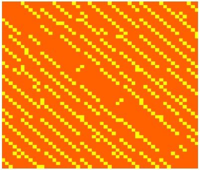
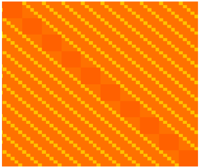
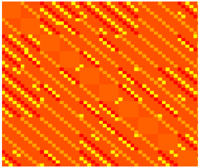
Spectral Methods Meet Asymmetry: Two Recent Stories

Yuxin Chen
Electrical Engineering, Princeton University

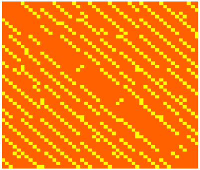
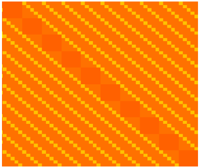
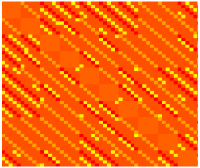
Spectral methods based on eigen-decomposition

$$\boldsymbol{M} \quad = \quad \underbrace{\mathbb{E}[\boldsymbol{M}]}_{\text{approx. low-rank}} \quad + \quad \boldsymbol{M} - \mathbb{E}\left[\boldsymbol{M}\right]$$

Methods based on *eigen-decomposition* of a certain data matrix $\boldsymbol{M}$ ...

Spectral methods based on eigen-decomposition

$$\boldsymbol{M} \quad = \quad \underbrace{\mathbb{E}[\boldsymbol{M}]}_{\text{approx. low-rank}} \quad + \quad \boldsymbol{M} - \mathbb{E}[\boldsymbol{M}]$$

Methods based on *eigen-decomposition* of a certain data matrix $\boldsymbol{M}$ ...

This talk: what happens if data matrix $\boldsymbol{M}$ is non-symmetric?

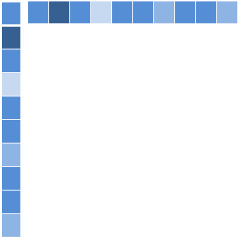
— *2 recent stories*

Asymmetry helps: eigenvalue and eigenvector analyses of asymmetrically perturbed low-rank matrices

Chen Cheng
Stanford Stats

Jianqing Fan
Princeton ORFE

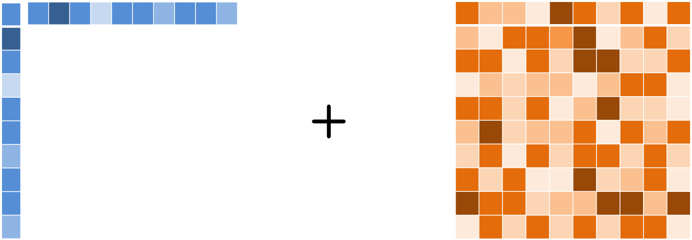
Eigenvalue / eigenvector estimation



$M^\star$: truth

- A rank-1 matrix: $M^\star = \lambda^\star u^\star u^{\star\top} \in \mathbb{R}^{n \times n}$

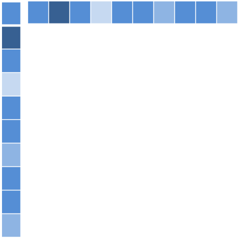
Eigenvalue / eigenvector estimation



$\boldsymbol{M}^{\star}$: truth $\qquad$ $\boldsymbol{H}$: noise

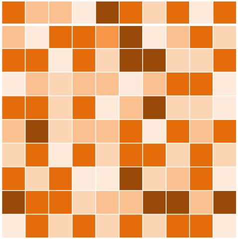
- A rank-1 matrix: $\boldsymbol{M}^{\star} = \lambda^{\star} \boldsymbol{u}^{\star} \boldsymbol{u}^{\star\top} \in \mathbb{R}^{n \times n}$
- Observed noisy data: $\boldsymbol{M} = \boldsymbol{M}^{\star} + \boldsymbol{H}$

Eigenvalue / eigenvector estimation



$\boldsymbol{M}^\star$: truth

$+$

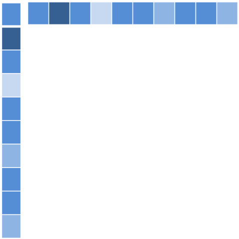


$\boldsymbol{H}$: noise

- A rank-1 matrix: $\boldsymbol{M}^\star = \lambda^\star \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} \in \mathbb{R}^{n\times n}$
- Observed noisy data: $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$
- **Goal:** estimate eigenvalue $\lambda^\star$ and eigenvector $\boldsymbol{u}^\star$

