Untold Gifts of Statistical Asymmetry to Eigen-analysis

Yuxin Chen

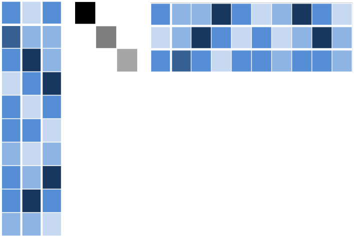
Electrical Engineering, Princeton University

Chen Cheng
Stanford

Yuting Wei
CMU

Jianqing Fan
Princeton

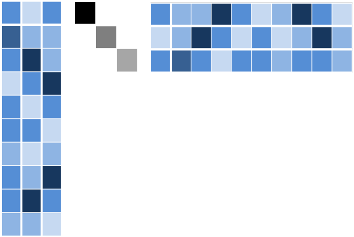
Eigen-analysis in high dimension



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

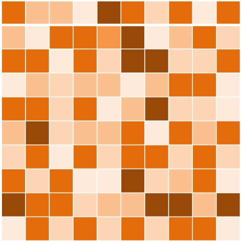
- Rank-r matrix: $\boldsymbol{M}^{\star} = \sum_{l=1}^{r} \lambda_l^{\star} \boldsymbol{u}_l^{\star} \boldsymbol{u}_l^{\star\top} \in \mathbb{R}^{n \times n}$

Eigen-analysis in high dimension



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

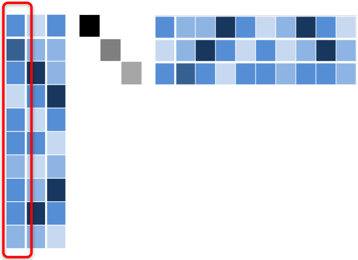
$+$



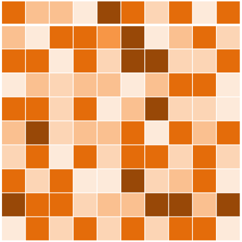
$\boldsymbol{H} = [H_{ij}]_{1 \le i,j \le n}$: independent noise

- 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 noisy data: $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$

Eigen-analysis in high dimension



$+$



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

$\boldsymbol{H} = [H_{ij}]_{1 \leq i,j \leq n}$: 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 noisy data: $\boldsymbol{M} = \boldsymbol{M}^{\star} + \boldsymbol{H}$
- **Goal:** estimate / infer unknown eigenvector $\boldsymbol{u}_l^{\star}$ $(1 \leq l \leq r)$

A small sample of prior work

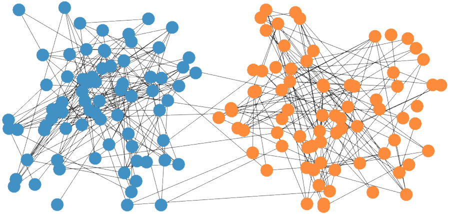
- Davis, Kahan '70
- Wedin '72
- Péché '06
- Vu '11
- Yu, Wang, Samworth '14
- Chen, Wainwright '15
- Wang '15
- Cai, Zhang '18
- Zhong '17
- Keshavan, Montanari, Oh '09
- O'Rourke, Vu, Wang '18
- Bryc, Silverstein '18
- Zhang, Cai, Wu '18
- Cai, Li, Chi, Poor, Chen '19
- $\cdots$

A large body of prior work focused on ℓ_2 analysis (e.g. $\|\boldsymbol{u}_l - \boldsymbol{u}_l^\star\|_2$)

Beyond ℓ_2 analysis: "fine-grained" statistical analysis

Entrywise eigenvector analysis:

— estimation and inference for each entry of $\boldsymbol{u}_l^\star$



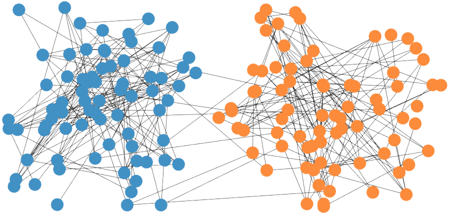
graph clustering (stochastic block model)

— *Abbe, Fan, Wang, Zhong '17*

Beyond ℓ_2 analysis: "fine-grained" statistical analysis

Entrywise eigenvector analysis:

— estimation and inference for each entry of $\boldsymbol{u}_l^\star$



graph clustering (stochastic block model)

— *Abbe, Fan, Wang, Zhong '17*

ranking from pairwise comparisons

— *Chen, Fan, Ma, Wang '17*

Beyond ℓ_2 analysis: "fine-grained" statistical analysis

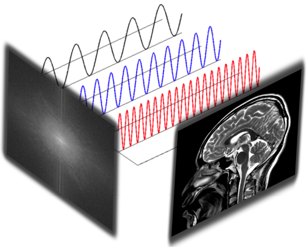
More generally, estimate & infer $\underbrace{\textbf{linear functions of eigenvectors}}_{\boldsymbol{a}^\top \boldsymbol{u}_l^\star,\text{ with } \boldsymbol{a} \text{ a fixed vector}}$

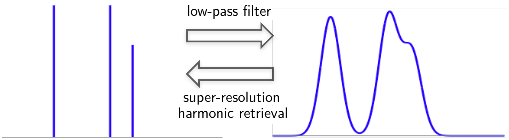
- e.g. entrywise analysis: $\boldsymbol{a} = \boldsymbol{e}_i$ (i.e. $\boldsymbol{a}^\top \boldsymbol{u}_l^\star = u_{l,i}^\star$)

Beyond ℓ_2 analysis: "fine-grained" statistical analysis

More generally, estimate & infer $\underbrace{\textbf{linear functions of eigenvectors}}_{\boldsymbol{a}^\top \boldsymbol{u}_l^\star, \text{ with } \boldsymbol{a} \text{ a fixed vector}}$

- e.g. entrywise analysis: $\boldsymbol{a} = \boldsymbol{e}_i$ (i.e. $\boldsymbol{a}^\top \boldsymbol{u}_l^\star = u_{l,i}^\star$)
- e.g. Fourier coefficients of eigenvectors: $\boldsymbol{a} = \boldsymbol{f}_i$ (Fourier basis)

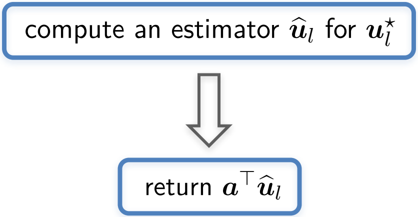




— *Hua, Sarkar '90, Candès, Fernandez-Granda '12, Chen, Chi '14*

Challenge: plug-in estimator?

— *goal: estimate* $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$



Challenge: plug-in estimator?

— *goal: estimate* $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$

compute an estimator $\widehat{\boldsymbol{u}}_l$ for $\boldsymbol{u}_l^\star$

⇓

return $\boldsymbol{a}^\top \widehat{\boldsymbol{u}}_l$

Issue: insufficiency of ℓ_2 analysis

$$\underbrace{|\boldsymbol{a}^\top \widehat{\boldsymbol{u}}_l - \boldsymbol{a}^\top \boldsymbol{u}_l^\star|}_{\text{target estimation risk}} \leq \|\boldsymbol{a}\|_2 \cdot \underbrace{\|\widehat{\boldsymbol{u}}_l - \boldsymbol{u}_l^\star\|_2}_{\text{invoke prior } \ell_2 \text{ bounds}}$$

Challenge: plug-in estimator?

