Asymmetry Helps: Estimation and Inference from Asymmetric and Heteroscedastic Noise

Chen Cheng

with Yuxin Chen (Princeton), Jianqing Fan (Princeton), Yuting Wei (CMU)

Department of Statistics, Stanford University

C. Cheng, Y. Wei, Y. Chen, "Inference for linear forms of eigenvectors under minimal eigenvalue separation: Asymmetry and heteroscedasticity", arXiv:2001.04620, 2020.

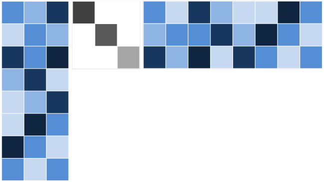
Y. Chen, C. Cheng, J. Fan, "Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices", arXiv:1811.12804, 2018. (alphabetical order)
— accepted to Annals of Statistics, 2020.

1. Introduction

2. Estimation

3. Inference

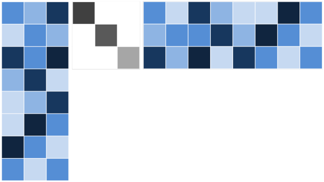
Problem: eigenvalue & eigenvector estimation



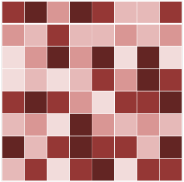
$M^\star$: symmetric low-rank matrix

- Rank-r matrix: $M^\star = \sum_{l=1}^{r} \lambda_l^\star u_l^\star u_l^{\star\top} \in \mathbb{R}^{n\times n}$.

Problem: eigenvalue & eigenvector estimation



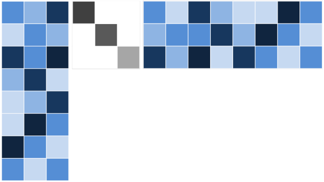
+



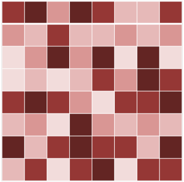
$\boldsymbol{M}^\star$: symmetric low-rank matrix $\qquad$ $\boldsymbol{H}$: $[H_{ij}]_{1\le i,j\le n}$ independent noise.

- Rank-r matrix: $\boldsymbol{M}^\star = \sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} \in \mathbb{R}^{n\times n}$.
- Observed data: $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$.

Problem: eigenvalue & eigenvector estimation



+

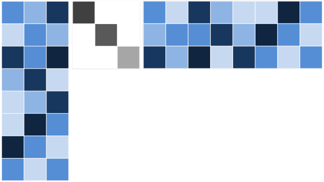


$M^\star$: symmetric low-rank matrix

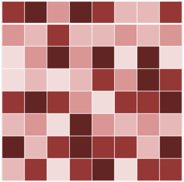
H: $[H_{ij}]_{1\le i,j\le n}$ independent noise.

- Rank-r matrix: $M^\star = \sum_{l=1}^{r} \lambda_l^\star u_l^\star u_l^{\star\top} \in \mathbb{R}^{n\times n}$.
- Observed data: $M = M^\star + H$.
- Applications:
  - Matrix denoising and completion.
  - Stochastic block model.
  - Ranking from pairwise comparisons.
  - ...

Problem: eigenvalue & eigenvector estimation



+

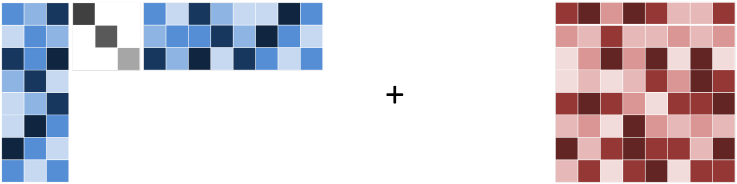


$\boldsymbol{M}^\star$: symmetric low-rank matrix

$\boldsymbol{H}$: $[H_{ij}]_{1\le i,j\le n}$ independent noise.

- Rank-r matrix: $\boldsymbol{M}^\star = \sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} \in \mathbb{R}^{n\times n}$.
- Observed data: $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$.
- **Goal**: retrieve eigenvalue & eigenvector information from $\boldsymbol{M}$.

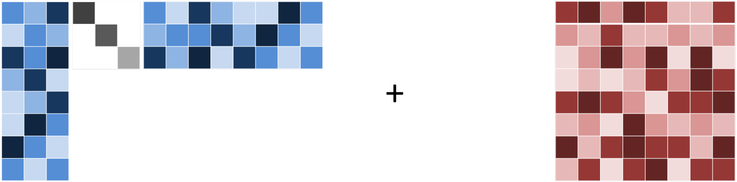
Problem: eigenvalue & eigenvector estimation



$\boldsymbol{M}^\star$: symmetric low-rank matrix $\qquad$ $\boldsymbol{H}$: $[H_{ij}]_{1\le i,j\le n}$ independent noise.

- Rank-r matrix: $\boldsymbol{M}^\star = \sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} \in \mathbb{R}^{n\times n}$.
- Observed data: $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$.
- **Goal**: retrieve eigenvalue & eigenvector information from $\boldsymbol{M}$.
- **Quantity of interest**: eigenvalue error; eigenvector ℓ_2 error, ℓ_∞ error, error for any linear function $\boldsymbol{a}^\top \boldsymbol{u}_l$.

Problem: eigenvalue & eigenvector estimation



$\boldsymbol{M}^\star$: symmetric low-rank matrix $\qquad$ $\boldsymbol{H}$: $[H_{ij}]_{1\le i,j\le n}$ independent noise.

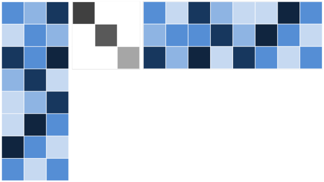
- Rank-r matrix: $\boldsymbol{M}^\star = \sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} \in \mathbb{R}^{n\times n}$.
- Observed data: $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$.
- **Goal**: retrieve eigenvalue & eigenvector information from $\boldsymbol{M}$.
- **Quantity of interest**: eigenvalue error; eigenvector ℓ_2 error, ℓ_∞ error, error for any linear function $\boldsymbol{a}^\top \boldsymbol{u}_l$.
- Strategy:
  - SVD on $\boldsymbol{M}$ or $\left(\boldsymbol{M} + \boldsymbol{M}^\top\right)/2$?

Problem: eigenvalue & eigenvector estimation

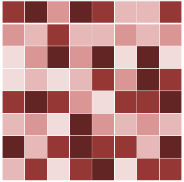
$\boldsymbol{M}^{\star}$: symmetric low-rank matrix $\qquad$ $\boldsymbol{H}$: $[H_{ij}]_{1\le i,j\le n}$ independent noise.

- Rank-r matrix: $\boldsymbol{M}^{\star} = \sum_{l=1}^{r} \lambda_l^{\star} \boldsymbol{u}_l^{\star} \boldsymbol{u}_l^{\star\top} \in \mathbb{R}^{n\times n}$.
- Observed data: $\boldsymbol{M} = \boldsymbol{M}^{\star} + \boldsymbol{H}$.
- **Goal**: retrieve eigenvalue & eigenvector information from $\boldsymbol{M}$.
- **Quantity of interest**: eigenvalue error; eigenvector ℓ_2 error, ℓ_∞ error, error for any linear function $\boldsymbol{a}^{\top}\boldsymbol{u}_l$.
- Strategy:
  - SVD on $\boldsymbol{M}$ or $\left(\boldsymbol{M} + \boldsymbol{M}^{\top}\right)/2$?
  - Eigen-decomposition on $\boldsymbol{M}$?

Problem: eigenvalue & eigenvector estimation



+



$\boldsymbol{M}^\star$: symmetric low-rank matrix

$\boldsymbol{H}$: $[H_{ij}]_{1 \le i,j \le n}$ independent noise.

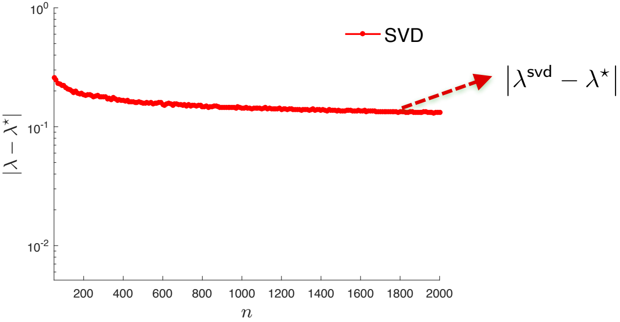
- Rank-r matrix: $\boldsymbol{M}^\star = \sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top} \in \mathbb{R}^{n \times n}$.
- Observed data: $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$.
- **Goal**: retrieve eigenvalue & eigenvector information from $\boldsymbol{M}$.
- **Quantity of interest**: eigenvalue error; eigenvector ℓ_2 error, ℓ_∞ error, error for any linear function $\boldsymbol{a}^\top \boldsymbol{u}_l$.
- Strategy:
  - SVD on $\boldsymbol{M}$ or $\left(\boldsymbol{M} + \boldsymbol{M}^\top\right)/2$? (Popular strategies)
  - Eigen-decomposition on $\boldsymbol{M}$? (Much less widely used)

A curious experiment: Gaussian noise

- $\boldsymbol{M} = \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} + \boldsymbol{H}$, $H_{i,j}$ i.i.d. $\mathcal{N}(0, \sigma^2)$, $\sigma = \frac{1}{\sqrt{n \log n}}$.
- Estimate the leading eigenvalue $\lambda^\star = 1$.
- SVD on $\boldsymbol{M}$ vs Eigen-decomposition on $\boldsymbol{M}$.