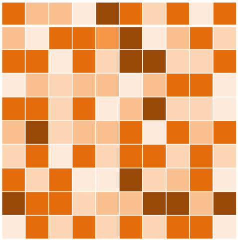
Non-symmetric noise matrix

$$\boldsymbol{M} =$$



$\boldsymbol{M}^\star = \lambda^\star \boldsymbol{u}^\star \boldsymbol{u}^{\star\top}$

$$+$$

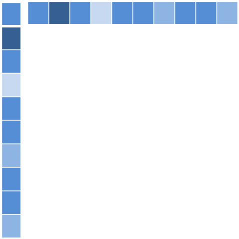


$\boldsymbol{H}$: *asymmetric* matrix

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

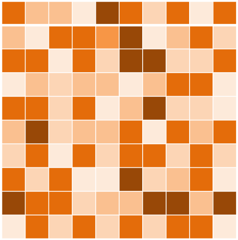
A natural estimation strategy: SVD

$$M =$$



$M^\star = \lambda^\star u^\star u^{\star\top}$

$$+$$

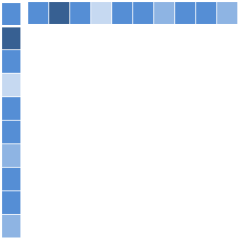


H: *asymmetric* matrix

- Use leading singular value λ^{svd} of M to estimate $\lambda^\star$
- Use leading left singular vector of M to estimate $u^\star$

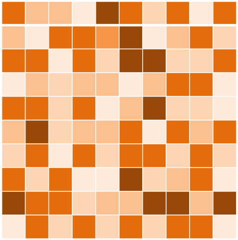
A less popular strategy: eigen-decomposition

$$\boldsymbol{M} =$$



$\boldsymbol{M}^\star = \lambda^\star \boldsymbol{u}^\star \boldsymbol{u}^{\star\top}$

$+$



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

- Use leading ~~singular value λ^{svd}~~ eigenvalue λ^{eigs} of $\boldsymbol{M}$ to estimate $\lambda^\star$
- Use leading ~~singular vector~~ eigenvector of $\boldsymbol{M}$ to estimate $\boldsymbol{u}^\star$

SVD vs. eigen-decomposition

For *asymmetric* matrices:

- Numerical stability

$$\text{SVD} \quad > \quad \text{eigen-decomposition}$$

SVD vs. eigen-decomposition

For *asymmetric* matrices:

- Numerical stability

$$\text{SVD} \quad > \quad \text{eigen-decomposition}$$

- **(Folklore?)** Statistical accuracy

$$\text{SVD} \quad \asymp \quad \text{eigen-decomposition}$$

SVD vs. eigen-decomposition

For *asymmetric* matrices:

- Numerical stability

$$\text{SVD} \quad > \quad \text{eigen-decomposition}$$

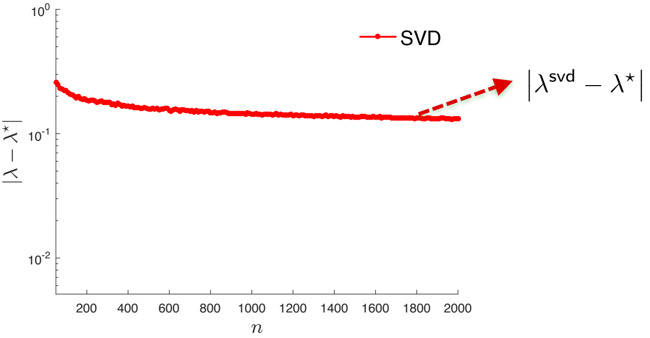
- **(Folklore?)** Statistical accuracy

$$\text{SVD} \quad \asymp \quad \text{eigen-decomposition}$$

Shall we always prefer SVD over eigen-decomposition?

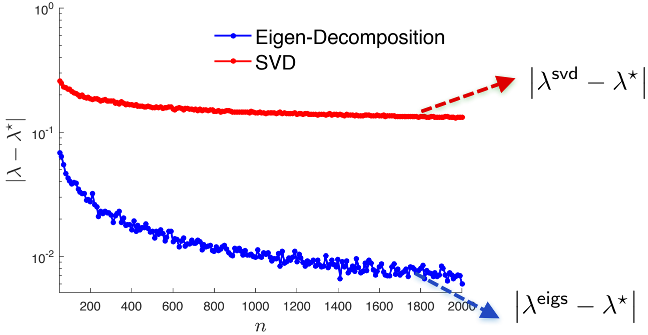
A curious numerical experiment: Gaussian noise

$$M = \underbrace{u^\star u^{\star\top}}_{M^\star} + H; \qquad \{H_{i,j}\}: \text{ i.i.d. } \mathcal{N}(0,\sigma^2), \sigma = \frac{1}{\sqrt{n\log n}}$$



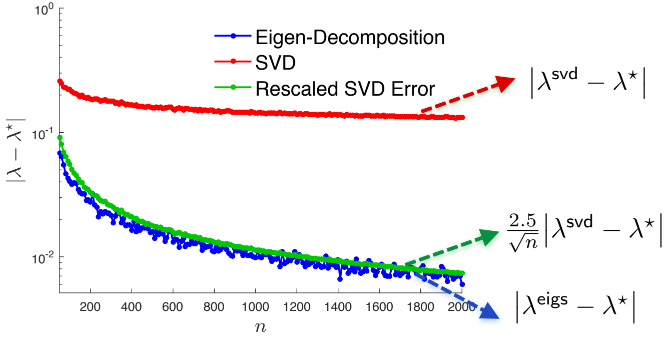
A curious numerical experiment: Gaussian noise

$$\boldsymbol{M} = \underbrace{\boldsymbol{u}^\star \boldsymbol{u}^{\star\top}}_{\boldsymbol{M}^\star} + \boldsymbol{H}; \qquad \{H_{i,j}\}: \text{ i.i.d. } \mathcal{N}(0, \sigma^2), \sigma = \frac{1}{\sqrt{n \log n}}$$