— *goal: estimate* $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$

compute an estimator $\widehat{\boldsymbol{u}}_l$ for $\boldsymbol{u}_l^\star$

⇓

return $\boldsymbol{a}^\top \widehat{\boldsymbol{u}}_l$

Issue: insufficiency of ℓ_2 analysis

$$\underbrace{|\boldsymbol{a}^\top \widehat{\boldsymbol{u}}_l - \boldsymbol{a}^\top \boldsymbol{u}_l^\star|}_{\text{target estimation risk}} \leq \|\boldsymbol{a}\|_2 \cdot \underbrace{\|\widehat{\boldsymbol{u}}_l - \boldsymbol{u}_l^\star\|_2}_{\text{invoke prior } \ell_2 \text{ bounds}}$$

- highly suboptimal (could be $\widetilde{O}(\sqrt{n})$ times larger than true risk)!

Challenge: plug-in estimator?

— *goal: estimate* $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$

compute an estimator $\widehat{\boldsymbol{u}}_l$ for $\boldsymbol{u}_l^\star$

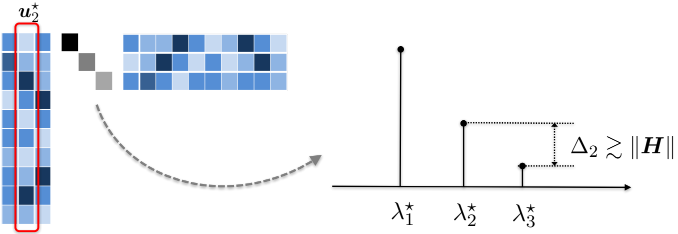
⇓

return $\boldsymbol{a}^\top \widehat{\boldsymbol{u}}_l$

Issue: non-negligible bias

- even when *each entry* of $\widehat{\boldsymbol{u}}_l$ is nearly unbiased, the plug-in estimator might suffer from systematic bias

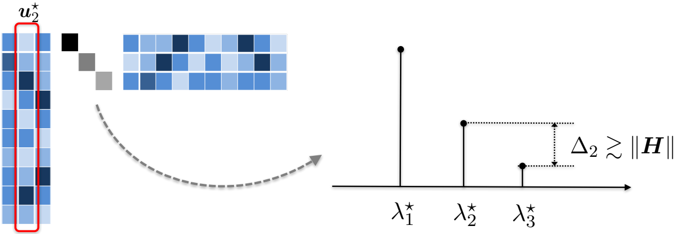
Challenge: stringent e-value separation requirement



To faithfully estimate $u_l^\star$, generic linear algebra typically requires

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

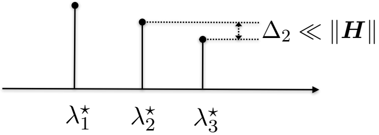
Challenge: stringent e-value separation requirement



To faithfully estimate $\boldsymbol{u}_l^\star$, generic linear algebra typically requires

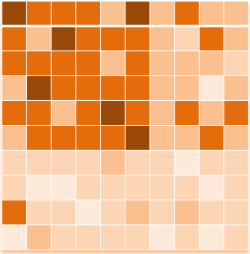
$$\text{(eigenvalue separation)} \quad \Delta_l := \min_{k:k\neq l} |\lambda_l^\star - \lambda_k^\star| \gtrsim \|\boldsymbol{H}\|$$

What if $\Delta_l \ll \|\boldsymbol{H}\|$?



Challenge: noise heteroscedasticity

$$\begin{bmatrix} \mathsf{Var}(H_{11}) & \cdots & \mathsf{Var}(H_{1n}) \\ \vdots & \ddots & \vdots \\ \mathsf{Var}(H_{n1}) & \cdots & \mathsf{Var}(H_{nn}) \end{bmatrix}$$



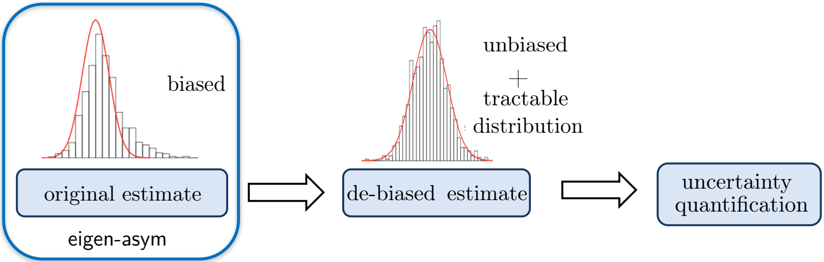
Noise variance might vary across locations, which are *a priori* unknown

This talk: estimation & inference for linear forms of eigenvectors (& eigenvalues) under independent noise

This talk: estimation & inference for linear forms of eigenvectors (& eigenvalues) under independent noise

- effective under minimal eigenvalue separation
- distribution-free
- adaptive to noise heteroscedasticity
- optimal (in some sense)

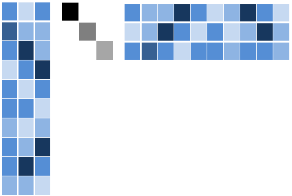
Outline



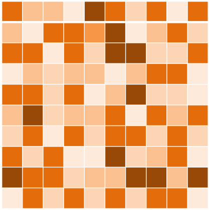
- **Estimation (eigen-decomposition w/o symmetrization)**
- Inference (de-biasing and confidence intervals)

Model: independent noise

$\boldsymbol{M} =$



$+$



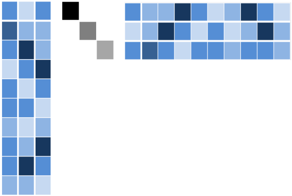
$$\underbrace{\boldsymbol{M}^\star = \sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top}}_{\text{symmetric low-rank}} \in \mathbb{R}^{n \times n}$$

$\boldsymbol{H}$: noise matrix

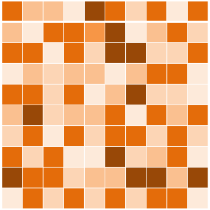
- **independent entries:** $\{H_{i,j}\}$ are independent
- **zero mean:** $\mathbb{E}[H_{i,j}] = 0$
- **variance:** $\sigma_{\min}^2 \le \mathsf{Var}(H_{i,j}) \le \sigma_{\max}^2$

Model: independent noise → asymmetric data

$\boldsymbol{M} =$



$+$



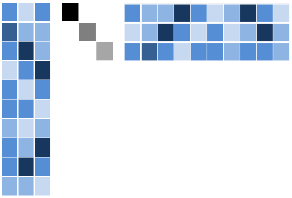
$\boldsymbol{H}$: *asymmetric* matrix

$$\underbrace{\boldsymbol{M}^\star = \sum_{l=1}^{r} \lambda_l^\star \boldsymbol{u}_l^\star \boldsymbol{u}_l^{\star\top}}_{\text{symmetric low-rank}} \in \mathbb{R}^{n\times n}$$

- **independent entries:** $\{H_{i,j}\}$ are independent
- **zero mean:** $\mathbb{E}[H_{i,j}] = 0$
- **variance:** $\sigma_{\min}^2 \leq \mathsf{Var}(H_{i,j}) \leq \sigma_{\max}^2$

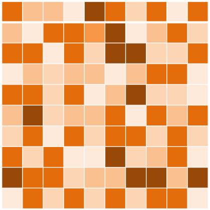
Model: independent noise $\rightarrow$ asymmetric data