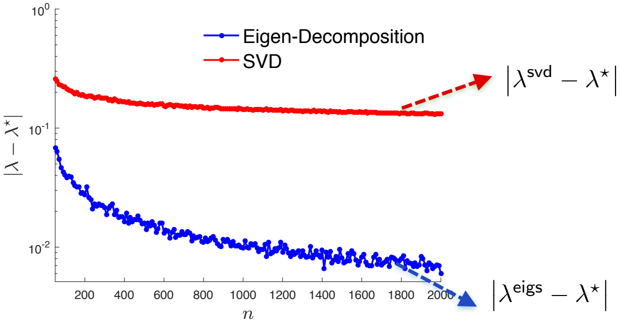
A curious experiment: Gaussian noise

- $\boldsymbol{M} = \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} + \boldsymbol{H}$, $H_{i,j}$ i.i.d. $\mathcal{N}(0, \sigma^2)$, $\sigma = \frac{1}{\sqrt{n \log n}}$.
- Estimate the leading eigenvalue $\lambda^\star = 1$.
- SVD on $\boldsymbol{M}$ vs Eigen-decomposition on $\boldsymbol{M}$.



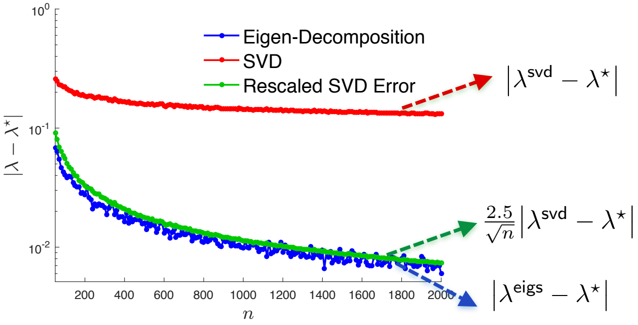
A curious experiment: Gaussian noise

- $\boldsymbol{M} = \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} + \boldsymbol{H}$, $H_{i,j}$ i.i.d. $\mathcal{N}(0, \sigma^2)$, $\sigma = \frac{1}{\sqrt{n \log n}}$.
- Estimate the leading eigenvalue $\lambda^\star = 1$.
- SVD on $\boldsymbol{M}$ vs Eigen-decomposition on $\boldsymbol{M}$.



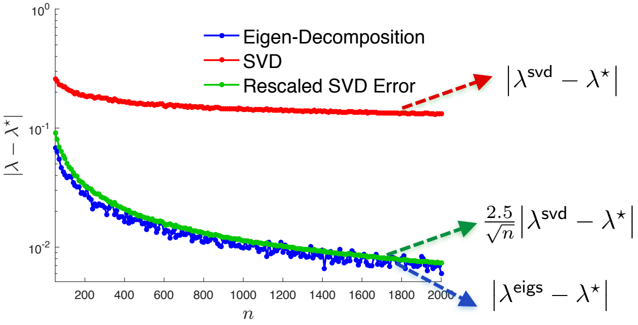
A curious experiment: Gaussian noise

- $\boldsymbol{M} = \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} + \boldsymbol{H}$, $H_{i,j}$ i.i.d. $\mathcal{N}(0, \sigma^2)$, $\sigma = \frac{1}{\sqrt{n \log n}}$.
- Estimate the leading eigenvalue $\lambda^\star = 1$.
- SVD on $\boldsymbol{M}$ vs Eigen-decomposition on $\boldsymbol{M}$.



A curious experiment: Gaussian noise

- $\boldsymbol{M} = \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} + \boldsymbol{H}$, $H_{i,j}$ i.i.d. $\mathcal{N}(0, \sigma^2)$, $\sigma = \frac{1}{\sqrt{n \log n}}$.
- Estimate the leading eigenvalue $\lambda^\star = 1$.
- SVD on $\boldsymbol{M}$ vs Eigen-decomposition on $\boldsymbol{M}$.



- Wait! But we should know everything under Gaussian noise!

A curious experiment: Gaussian noise

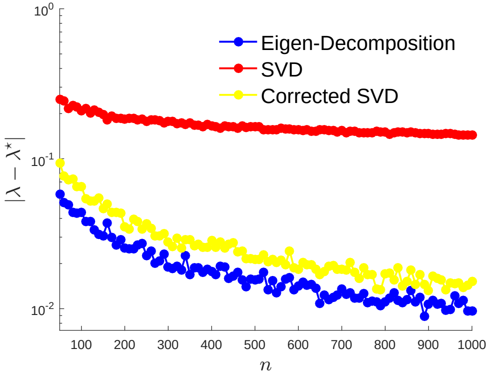
- Indeed, for SVD from i.i.d. Gaussian noise, one can use a corrected singular value (Benaych-Georges and Nadakuditi, 2012)

$$\lambda^{\text{svd,c}} = \lambda^{\text{svd}} - n\sigma^2 = f(\sigma, \lambda^{\text{svd}}).$$

A curious experiment: Gaussian noise

- Indeed, for SVD from i.i.d. Gaussian noise, one can use a corrected singular value (Benaych-Georges and Nadakuditi, 2012)

$$\lambda^{\text{svd,c}} = \lambda^{\text{svd}} - n\sigma^2 = f(\sigma, \lambda^{\text{svd}}).$$



A curious experiment: Gaussian noise

- Indeed, for SVD from i.i.d. Gaussian noise, one can use a corrected singular value (Benaych-Georges and Nadakuditi, 2012)

$$\lambda^{\mathrm{svd,c}} = \lambda^{\mathrm{svd}} - n\sigma^2 = f(\sigma, \lambda^{\mathrm{svd}}).$$

- For heteroscedastic Gaussian noise, the correction formula is far more complicated (Bryc et al., 2018)

The expression for the constant shift depends on the variances of the entries of $\mathbf{C}$. To write the expression we introduce a diagonal $N \times N$ matrix $\mathbf{\Delta}_R = \mathbb{E}(\mathbf{C}^*\mathbf{C})/M$ and a diagonal $M \times M$ matrix $\mathbf{\Delta}_S = \mathbb{E}(\mathbf{C}\mathbf{C}^*)/N$. The diagonal entries of these matrices are

$$[\mathbf{\Delta}_R]_{j,j} = \frac{1}{M}\sum_{i=1}^{M} \sigma_{i,j}^2, \quad [\mathbf{\Delta}_S]_{i,i} = \frac{1}{N}\sum_{j=1}^{N} \sigma_{i,j}^2.$$

Let $\mathbf{\Sigma}_r = \mathbf{\Sigma}_R^{(N)}$ and $\mathbf{\Sigma}_S = \mathbf{\Sigma}_S^{(N)}$ be deterministic $K \times K$ matrices given by

(1.11) $$\mathbf{\Sigma}_R = \mathbf{G}^*\mathbf{\Delta}_R\mathbf{G}$$

and

(1.12) $$\mathbf{\Sigma}_S = \mathbf{F}^*\mathbf{\Delta}_S\mathbf{F}$$

Define

(1.13) $$m_r^{(N)} = \frac{1}{2}\left[\frac{\sqrt{c}}{\gamma_r^3 MN}\mathbf{\Sigma}_R + \frac{1}{\sqrt{c}\gamma_r}\mathbf{\Sigma}_S\right]_{r,r}.$$

Theorem 1.1. *With the above notation, there exist* $\varepsilon_1^{(N)} \to 0, \dots, \varepsilon_K^{(N)} \to 0$ *in probability such that for* $1 \le r \le K$ *we have*

(1.14) $$\lambda_r = \rho_r^{(N)} + Z_r^{(N)} + m_r^{(N)} + \varepsilon_r^{(N)}.$$

Another experiment: matrix completion

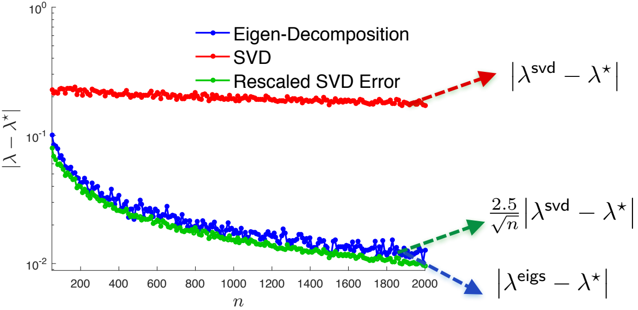
- What if the noise is heteroscedastic we do not have prior knowledge about?
- $\boldsymbol{M}^\star = \boldsymbol{u}^\star \boldsymbol{u}^{\star\top}$, $M_{ij} = \begin{cases} \frac{1}{p} M_{ij}^\star, & \text{with prob. } p, \\ 0, & \text{else,} \end{cases}$ $p = \frac{3\log n}{n}$.

