbol{W}$.

Algorithmic heuristics

Let us focus on the spiked Wigner model $\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} x x^\top + \boldsymbol{W}$.

PCA: determine spiked/non-spiked by $\|\boldsymbol{X}\|_{\mathrm{op}}$. Trace method:

Algorithmic heuristics

Let us focus on the spiked Wigner model $\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} x x^\top + \boldsymbol{W}$.

PCA: determine spiked/non-spiked by $\|\boldsymbol{X}\|_{\mathrm{op}}$. Trace method:

$$\|\boldsymbol{X}\|_{\mathrm{op}}^{2k} - \|\boldsymbol{W}\|_{\mathrm{op}}^{2k} \overset{k \gg (\log n)}{\approx} \mathrm{tr}(\boldsymbol{X}^{2k}) - \mathrm{tr}(\boldsymbol{W}^{2k})$$
$$= \sum_{i_1,\dots,i_{2k}} \boldsymbol{X}_{i_1,i_2} \dots \boldsymbol{X}_{i_{2k},i_1} - \boldsymbol{W}_{i_1,i_2} \dots \boldsymbol{W}_{i_{2k},i_1} \quad (\star)$$

Algorithmic heuristics

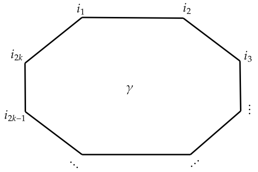
Let us focus on the spiked Wigner model $\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} xx^\top + \boldsymbol{W}$.

PCA: determine spiked/non-spiked by $\|\boldsymbol{X}\|_{\mathrm{op}}$. Trace method:

$$\|\boldsymbol{X}\|_{\mathrm{op}}^{2k} - \|\boldsymbol{W}\|_{\mathrm{op}}^{2k} \overset{k \gg (\log n)}{\approx} \mathrm{tr}(\boldsymbol{X}^{2k}) - \mathrm{tr}(\boldsymbol{W}^{2k})$$
$$= \sum_{i_1,\dots,i_{2k}} \boldsymbol{X}_{i_1,i_2} \dots \boldsymbol{X}_{i_{2k},i_1} - \boldsymbol{W}_{i_1,i_2} \dots \boldsymbol{W}_{i_{2k},i_1} \quad (\star)$$

Observation: $\gamma = (i_1, \dots, i_{2k}, i_1)$ forming a "self-avoided" cycle is the main contribution of $(\star)$.

$$(\star) \approx \sum_{\gamma \text{ self-avoided}} \boldsymbol{X}_\gamma - \boldsymbol{W}_\gamma$$
$$\|\boldsymbol{X}\|_{\mathrm{op}}^{2k} \approx \sum_{\gamma \text{ self-avoided}} \boldsymbol{X}_\gamma$$



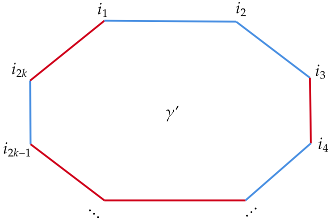
Algorithmic heuristics

Now we are given two matrices $\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} x x^\top + \boldsymbol{W}$, $\boldsymbol{Y} = \frac{\mu}{\sqrt{n}} y y^\top + \boldsymbol{Z}$.

Algorithmic heuristics

Now we are given two matrices $\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} xx^\top + \boldsymbol{W}$, $\boldsymbol{Y} = \frac{\mu}{\sqrt{n}} yy^\top + \boldsymbol{Z}$.

Natural idea: cycles $\Longrightarrow$ decorated (edge-colored) cycles

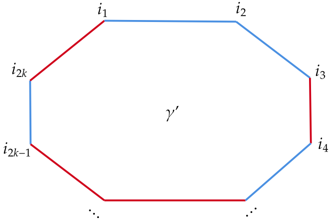


$$(\boldsymbol{X}, \boldsymbol{Y})_{\gamma'} := \boldsymbol{X}_{i_1,i_2} \boldsymbol{X}_{i_2,i_3} \boldsymbol{Y}_{i_3,i_4} \dots \boldsymbol{X}_{i_{2k-1},i_{2k}} \boldsymbol{Y}_{i_{2k},i_1}$$

Algorithmic heuristics

Now we are given two matrices $\boldsymbol{X} = \frac{\lambda}{\sqrt{n}} xx^\top + \boldsymbol{W}$, $\boldsymbol{Y} = \frac{\mu}{\sqrt{n}} yy^\top + \boldsymbol{Z}$.