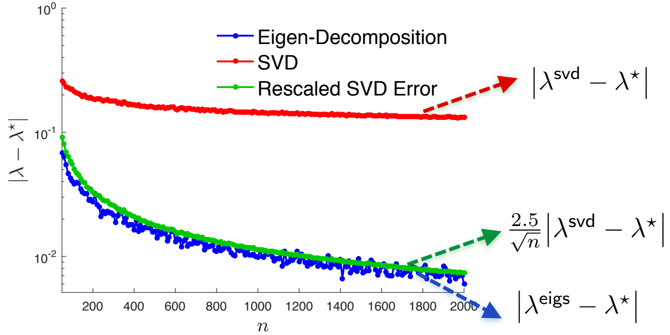
A curious numerical experiment: Gaussian noise

$$\boldsymbol{M} = \underbrace{\boldsymbol{u}^\star \boldsymbol{u}^{\star\top}}_{\boldsymbol{M}^\star} + \boldsymbol{H}; \qquad \{H_{i,j}\}: \text{ i.i.d. } \mathcal{N}(0,\sigma^2), \sigma = \frac{1}{\sqrt{n\log n}}$$



A curious numerical experiment: Gaussian noise

$$\boldsymbol{M} = \underbrace{\boldsymbol{u}^\star \boldsymbol{u}^{\star\top}}_{\boldsymbol{M}^\star} + \boldsymbol{H}; \qquad \{H_{i,j}\} : \text{ i.i.d. } \mathcal{N}(0, \sigma^2), \sigma = \frac{1}{\sqrt{n \log n}}$$



empirically, $\left|\lambda^{\mathsf{eigs}} - \lambda^\star\right| \approx \frac{2.5}{\sqrt{n}}\left|\lambda^{\mathsf{svd}} - \lambda^\star\right|$

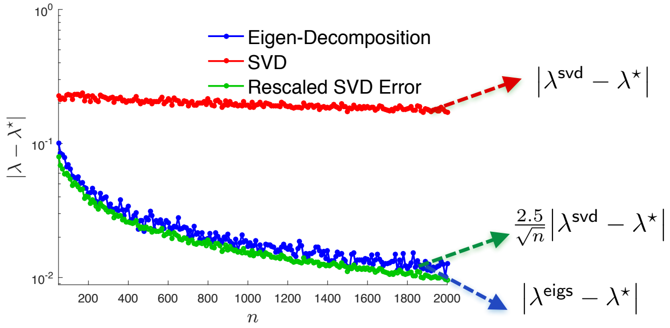
Another numerical experiment: matrix completion

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

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

Another numerical experiment: matrix completion

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



empirically, $\left|\lambda^{\mathsf{eigs}} - \lambda^\star\right| \approx \frac{2.5}{\sqrt{n}}\left|\lambda^{\mathsf{svd}} - \lambda^\star\right|$

Why does eigen-decomposition work so much better than SVD?

Problem setup

$$\boldsymbol{M} = \underbrace{\boldsymbol{u}^{\star}\boldsymbol{u}^{\star\top}}_{\boldsymbol{M}^{\star}} + \boldsymbol{H} \in \mathbb{R}^{n \times n}$$

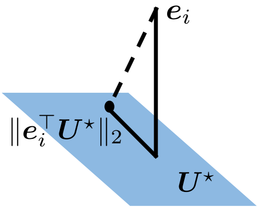
- $\boldsymbol{H}$: noise matrix
    - **independent entries:** $\{H_{i,j}\}$ are independent
    - **zero mean:** $\mathbb{E}[H_{i,j}] = 0$
    - **variance:** $\mathsf{Var}(H_{i,j}) \leq \sigma^2$
    - **magnitudes:** $\mathbb{P}\{|H_{i,j}| \geq B\} \lesssim n^{-12}$

Problem setup

$$\boldsymbol{M} = \underbrace{\boldsymbol{u}^\star \boldsymbol{u}^{\star\top}}_{\boldsymbol{M}^\star} + \boldsymbol{H} \in \mathbb{R}^{n\times n}$$

- $\boldsymbol{H}$: noise matrix
  - **independent entries:** $\{H_{i,j}\}$ are independent
  - **zero mean:** $\mathbb{E}[H_{i,j}] = 0$
  - **variance:** $\mathsf{Var}(H_{i,j}) \le \sigma^2$
  - **magnitudes:** $\mathbb{P}\{|H_{i,j}| \ge B\} \lesssim n^{-12}$
- $\boldsymbol{M}^\star$ obeys incoherence condition

$$\max_{1\le i\le n} \left|\boldsymbol{e}_i^\top \boldsymbol{u}^\star\right| \le \sqrt{\frac{\mu}{n}}$$



Classical linear algebra results

$$|\lambda^{\mathsf{svd}} - \lambda^\star| \le \|\boldsymbol{H}\| \qquad \text{(Weyl)}$$
$$|\lambda^{\mathsf{eigs}} - \lambda^\star| \le \|\boldsymbol{H}\| \qquad \text{(Bauer-Fike)}$$