$\boldsymbol{M} =$



$$\underbrace{\boldsymbol{M}^{\star} = \sum_{l=1}^{r} \lambda_l^{\star} \boldsymbol{u}_l^{\star} \boldsymbol{u}_l^{\star\top}}_{\text{symmetric low-rank}} \in \mathbb{R}^{n \times n}$$

$+$

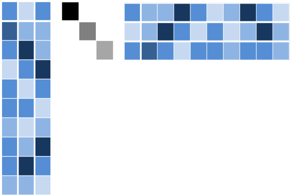


$\boldsymbol{H}$: *asymmetric* matrix

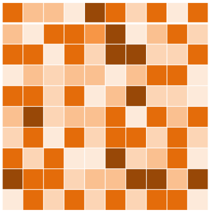
This may arise when, e.g., we have 2 samples for each entry of $\boldsymbol{M}^{\star}$ and arrange them asymmetrically

A natural strategy: symmetrization + eigen-decomposition

$\boldsymbol{M} =$



$+$



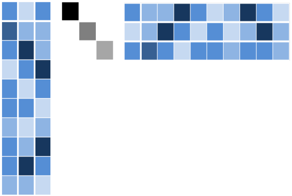
$\boldsymbol{H}$: *asymmetric* matrix

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

- Use l^{th} eigenvector of $\frac{1}{2}(\boldsymbol{M} + \boldsymbol{M}^\top)$ to estimate $\boldsymbol{u}_l^\star$
- Use l^{th} eigenvalue of $\frac{1}{2}(\boldsymbol{M} + \boldsymbol{M}^\top)$ to estimate $\lambda_l^\star$

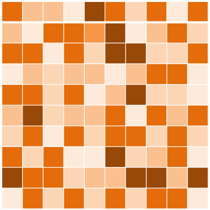
A less popular strategy: eigen-decomposition w/o symmetrization

$\boldsymbol{M} =$



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

$+$



$\boldsymbol{H}$: *asymmetric* matrix

- Use l^{th} eigenvector of ~~$\frac{1}{2}(\boldsymbol{M}+\boldsymbol{M}^\top)$~~ $\boldsymbol{M}$ to estimate $\boldsymbol{u}_l^\star$
- Use l^{th} eigenvalue of ~~$\frac{1}{2}(\boldsymbol{M}+\boldsymbol{M}^\top)$~~ $\boldsymbol{M}$ to estimate $\lambda_l^\star$

Symmetrize or not?

eigen-sym: eigen-decomposition w/ symmetrization

eigen-asym: eigen-decomposition w/o symmetrization

- Numerical stability

$$\text{eigen-sym} \quad > \quad \text{eigen-asym}$$

Symmetrize or not?

eigen-sym: eigen-decomposition w/ symmetrization

eigen-asym: eigen-decomposition w/o symmetrization

- Numerical stability

$$\text{eigen-sym} \quad > \quad \text{eigen-asym}$$

- **(Folklore?)** Statistical accuracy

$$\text{eigen-sym} \quad \asymp \quad \text{eigen-asym}$$

Symmetrize or not?

eigen-sym: eigen-decomposition w/ symmetrization

eigen-asym: eigen-decomposition w/o symmetrization

- Numerical stability

$$\text{eigen-sym} \quad > \quad \text{eigen-asym}$$

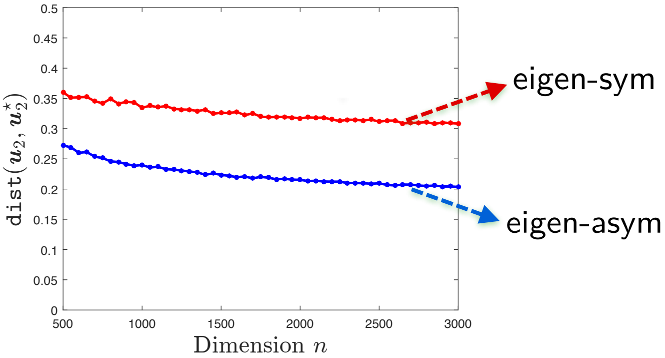
- **(Folklore?)** Statistical accuracy

$$\text{eigen-sym} \quad \asymp \quad \text{eigen-asym}$$

Shall we always symmetrize data before eigen-decomposition?

Numerical experiments: heteroscadestic Gaussian noise

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



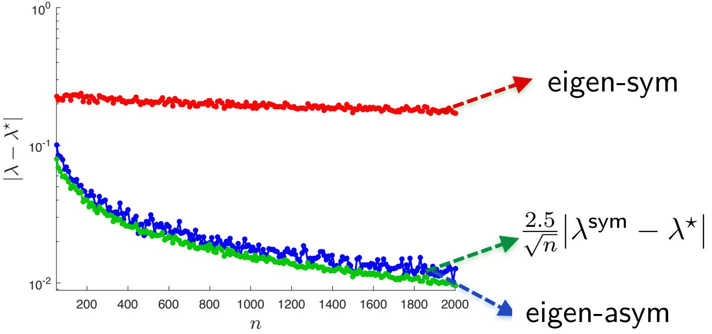
Numerical experiments: matrix completion

$$\boldsymbol{M}^{\star} = \boldsymbol{u}^{\star}\boldsymbol{u}^{\star\top}; \qquad M_{i,j} = \begin{cases} \frac{1}{p} M^{\star}_{i,j} & \text{with prob. } p, \\ 0, & \text{else}, \end{cases} \qquad p = \frac{3 \log n}{n}$$

$$\begin{bmatrix} \checkmark & ? & ? & ? & \checkmark & ? \\ ? & ? & \checkmark & \checkmark & ? & ? \\ \checkmark & ? & ? & \checkmark & ? & ? \\ ? & ? & \checkmark & ? & ? & \checkmark \\ \checkmark & ? & ? & ? & ? & ? \\ ? & \checkmark & ? & ? & \checkmark & ? \end{bmatrix}$$

Numerical experiments: matrix completion

$$M^\star = u^\star u^{\star\top}; \qquad M_{i,j} = \begin{cases} \frac{1}{p} M^\star_{i,j} & \text{with prob. } p, \\ 0, & \text{else,} \end{cases} \qquad p = \frac{3 \log n}{n}$$



Why does eigen-decomposition w/o symmetrization work better?

Problem setup

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

- $\lambda_1^\star \geq \lambda_2^\star \geq \cdots \geq \lambda_r^\star$ and $|\lambda_i^\star| \geq \lambda_{\min}^\star$

Problem setup

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

- $\lambda_1^\star \geq \lambda_2^\star \geq \cdots \geq \lambda_r^\star$ and $|\lambda_i^\star| \geq \lambda_{\min}^\star$
- $\boldsymbol{H}$: noise matrix
  - **independent entries:** $\{H_{i,j}\}$ are independent
  - **zero mean:** $\mathbb{E}[H_{i,j}] = 0$
  - **variance:** $\mathsf{Var}(H_{i,j}) \leq \sigma_{\max}^2 \ll \frac{(\lambda_{\min}^\star)^2}{n \log n}$ (s.t. $\|\boldsymbol{H}\| \ll \lambda_{\min}^\star$)
  - **magnitudes:** $|H_{i,j}| \leq \sigma_{\max}\sqrt{n/\log n}$ with high prob.