Natural idea: cycles $\Longrightarrow$ decorated (edge-colored) cycles



$$(\boldsymbol{X}, \boldsymbol{Y})_{\gamma'} := \boldsymbol{X}_{i_1,i_2} \boldsymbol{X}_{i_2,i_3} \boldsymbol{Y}_{i_3,i_4} \ldots \boldsymbol{X}_{i_{2k-1},i_{2k}} \boldsymbol{Y}_{i_{2k},i_1}$$

Approach: consider a suitable **(weighted)** linear combination

$$\sum_{\gamma' : \text{decorated cycle}} \Lambda(\gamma') \cdot (\boldsymbol{X}, \boldsymbol{Y})_{\gamma'} .$$

Algorithmic heuristics

Algorithmic heuristics

- Why $(\star) = \sum_{\gamma':\text{decorated cycle}} \Lambda(\gamma') \cdot (\boldsymbol{X}, \boldsymbol{Y})_{\gamma'}$ outperforms the PLS method/ BBP-threshold?

Algorithmic heuristics

- Why $(\star) = \sum_{\gamma' \text{:decorated cycle}} \Lambda(\gamma') \cdot (\boldsymbol{X}, \boldsymbol{Y})_{\gamma'}$ outperforms the PLS method/ BBP-threshold?

High-level answer: PLS method (studying $\boldsymbol{X}\boldsymbol{Y}^\top$)/BBP-threshold (studying $\boldsymbol{X}$) correspond to two **sub-classes**

$$\left\{ \boldsymbol{X}_{i_1,i_2} \boldsymbol{Y}_{i_2,i_3} \boldsymbol{X}_{i_3,i_4} \dots \boldsymbol{Y}_{i_{2k},i_1} \right\}, \quad \left\{ \boldsymbol{X}_{i_1,i_2} \boldsymbol{X}_{i_2,i_3} \dots \boldsymbol{X}_{i_{2k},i_1} \right\}.$$

Algorithmic heuristics

• Why $(\star) = \sum_{\gamma':\text{decorated cycle}} \Lambda(\gamma') \cdot (\boldsymbol{X}, \boldsymbol{Y})_{\gamma'}$ outperforms the PLS method/ BBP-threshold?

High-level answer: PLS method (studying $\boldsymbol{X}\boldsymbol{Y}^\top$)/BBP-threshold (studying $\boldsymbol{X}$) correspond to two **sub-classes**

$$\left\{ \boldsymbol{X}_{i_1,i_2} \boldsymbol{Y}_{i_2,i_3} \boldsymbol{X}_{i_3,i_4} \ldots \boldsymbol{Y}_{i_{2k},i_1} \right\}, \quad \left\{ \boldsymbol{X}_{i_1,i_2} \boldsymbol{X}_{i_2,i_3} \ldots \boldsymbol{X}_{i_{2k},i_1} \right\}.$$

• Can this statistic be efficiently calculated?

Algorithmic heuristics

- Why $(\star) = \sum_{\gamma':\text{decorated cycle}} \Lambda(\gamma') \cdot (\boldsymbol{X}, \boldsymbol{Y})_{\gamma'}$ outperforms the PLS method/ BBP-threshold?

High-level answer: PLS method (studying $\boldsymbol{X}\boldsymbol{Y}^\top$)/BBP-threshold (studying $\boldsymbol{X}$) correspond to two **sub-classes**

$$\left\{ \boldsymbol{X}_{i_1,i_2} \boldsymbol{Y}_{i_2,i_3} \boldsymbol{X}_{i_3,i_4} \ldots \boldsymbol{Y}_{i_{2k},i_1} \right\}, \quad \left\{ \boldsymbol{X}_{i_1,i_2} \boldsymbol{X}_{i_2,i_3} \ldots \boldsymbol{X}_{i_{2k},i_1} \right\}.$$

- Can this statistic be efficiently calculated?
  - γ' has length $\Theta(\log n)$: brutal-force enumeration takes time $n^{\Theta(\log n)}$ (pseudo-polynomial).

Algorithmic heuristics

- Why $(\star) = \sum_{\gamma':\text{decorated cycle}} \Lambda(\gamma') \cdot (\boldsymbol{X}, \boldsymbol{Y})_{\gamma'}$ outperforms the PLS method/ BBP-threshold?