  $\boldsymbol{H} = \boldsymbol{M} - \boldsymbol{M}^\star$.

Another experiment: matrix completion

- What if the noise is heteroscedastic we do not have prior knowledge about?
- $\boldsymbol{M}^\star = \boldsymbol{u}^\star\boldsymbol{u}^{\star\top}$, $M_{ij} = \begin{cases} \frac{1}{p}M^\star_{ij}, & \text{with prob. } p, \\ 0, & \text{else,} \end{cases}$ $p = \frac{3\log n}{n}$.

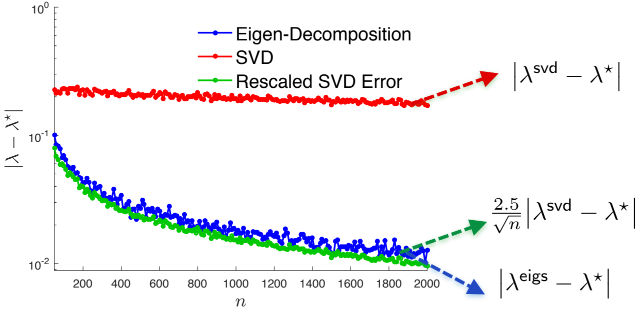
  $\boldsymbol{H} = \boldsymbol{M} - \boldsymbol{M}^\star$.



Another experiment: matrix completion

- What if the noise is heteroscedastic we do not have prior knowledge about?
- $\boldsymbol{M}^\star = \boldsymbol{u}^\star \boldsymbol{u}^{\star\top}$, $M_{ij} = \begin{cases} \frac{1}{p} M^\star_{ij}, & \text{with prob. } p, \\ 0, & \text{else,} \end{cases}$ $p = \frac{3 \log n}{n}$.

  $\boldsymbol{H} = \boldsymbol{M} - \boldsymbol{M}^\star$.



- Eigen-decomposition is nearly unbiased regardless of the noise distribution!

One more experiment: heteroscedastic Gaussian noise

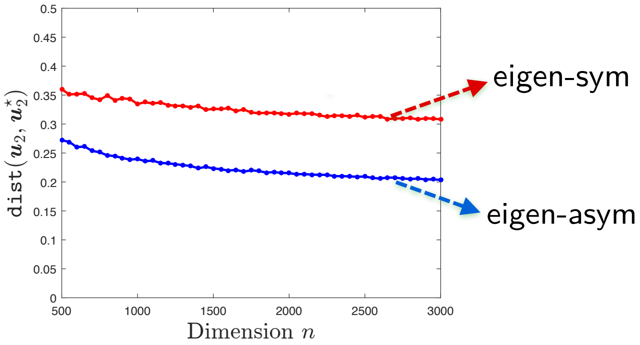
- $\boldsymbol{M} = \boldsymbol{u}_1^\star \boldsymbol{u}_1^{\star\top} + 0.95\boldsymbol{u}_2^\star \boldsymbol{u}_2^{\star\top} + \boldsymbol{H}$, $\boldsymbol{u}_1^\star = \frac{1}{\sqrt{n}}\begin{bmatrix} \mathbf{1}_{n/2} \\ \mathbf{1}_{n/2} \end{bmatrix}$; $\boldsymbol{u}_2^\star = \frac{1}{\sqrt{n}}\begin{bmatrix} \mathbf{1}_{n/2} \\ -\mathbf{1}_{n/2} \end{bmatrix}$
- $[\mathsf{Var}(H_{ij})]_{i,j} \approx \frac{1}{n \log n}\left(\begin{bmatrix} \mathbf{1}\mathbf{1}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \frac{1}{100}\mathbf{1}\mathbf{1}^\top\right)$

One more experiment: heteroscedastic Gaussian noise

- $\boldsymbol{M} = \boldsymbol{u}_1^\star \boldsymbol{u}_1^{\star\top} + 0.95\boldsymbol{u}_2^\star \boldsymbol{u}_2^{\star\top} + \boldsymbol{H}$, $\boldsymbol{u}_1^\star = \frac{1}{\sqrt{n}}\begin{bmatrix} \mathbf{1}_{n/2} \\ \mathbf{1}_{n/2} \end{bmatrix}$; $\boldsymbol{u}_2^\star = \frac{1}{\sqrt{n}}\begin{bmatrix} \mathbf{1}_{n/2} \\ -\mathbf{1}_{n/2} \end{bmatrix}$
- $[\mathsf{Var}(H_{ij})]_{i,j} \approx \frac{1}{n \log n}\left(\begin{bmatrix} \mathbf{1}\mathbf{1}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \frac{1}{100}\mathbf{1}\mathbf{1}^\top\right)$
- Estimate $\boldsymbol{u}_2^\star$ by eigen-decomposition on the symmetrized data $(\boldsymbol{M} + \boldsymbol{M}^\top)/2$ and the original data $\boldsymbol{M}$.

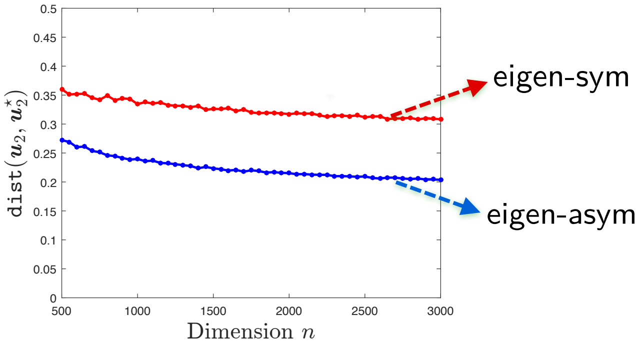
One more experiment: heteroscedastic Gaussian noise

- $\boldsymbol{M} = \boldsymbol{u}_1^\star \boldsymbol{u}_1^{\star\top} + 0.95\boldsymbol{u}_2^\star \boldsymbol{u}_2^{\star\top} + \boldsymbol{H}$, $\boldsymbol{u}_1^\star = \frac{1}{\sqrt{n}}\begin{bmatrix} \mathbf{1}_{n/2} \\ \mathbf{1}_{n/2} \end{bmatrix}$; $\boldsymbol{u}_2^\star = \frac{1}{\sqrt{n}}\begin{bmatrix} \mathbf{1}_{n/2} \\ -\mathbf{1}_{n/2} \end{bmatrix}$
- $[\mathsf{Var}(H_{ij})]_{i,j} \approx \frac{1}{n \log n}\left(\begin{bmatrix} \mathbf{1}\mathbf{1}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \frac{1}{100}\mathbf{1}\mathbf{1}^\top\right)$
- Estimate $\boldsymbol{u}_2^\star$ by eigen-decomposition on the symmetrized data $(\boldsymbol{M} + \boldsymbol{M}^\top)/2$ and the original data $\boldsymbol{M}$.



One more experiment: heteroscedastic Gaussian noise

- $\boldsymbol{M} = \boldsymbol{u}_1^\star \boldsymbol{u}_1^{\star\top} + 0.95\boldsymbol{u}_2^\star \boldsymbol{u}_2^{\star\top} + \boldsymbol{H}$, $\boldsymbol{u}_1^\star = \frac{1}{\sqrt{n}}\begin{bmatrix} \mathbf{1}_{n/2} \\ \mathbf{1}_{n/2} \end{bmatrix}$; $\boldsymbol{u}_2^\star = \frac{1}{\sqrt{n}}\begin{bmatrix} \mathbf{1}_{n/2} \\ -\mathbf{1}_{n/2} \end{bmatrix}$
- $[\mathsf{Var}(H_{ij})]_{i,j} \approx \frac{1}{n\log n}\left(\begin{bmatrix} \mathbf{1}\mathbf{1}^\top & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} + \frac{1}{100}\mathbf{1}\mathbf{1}^\top\right)$
- Estimate $\boldsymbol{u}_2^\star$ by eigen-decomposition on the symmetrized data $(\boldsymbol{M}+\boldsymbol{M}^\top)/2$ and the original data $\boldsymbol{M}$.



Symmetrization for heteroscedastic data seems suboptimal!

Why does eigen-decomposition work so well on asymmetric data?