Classical linear algebra results

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

$$\Downarrow \quad \text{matrix Bernstein inequality}$$

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

Classical linear algebra results

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

$$|\lambda^{\mathsf{eigs}} - \lambda^\star| \leq \|\boldsymbol{H}\| \qquad \text{(Bauer-Fike)}$$

$$\Downarrow \quad \text{matrix Bernstein inequality}$$

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

$$|\lambda^{\mathsf{eigs}} - \lambda^\star| \lesssim \sigma\sqrt{n\log n} + B\log n$$

Classical linear algebra results

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

$\Downarrow$ matrix Bernstein inequality

$$\left|\lambda^{\mathsf{svd}} - \lambda^{\star}\right| \lesssim \sigma\sqrt{n\log n} + B\log n \qquad \text{(reasonably tight if } \|\boldsymbol{H}\| \text{ is large)}$$
$$\left|\lambda^{\mathsf{eigs}} - \lambda^{\star}\right| \lesssim \sigma\sqrt{n\log n} + B\log n \qquad (\textit{can be significantly improved})$$

Main results: eigenvalue perturbation

Theorem 1 (Chen, Cheng, Fan '18)

With high prob., leading eigenvalue λ^{eigs} *of* $\boldsymbol{M}$ *obeys*

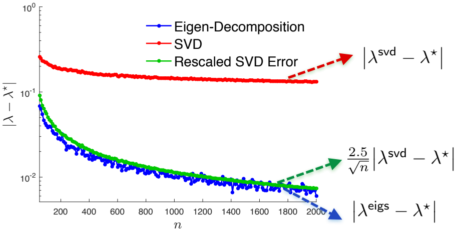
$$|\lambda^{\mathsf{eigs}} - \lambda^{\star}| \lesssim \sqrt{\frac{\mu}{n}}\left(\sigma\sqrt{n\log n} + B\log n\right)$$

Main results: eigenvalue perturbation

Theorem 1 (Chen, Cheng, Fan '18)

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

Main results: entrywise eigenvector perturbation

Theorem 2 (Chen, Cheng, Fan '18)

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

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

Main results: 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

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

Main results: 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

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

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

Main results: 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

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

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

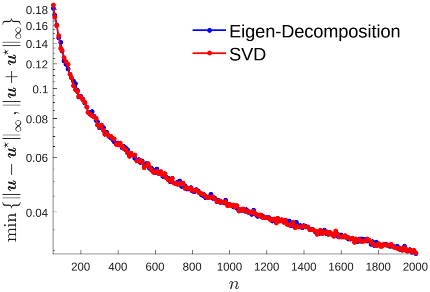
- entrywise eigenvector perturbation is well-controlled

Main results: entrywise eigenvector perturbation

Theorem 2 (Chen, Cheng, Fan '18)

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

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



$\{H_{i,j}\}$: i.i.d. $\mathcal{N}(0,\sigma^2)$; $\sigma^2 = \frac{1}{n\log n}$

Main results: 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*

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

Main results: 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

Intuition: asymmetry reduces bias

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

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

Intuition: asymmetry reduces bias

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

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

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

Intuition: asymmetry reduces bias

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

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

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

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

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

Intuition: asymmetry reduces bias

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

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

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

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

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

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

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

What happens if $M^\star$ is also not symmetric?

- A rank-1 matrix: $\boldsymbol{M}^\star = \lambda^\star \boldsymbol{u}^\star \boldsymbol{v}^{\star\top} \in \mathbb{R}^{n_1 \times n_2}$
- Suppose we observe 2 independent noisy copies

$$\boldsymbol{M}_1 = \boldsymbol{M}^\star + \boldsymbol{H}_1, \qquad \boldsymbol{M}_2 = \boldsymbol{M}^\star + \boldsymbol{H}_2$$

- **Goal:** estimate $\lambda^\star$, $\boldsymbol{u}^\star$ and $\boldsymbol{v}^\star$

Asymmetrization + dilation

Compute leading eigenvalue / eigenvector of

$$\left[\begin{array}{cc} \mathbf{0} & \boldsymbol{M}_1 \\ \boldsymbol{M}_2^{\top} & \mathbf{0} \end{array}\right] = \left[\begin{array}{cc} \mathbf{0} & \boldsymbol{M}^{\star} + \boldsymbol{H}_1 \\ \boldsymbol{M}^{\star\top} + \boldsymbol{H}_2^{\top} & \mathbf{0} \end{array}\right]$$

- Our findings (eigenvalue / eigenvector perturbation) continue to hold for this case!