High-level answer: PLS method (studying $\boldsymbol{X}\boldsymbol{Y}^\top$)/BBP-threshold (studying $\boldsymbol{X}$) correspond to two **sub-classes**

$$\left\{ \boldsymbol{X}_{i_1,i_2} \boldsymbol{Y}_{i_2,i_3} \boldsymbol{X}_{i_3,i_4} \dots \boldsymbol{Y}_{i_{2k},i_1} \right\}, \quad \left\{ \boldsymbol{X}_{i_1,i_2} \boldsymbol{X}_{i_2,i_3} \dots \boldsymbol{X}_{i_{2k},i_1} \right\}.$$

- Can this statistic be efficiently calculated?
  - γ' has length $\Theta(\log n)$: brutal-force enumeration takes time $n^{\Theta(\log n)}$ (pseudo-polynomial).
  - Use the idea of color-coding [Alon-Yuster-Zwick 1995], [Hopkins-Steurer 2017], $(\star)$ can be approximately computed efficiently.

Algorithmic heuristics

- Why $(\star) = \sum_{\gamma':\text{decorated cycle}} \Lambda(\gamma') \cdot (\boldsymbol{X}, \boldsymbol{Y})_{\gamma'}$ outperforms the PLS method/ BBP-threshold?

High-level answer: PLS method (studying $\boldsymbol{X}\boldsymbol{Y}^\top$)/BBP-threshold (studying $\boldsymbol{X}$) correspond to two **sub-classes**

$$\left\{\boldsymbol{X}_{i_1,i_2}\boldsymbol{Y}_{i_2,i_3}\boldsymbol{X}_{i_3,i_4}\ldots\boldsymbol{Y}_{i_{2k},i_1}\right\}, \quad \left\{\boldsymbol{X}_{i_1,i_2}\boldsymbol{X}_{i_2,i_3}\ldots\boldsymbol{X}_{i_{2k},i_1}\right\}.$$

- Can this statistic be efficiently calculated?
  - γ' has length $\Theta(\log n)$: brutal-force enumeration takes time $n^{\Theta(\log n)}$ (pseudo-polynomial).
  - Use the idea of color-coding [Alon-Yuster-Zwick 1995], [Hopkins-Steurer 2017], $(\star)$ can be approximately computed efficiently.
- Detection $\implies$ recovery?

Algorithmic heuristics

- Why $(\star) = \sum_{\gamma':\text{decorated cycle}} \Lambda(\gamma') \cdot (\boldsymbol{X}, \boldsymbol{Y})_{\gamma'}$ outperforms the PLS method/ BBP-threshold?

High-level answer: PLS method (studying $\boldsymbol{X}\boldsymbol{Y}^\top$)/BBP-threshold (studying $\boldsymbol{X}$) correspond to two **sub-classes**

$$\Big\{ \boldsymbol{X}_{i_1,i_2} \boldsymbol{Y}_{i_2,i_3} \boldsymbol{X}_{i_3,i_4} \ldots \boldsymbol{Y}_{i_{2k},i_1} \Big\}, \quad \Big\{ \boldsymbol{X}_{i_1,i_2} \boldsymbol{X}_{i_2,i_3} \ldots \boldsymbol{X}_{i_{2k},i_1} \Big\}.$$

- Can this statistic be efficiently calculated?
  - γ' has length $\Theta(\log n)$: brutal-force enumeration takes time $n^{\Theta(\log n)}$ (pseudo-polynomial).
  - Use the idea of color-coding [Alon-Yuster-Zwick 1995], [Hopkins-Steurer 2017], $(\star)$ can be approximately computed efficiently.

- Detection $\Longrightarrow$ recovery?

High-level answer: decorated cycles $\Longrightarrow$ decorated paths.

Computational lower bounds

Question: is our algorithm optimal?

Computational lower bounds

Question: is our algorithm optimal?

• We cannot deduce the optimality from statistical lower bounds even when the correlation $\rho = 0$:

- $F < 1 \iff$ BBP threshold;
- For certain choice of (sparse) priors π detection/recovery is possible below BBP threshold.

Computational lower bounds

Question: is our algorithm optimal?

• We cannot deduce the optimality from statistical lower bounds even when the correlation $\rho = 0$:

- $F < 1 \iff$ BBP threshold;
- For certain choice of (sparse) priors π detection/recovery is possible below BBP threshold.

Statistical-computational gap: a major challenge in high-dimensional statistics.

Computational lower bounds

Question: is our algorithm optimal?

• We cannot deduce the optimality from statistical lower bounds even when the correlation $\rho = 0$:

- $F < 1 \iff$ BBP threshold;
- For certain choice of (sparse) priors π detection/recovery is possible below BBP threshold.

Statistical-computational gap: a major challenge in high-dimensional statistics.

• Instead, can we show all poly-time algorithms fail when $F < 1$?