Problem setup

$$\boldsymbol{M} = \underbrace{\sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top}}_{\boldsymbol{M}^\star} + \boldsymbol{H} \in \mathbb{R}^{n\times n}$$

Problem setup

$$\boldsymbol{M} = \underbrace{\sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top}}_{\boldsymbol{M}^\star} + \boldsymbol{H} \in \mathbb{R}^{n\times n}$$

- $\boldsymbol{M}^\star$: rank-r ground-truth, $|\lambda_1^\star| \geq \cdots \geq |\lambda_r^\star| > 0$.
- $\boldsymbol{H}$: noise matrix
  - **independent entries:** $\{H_{i,j}\}$ are independent
  - **zero mean:** $\mathbb{E}[H_{i,j}] = 0$
  - **variance:** $\mathsf{Var}(H_{i,j}) \leq \sigma^2$
  - **magnitudes:** $\mathbb{P}\{|H_{i,j}| \geq B\} \lesssim n^{-12}$

Problem setup

$$\boldsymbol{M} = \underbrace{\sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top}}_{\boldsymbol{M}^\star} + \boldsymbol{H} \in \mathbb{R}^{n\times n}$$

- $\boldsymbol{M}^\star$: rank-r ground-truth, $|\lambda_1^\star| \geq \cdots \geq |\lambda_r^\star| > 0$.
- $\boldsymbol{H}$: noise matrix
  - **independent entries:** $\{H_{i,j}\}$ are independent
  - **zero mean:** $\mathbb{E}[H_{i,j}] = 0$
  - **variance:** $\mathsf{Var}(H_{i,j}) \leq \sigma^2$
  - **magnitudes:** $\mathbb{P}\{|H_{i,j}| \geq B\} \lesssim n^{-12}$
- $\boldsymbol{M}^\star$ obeys incoherence condition

$$\|\boldsymbol{u}_k^\star\|_\infty \leq \sqrt{\frac{\mu}{n}}$$

Classical linear algebra for eigenvalue

$$\left|\lambda_l^{\mathsf{svd}} - \lambda_l^{\star}\right| \leq \|\boldsymbol{H}\| \qquad \text{(Weyl)}$$

$$\left|\lambda_l^{\mathsf{eigs}} - \lambda_j^{\star}\right| \leq \|\boldsymbol{H}\| \qquad \text{(Bauer-Fike)}$$

Classical linear algebra for eigenvalue

$$\left|\lambda_l^{\mathsf{svd}} - \lambda_l^{\star}\right| \leq \|\boldsymbol{H}\| \qquad \text{(Weyl)}$$

$$\left|\lambda_l^{\mathsf{eigs}} - \lambda_j^{\star}\right| \leq \|\boldsymbol{H}\| \qquad \text{(Bauer-Fike)}$$

$\Downarrow$ matrix Bernstein inequality

$$\left|\lambda_l^{\mathsf{svd}} - \lambda_l^{\star}\right| \lesssim \sigma\sqrt{n \log n} + B \log n$$

$$\left|\lambda_l^{\mathsf{eigs}} - \lambda_j^{\star}\right| \lesssim \sigma\sqrt{n \log n} + B \log n$$

Classical linear algebra for eigenvalue

$$\left|\lambda_l^{\mathsf{svd}} - \lambda_l^\star\right| \le \|\boldsymbol{H}\| \qquad \text{(Weyl)}$$

$$\left|\lambda_l^{\mathsf{eigs}} - \lambda_j^\star\right| \le \|\boldsymbol{H}\| \qquad \text{(Bauer-Fike)}$$

$\Downarrow$ matrix Bernstein inequality

$$\left|\lambda_l^{\mathsf{svd}} - \lambda_l^\star\right| \lesssim \sigma\sqrt{n\log n} + B\log n \qquad \text{(reasonably tight if } \|\boldsymbol{H}\| \text{ is large)}$$

$$\left|\lambda_l^{\mathsf{eigs}} - \lambda_j^\star\right| \lesssim \sigma\sqrt{n\log n} + B\log n$$

Classical linear algebra for eigenvalue

$$\left|\lambda_l^{\mathsf{svd}} - \lambda_l^\star\right| \leq \|\boldsymbol{H}\| \qquad \text{(Weyl)}$$

$$\left|\lambda_l^{\mathsf{eigs}} - \lambda_j^\star\right| \leq \|\boldsymbol{H}\| \qquad \text{(Bauer-Fike)}$$

$\Downarrow$ matrix Bernstein inequality

$$\left|\lambda_l^{\mathsf{svd}} - \lambda_l^\star\right| \lesssim \sigma\sqrt{n\log n} + B\log n \qquad \text{(reasonably tight if } \|\boldsymbol{H}\| \text{ is large)}$$

$$\left|\lambda_l^{\mathsf{eigs}} - \lambda_j^\star\right| \lesssim \sigma\sqrt{n\log n} + B\log n \qquad \text{(\textit{can be significantly improved})}$$

Rank-1: eigenvalue perturbation

Theorem 1 (Chen, Cheng, Fan '18)

Assume $\sigma\sqrt{n\log n} + B\log n \leq c|\lambda^\star|$ *for some constant* c*. With high prob., leading eigenvalue* λ^{eigs} *of* $\boldsymbol{M}$ *obeys*

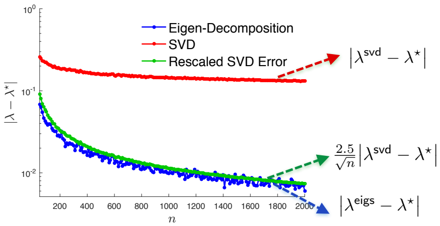
$$\left|\lambda^{\mathsf{eigs}} - \lambda^\star\right| \lesssim \sqrt{\frac{\mu}{n}}\left(\sigma\sqrt{n\log n} + B\log n\right)$$

Rank-1: eigenvalue perturbation

Theorem 1 (Chen, Cheng, Fan '18)

Assume $\sigma\sqrt{n\log n} + B\log n \leq c|\lambda^\star|$ *for some constant* c. *With high prob., leading eigenvalue* λ^{eigs} *of* $\boldsymbol{M}$ *obeys*

$$\left|\lambda^{\mathsf{eigs}} - \lambda^\star\right| \lesssim \sqrt{\frac{\mu}{n}}\left(\sigma\sqrt{n\log n} + B\log n\right)$$



- Eigen-decomposition is $\sqrt{\frac{n}{\mu}}$ *times better* than SVD!

— recall $\left|\lambda^{\mathsf{svd}} - \lambda^\star\right| \lesssim \sigma\sqrt{n\log n} + B\log n$

Rank-1: entrywise eigenvector perturbation

Theorem 2 (Chen, Cheng, Fan '18)

With high prob., leading eigenvector $\boldsymbol{u}$ of $\boldsymbol{M}$ obeys

$$\min \left\| \boldsymbol{u} \pm \boldsymbol{u}^{\star} \right\|_{\infty} \lesssim \sqrt{\frac{\mu}{n}} \left(\sigma \sqrt{n \log n} + B \log n \right)$$

Rank-1: entrywise eigenvector perturbation

Theorem 2 (Chen, Cheng, Fan '18)

With high prob., leading eigenvector $\boldsymbol{u}$ of $\boldsymbol{M}$ obeys

$$\min \left\| \boldsymbol{u} \pm \boldsymbol{u}^\star \right\|_\infty \lesssim \sqrt{\frac{\mu}{n}} \left(\sigma \sqrt{n \log n} + B \log n \right)$$

- if $\|\boldsymbol{H}\| \ll \left|\lambda^\star\right|$, then

$$\min \left\| \boldsymbol{u} \pm \boldsymbol{u}^\star \right\|_2 \ll \left\| \boldsymbol{u}^\star \right\|_2 \qquad \text{(classical bound)}$$

Rank-1: entrywise eigenvector perturbation

Theorem 2 (Chen, Cheng, Fan '18)

With high prob., leading eigenvector $\boldsymbol{u}$ of $\boldsymbol{M}$ obeys

$$\min \left\| \boldsymbol{u} \pm \boldsymbol{u}^\star \right\|_\infty \lesssim \sqrt{\frac{\mu}{n}} \big(\sigma \sqrt{n \log n} + B \log n \big)$$

- if $\|\boldsymbol{H}\| \ll |\lambda^\star|$, then

$$\min \left\| \boldsymbol{u} \pm \boldsymbol{u}^\star \right\|_2 \ll \left\| \boldsymbol{u}^\star \right\|_2 \qquad \text{(classical bound)}$$

$$\min \left\| \boldsymbol{u} \pm \boldsymbol{u}^\star \right\|_\infty \ll \left\| \boldsymbol{u}^\star \right\|_\infty \qquad \text{(our bound)}$$

Rank-1: entrywise eigenvector perturbation

Theorem 2 (Chen, Cheng, Fan '18)

With high prob., leading eigenvector $\boldsymbol{u}$ of $\boldsymbol{M}$ obeys

$$\min \left\| \boldsymbol{u} \pm \boldsymbol{u}^{\star} \right\|_{\infty} \lesssim \sqrt{\frac{\mu}{n}} \left(\sigma \sqrt{n \log n} + B \log n \right)$$

- if $\|\boldsymbol{H}\| \ll \left|\lambda^{\star}\right|$, then

$$\min \left\| \boldsymbol{u} \pm \boldsymbol{u}^{\star} \right\|_{2} \ll \left\| \boldsymbol{u}^{\star} \right\|_{2} \qquad \text{(classical bound)}$$
$$\min \left\| \boldsymbol{u} \pm \boldsymbol{u}^{\star} \right\|_{\infty} \ll \left\| \boldsymbol{u}^{\star} \right\|_{\infty} \qquad \text{(our bound)}$$