Rank-r case

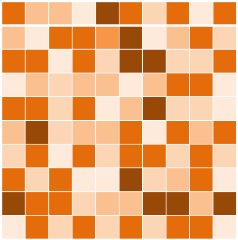
$\boldsymbol{M}^{\star}$: truth

- A rank-r and well-conditioned matrix: $\boldsymbol{M}^{\star} = \sum_{i=1}^{r} \lambda_i^{\star} \boldsymbol{u}_i^{\star} \boldsymbol{u}_i^{\star\top}$
- Observed noisy data: $\boldsymbol{M} = \boldsymbol{M}^{\star} + \boldsymbol{H}$, where $\{H_{i,j}\}$ are independent
- **Goal:** estimate $\lambda^{\star}$

Rank-r case

$M^\star$: truth

$+$



H: noise

- A rank-r and well-conditioned matrix: $\boldsymbol{M}^\star = \sum_{i=1}^{r} \lambda_i^\star \boldsymbol{u}_i^\star \boldsymbol{u}_i^{\star\top}$
- Observed noisy data: $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{H}$, where $\{H_{i,j}\}$ are independent
- **Goal:** estimate $\lambda^\star$

Eigenvalue perturbation: rank-r case

Theorem 4 (Chen, Cheng, Fan '18)

With high prob., ith largest eigenvalue λ_i ($1 \le i \le r$) of $\boldsymbol{M}$ obeys

$$|\lambda_i - \lambda_j^\star| \lesssim \sqrt{\frac{\mu r^2}{n}} \left(\sigma \sqrt{n \log n} + B \log n\right)$$

for some $1 \le j \le r$

Eigenvalue perturbation: rank-r case

Theorem 4 (Chen, Cheng, Fan '18)

With high prob., ith largest eigenvalue λ_i $(1 \leq i \leq r)$ of $\boldsymbol{M}$ obeys

$$|\lambda_i - \lambda_j^\star| \lesssim \sqrt{\frac{\mu r^2}{n}}\left(\sigma\sqrt{n\log n} + B\log n\right)$$

for some $1 \leq j \leq r$

- Eigen-decomposition is $\sqrt{\frac{n}{\mu r^2}}$ times better than SVD!

Eigenvalue perturbation: rank-r case

Theorem 4 (Chen, Cheng, Fan '18)

With high prob., ith largest eigenvalue λ_i ($1 \leq i \leq r$) of $\boldsymbol{M}$ obeys

$$|\lambda_i - \lambda_j^\star| \lesssim \sqrt{\frac{\mu r^2}{n}} \left(\sigma \sqrt{n \log n} + B \log n\right)$$

for some $1 \leq j \leq r$

- Eigen-decomposition is $\sqrt{\frac{n}{\mu r^2}}$ times better than SVD!
- Might be improvable to $\sqrt{\frac{\mu r}{n}} \left(\sigma \sqrt{n \log n} + B \log n\right)$?

Summary for this part

Eigen-decomposition could be much more powerful than SVD when dealing with non-symmetric data matrices

Summary for this part

Eigen-decomposition could be much more powerful than SVD when dealing with non-symmetric data matrices

Future directions:

- Eigenvector perturbation for rank-r case
- Beyond i.i.d. noise

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

Spectral Methods are Optimal for Top-K Ranking

Cong Ma
Princeton ORFE

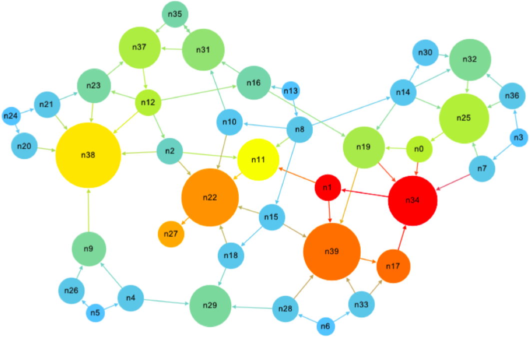
Kaizheng Wang
Princeton ORFE

Jianqing Fan
Princeton ORFE

Ranking

A fundamental problem in a wide range of contexts

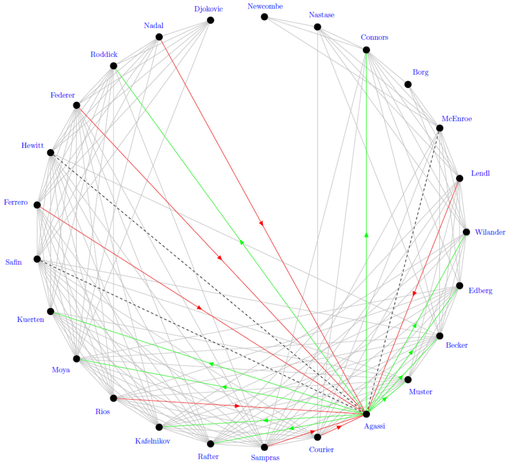
- web search, recommendation systems, admissions, sports competitions, voting, ...