Computational lower bounds

Computational lower bounds

- Traditionally, complexity theory studies hardness of computational problems for worst-case instance.

Computational lower bounds

- Traditionally, complexity theory studies hardness of computational problems for worst-case instance.
- Usually certify hardness by reduction: solving this problem $\Longrightarrow$ solving some well-known hard problems.

Computational lower bounds

- Traditionally, complexity theory studies hardness of computational problems for worst-case instance.
- Usually certify hardness by reduction: solving this problem $\Longrightarrow$ solving some well-known hard problems.
- For problems with random instance, we care about the hardness for a typical instance $\Longrightarrow$ reduction is much more difficult than in worst-case instance.

Computational lower bounds

- Traditionally, complexity theory studies hardness of computational problems for worst-case instance.
- Usually certify hardness by reduction: solving this problem $\Longrightarrow$ solving some well-known hard problems.
- For problems with random instance, we care about the hardness for a typical instance $\Longrightarrow$ reduction is much more difficult than in worst-case instance.
- Another popular **evidence** of average-case hardness: show that a wide class of algorithms fail to solve the problem, e.g.,

Computational lower bounds

- Traditionally, complexity theory studies hardness of computational problems for worst-case instance.
- Usually certify hardness by reduction: solving this problem $\Longrightarrow$ solving some well-known hard problems.
- For problems with random instance, we care about the hardness for a typical instance $\Longrightarrow$ reduction is much more difficult than in worst-case instance.
- Another popular **evidence** of average-case hardness: show that a wide class of algorithms fail to solve the problem, e.g.,
  - local algorithms/MCMC;

Computational lower bounds

- Traditionally, complexity theory studies hardness of computational problems for worst-case instance.
- Usually certify hardness by reduction: solving this problem $\Longrightarrow$ solving some well-known hard problems.
- For problems with random instance, we care about the hardness for a typical instance $\Longrightarrow$ reduction is much more difficult than in worst-case instance.
- Another popular **evidence** of average-case hardness: show that a wide class of algorithms fail to solve the problem, e.g.,
    - local algorithms/MCMC;
    - smooth algorithms;

Computational lower bounds

- Traditionally, complexity theory studies hardness of computational problems for worst-case instance.
- Usually certify hardness by reduction: solving this problem $\Longrightarrow$ solving some well-known hard problems.
- For problems with random instance, we care about the hardness for a typical instance $\Longrightarrow$ reduction is much more difficult than in worst-case instance.
- Another popular **evidence** of average-case hardness: show that a wide class of algorithms fail to solve the problem, e.g.,
    - local algorithms/MCMC;
    - smooth algorithms;
    - sum-of-squares method;

Computational lower bounds

- Traditionally, complexity theory studies hardness of computational problems for worst-case instance.
- Usually certify hardness by reduction: solving this problem $\Longrightarrow$ solving some well-known hard problems.
- For problems with random instance, we care about the hardness for a typical instance $\Longrightarrow$ reduction is much more difficult than in worst-case instance.
- Another popular **evidence** of average-case hardness: show that a wide class of algorithms fail to solve the problem, e.g.,
  - local algorithms/MCMC;
  - smooth algorithms;
  - sum-of-squares method;
  - statistical physics method/belief propagation;

Computational lower bounds

- Traditionally, complexity theory studies hardness of computational problems for worst-case instance.
- Usually certify hardness by reduction: solving this problem $\Longrightarrow$ solving some well-known hard problems.
- For problems with random instance, we care about the hardness for a typical instance $\Longrightarrow$ reduction is much more difficult than in worst-case instance.
- Another popular **evidence** of average-case hardness: show that a wide class of algorithms fail to solve the problem, e.g.,
  - local algorithms/MCMC;
  - smooth algorithms;
  - sum-of-squares method;
  - statistical physics method/belief propagation;
  - this talk: low-degree polynomials; ...