- entrywise eigenvector perturbation is well-controlled

Rank-1: perturbation of linear forms of eigenvectors

Theorem 3 (Chen, Cheng, Fan '18)

Fix any unit vector $\boldsymbol{a}$. With high prob., leading eigenvector $\boldsymbol{u}$ of $\boldsymbol{M}$ obeys

$$\min\left\{\left|\boldsymbol{a}^\top(\boldsymbol{u}\pm\boldsymbol{u}^\star)\right|\right\} \lesssim \max\left\{\left|\boldsymbol{a}^\top\boldsymbol{u}^\star\right|, \sqrt{\frac{\mu}{n}}\right\}\left(\sigma\sqrt{n\log n}+B\log n\right)$$

Rank-1: perturbation of linear forms of eigenvectors

Theorem 3 (Chen, Cheng, Fan '18)

Fix any unit vector $\boldsymbol{a}$. With high prob., leading eigenvector $\boldsymbol{u}$ of $\boldsymbol{M}$ obeys

$$\min\left\{\left|\boldsymbol{a}^{\top}(\boldsymbol{u}\pm\boldsymbol{u}^{\star})\right|\right\}\lesssim\max\left\{\left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right|,\sqrt{\frac{\mu}{n}}\right\}\left(\sigma\sqrt{n\log n}+B\log n\right)$$

- if $\|\boldsymbol{H}\|\ll\left|\lambda^{\star}\right|$, then

$$\min\left\{\left|\boldsymbol{a}^{\top}(\boldsymbol{u}\pm\boldsymbol{u}^{\star})\right|\right\}\ll\max\left\{\left|\boldsymbol{a}^{\top}\boldsymbol{u}^{\star}\right|,\|\boldsymbol{u}^{\star}\|_{\infty}\right\}$$

- perturbation of an arbitrary linear form of leading eigenvector is well-controlled.

Rank-1: perturbation of linear forms of eigenvectors

Theorem 3 (Chen, Cheng, Fan '18)

Fix any unit vector $\boldsymbol{a}$. *With high prob., leading eigenvector* $\boldsymbol{u}$ *of* $\boldsymbol{M}$ *obeys*

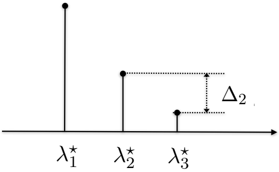
$$\min\left\{\left|\boldsymbol{a}^\top(\boldsymbol{u}\pm\boldsymbol{u}^\star)\right|\right\} \lesssim \max\left\{\left|\boldsymbol{a}^\top\boldsymbol{u}^\star\right|, \sqrt{\frac{\mu}{n}}\right\}\left(\sigma\sqrt{n\log n}+B\log n\right)$$

- if $\|\boldsymbol{H}\| \ll \left|\lambda^\star\right|$, then

$$\min\left\{\left|\boldsymbol{a}^\top(\boldsymbol{u}\pm\boldsymbol{u}^\star)\right|\right\} \ll \max\left\{\left|\boldsymbol{a}^\top\boldsymbol{u}^\star\right|, \|\boldsymbol{u}^\star\|_\infty\right\}$$

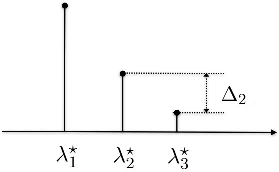
- perturbation of an arbitrary linear form of leading eigenvector is well-controlled.
- very few results are available for symmetric noise.

Classical linear algebra for eigenvector



(eigenvalue separation) $\Delta_l := \min_{k:k\neq l} |\lambda_l^\star - \lambda_k^\star|$

Classical linear algebra for eigenvector

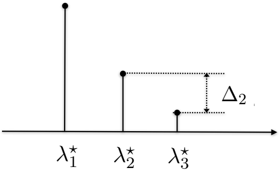


(eigenvalue separation) $\Delta_l := \min_{k:k\neq l} |\lambda_l^\star - \lambda_k^\star|$

$$\min \left\| \boldsymbol{u}_l^{\mathsf{svd}} \pm \boldsymbol{u}_l^\star \right\|_2 \lesssim \frac{\|\boldsymbol{H}\|}{\Delta_l} \qquad \text{(Davis-Kahan)}$$

$$\min \left\| \boldsymbol{u}_l^{\mathsf{eigs}} \pm \boldsymbol{u}_l^\star \right\|_2 \lesssim \; ??$$

Classical linear algebra for eigenvector



(eigenvalue separation) $\Delta_l := \min_{k:k\neq l} |\lambda_l^\star - \lambda_k^\star|$

$$\min \left\| \boldsymbol{u}_l^{\mathsf{svd}} \pm \boldsymbol{u}_l^\star \right\|_2 \lesssim \frac{\|\boldsymbol{H}\|}{\Delta_l} \qquad \text{(Davis-Kahan)}$$

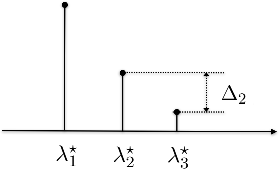
$$\min \left\| \boldsymbol{u}_l^{\mathsf{eigs}} \pm \boldsymbol{u}_l^\star \right\|_2 \lesssim ??$$

$\Downarrow$ matrix concentration inequality

$$\min \left\| \boldsymbol{u}_l^{\mathsf{svd}} \pm \boldsymbol{u}_l^\star \right\|_2 \lesssim \frac{\sigma\sqrt{n}}{\Delta_l}$$ (requires $\Delta_l \gtrsim \|\boldsymbol{H}\|$, and B is sufficiently small)

$$\min \left\| \boldsymbol{u}_l^{\mathsf{eigs}} \pm \boldsymbol{u}_l^\star \right\|_2 \lesssim ??$$

Rank-r: eigenvalue / eigenvector perturbation



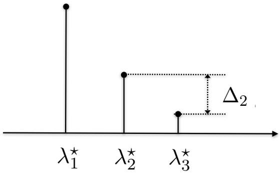
(eigenvalue separation) $\Delta_l := \min_{k:k\neq l} |\lambda_l^\star - \lambda_k^\star|$

Theorem 4 (Cheng, Wei, Chen '20)

Suppose $\boldsymbol{M}^\star$ is well-conditioned, incoherent, and $r = O(1)$. Assume

$$\Delta_l > 2c_0\sigma\sqrt{\log n} \qquad \text{for some const } c_0 > 0 \tag{1}$$

Rank-r: eigenvalue / eigenvector perturbation



(eigenvalue separation) $\Delta_l := \min_{k:k\neq l} |\lambda_l^\star - \lambda_k^\star|$

Theorem 4 (Cheng, Wei, Chen '20)

Suppose $M^\star$ is well-conditioned, incoherent, and $r = O(1)$. Assume

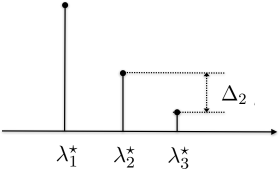
$$\Delta_l > 2c_0\sigma\sqrt{\log n} \qquad \text{for some const } c_0 > 0 \tag{1}$$

With high prob., l^{th} largest e-value $\lambda_l^{\mathsf{eigs}}$ & e-vector $\boldsymbol{u}_l^{\mathsf{eigs}}$ of $\boldsymbol{M}$ obey

$$\left|\lambda_l^{\mathsf{eigs}} - \lambda_l^\star\right| \leq c_0\sigma\sqrt{\log n}$$

$$\min \|\boldsymbol{u}_l^{\mathsf{eigs}} \pm \boldsymbol{u}_l^\star\|_2 \lesssim \frac{\sigma\sqrt{\log n}}{\Delta_l^\star} + \frac{\sigma\sqrt{n\log n}}{\|\boldsymbol{M}^\star\|}$$

Rank-r: eigenvalue / eigenvector perturbation



(eigenvalue separation) $\Delta_l := \min_{k:k\neq l} |\lambda_l^\star - \lambda_k^\star|$

Theorem 4 (Cheng, Wei, Chen '20)

Suppose $\boldsymbol{M}^\star$ is well-conditioned, incoherent, and $r = O(1)$. Assume

$$\Delta_l > 2c_0\sigma\sqrt{\log n} \qquad \textit{for some const } c_0 > 0 \tag{1}$$

With high prob., l^{th} largest e-value λ_l^{eigs} & e-vector $\boldsymbol{u}_l^{\text{eigs}}$ of $\boldsymbol{M}$ obey

$$\left|\lambda_l^{\text{eigs}} - \lambda_l^\star\right| \leq c_0\sigma\sqrt{\log n}$$

$$\min \|\boldsymbol{u}_l^{\text{eigs}} \pm \boldsymbol{u}_l^\star\|_2 \lesssim \frac{\sigma\sqrt{\log n}}{\Delta_l^\star} + \frac{\sigma\sqrt{n\log n}}{\|\boldsymbol{M}^\star\|}$$

Similar bounds for entrywise perturbation and linear forms perturbation.