Problem setup

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

- $\lambda_1^\star \geq \lambda_2^\star \geq \cdots \geq \lambda_r^\star$ and $|\lambda_i^\star| \geq \lambda_{\min}^\star$
- $\boldsymbol{H}$: noise matrix
  - **independent entries:** $\{H_{i,j}\}$ are independent
  - **zero mean:** $\mathbb{E}[H_{i,j}] = 0$
  - **variance:** $\mathsf{Var}(H_{i,j}) \leq \sigma_{\max}^2 \ll \frac{(\lambda_{\min}^\star)^2}{n \log n}$ (s.t. $\|\boldsymbol{H}\| \ll \lambda_{\min}^\star$)
  - **magnitudes:** $|H_{i,j}| \leq \sigma_{\max}\sqrt{n/\log n}$ with high prob.
- $\boldsymbol{M}^\star$ obeys incoherence condition

$$\max_{1\leq l\leq r} \|\boldsymbol{u}_l^\star\|_\infty \leq \sqrt{\mu/n}$$

Review: classical *eigenvalue* perturbation results

$$|\lambda_l^{\mathsf{sym}} - \lambda_l^{\star}| \leq \|\tfrac{1}{2}(\boldsymbol{H} + \boldsymbol{H}^{\top})\| \qquad \text{(Weyl)}$$
$$|\lambda_l^{\mathsf{asym}} - \lambda_l^{\star}| \leq \|\boldsymbol{H}\| \qquad \text{(Bauer-Fike)}$$

Review: classical *eigenvalue* perturbation results

$$|\lambda_l^{\mathsf{sym}} - \lambda_l^\star| \leq \|\tfrac{1}{2}(\boldsymbol{H} + \boldsymbol{H}^\top)\| \qquad \text{(Weyl)}$$
$$|\lambda_l^{\mathsf{asym}} - \lambda_l^\star| \leq \|\boldsymbol{H}\| \qquad \text{(Bauer-Fike)}$$

$\Downarrow$ matrix concentration inequality

$$|\lambda_l^{\mathsf{sym}} - \lambda_l^\star| \lesssim \sigma_{\max}\sqrt{n}$$
$$|\lambda_l^{\mathsf{asym}} - \lambda_l^\star| \lesssim \sigma_{\max}\sqrt{n}$$

Review: classical *eigenvalue* perturbation results

$$|\lambda_l^{\mathsf{sym}} - \lambda_l^\star| \le \|\tfrac{1}{2}(\boldsymbol{H} + \boldsymbol{H}^\top)\| \qquad \text{(Weyl)}$$
$$|\lambda_l^{\mathsf{asym}} - \lambda_l^\star| \le \|\boldsymbol{H}\| \qquad \text{(Bauer-Fike)}$$

$\Downarrow$ matrix concentration inequality

$$|\lambda_l^{\mathsf{sym}} - \lambda_l^\star| \lesssim \sigma_{\max}\sqrt{n} \qquad \text{(reasonably tight if } \|\boldsymbol{H}\| \text{ is large)}$$
$$|\lambda_l^{\mathsf{asym}} - \lambda_l^\star| \lesssim \sigma_{\max}\sqrt{n}$$

Review: classical *eigenvalue* perturbation results

$$|\lambda_l^{\mathsf{sym}} - \lambda_l^\star| \leq \|\tfrac{1}{2}(\boldsymbol{H} + \boldsymbol{H}^\top)\| \qquad \text{(Weyl)}$$
$$|\lambda_l^{\mathsf{asym}} - \lambda_l^\star| \leq \|\boldsymbol{H}\| \qquad \text{(Bauer-Fike)}$$

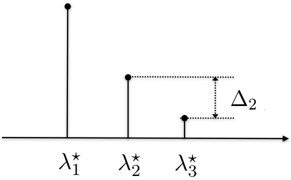
$\Downarrow$ matrix concentration inequality

$$|\lambda_l^{\mathsf{sym}} - \lambda_l^\star| \lesssim \sigma_{\max}\sqrt{n}$$ (reasonably tight if $\|\boldsymbol{H}\|$ is large)

$$|\lambda_l^{\mathsf{asym}} - \lambda_l^\star| \lesssim \sigma_{\max}\sqrt{n}$$ (*can be significantly improved*)

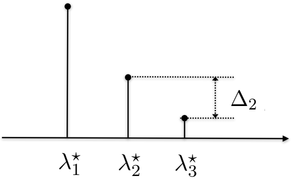
Review: classical *eigenvector* perturbation results

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



Review: classical *eigenvector* perturbation results

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

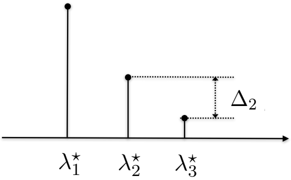


$$\min \left\| \boldsymbol{u}_l^{\mathsf{sym}} \pm \boldsymbol{u}_l^\star \right\|_2 \lesssim \frac{\|\boldsymbol{H} + \boldsymbol{H}^\top\|}{\Delta_l} \qquad \text{(Davis-Kahan)}$$

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

Review: classical *eigenvector* perturbation results

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



$$\min \|\boldsymbol{u}_l^{\textsf{sym}} \pm \boldsymbol{u}_l^\star\|_2 \lesssim \frac{\|\boldsymbol{H} + \boldsymbol{H}^\top\|}{\Delta_l} \qquad \text{(Davis-Kahan)}$$

$$\min \|\boldsymbol{u}_l^{\textsf{asym}} \pm \boldsymbol{u}_l^\star\|_2 \lesssim ??$$

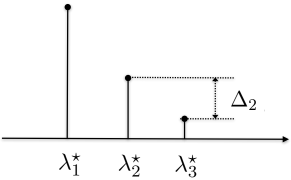
$\Downarrow$ matrix concentration inequality

$$\min \|\boldsymbol{u}_l^{\textsf{sym}} \pm \boldsymbol{u}_l^\star\|_2 \lesssim \frac{\sigma_{\max}\sqrt{n}}{\Delta_l} \qquad \text{(requires } \Delta_l \gtrsim \|\boldsymbol{H} + \boldsymbol{H}^\top\|\text{)}$$

$$\min \|\boldsymbol{u}_l^{\textsf{asym}} \pm \boldsymbol{u}_l^\star\|_2 \lesssim ??$$

Main results: eigenvalue / eigenvector perturbation

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



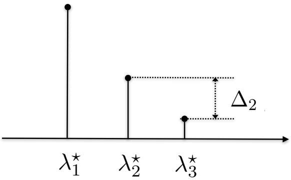
Theorem 1 (Cheng, Wei, Chen '20)

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

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

Main results: eigenvalue / eigenvector perturbation

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



Theorem 1 (Cheng, Wei, Chen '20)

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

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

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

$$\left|\lambda_l^{\mathsf{asym}} - \lambda_l^\star\right| \leq c_0\sigma_{\max}\sqrt{\log n}$$

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

Eigen-sym vs. eigen-asym

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

$$|\lambda_l^{\mathsf{sym}} - \lambda_l^\star| \lesssim \sigma_{\max}\sqrt{n} \qquad \text{(Weyl)}$$
$$|\lambda_l^{\mathsf{asym}} - \lambda_l^\star| \lesssim \sigma_{\max}\sqrt{\log n} \qquad \text{(ours)}$$