The low-degree method

The low-degree method

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

- Null distribution $\mathbb{Q}$ (i.e., two independent noise matrices).
- Planted distribution $\mathbb{P}$ (i.e., two matrices with correlated spikes);

The low-degree method

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

- Null distribution $\mathbb{Q}$ (i.e., two independent noise matrices).
- Planted distribution $\mathbb{P}$ (i.e., two matrices with correlated spikes);

Degree-D test: multivariate polynomials $f : (\boldsymbol{X}, \boldsymbol{Y}) \longrightarrow \mathbb{R}$ of degree $D = D_n$

The low-degree method

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

- Null distribution $\mathbb{Q}$ (i.e., two independent noise matrices).
- Planted distribution $\mathbb{P}$ (i.e., two matrices with correlated spikes);

Degree-D test: multivariate polynomials $f : (\boldsymbol{X}, \boldsymbol{Y}) \longrightarrow \mathbb{R}$ of degree $D = D_n$

"Success": $f = f_n$ separates $\mathbb{P}$ and $\mathbb{Q}$ if

$$\sqrt{\max\{\mathrm{Var}_{\mathbb{P}}(f), \mathrm{Var}_{\mathbb{Q}}(f)\}} = o(1) \cdot \left|\mathbb{E}_{\mathbb{P}}[f] - \mathbb{E}_{\mathbb{Q}}[f]\right| .$$

The low-degree method

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

- Null distribution $\mathbb{Q}$ (i.e., two independent noise matrices).
- Planted distribution $\mathbb{P}$ (i.e., two matrices with correlated spikes);

Degree-D test: multivariate polynomials $f : (\boldsymbol{X}, \boldsymbol{Y}) \longrightarrow \mathbb{R}$ of degree $D = D_n$

"Success": $f = f_n$ separates $\mathbb{P}$ and $\mathbb{Q}$ if

$$\sqrt{\max\{\mathrm{Var}_{\mathbb{P}}(f), \mathrm{Var}_{\mathbb{Q}}(f)\}} = o(1) \cdot \left|\mathbb{E}_{\mathbb{P}}[f] - \mathbb{E}_{\mathbb{Q}}[f]\right|.$$

Heuristics: failure of degree-D polynomials $\stackrel{\text{evidence}}{\Longrightarrow}$ failure of algorithms with running time $n^{D/\operatorname{polylog}(n)}$.

The low-degree method

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

- Null distribution $\mathbb{Q}$ (i.e., two independent noise matrices).
- Planted distribution $\mathbb{P}$ (i.e., two matrices with correlated spikes);

Degree-D test: multivariate polynomials $f : (\boldsymbol{X}, \boldsymbol{Y}) \longrightarrow \mathbb{R}$ of degree $D = D_n$

"Success": $f = f_n$ separates $\mathbb{P}$ and $\mathbb{Q}$ if

$$\sqrt{\max\{\mathrm{Var}_{\mathbb{P}}(f), \mathrm{Var}_{\mathbb{Q}}(f)\}} = o(1) \cdot \left|\mathbb{E}_{\mathbb{P}}[f] - \mathbb{E}_{\mathbb{Q}}[f]\right| .$$

Heuristics: failure of degree-D polynomials $\overset{\text{evidence}}{\Longrightarrow}$ failure of algorithms with running time $n^{D/\operatorname{polylog}(n)}$.

Usually prove the "failure" of degree-D polynomials by showing the following bound on the low-degree advantage:

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

The low-degree method

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

likelihood ratio: $L = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$

The low-degree method

$$\mathrm{Adv}_{\leq D}(\mathbb{P}, \mathbb{Q}) := \max_{\deg(f) \leq D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \quad \text{likelihood ratio: } L = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$$

$$= \max_{\deg(f) \leq D} \frac{\mathbb{E}_{\mathbb{Q}}[L \cdot f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \quad \langle f, g \rangle := \mathbb{E}_{\mathbb{Q}}[f \cdot g]$$

The low-degree method

$$\mathrm{Adv}_{\leq D}(\mathbb{P},\mathbb{Q}) := \max_{\deg(f)\leq D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}}$$ likelihood ratio: $L = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$

$$= \max_{\deg(f)\leq D} \frac{\mathbb{E}_{\mathbb{Q}}[L\cdot f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}}$$ $\langle f, g\rangle := \mathbb{E}_{\mathbb{Q}}[f\cdot g]$

$$= \max_{\deg(f)\leq D} \frac{\langle L,f\rangle}{\|f\|}$$ $\|f\| = \sqrt{\langle f, f\rangle}$

The low-degree method

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

likelihood ratio: $L = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$

$$= \max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{Q}}[L\cdot f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}}$$

$$\langle f, g\rangle := \mathbb{E}_{\mathbb{Q}}[f \cdot g]$$

$$= \max_{\deg(f)\le D} \frac{\langle L,f\rangle}{\|f\|}$$

$$\|f\| = \sqrt{\langle f, f\rangle}$$

$$= \|L_{\le D}\| \,.$$

The low-degree method

$$\mathrm{Adv}_{\leq D}(\mathbb{P},\mathbb{Q}) := \max_{\deg(f)\leq D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \text{likelihood ratio: } L = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$$

$$= \max_{\deg(f)\leq D} \frac{\mathbb{E}_{\mathbb{Q}}[L\cdot f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \langle f, g\rangle := \mathbb{E}_{\mathbb{Q}}[f\cdot g]$$

$$= \max_{\deg(f)\leq D} \frac{\langle L, f\rangle}{\|f\|} \qquad \|f\| = \sqrt{\langle f, f\rangle}$$

$$= \|L_{\leq D}\| .$$

- Maximizer: $f = L_{\leq D} :=$ projection of L onto degree-D subspace.