SVD vs. Eigen-decomposition

1. **eigenvalue estimation:** Eigen-decomposition is $O(\sqrt{n})$ times more accurate

$$\left|\lambda_l^{\mathsf{svd}} - \lambda_l^\star\right| \lesssim \sigma\sqrt{n} \qquad \text{(Weyl)}$$

$$\left|\lambda_l^{\mathsf{eigs}} - \lambda_l^\star\right| \lesssim \sigma\sqrt{\log n} \qquad \text{(ours)}$$

SVD vs. Eigen-decomposition

1. **eigenvalue estimation:** Eigen-decomposition is $O(\sqrt{n})$ times more accurate

$$\left|\lambda_l^{\mathsf{svd}} - \lambda_l^\star\right| \lesssim \sigma\sqrt{n} \qquad \text{(Weyl)}$$
$$\left|\lambda_l^{\mathsf{eigs}} - \lambda_l^\star\right| \lesssim \sigma\sqrt{\log n} \qquad \text{(ours)}$$

2. **eigenvector estimation:** Eigen-decomposition works under $O(\sqrt{n})$ times smaller eigenvalue separation

$$\min\left\|\boldsymbol{u}_l^{\mathsf{svd}} \pm \boldsymbol{u}_l^\star\right\| = o(1) \quad \text{if } \Delta_l \gtrsim \sigma\sqrt{n} \qquad \text{(Davis-Kahan)}$$
$$\min\left\|\boldsymbol{u}_l^{\mathsf{eigs}} \pm \boldsymbol{u}_l^\star\right\| = o(1) \quad \text{if } \Delta_l \gtrsim \sigma\sqrt{\log n} \qquad \text{(ours)}$$

SVD vs. Eigen-decomposition

1. **eigenvalue estimation:** Eigen-decomposition is $O(\sqrt{n})$ times more accurate

$$\left|\lambda_l^{\mathsf{svd}} - \lambda_l^{\star}\right| \lesssim \sigma\sqrt{n} \qquad \text{(Weyl)}$$
$$\left|\lambda_l^{\mathsf{eigs}} - \lambda_l^{\star}\right| \lesssim \sigma\sqrt{\log n} \qquad \text{(ours)}$$

2. **eigenvector estimation:** Eigen-decomposition works under $O(\sqrt{n})$ times smaller eigenvalue separation

$$\min \left\|\boldsymbol{u}_l^{\mathsf{svd}} \pm \boldsymbol{u}_l^{\star}\right\| = o(1) \quad \text{if } \Delta_l \gtrsim \sigma\sqrt{n} \qquad \text{(Davis-Kahan)}$$
$$\min \left\|\boldsymbol{u}_l^{\mathsf{eigs}} \pm \boldsymbol{u}_l^{\star}\right\| = o(1) \quad \text{if } \Delta_l \gtrsim \sigma\sqrt{\log n} \qquad \text{(ours)}$$

(The same bound holds for symmetrized eigen-decomposition on $(\boldsymbol{M} + \boldsymbol{M}^{\top})/2$ as SVD on $\boldsymbol{M}$)

Summary of estimation from eigen-decomposition on asymmetric noise

- no need of bias correction
- faithful eigenvector estimation under much smaller eigenvalue separation
- distribution-free
- adaptive to heteroscedastic noise

Summary of estimation from eigen-decomposition on asymmetric noise

- no need of bias correction
- faithful eigenvector estimation under much smaller eigenvalue separation
- distribution-free
- adaptive to heteroscedastic noise

Cool!

Summary of estimation from eigen-decomposition on asymmetric noise

- no need of bias correction
- faithful eigenvector estimation under much smaller eigenvalue separation
- distribution-free
- adaptive to heteroscedastic noise

Cool! Can we do more?

Problem setup

$$\boldsymbol{M} = \underbrace{\sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top}}_{\boldsymbol{M}^\star} + \boldsymbol{H} \in \mathbb{R}^{n\times n}$$

Problem setup

$$\boldsymbol{M} = \underbrace{\sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top}}_{\boldsymbol{M}^\star} + \boldsymbol{H} \in \mathbb{R}^{n\times n}$$

- **Goal**: Infer eigenvalues $\lambda_l^\star$ and linear forms $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$.

Problem setup

$$\boldsymbol{M} = \underbrace{\sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top}}_{\boldsymbol{M}^\star} + \boldsymbol{H} \in \mathbb{R}^{n\times n}$$

- **Goal**: Infer eigenvalues $\lambda_l^\star$ and linear forms $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$.
- $\boldsymbol{H}$: noise matrix
  - **independent entries:** $\{H_{i,j}\}$ are independent
  - **zero mean:** $\mathbb{E}[H_{i,j}] = 0$
  - **variance:** $\sigma_{\min}^2 \leq \mathsf{Var}(H_{i,j}) \leq \sigma_{\max}^2 \ll \frac{(\lambda_{\min}^\star)^2}{n \log n}$ with $\frac{\sigma_{\max}}{\sigma_{\min}} = O(1)$
  - **magnitudes:** $|H_{i,j}| \leq \sigma_{\max}\sqrt{n/\log n}$ with high prob.
- $\boldsymbol{M}^\star$ obeys incoherence condition

$$\|\boldsymbol{u}_k^\star\|_\infty \leq \sqrt{\frac{\mu}{n}}$$

Problem setup

$$\boldsymbol{M} = \underbrace{\sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top}}_{\boldsymbol{M}^\star} + \boldsymbol{H} \in \mathbb{R}^{n\times n}$$

- **Goal**: Infer eigenvalues $\lambda_l^\star$ and linear forms $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$.
- $\boldsymbol{H}$: noise matrix
  - **independent entries:** $\{H_{i,j}\}$ are independent
  - **zero mean:** $\mathbb{E}[H_{i,j}] = 0$
  - **variance:** $\sigma_{\min}^2 \le \mathsf{Var}(H_{i,j}) \le \sigma_{\max}^2 \ll \frac{(\lambda_{\min}^\star)^2}{n \log n}$ with $\frac{\sigma_{\max}}{\sigma_{\min}} = O(1)$
  - **magnitudes:** $|H_{i,j}| \le \sigma_{\max}\sqrt{n/\log n}$ with high prob.
- $\boldsymbol{M}^\star$ obeys incoherence condition

$$\|\boldsymbol{u}_k^\star\|_\infty \le \sqrt{\frac{\mu}{n}}$$

- Can we quantify the uncertainty of the proposed estimators? Do they achieve statistical optimality?

Which estimator shall we use?

A natural start point: λ_l and $\boldsymbol{a}^\top \boldsymbol{u}_l$ (or $\boldsymbol{a}^\top \boldsymbol{w}_l$)

Which estimator shall we use?

A natural start point: λ_l and $\boldsymbol{a}^\top \boldsymbol{u}_l$ (or $\boldsymbol{a}^\top \boldsymbol{w}_l$)

- λ_l: Good enough unbiased estimator for $\lambda_l^\star$!
- $\boldsymbol{a}^\top \boldsymbol{u}_l$ (or $\boldsymbol{a}^\top \boldsymbol{w}_l$): Not so good for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$.

Which estimator shall we use?

A natural start point: λ_l and $\boldsymbol{a}^\top \boldsymbol{u}_l$ (or $\boldsymbol{a}^\top \boldsymbol{w}_l$)

- λ_l: Good enough unbiased estimator for $\lambda_l^\star$!
- $\boldsymbol{a}^\top \boldsymbol{u}_l$ (or $\boldsymbol{a}^\top \boldsymbol{w}_l$): Not so good for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$.

Issues:

- **bias aggregation:** even though $\boldsymbol{u}_l$ is nearly unbiased estimate of $\boldsymbol{u}_l^\star$ in every entry, it does NOT mean $\boldsymbol{a}^\top \boldsymbol{u}_l$ is nearly unbiased

Which estimator shall we use?

A natural start point: λ_l and $\boldsymbol{a}^\top \boldsymbol{u}_l$ (or $\boldsymbol{a}^\top \boldsymbol{w}_l$)

- λ_l: Good enough unbiased estimator for $\lambda_l^\star$!
- $\boldsymbol{a}^\top \boldsymbol{u}_l$ (or $\boldsymbol{a}^\top \boldsymbol{w}_l$): Not so good for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$.