Eigen-sym vs. eigen-asym

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

$$|\lambda_l^{\mathsf{sym}} - \lambda_l^\star| \lesssim \sigma_{\max}\sqrt{n} \qquad \text{(Weyl)}$$
$$|\lambda_l^{\mathsf{asym}} - \lambda_l^\star| \lesssim \sigma_{\max}\sqrt{\log n} \qquad \text{(ours)}$$

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

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

Main results: entrywise eigenvector perturbation

Theorem 2 (Cheng, Wei, Chen '20)

Under same assumptions as in Theorem 1, with high prob. one has

$$\min \left\| \boldsymbol{u}_l^{\mathsf{asym}} \pm \boldsymbol{u}_l^\star \right\|_\infty \lesssim \frac{\sigma_{\max}\sqrt{\log n}}{\|\boldsymbol{M}^\star\|} + \frac{\sigma_{\max}}{\Delta_l}\sqrt{\frac{\log n}{n}}$$

Main results: entrywise eigenvector perturbation

Theorem 2 (Cheng, Wei, Chen '20)

Under same assumptions as in Theorem 1, with high prob. one has

$$\min \|\boldsymbol{u}_l^{\mathsf{asym}} \pm \boldsymbol{u}_l^\star\|_\infty \lesssim \frac{\sigma_{\max}\sqrt{\log n}}{\|\boldsymbol{M}^\star\|} + \frac{\sigma_{\max}}{\Delta_l}\sqrt{\frac{\log n}{n}}$$

- ℓ_∞ perturbation is well-controlled (i.e. $\frac{\min \|\boldsymbol{u}_l^{\mathsf{asym}} \pm \boldsymbol{u}_l^\star\|_\infty}{\|\boldsymbol{u}_l^\star\|_\infty} \ll 1$) even under very small eigenvalue separation (i.e. $\Delta_l \asymp \sigma_{\max}\sqrt{\log n}$)

Main results: entrywise eigenvector perturbation

Theorem 2 (Cheng, Wei, Chen '20)

Under same assumptions as in Theorem 1, with high prob. one has

$$\min \|\boldsymbol{u}_l^{\mathsf{asym}} \pm \boldsymbol{u}_l^\star\|_\infty \lesssim \frac{\sigma_{\max}\sqrt{\log n}}{\|\boldsymbol{M}^\star\|} + \frac{\sigma_{\max}}{\Delta_l}\sqrt{\frac{\log n}{n}}$$

- ℓ_∞ perturbation is well-controlled (i.e. $\frac{\min \|\boldsymbol{u}_l^{\mathsf{asym}} \pm \boldsymbol{u}_l^\star\|_\infty}{\|\boldsymbol{u}_l^\star\|_\infty} \ll 1$) even under very small eigenvalue separation (i.e. $\Delta_l \asymp \sigma_{\max}\sqrt{\log n}$)
- no available ℓ_∞ guarantees for $\boldsymbol{u}_l^{\mathsf{sym}}$ if $\Delta_l \ll \sigma_{\max}\sqrt{n}$

Intuition: asymmetry reduces bias (rank-1, i.i.d. noise)

From $\underbrace{\text{Neumann series}}_{\text{some sort of Taylor expansion}}$ one can verify

$$|\lambda_1 - \lambda_1^\star| \asymp \left| \frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda} + \frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H}^2 \boldsymbol{u}_1^\star}{\lambda^2} + \frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H}^3 \boldsymbol{u}_1^\star}{\lambda^3} + \cdots \right|$$

Intuition: asymmetry reduces bias (rank-1, i.i.d. noise)

From $\underbrace{\text{Neumann series}}_{\text{some sort of Taylor expansion}}$ one can verify

$$|\lambda_1 - \lambda_1^\star| \asymp \left| \frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda} + \boxed{\frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H}^2 \boldsymbol{u}_1^\star}{\lambda^2}} + \frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H}^3 \boldsymbol{u}_1^\star}{\lambda^3} + \cdots \right|$$

To develop some intuition, let's look at 2nd order term

Intuition: asymmetry reduces bias (rank-1, i.i.d. noise)

From $\underbrace{\text{Neumann series}}_{\text{some sort of Taylor expansion}}$ one can verify

$$|\lambda_1 - \lambda_1^\star| \asymp \left| \frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda} + \boxed{\frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H}^2 \boldsymbol{u}_1^\star}{\lambda^2}} + \frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H}^3 \boldsymbol{u}_1^\star}{\lambda^3} + \cdots \right|$$

To develop some intuition, let's look at 2nd order term

- if $\boldsymbol{H}$ is symmetric,

$$\mathbb{E}[\boldsymbol{u}_1^{\star\top} \boldsymbol{H}^2 \boldsymbol{u}_1^\star] = \mathbb{E}[\|\boldsymbol{H}\boldsymbol{u}_1^\star\|_2^2] = n\sigma^2$$

Intuition: asymmetry reduces bias (rank-1, i.i.d. noise)

From $\underbrace{\text{Neumann series}}_{\text{some sort of Taylor expansion}}$ one can verify

$$|\lambda_1 - \lambda_1^\star| \asymp \left| \frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda} + \boxed{\frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H}^2 \boldsymbol{u}_1^\star}{\lambda^2}} + \frac{\boldsymbol{u}_1^{\star\top} \boldsymbol{H}^3 \boldsymbol{u}_1^\star}{\lambda^3} + \cdots \right|$$

To develop some intuition, let's look at 2nd order term

- if $\boldsymbol{H}$ is symmetric,

$$\mathbb{E}[\boldsymbol{u}_1^{\star\top} \boldsymbol{H}^2 \boldsymbol{u}_1^\star] = \mathbb{E}[\|\boldsymbol{H} \boldsymbol{u}_1^\star\|_2^2] = n\sigma^2$$

- if $\boldsymbol{H}$ is asymmetric,

$$\underbrace{\mathbb{E}[\boldsymbol{u}_1^{\star\top} \boldsymbol{H}^2 \boldsymbol{u}_1^\star] = \mathbb{E}[\langle \boldsymbol{H}^\top \boldsymbol{u}_1^\star, \boldsymbol{H} \boldsymbol{u}_1^\star \rangle] = \sigma^2}_{\text{significantly smaller bias than symmetric case}}$$

Summary of main features of eigen-asym

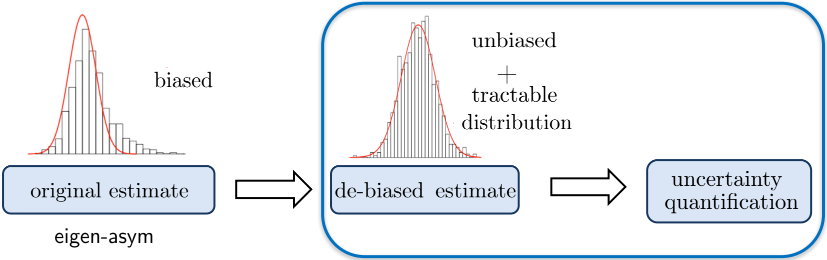
- much higher e-value estimation accuracy
- faithful e-vector estimation under much smaller e-value separation
- distribution-free
- adaptive to heteroscedastic noise

Summary of main features of eigen-asym

- much higher e-value estimation accuracy
- faithful e-vector estimation under much smaller e-value separation
- distribution-free
- adaptive to heteroscedastic noise

Statistical asymmetry implicitly **suppresses estimation bias**, which empowers eigen-decomposition!