The low-degree method

$$\mathrm{Adv}_{\le D}(\mathbb{P},\mathbb{Q}) := \max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{P}}[f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \text{likelihood ratio: } L = \frac{\mathrm{d}\mathbb{P}}{\mathrm{d}\mathbb{Q}}$$

$$= \max_{\deg(f)\le D} \frac{\mathbb{E}_{\mathbb{Q}}[L\cdot f]}{\sqrt{\mathbb{E}_{\mathbb{Q}}[f^2]}} \qquad \langle f, g\rangle := \mathbb{E}_{\mathbb{Q}}[f\cdot g]$$

$$= \max_{\deg(f)\le D} \frac{\langle L, f\rangle}{\|f\|} \qquad \|f\| = \sqrt{\langle f, f\rangle}$$

$$= \|L_{\le D}\| \,.$$

- Maximizer: $f = L_{\le D} :=$ projection of L onto degree-D subspace.
- Since $\mathbb{Q}$ is a product measure, we can directly calculate the projection using orthogonal polynomials w.r.t. $\mathbb{Q}$.

Advantages of low-degree method

Advantages of low-degree method

- Possible to calculate $\|L_{\leq D}\|$ in many problems.

Advantages of low-degree method

- Possible to calculate $\|L_{\leq D}\|$ in many problems.
- Predictions seem "correct"! e.g., for planted clique, sparse/tensor PCA, community detection, etc. It is the case that
  - The best known poly-time algorithms are captured by low-degree polynomials (spectral/AMP/...);
  - Low-degree polynomials fail in the conjectured "hard" regime.

Advantages of low-degree method

- Possible to calculate $\|L_{\leq D}\|$ in many problems.
- Predictions seem "correct"! e.g., for planted clique, sparse/tensor PCA, community detection, etc. It is the case that
  - The best known poly-time algorithms are captured by low-degree polynomials (spectral/AMP/...);
  - Low-degree polynomials fail in the conjectured "hard" regime.
- (Relatively) simple:
  - Much simpler than sum-of-squares lower bound;
  - Rules out detection rather than certification.

Advantages of low-degree method

- Possible to calculate $\|L_{\leq D}\|$ in many problems.
- Predictions seem "correct"! e.g., for planted clique, sparse/tensor PCA, community detection, etc. It is the case that
  - The best known poly-time algorithms are captured by low-degree polynomials (spectral/AMP/...);
  - Low-degree polynomials fail in the conjectured "hard" regime.
- (Relatively) simple:
  - Much simpler than sum-of-squares lower bound;
  - Rules out detection rather than certification.
- General: not much assumptions on $\mathbb{P}, \mathbb{Q}$.

Advantages of low-degree method

- Possible to calculate $\|L_{\leq D}\|$ in many problems.
- Predictions seem "correct"! e.g., for planted clique, sparse/tensor PCA, community detection, etc. It is the case that
  - The best known poly-time algorithms are captured by low-degree polynomials (spectral/AMP/...);
  - Low-degree polynomials fail in the conjectured "hard" regime.
- (Relatively) simple:
  - Much simpler than sum-of-squares lower bound;
  - Rules out detection rather than certification.
- General: not much assumptions on $\mathbb{P}, \mathbb{Q}$.
- By varying D, can explore running times other than polynomial.

Advantages of low-degree method

- Possible to calculate $\|L_{\leq D}\|$ in many problems.
- Predictions seem "correct"! e.g., for planted clique, sparse/tensor PCA, community detection, etc. It is the case that
  - The best known poly-time algorithms are captured by low-degree polynomials (spectral/AMP/...);
  - Low-degree polynomials fail in the conjectured "hard" regime.
- (Relatively) simple:
  - Much simpler than sum-of-squares lower bound;
  - Rules out detection rather than certification.
- General: not much assumptions on $\mathbb{P}, \mathbb{Q}$.
- By varying D, can explore running times other than polynomial.
- No ingenuity required.