PageRank

figure credit: Dzenan Hamzic

Rank aggregation from pairwise comparisons

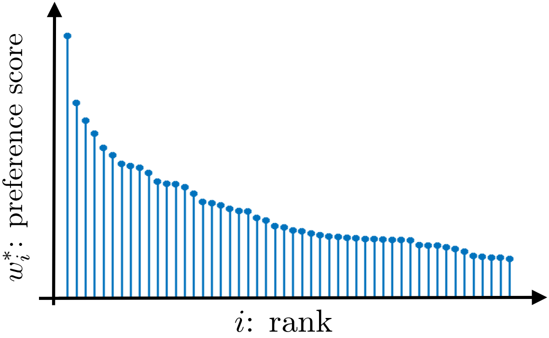


pairwise comparisons for ranking top tennis players

figure credit: Bozóki, Csató, Temesi

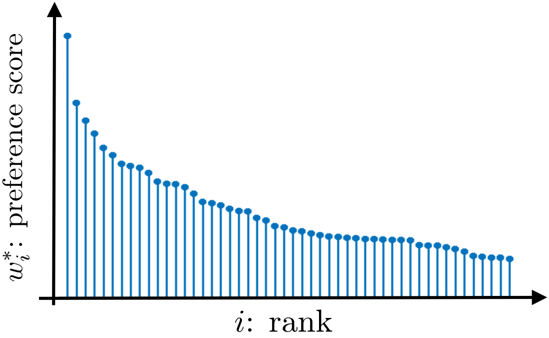
Parametric models

Assign latent score to each of n items $\boldsymbol{w}^* = [w_1^*, \cdots, w_n^*]$



Parametric models

Assign latent score to each of n items $\boldsymbol{w}^* = [w_1^*, \cdots, w_n^*]$



- **This work:** Bradley-Terry-Luce (logistic) model

$$\mathbb{P}\left\{\text{item } j \text{ beats item } i\right\} = \frac{w_j^*}{w_i^* + w_j^*}$$

- *Other models: Thurstone model, low-rank model, ...*

Typical ranking procedures

Estimate latent scores

$\longrightarrow$ rank items based on score estimates

Top-K ranking

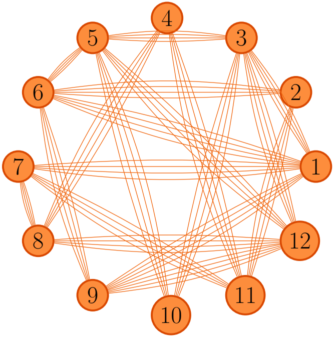
Estimate latent scores

$\longrightarrow$ rank items based on score estimates

Goal: identify the set of top-K items under minimal sample size

Model: random sampling

- Comparison graph: Erdős–Renyi graph $\mathcal{G} \sim \mathcal{G}(n, p)$

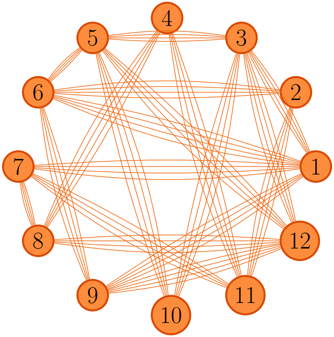


- For each $(i, j) \in \mathcal{G}$, obtain L paired comparisons

$$y_{i,j}^{(l)} \overset{\text{ind.}}{=} \begin{cases} 1, & \text{with prob. } \frac{w_j^*}{w_i^* + w_j^*} \\ 0, & \text{else} \end{cases} \qquad 1 \le l \le L$$

Model: random sampling

- Comparison graph: Erdős–Renyi graph $\mathcal{G} \sim \mathcal{G}(n, p)$



- For each $(i, j) \in \mathcal{G}$, obtain L paired comparisons

$$y_{i,j} = \frac{1}{L} \sum_{l=1}^{L} y_{i,j}^{(l)} \qquad \text{(sufficient statistic)}$$

Prior art

	mean square error for estimating scores	top-K ranking accuracy	
Spectral method	✔	?	Negahban et al. ‘12
MLE	✔	?	Negahban et al. ’12 Hajek et al. ‘14
Spectral MLE	✔	✔	Chen & Suh. ’15

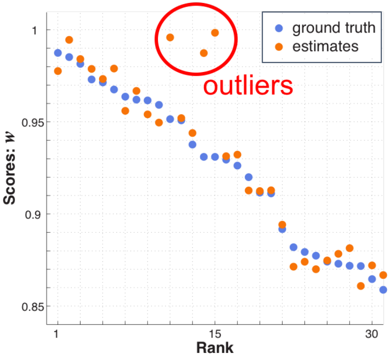
Prior art

"meta metric"

⇧

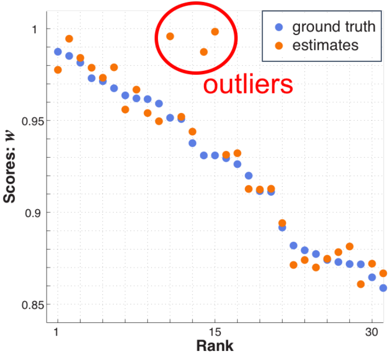
	mean square error for estimating scores	top-K ranking accuracy	
Spectral method	✔	?	Negahban et al. '12
MLE	✔	?	Negahban et al. '12 Hajek et al. '14
Spectral MLE	✔	✔	Chen & Suh. '15