Outline



- Estimation (eigen-decomposition w/o symmetrization)
- **Inference (de-biasing and confidence intervals)**

Notation

— let's drop superscript "asym" to simplify notation

- λ_l: l^{th} eigenvalue of $\boldsymbol{M}$
- $\boldsymbol{u}_l$: l^{th} right eigenvector of $\boldsymbol{M}$ obeying $\boldsymbol{u}_l^\top \boldsymbol{u}_l^\star \geq 0$
- $\boldsymbol{w}_l$: l^{th} left eigenvector of $\boldsymbol{M}$ obeying $\boldsymbol{w}_l^\top \boldsymbol{u}_l^\star \geq 0$

As we shall see, it's crucial to employ $\boldsymbol{u}_l$ and $\boldsymbol{w}_l$ simultaneously

Which estimator shall we use?

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

Which estimator shall we use?

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

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: $\boldsymbol{a}^\top \boldsymbol{u}_l$ (or $\boldsymbol{a}^\top \boldsymbol{w}_l$)

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 observations (rank-1 case)

From Neumann's series:

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1) \left\{ \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \tfrac{1}{\lambda_1^\star} \boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star + \tfrac{1}{\lambda_1^{\star 2}} \boldsymbol{a}^\top \boldsymbol{H}^2 \boldsymbol{u}_1^\star + \cdots \right\}$$

Key observations (rank-1 case)

From Neumann's series:

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1) \left\{ \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \frac{1}{\lambda_1^\star} \boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star + \frac{1}{\lambda_1^{\star 2}} \boldsymbol{a}^\top \boldsymbol{H}^2 \boldsymbol{u}_1^\star + \cdots \right\}$$

Key observations (rank-1 case)

From Neumann's series:

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1) \left\{ \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \tfrac{1}{\lambda_1^\star} \boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star \right\}$$

$$\Downarrow \qquad (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1 \approx 1)$$

- If $\boldsymbol{a}^\top \boldsymbol{u}_1^\star$ is small, then

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda_1^\star} + (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1 - 1) \left\{ \boldsymbol{a}^\top \boldsymbol{u}^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda_1^\star} \right\}$$

Key observations (rank-1 case)

From Neumann's series:

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx (\boldsymbol{u}_1^{\star\top}\boldsymbol{u}_1)\left\{\boldsymbol{a}^\top \boldsymbol{u}_1^\star + \tfrac{1}{\lambda_1^\star}\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_1^\star\right\}$$

$$\Downarrow \qquad (\boldsymbol{u}_1^{\star\top}\boldsymbol{u}_1 \approx 1)$$

- If $\boldsymbol{a}^\top \boldsymbol{u}_1^\star$ is small, then

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_1^\star}{\lambda_1^\star} + \cancel{(\boldsymbol{u}_1^{\star\top}\boldsymbol{u}_1 - 1)\left\{\boldsymbol{a}^\top \boldsymbol{u}_1^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H}\boldsymbol{u}_1^\star}{\lambda_1^\star}\right\}}$$

$\Longrightarrow$ $\boldsymbol{a}^\top \boldsymbol{u}_1$ is nearly unbiased estimate of $\boldsymbol{a}^\top \boldsymbol{u}_1^\star$

Key observations (rank-1 case)

From Neumann's series:

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1) \left\{ \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \tfrac{1}{\lambda_1^\star} \boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star \right\}$$

$$\Downarrow \qquad (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1 \approx 1)$$

- If $\boldsymbol{a}^\top \boldsymbol{u}_1^\star$ is large, then we can't simply replace $\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1$ by 1

$$\tfrac{1}{\color{red}{\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1}} \boldsymbol{a}^\top \boldsymbol{u}_1 - \boldsymbol{a}^\top \boldsymbol{u}_1^\star \approx \tfrac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda_1^\star}$$

Key observations (rank-1 case)

From Neumann's series:

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1) \left\{ \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \tfrac{1}{\lambda_1^\star} \boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star \right\}$$

$$\Downarrow \qquad (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1 \approx 1)$$

- If $\boldsymbol{a}^\top \boldsymbol{u}_1^\star$ is large, then we can't simply replace $\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1$ by 1

$$\frac{1}{\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1} \boldsymbol{a}^\top \boldsymbol{u}_1 - \boldsymbol{a}^\top \boldsymbol{u}_1^\star \approx \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda_1^\star}$$

  - **problem:** $\boldsymbol{a}^\top \boldsymbol{u}_1$ significantly under-estimates $\boldsymbol{a}^\top \boldsymbol{u}_1^\star$;

Key observations (rank-1 case)

From Neumann's series:

$$\boldsymbol{a}^\top \boldsymbol{u}_1 \approx (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1) \left\{ \boldsymbol{a}^\top \boldsymbol{u}_1^\star + \tfrac{1}{\lambda_1^\star} \boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star \right\}$$

$$\Downarrow \qquad (\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1 \approx 1)$$

- If $\boldsymbol{a}^\top \boldsymbol{u}_1^\star$ is large, then we can't simply replace $\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1$ by 1

$$\frac{1}{\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1} \boldsymbol{a}^\top \boldsymbol{u}_1 - \boldsymbol{a}^\top \boldsymbol{u}_1^\star \approx \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_1^\star}{\lambda_1^\star}$$

  - **problem:** $\boldsymbol{a}^\top \boldsymbol{u}_1$ significantly under-estimates $\boldsymbol{a}^\top \boldsymbol{u}_1^\star$;
  - **bias correction:** estimate $\boldsymbol{u}_1^{\star\top} \boldsymbol{u}_1$ and use it to adjust $\boldsymbol{a}^\top \boldsymbol{u}_1$

Our estimator for $a^\top u_l^\star$

$$\begin{cases} \boldsymbol{a}^\top \boldsymbol{u}_l & \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_l^\star}{\lambda_l^\star} \\ & \end{cases}$$

When $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ is small:

$$\widehat{u}_{\boldsymbol{a},l} = \boldsymbol{a}^\top \boldsymbol{u}_l$$

Our estimator for $a^\top u_l^\star$

$$\begin{cases} \boldsymbol{a}^\top \boldsymbol{u}_l & \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_l^\star}{\lambda_l^\star} \\ \boldsymbol{a}^\top \boldsymbol{w}_l & \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H}^\top \boldsymbol{u}_l^\star}{\lambda_l^\star} \end{cases}$$

When $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ is small:

$$\widehat{u}_{\boldsymbol{a},l} = \boldsymbol{a}^\top \boldsymbol{u}_l \quad \text{or} \quad \boldsymbol{a}^\top \boldsymbol{w}_l$$

Our estimator for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$

$$\begin{cases} \boldsymbol{a}^\top \boldsymbol{u}_l & \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H} \boldsymbol{u}_l^\star}{\lambda_l^\star} \\ \boldsymbol{a}^\top \boldsymbol{w}_l & \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top \boldsymbol{H}^\top \boldsymbol{u}_l^\star}{\lambda_l^\star} \end{cases}$$

When $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ is small:

$$\widehat{u}_{\boldsymbol{a},l} = \underbrace{\frac{1}{2}\left(\boldsymbol{a}^\top \boldsymbol{u}_l + \boldsymbol{a}^\top \boldsymbol{w}_l\right)}_{\text{average them to reduce uncertainty}}$$