Advantages of low-degree method

- Possible to calculate $\|L_{\le D}\|$ in many problems.
- Predictions seem "correct"! e.g., for planted clique, sparse/tensor PCA, community detection, etc. It is the case that
  - The best known poly-time algorithms are captured by low-degree polynomials (spectral/AMP/...);
  - Low-degree polynomials fail in the conjectured "hard" regime.
- (Relatively) simple:
  - Much simpler than sum-of-squares lower bound;
  - Rules out detection rather than certification.
- General: not much assumptions on $\mathbb{P}, \mathbb{Q}$.
- By varying D, can explore running times other than polynomial.
- No ingenuity required.
- Interpretable.

Our results

Theorem (L.25)

(1) For the correlated spiked Wigner model, suppose that $F(\lambda, \mu, \rho, 1) < 1$*. Then all degree* $D = n^{o(1)}$ *polynomials fails to strongly separate* $\mathbb{P}$ *and* $\mathbb{Q}$*.*
(2) For the correlated spiked Wishart model, suppose that $\frac{n}{N} = \gamma$ *for some* $\gamma = \Theta(1)$ *and* $F(\lambda, \mu, \rho, \gamma) < 1$*. Then all degree* $D = n^{o(1)}$ *polynomials fails to strongly separate* $\mathbb{P}$ *and* $\mathbb{Q}$*.*

Our results

Theorem (L.25)

(1) For the correlated spiked Wigner model, suppose that $F(\lambda, \mu, \rho, 1) < 1$. Then all degree $D = n^{o(1)}$ polynomials fails to strongly separate $\mathbb{P}$ and $\mathbb{Q}$.
(2) For the correlated spiked Wishart model, suppose that $\frac{n}{N} = \gamma$ for some $\gamma = \Theta(1)$ and $F(\lambda, \mu, \rho, \gamma) < 1$. Then all degree $D = n^{o(1)}$ polynomials fails to strongly separate $\mathbb{P}$ and $\mathbb{Q}$.

Suggests that detection requires stretched-exponential time below the threshold $F = 1$.

Our results

Theorem (L.25)

(1) For the correlated spiked Wigner model, suppose that $F(\lambda, \mu, \rho, 1) < 1$. Then all degree $D = n^{o(1)}$ polynomials fails to strongly separate $\mathbb{P}$ and $\mathbb{Q}$.
(2) For the correlated spiked Wishart model, suppose that $\frac{n}{N} = \gamma$ for some $\gamma = \Theta(1)$ and $F(\lambda, \mu, \rho, \gamma) < 1$. Then all degree $D = n^{o(1)}$ polynomials fails to strongly separate $\mathbb{P}$ and $\mathbb{Q}$.

Suggests that detection requires stretched-exponential time below the threshold $F = 1$.

Recovery should be harder than detection $\Longrightarrow$ recovery also requires stretched-exponential time.

Our results

Theorem (L.25)

(1) For the correlated spiked Wigner model, suppose that $F(\lambda, \mu, \rho, 1) < 1$. Then all degree $D = n^{o(1)}$ polynomials fails to strongly separate $\mathbb{P}$ and $\mathbb{Q}$.
(2) For the correlated spiked Wishart model, suppose that $\frac{n}{N} = \gamma$ for some $\gamma = \Theta(1)$ and $F(\lambda, \mu, \rho, \gamma) < 1$. Then all degree $D = n^{o(1)}$ polynomials fails to strongly separate $\mathbb{P}$ and $\mathbb{Q}$.

Suggests that detection requires stretched-exponential time below the threshold $F = 1$.

Recovery should be harder than detection $\Longrightarrow$ recovery also requires stretched-exponential time.

In conclusion, suggests that $F = 1$ is the exact computational threshold (and thus our algorithm is optimal).

Proof sketch

Standard calculation yields that

$$\begin{aligned}\mathsf{Adv}_{\leq D}(\mathbb{P}, \mathbb{Q}) &= \|L_{\leq D}\| \\ &= \mathbb{E}_{(x,y),(x',y')\sim\pi}\left[\exp_{\leq D}\left(\frac{\lambda^2}{2}\Big(\frac{\langle x,x'\rangle}{\sqrt{n}}\Big)^2 + \frac{\mu^2}{2}\Big(\frac{\langle y,y'\rangle}{\sqrt{n}}\Big)^2\right)\right].\end{aligned}$$