Issues:

- **bias aggregation:** even though $\boldsymbol{u}_l$ is nearly unbiased estimate of $\boldsymbol{u}_l^\star$ in every entry, it does NOT mean $\boldsymbol{a}^\top \boldsymbol{u}_l$ is nearly unbiased
- **optimality?** it is unclear whether $\boldsymbol{a}^\top \boldsymbol{u}_l$ incurrs minimal uncertainty

Key observation: rank-1 case

Neumann series imply

$$\boldsymbol{u}_1 = \frac{\lambda_1^\star \boldsymbol{u}_1^{\star\top} \boldsymbol{u}}{\lambda_1} \sum_{s=0}^{+\infty} \left(\frac{\boldsymbol{H}}{\lambda_1} \right)^s \boldsymbol{u}^\star.$$

Key observation: rank-1 case

Neumann series imply

$$\boldsymbol{u}_1 = \frac{\lambda_1^\star \boldsymbol{u}_1^{\star\top} \boldsymbol{u}}{\lambda_1} \sum_{s=0}^{+\infty} \left(\frac{\boldsymbol{H}}{\lambda_1} \right)^s \boldsymbol{u}^\star.$$

Hence

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx \left(\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1 \right) \left\{ \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda_1^\star} \right\}.$$

Key observation: rank-1 case

Neumann series imply

$$\boldsymbol{u}_1 = \frac{\lambda_1^\star \boldsymbol{u}_1^{\star\top} \boldsymbol{u}}{\lambda_1} \sum_{s=0}^{+\infty} \left(\frac{\boldsymbol{H}}{\lambda_1} \right)^s \boldsymbol{u}^\star.$$

Hence

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx \left(\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1 \right) \left\{ \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda_1^\star} \right\}.$$

The plug-in estimator $\boldsymbol{a}^\top \boldsymbol{u}_1$ is an underestimate by approximately $\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1$.

Key observation: rank-1 case

Neumann series imply

$$\boldsymbol{u}_1 = \frac{\lambda_1^\star \boldsymbol{u}_1^{\star\top} \boldsymbol{u}}{\lambda_1} \sum_{s=0}^{+\infty} \left(\frac{\boldsymbol{H}}{\lambda_1} \right)^s \boldsymbol{u}^\star.$$

Hence

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx \left(\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1 \right) \left\{ \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda_1^\star} \right\}.$$

The plug-in estimator $\boldsymbol{a}^\top \boldsymbol{u}_1$ is an underestimate by approximately $\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1$.

How can we reduce the bias?

Key observation: rank-1 case

Neumann series imply

$$\boldsymbol{u}_1 = \frac{\lambda_1^\star \boldsymbol{u}_1^{\star\top} \boldsymbol{u}}{\lambda_1} \sum_{s=0}^{+\infty} \left(\frac{\boldsymbol{H}}{\lambda_1} \right)^s \boldsymbol{u}^\star.$$

Hence

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx \left(\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1 \right) \left\{ \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda_1^\star} \right\}.$$

The plug-in estimator $\boldsymbol{a}^\top \boldsymbol{u}_1$ is an underestimate by approximately $\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1$.

How can we reduce the bias?

$$\hat{u}_{\boldsymbol{a},l} = \sqrt{\frac{1}{\boldsymbol{u}_l^\top \boldsymbol{w}_l} \left(\boldsymbol{a}^\top \boldsymbol{u}_l \right) \left(\boldsymbol{a}^\top \boldsymbol{w}_l \right)}.$$

Key observation: rank-1 case

Neumann series imply

$$\boldsymbol{u}_1 = \frac{\lambda_1^\star \boldsymbol{u}_1^{\star\top} \boldsymbol{u}}{\lambda_1} \sum_{s=0}^{+\infty} \left(\frac{\boldsymbol{H}}{\lambda_1} \right)^s \boldsymbol{u}^\star.$$

Hence

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx \left(\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1 \right) \left\{ \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda_1^\star} \right\}.$$

The plug-in estimator $\boldsymbol{a}^\top \boldsymbol{u}_1$ is an underestimate by approximately $\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1$.

How can we reduce the bias?

$$\hat{u}_{\boldsymbol{a},l} = \sqrt{\frac{1}{\boldsymbol{u}_l^\top \boldsymbol{w}_l} \left(\boldsymbol{a}^\top \boldsymbol{u}_l \right) \left(\boldsymbol{a}^\top \boldsymbol{w}_l \right)}.$$

The bias term has been canceled out!

Rank-r: distributional characterization

- $\boldsymbol{M}^\star$ is well-conditioned, incoherent, and $r = O(1)$
- $$\begin{cases} \frac{1}{\|\boldsymbol{a}\|_2} \left| \boldsymbol{a}^\top \boldsymbol{u}_l^\star \right| = o\Big(\frac{1}{\sqrt{\log n}} \min\left\{ \frac{\Delta_l^\star}{|\lambda_l^\star|}, 1 \right\} \Big) & \text{(size of target quantity)} \\ \frac{1}{\|\boldsymbol{a}\|_2} \left| \boldsymbol{a}^\top \boldsymbol{u}_k^\star \right| = o\Big(\frac{1}{\sqrt{\log n}} \frac{|\lambda_l^\star - \lambda_k^\star|}{|\lambda_l^\star|} \Big), \quad \forall k \neq l & \text{(size of ``interferers'')} \end{cases}$$
- $\sigma_{\max} \log n = o(\Delta_l^\star)$ (minimal e-value separation)

Rank-r: distributional characterization

- $\boldsymbol{M}^\star$ is well-conditioned, incoherent, and $r = O(1)$
- $\begin{cases} \frac{1}{\|\boldsymbol{a}\|_2}\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| = o\left(\frac{1}{\sqrt{\log n}} \min\left\{\frac{\Delta_l^\star}{|\lambda_l^\star|}, 1\right\}\right) & \text{(size of target quantity)} \\ \frac{1}{\|\boldsymbol{a}\|_2}\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right| = o\left(\frac{1}{\sqrt{\log n}} \frac{|\lambda_l^\star - \lambda_k^\star|}{|\lambda_l^\star|}\right), \quad \forall k \neq l & \text{(size of "interferers")} \end{cases}$
- $\sigma_{\max} \log n = o(\Delta_l^\star)$ (minimal e-value separation)

Theorem 5 (Cheng, Wei, Chen '20)

Under above assumptions, with high prob. one has

$$\widehat{u}_{\boldsymbol{a},l} \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{1}{2\lambda_l^\star}\boldsymbol{a}^\top(\boldsymbol{H} + \boldsymbol{H}^\top)\boldsymbol{u}_l^\star \rightsquigarrow \mathcal{N}(\boldsymbol{a}^\top \boldsymbol{u}_l^\star, v_{\boldsymbol{a},l}^\star)$$

Rank-r: distributional characterization

- $\boldsymbol{M}^\star$ is well-conditioned, incoherent, and $r = O(1)$
- $\begin{cases} \frac{1}{\|\boldsymbol{a}\|_2}\left|\boldsymbol{a}^\top \boldsymbol{u}_l^\star\right| = o\left(\frac{1}{\sqrt{\log n}} \min\left\{\frac{\Delta_l^\star}{|\lambda_l^\star|}, 1\right\}\right) & \text{(size of target quantity)} \\ \frac{1}{\|\boldsymbol{a}\|_2}\left|\boldsymbol{a}^\top \boldsymbol{u}_k^\star\right| = o\left(\frac{1}{\sqrt{\log n}} \frac{|\lambda_l^\star - \lambda_k^\star|}{|\lambda_l^\star|}\right), \quad \forall k \neq l & \text{(size of "interferers")} \end{cases}$
- $\sigma_{\max} \log n = o(\Delta_l^\star)$ (minimal e-value separation)

THEOREM 5 (CHENG, WEI, CHEN '20)

Under above assumptions, with high prob. one has

$$\widehat{u}_{\boldsymbol{a},l} \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{1}{2\lambda_l^\star} \boldsymbol{a}^\top (\boldsymbol{H} + \boldsymbol{H}^\top) \boldsymbol{u}_l^\star \rightsquigarrow \mathcal{N}(\boldsymbol{a}^\top \boldsymbol{u}_l^\star, v_{\boldsymbol{a},l}^\star)$$

THEOREM 6 (CHENG, WEI, CHEN '20)

Under above assumptions, with high prob. one has

$$\lambda_l \approx \lambda_l^\star + \boldsymbol{u}_l^{\star\top} \boldsymbol{H} \boldsymbol{u}_l^\star \rightsquigarrow \mathcal{N}(\lambda_l^\star, v_{\lambda,l}^\star)$$

Rank-r: confidence intervals & optimality

- $v^{\star}_{\boldsymbol{a},l}$ and $v^{\star}_{\lambda,l}$ can both be faithfully estimated.

Rank-r: confidence intervals & optimality

- $v^{\star}_{\boldsymbol{a},l}$ and $v^{\star}_{\lambda,l}$ can both be faithfully estimated.
- yields $(1-\alpha)$-confidence intervals

$$\left[\widehat{u}_{\boldsymbol{a},l} \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{v_{\boldsymbol{a},l}}}\right]$$

$$\left[\lambda_l \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{v_{\lambda,l}}}\right]$$

Rank-r: confidence intervals & optimality

- $v^{\star}_{\boldsymbol{a},l}$ and $v^{\star}_{\lambda,l}$ can both be faithfully estimated.
- yields $(1-\alpha)$-confidence intervals

$$\left[\widehat{u}_{\boldsymbol{a},l} \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{v_{\boldsymbol{a},l}}}\right]$$
$$\left[\lambda_l \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{v_{\lambda,l}}}\right]$$

- Can they be further shortened?