Our estimator for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$

$$\begin{cases} \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\boldsymbol{a}^\top\boldsymbol{u}_l & \approx \boldsymbol{a}^\top\boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_1^\star} \\ \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}\boldsymbol{a}^\top\boldsymbol{w}_l & \approx \boldsymbol{a}^\top\boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top\boldsymbol{H}^\top\boldsymbol{u}_l^\star}{\lambda_1^\star} \end{cases}$$

When $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$ is large, our estimator is

$$\widehat{u}_{\boldsymbol{a},l} = \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\boldsymbol{a}^\top\boldsymbol{u}_l \quad \text{or} \quad \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}\boldsymbol{a}^\top\boldsymbol{w}_l$$

Our estimator for $a^\top \boldsymbol{u}_l^\star$

$$\begin{cases} \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\boldsymbol{a}^\top\boldsymbol{u}_l & \approx \boldsymbol{a}^\top\boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_1^\star} \\ \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}\boldsymbol{a}^\top\boldsymbol{w}_l & \approx \boldsymbol{a}^\top\boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top\boldsymbol{H}^\top\boldsymbol{u}_l^\star}{\lambda_1^\star} \end{cases}$$

When $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$ is large, our estimator is

$$\widehat{u}_{\boldsymbol{a},l} = \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\boldsymbol{a}^\top\boldsymbol{u}_l \text{ or } \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}\boldsymbol{a}^\top\boldsymbol{w}_l \quad \text{(un-implementable)}$$

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

Our estimator for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$

$$\begin{cases} \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\boldsymbol{a}^\top\boldsymbol{u}_l & \approx \boldsymbol{a}^\top\boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_1^\star} \\ \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}\boldsymbol{a}^\top\boldsymbol{w}_l & \approx \boldsymbol{a}^\top\boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top\boldsymbol{H}^\top\boldsymbol{u}_l^\star}{\lambda_1^\star} \end{cases}$$

When $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$ is large, our estimator is

~~$\widehat{u}_{\boldsymbol{a},l} = \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\boldsymbol{a}^\top\boldsymbol{u}_l$ or $\frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}\boldsymbol{a}^\top\boldsymbol{w}_l$~~ (un-implementable)

~~$\widehat{u}_{\boldsymbol{a},l} = \sqrt{\frac{1}{(\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l)(\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l)}(\boldsymbol{a}^\top\boldsymbol{u}_l)(\boldsymbol{a}^\top\boldsymbol{w}_l)}$~~ (un-implementable)

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

Our estimator for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$

$$\begin{cases} \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\boldsymbol{a}^\top\boldsymbol{u}_l & \approx \boldsymbol{a}^\top\boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top\boldsymbol{H}\boldsymbol{u}_l^\star}{\lambda_1^\star} \\ \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}\boldsymbol{a}^\top\boldsymbol{w}_l & \approx \boldsymbol{a}^\top\boldsymbol{u}_l^\star + \frac{\boldsymbol{a}^\top\boldsymbol{H}^\top\boldsymbol{u}_l^\star}{\lambda_1^\star} \end{cases}$$

When $\boldsymbol{a}^\top\boldsymbol{u}_l^\star$ is large, our estimator is

$$\widehat{u}_{\boldsymbol{a},l} = \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l}\boldsymbol{a}^\top\boldsymbol{u}_l \quad \text{or} \quad \frac{1}{\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l}\boldsymbol{a}^\top\boldsymbol{w}_l \qquad \text{(un-implementable)}$$

$$\widehat{u}_{\boldsymbol{a},l} = \sqrt{\frac{1}{(\boldsymbol{u}_l^{\star\top}\boldsymbol{u}_l)(\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l)}(\boldsymbol{a}^\top\boldsymbol{u}_l)(\boldsymbol{a}^\top\boldsymbol{w}_l)} \qquad \text{(un-implementable)}$$

$$\widehat{u}_{\boldsymbol{a},l} = \sqrt{\frac{1}{\boldsymbol{u}_l^\top\boldsymbol{w}_l}(\boldsymbol{a}^\top\boldsymbol{u}_l)(\boldsymbol{a}^\top\boldsymbol{w}_l)} \qquad \text{(use } \boldsymbol{u}_l^\top\boldsymbol{w}_l \text{ to estimate } (\boldsymbol{u}_l^\star\boldsymbol{u}_l^\top)(\boldsymbol{u}_l^{\star\top}\boldsymbol{w}_l))$$

Problem setup

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

- $\lambda_1^\star \geq \lambda_2^\star \geq \cdots \geq \lambda_r^\star$ and $|\lambda_i^\star| \geq \lambda_{\min}^\star$

Problem setup

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

- $\lambda_1^\star \geq \lambda_2^\star \geq \cdots \geq \lambda_r^\star$ and $|\lambda_i^\star| \geq \lambda_{\min}^\star$
- $\boldsymbol{H}$: 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.

Problem setup

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

- $\lambda_1^\star \geq \lambda_2^\star \geq \cdots \geq \lambda_r^\star$ and $|\lambda_i^\star| \geq \lambda_{\min}^\star$
- $\boldsymbol{H}$: 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

$$\max_{1\leq l\leq r} \|\boldsymbol{u}_l^\star\|_\infty \leq \sqrt{\mu/n}$$

Main results: distributional theory

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

Theorem 3 (Cheng, Wei, Chen '20)

Under above assumptions, with high prob. one has

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

Distributional theory $\longrightarrow$ confidence intervals

$$\widehat{u}_{\boldsymbol{a},l} \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \underbrace{\frac{1}{2\lambda_l^\star}\boldsymbol{a}^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}_{\text{approximately } \mathcal{N}(0, v_{\boldsymbol{a},l}^\star)}$$

- $\widehat{u}_{\boldsymbol{a},l}$: unbiased estimator for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$

Distributional theory $\longrightarrow$ confidence intervals

$$\widehat{u}_{\boldsymbol{a},l} \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \underbrace{\tfrac{1}{2\lambda_l^\star}\boldsymbol{a}^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}_{\text{approximately } \mathcal{N}(0,\, v_{\boldsymbol{a},l}^\star)}$$

- $\widehat{u}_{\boldsymbol{a},l}$: unbiased estimator for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$
- estimation error is approximately Gaussian, whose variance $v_{\boldsymbol{a},l}^\star$ can be faithfully estimated

Distributional theory $\longrightarrow$ confidence intervals

$$\widehat{u}_{\boldsymbol{a},l} \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \underbrace{\tfrac{1}{2\lambda_l^\star}\boldsymbol{a}^\top(\boldsymbol{H}+\boldsymbol{H}^\top)\boldsymbol{u}_l^\star}_{\text{approximately } \mathcal{N}(0,\, v^\star_{\boldsymbol{a},l})}$$

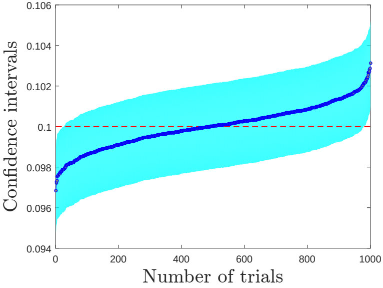
- $\widehat{u}_{\boldsymbol{a},l}$: unbiased estimator for $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$
- estimation error is approximately Gaussian, whose variance $v^\star_{\boldsymbol{a},l}$ can be faithfully estimated
- yields $(1-\alpha)$-confidence interval:

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

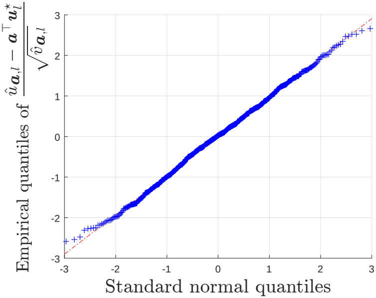
Numerics: estimating $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$

- rank-2: $\lambda_1^\star = 1$, $\lambda_2^\star = 0.95$, $\boldsymbol{a}^\top \boldsymbol{u}_1^\star = 0$, $\boldsymbol{a}^\top \boldsymbol{u}_2^\star = 0.1$
- heteroscadestic 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

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

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

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

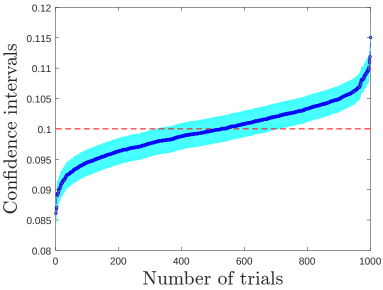
Recall that our theory requires control of the "interferers" $\{\boldsymbol{a}^\top \boldsymbol{u}_k^\star\}_{k\neq l}$

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

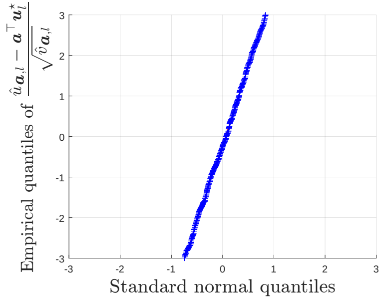
Numerics: estimating $\boldsymbol{a}^\top \boldsymbol{u}_2^\star$

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

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



95% confidence intervals



Q-Q (quantile-quantile) plot

Distributional guarantees for eigenvalues

Proposed estimator for $\lambda_l^\star$: l^{th} eigenvalue λ_l of $\boldsymbol{M}$

Distributional guarantees for eigenvalues

Proposed estimator for $\lambda_l^\star$: l^{th} eigenvalue λ_l of $\boldsymbol{M}$

- $\boldsymbol{M}^\star$ is well-conditioned, incoherent, and $r = O(1)$
- $\sigma_{\max} \log n = o(\Delta_l^\star)$ (minimal e-value separation)

Theorem 4 (Cheng, Wei, Chen '20)

Under above assumptions, with high prob. one has

$$\lambda_l \approx \lambda_l^\star + \boldsymbol{u}_l^{\star\top} \boldsymbol{H} \boldsymbol{u}_l^\star$$

Distributional guarantees for eigenvalues

$$\lambda_l \approx \lambda_l^\star + \underbrace{\boldsymbol{u}_l^{\star\top} \boldsymbol{H} \boldsymbol{u}_l^\star}_{\text{approximately } \mathcal{N}(0, v_{\lambda,l}^\star)}$$

- λ_l is unbiased estimator of $\lambda_l^\star$
  — *no need of bias correction!*

Distributional guarantees for eigenvalues

$$\lambda_l \approx \lambda_l^\star + \underbrace{\boldsymbol{u}_l^{\star\top} \boldsymbol{H} \boldsymbol{u}_l^\star}_{\text{approximately } \mathcal{N}(0, v_{\lambda,l}^\star)}$$

- λ_l is unbiased estimator of $\lambda_l^\star$
  — *no need of bias correction!*
- estimation error is approximately Gaussian, whose variance $v_{\lambda,l}^\star$ can be faithfully estimated

Distributional guarantees for eigenvalues

$$\lambda_l \approx \lambda_l^\star + \underbrace{\boldsymbol{u}_l^{\star\top} \boldsymbol{H} \boldsymbol{u}_l^\star}_{\text{approximately } \mathcal{N}(0, v_{\lambda,l}^\star)}$$

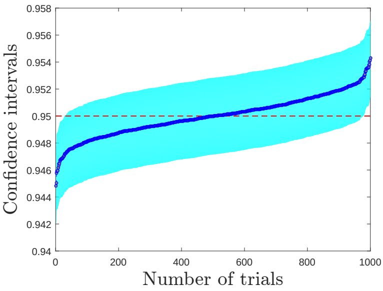
- λ_l is unbiased estimator of $\lambda_l^\star$
  — *no need of bias correction!*
- estimation error is approximately Gaussian, whose variance $v_{\lambda,l}^\star$ can be faithfully estimated
- yields $(1-\alpha)$-confidence interval

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

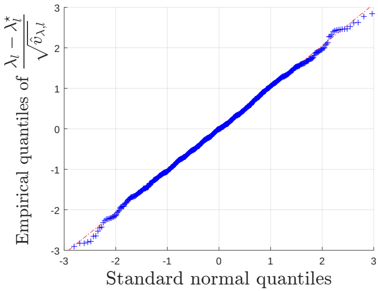
Numerics: estimating $\lambda_2^\star$

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

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



numerical coverage



Q-Q (quantile-quantile) plot

Can we further shorten our confidence intervals?

Cramer-Rao lower bound

- $H_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$
- $\boldsymbol{a}^\top \boldsymbol{u}_l^\star = o(1)$

Theorem 5 (Cheng, Wei, Chen '20)

Any unbiased estimator $\widehat{U}_{\boldsymbol{a}}$ *of* $\boldsymbol{a}^\top \boldsymbol{u}_l^\star$ *obeys*

$$\mathsf{Var}[\widehat{U}_{\boldsymbol{a}}] \geq (1 - o(1))\mathsf{Var}\Big(\frac{1}{2\lambda_l^\star}\boldsymbol{a}(\boldsymbol{H} + \boldsymbol{H}^\top)\boldsymbol{u}_l^\star\Big)$$

- in comparison, our estimator obeys

$$\widehat{u}_{\boldsymbol{a},l} \approx \boldsymbol{a}^\top \boldsymbol{u}_l^\star + \frac{1}{2\lambda_l^\star}\boldsymbol{a}(\boldsymbol{H} + \boldsymbol{H}^\top)\boldsymbol{u}_l^\star \quad \longrightarrow \quad \underbrace{\text{optimal!}}_{\text{including pre-constant}}$$

Cramer-Rao lower bound

- $H_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$
- $\boldsymbol{a}^\top \boldsymbol{u}_l^\star = o(1)$

Theorem 5 (Cheng, Wei, Chen '20)

Any unbiased estimator $\widehat{\Lambda}$ *of* $\lambda_l^\star$ *obeys*

$$\mathsf{Var}[\widehat{\Lambda}] \geq (1 - o(1))\mathsf{Var}\big(\boldsymbol{u}_l^{\star\top} \boldsymbol{H} \boldsymbol{u}_l^\star\big)$$

- in comparison, our estimator obeys

$$\lambda_l \approx \lambda_l^\star + \boldsymbol{u}_l^{\star\top} \boldsymbol{H} \boldsymbol{u}_l^\star \quad \longrightarrow \quad \underbrace{\text{optimal!}}_{\text{including pre-constant}}$$

Concluding remarks

Eigen-decomposition (without symmetrization) could be very powerful when dealing with non-symmetric data matrices

Concluding remarks

Eigen-decomposition (without symmetrization) could be very powerful when dealing with non-symmetric data matrices

- 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