Small ℓ_2 loss $\neq$ high ranking accuracy

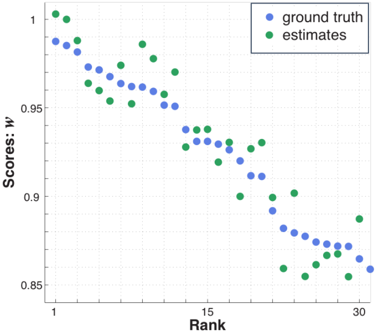


Top 3 : $\{15, 11, 2\}$

Small ℓ_2 loss $\neq$ high ranking accuracy

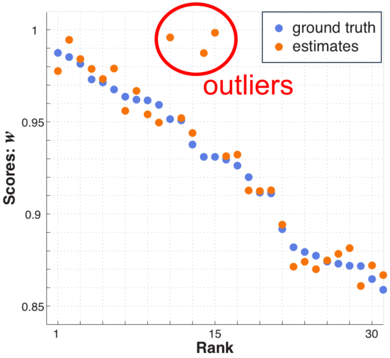


Top 3 : $\{15, 11, 2\}$

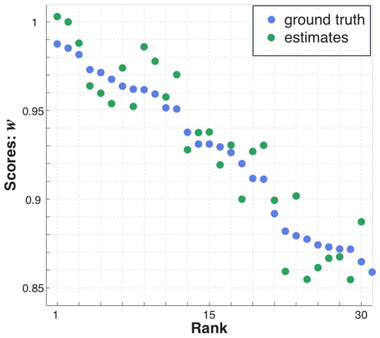


Top 3: $\{1, 2, 3\}$

Small ℓ_2 loss $\neq$ high ranking accuracy



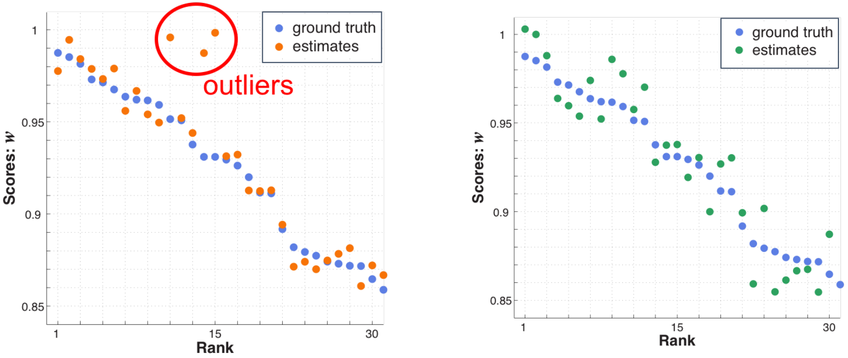
Top 3 : $\{15, 11, 2\}$



Top 3: $\{1, 2, 3\}$

These two estimates have same ℓ_2 loss, but output different rankings

Small ℓ_2 loss $\neq$ high ranking accuracy



Top 3 : $\{15, 11, 2\}$

Top 3: $\{1, 2, 3\}$

These two estimates have same ℓ_2 loss, but output different rankings

Need to control entrywise error!

Optimality?

Is spectral method alone optimal for top-K ranking?

Optimality?

Is spectral method alone optimal for top-K ranking?

Partial answer (Jang et al '16):

spectral method works if comparison graph is sufficiently dense

Optimality?

Is spectral method alone optimal for top-K ranking?

Partial answer (Jang et al '16):

spectral method works if comparison graph is sufficiently dense

This work: affirmative answer + $\underbrace{\textbf{entire regime}}_{\textbf{inc. sparse graphs}}$

Spectral method (Rank Centrality)

Negahban, Oh, Shah '12

- Construct a (highly asymmetric) probability transition matrix $\boldsymbol{P}$, whose off-diagonal entries obey

$$P_{i,j} \;\propto\; \begin{cases} y_{i,j}, & \text{if } (i,j) \in \mathcal{G} \\ 0, & \text{if } (i,j) \notin \mathcal{G} \end{cases}$$

- Return score estimate as leading left eigenvector of $\boldsymbol{P}$

Rationale behind spectral method

In large-sample case, $\boldsymbol{P} \rightarrow \boldsymbol{P}^*$, whose off-diagonal entries obey

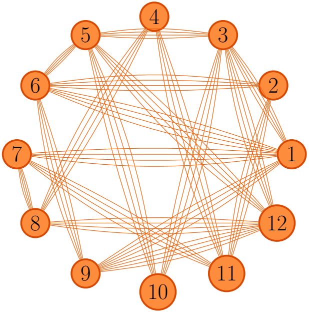
$$P^*_{i,j} \;\propto\; \begin{cases} \frac{w_j^*}{w_i^* + w_j^*}, & \text{if } (i,j) \in \mathcal{G} \\ 0, & \text{if } (i,j) \notin \mathcal{G} \end{cases}$$

- Stationary distribution of $\underbrace{\text{reversible}}_{\text{check detailed balance}}$ $\boldsymbol{P}^*$

$$\boldsymbol{\pi}^* \;\propto\; \underbrace{[w_1^*, w_2^*, \ldots, w_n^*]}_{\text{true score}}$$