Proof sketch

Standard calculation yields that

$$\begin{aligned}\mathsf{Adv}_{\le D}(\mathbb{P},\mathbb{Q}) &= \|L_{\le D}\| \\ &= \mathbb{E}_{(x,y),(x',y')\sim\pi}\left[\exp_{\le D}\left(\frac{\lambda^2}{2}\Big(\frac{\langle x,x'\rangle}{\sqrt{n}}\Big)^2+\frac{\mu^2}{2}\Big(\frac{\langle y,y'\rangle}{\sqrt{n}}\Big)^2\right)\right].\end{aligned}$$

Our intuition in bounding $(\star)$: by CLT

$$\left(\frac{\langle x,x'\rangle}{\sqrt{n}},\frac{\langle y,y'\rangle}{\sqrt{n}}\right)\approx(U,V)\sim\mathcal{N}\left(0,\begin{pmatrix}1&\rho^2\\ \rho^2&1\end{pmatrix}\right).$$

So $(\star)\approx\mathbb{E}[\exp_{\le D}(\frac{\lambda^2U^2+\mu^2V^2}{2})]$, which is $O(1)$ if and only if $F(\lambda,\mu,\rho)\le 1$.

Proof sketch

Standard calculation yields that

$$\begin{aligned}\mathsf{Adv}_{\le D}(\mathbb{P},\mathbb{Q}) &= \|L_{\le D}\| \\ &= \mathbb{E}_{(x,y),(x',y')\sim\pi}\left[\exp_{\le D}\left(\frac{\lambda^2}{2}\left(\frac{\langle x,x'\rangle}{\sqrt{n}}\right)^2 + \frac{\mu^2}{2}\left(\frac{\langle y,y'\rangle}{\sqrt{n}}\right)^2\right)\right].\end{aligned}$$

Our intuition in bounding $(\star)$: by CLT

$$\left(\frac{\langle x,x'\rangle}{\sqrt{n}}, \frac{\langle y,y'\rangle}{\sqrt{n}}\right) \approx (U,V) \sim \mathcal{N}\left(0, \begin{pmatrix} 1 & \rho^2 \\ \rho^2 & 1 \end{pmatrix}\right).$$

So $(\star) \approx \mathbb{E}[\exp_{\le D}(\frac{\lambda^2 U^2 + \mu^2 V^2}{2})]$, which is $O(1)$ if and only if $F(\lambda,\mu,\rho) \le 1$.

Approach: verify such Gaussian approximation using a careful Lindeberg's argument.

Summary and future perspectives

Take-home message:

Summary and future perspectives

Take-home message:

- For the correlated spiked model, we determine the precise algorithmic phase transition (in the low-degree polynomial framework).

Summary and future perspectives

Take-home message:

- For the correlated spiked model, we determine the precise algorithmic phase transition (in the low-degree polynomial framework).
- Our algorithmic approach exploits the power of counting decorated cycles/paths.

Summary and future perspectives

Take-home message:

- For the correlated spiked model, we determine the precise algorithmic phase transition (in the low-degree polynomial framework).
- Our algorithmic approach exploits the power of counting decorated cycles/paths.
- Our computational lower bound use a interpolation argument in the low-degree analysis.

Summary and future perspectives

Take-home message:

- For the correlated spiked model, we determine the precise algorithmic phase transition (in the low-degree polynomial framework).
- Our algorithmic approach exploits the power of counting decorated cycles/paths.
- Our computational lower bound use a interpolation argument in the low-degree analysis.

Open problems:

Summary and future perspectives

Take-home message:

- For the correlated spiked model, we determine the precise algorithmic phase transition (in the low-degree polynomial framework).
- Our algorithmic approach exploits the power of counting decorated cycles/paths.
- Our computational lower bound use a interpolation argument in the low-degree analysis.

Open problems:

- Practical algorithms;

Summary and future perspectives

Take-home message:

- For the correlated spiked model, we determine the precise algorithmic phase transition (in the low-degree polynomial framework).
- Our algorithmic approach exploits the power of counting decorated cycles/paths.
- Our computational lower bound use a interpolation argument in the low-degree analysis.

Open problems:

- Practical algorithms;
- General-rank model and other multi-modal inference problems.

Summary and future perspectives

Take-home message:

- For the correlated spiked model, we determine the precise algorithmic phase transition (in the low-degree polynomial framework).
- Our algorithmic approach exploits the power of counting decorated cycles/paths.
- Our computational lower bound use a interpolation argument in the low-degree analysis.

Open problems:

- Practical algorithms;
- General-rank model and other multi-modal inference problems.

Reference:

Zhangsong Li. The Algorithmic Phase Transition for Correlated Spiked Models. arXiv:2511.06040.