Rank-r: confidence intervals & optimality

- $v^\star_{\boldsymbol{a},l}$ and $v^\star_{\lambda,l}$ can both be faithfully estimated.
- yields $(1-\alpha)$-confidence intervals

$$\left[\widehat{u}_{\boldsymbol{a},l} \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{v_{\boldsymbol{a},l}}}\right]$$
$$\left[\lambda_l \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{v_{\lambda,l}}}\right]$$

- Can they be further shortened?
- $H_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2), \boldsymbol{a}^\top \boldsymbol{u}_l = o(1)$, Cramer-Rao lower bounds follow as

Theorem 7 (Cheng, Wei, Chen '20)

Any unbiased estimator $\widehat{U}_{\boldsymbol{a}}$ (resp. $\widehat{\Lambda}$) of $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ (resp. $\lambda_l^\star$) obeys

$$\mathsf{Var}[\widehat{U}_{\boldsymbol{a}}] \geq (1-o(1))\mathsf{Var}\Big(\tfrac{1}{2\lambda_l^\star}\boldsymbol{a}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\Big) = (1-o(1))v^\star_{\boldsymbol{a},l}$$
$$\mathsf{Var}[\widehat{\Lambda}] \geq (1-o(1))\mathsf{Var}\big(\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star\big) = (1-o(1))v^\star_{\lambda,l}$$

Rank-r: confidence intervals & optimality

- $v^\star_{\boldsymbol{a},l}$ and $v^\star_{\lambda,l}$ can both be faithfully estimated.
- yields $(1-\alpha)$-confidence intervals

$$\left[\widehat{u}_{\boldsymbol{a},l} \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{v_{\boldsymbol{a},l}}}\right]$$
$$\left[\lambda_l \pm \Phi^{-1}(1-\alpha/2)\sqrt{\hat{v_{\lambda,l}}}\right]$$

- Can they be further shortened?  Optimal!
- $H_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma^2), \boldsymbol{a}^\top \boldsymbol{u}_l = o(1)$, Cramer-Rao lower bounds follow as

Theorem 7 (Cheng, Wei, Chen '20)

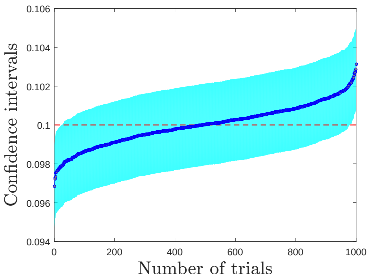
Any unbiased estimator $\widehat{U}_{\boldsymbol{a}}$ (resp. $\widehat{\Lambda}$) of $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ (resp. $\lambda_l^\star$) obeys

$$\mathsf{Var}[\widehat{U}_{\boldsymbol{a}}] \geq (1-o(1))\mathsf{Var}\left(\tfrac{1}{2\lambda_l^\star}\boldsymbol{a}(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star\right) = (1-o(1))v^\star_{\boldsymbol{a},l}$$
$$\mathsf{Var}[\widehat{\Lambda}] \geq (1-o(1))\mathsf{Var}\left(\boldsymbol{u}_l^{\star\top}\boldsymbol{H}\boldsymbol{u}_l^\star\right) = (1-o(1))v^\star_{\lambda,l}$$

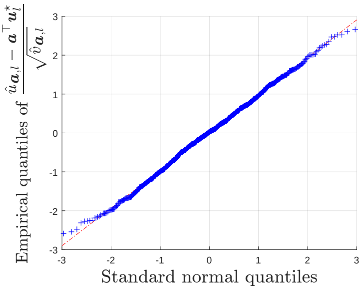
Experiment: estimating $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$

- rank-2: $\lambda_1^\star = 1$, $\lambda_2^\star = 0.95$, $\boldsymbol{a}^\top \boldsymbol{u}_1^\star = 0$, $\boldsymbol{a}^\top \boldsymbol{u}_2^\star = 0.1$
- heteroscedastic Gaussian noise; $\left[\mathsf{Var}(H_{ij})\right]_{i,j} = \begin{bmatrix} \sigma_1^2 \\ (\sigma_1 + \delta_\sigma)^2 \\ \vdots \\ (\sigma_1 + (n-1)\delta_\sigma)^2 \end{bmatrix} \mathbf{1}_n^\top$

$\sigma_1 = \frac{0.1}{\sqrt{n \log n}}$, $\delta_\sigma = \frac{0.4}{n\sqrt{n \log n}}$



95% confidence intervals



Q-Q (quantile-quantile) plot

Experiment: estimating $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$

Recall that our theory requires control of the "interferers" $\{\boldsymbol{a}^\top \boldsymbol{u}_k^\star\}_{k \neq l}$

Experiment: estimating $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$

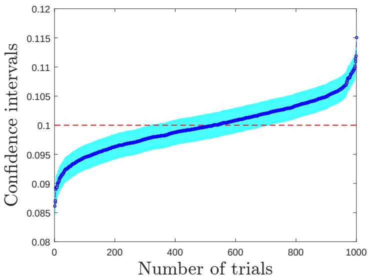
Recall that our theory requires control of the "interferers" $\{\boldsymbol{a}^\top \boldsymbol{u}_k^\star\}_{k \neq l}$

Numerically, it does seem that these "interferers" cannot be too large

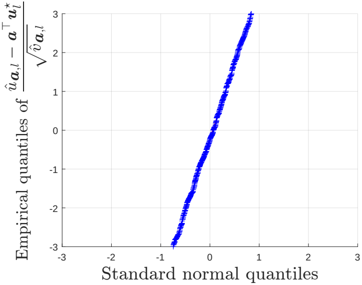
Experiment: estimating $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$

- rank-2: $\lambda_1^\star = 1$, $\lambda_2^\star = 0.95$, $\boldsymbol{a}^\top \boldsymbol{u}_1^\star = 0.2$, $\boldsymbol{a}^\top \boldsymbol{u}_2^\star = 0.1$
- heteroscedastic Gaussian noise; $\left[\mathsf{Var}(H_{ij})\right]_{i,j} = \begin{bmatrix} \sigma_1^2 \\ (\sigma_1 + \delta_\sigma)^2 \\ \vdots \\ (\sigma_1 + (n-1)\delta_\sigma)^2 \end{bmatrix} \mathbf{1}_n^\top$

$\sigma_1 = \frac{0.1}{\sqrt{n \log n}}$, $\delta_\sigma = \frac{0.4}{n\sqrt{n \log n}}$



95% confidence intervals

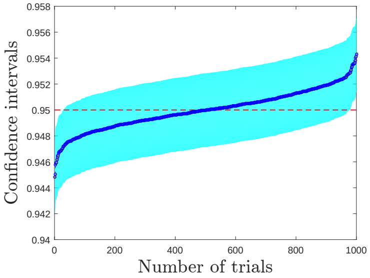


Q-Q (quantile-quantile) plot

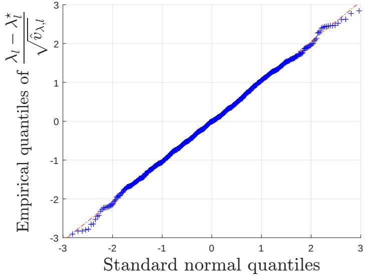
Experiment: estimating $\lambda_2^\star$

- rank-2: $\lambda_1^\star = 1$, $\lambda_2^\star = 0.95$, $\boldsymbol{a}^\top \boldsymbol{u}_1^\star = 0$, $\boldsymbol{a}^\top \boldsymbol{u}_1^\star = 0.1$
- heteroscedastic Gaussian noise; $\left[\mathsf{Var}(H_{ij})\right]_{i,j} = \begin{bmatrix} \sigma_1^2 \\ (\sigma_1 + \delta_\sigma)^2 \\ \vdots \\ (\sigma_1 + (n-1)\delta_\sigma)^2 \end{bmatrix} \mathbf{1}_n^\top$

$\sigma_1 = \frac{0.1}{\sqrt{n \log n}}$, $\delta_\sigma = \frac{0.4}{n\sqrt{n \log n}}$



95% confidence intervals



Q-Q (quantile-quantile) plot

Experiment: other settings

Settings	Target	Numerical coverage rates
heteroscedastic Gaussian noise	linear form $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$	0.953
	eigenvalue $\lambda_2^\star$	0.950
heteroscedastic Bernoulli noise	linear form $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$	0.955
	eigenvalue $\lambda_2^\star$	0.942
missing data + noise	linear form $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$	0.947
	eigenvalue $\lambda_2^\star$	0.954

Table 1: Numerical coverage rates for our 95% confidence intervals over 1000 independent trials.

- Our theory is corroborated by experiments!

Conclusions

Eigen-decomposition without symmetrization could be very powerful

Conclusions

Eigen-decomposition without symmetrization could be very powerful

- effective under minimal eigenvalue separation
- distribution-free
- adaptive to heteroscedastic noise
- enables "fine-grained" inference
- statistically optimal

C. Cheng, Y. Wei, Y. Chen, "Inference for linear forms of eigenvectors under minimal eigenvalue separation: Asymmetry and heteroscedasticity", arXiv:2001.04620, 2020

Y. Chen, C. Cheng, J. Fan, "Asymmetry helps: Eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices", arXiv:1811.12804, 2018 (alphabetical order)

— accepted to Annals of Statistics, 2020