Main result

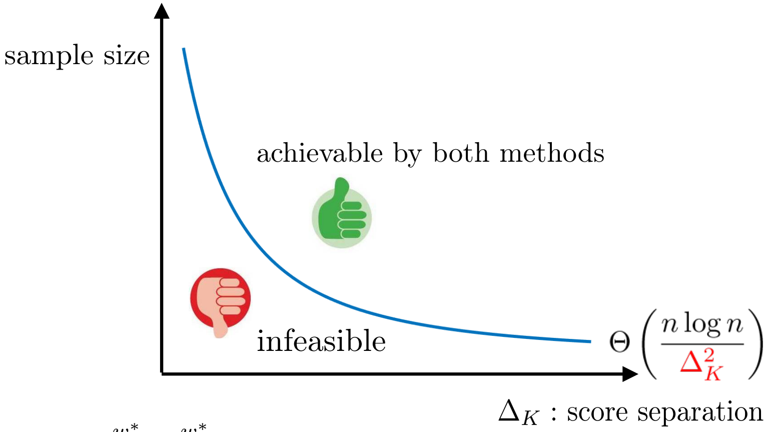
comparison graph $\mathcal{G}(n, p)$; sample size $\asymp pn^2L$



Theorem 5 (Chen, Fan, Ma, Wang '17)

When $p \gtrsim \frac{\log n}{n}$, spectral methods achieve optimal sample complexity for top-K ranking!

Main result



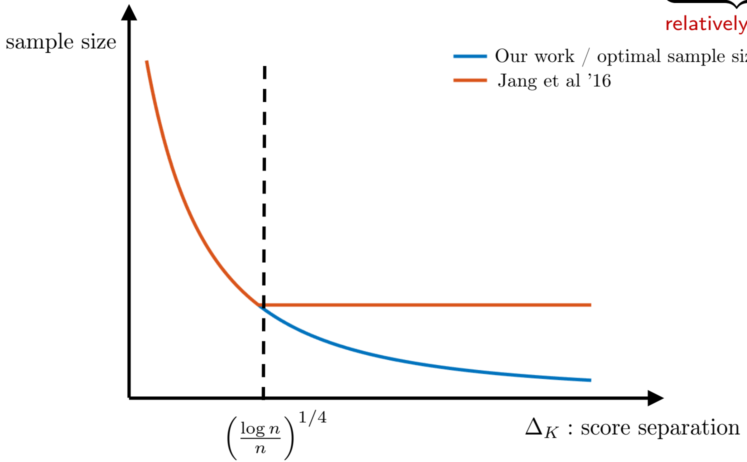
- $\Delta_K := \frac{w^*_{(K)} - w^*_{(K+1)}}{\|\boldsymbol{w}^*\|_\infty}$: score separation

Comparison with Jang et al '16

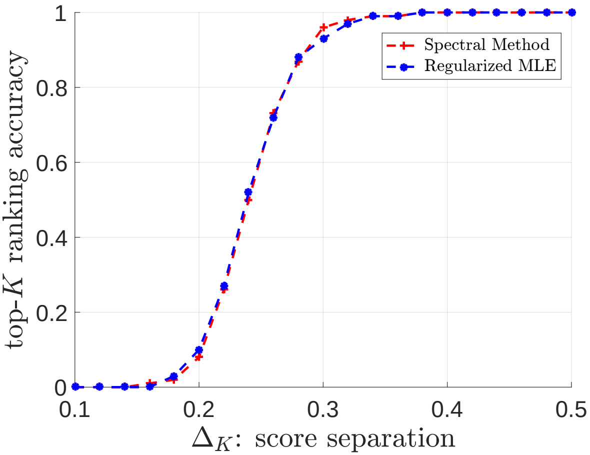
Jang et al '16: spectral method controls entrywise error if $\underbrace{p \gtrsim \sqrt{\frac{\log n}{n}}}_{\text{relatively dense}}$

Comparison with Jang et al '16

Jang et al '16: spectral method controls entrywise error if $\underbrace{p \gtrsim \sqrt{\frac{\log n}{n}}}_{\text{relatively dense}}$

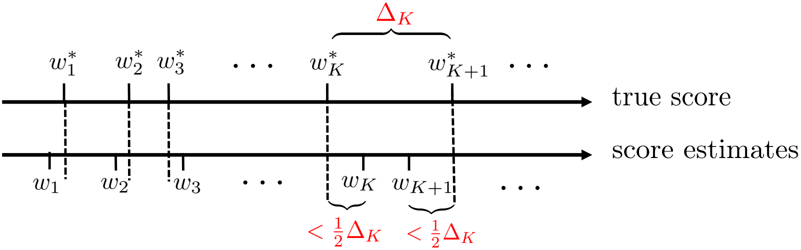


Empirical top-K ranking accuracy



$$n = 200, \; p = 0.25, \; L = 20$$

Optimal control of entrywise error



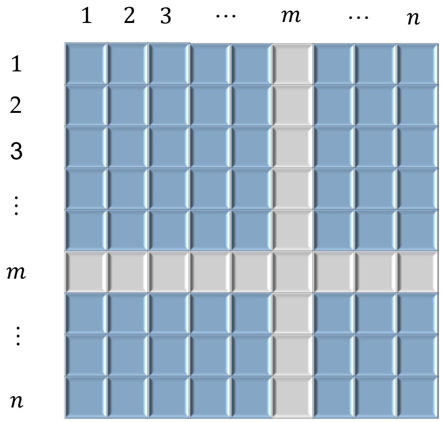
Theorem 6

Suppose $p \gtrsim \frac{\log n}{n}$ *and sample size* $\gtrsim \frac{n \log n}{\Delta_K^2}$*. Then with high prob., the estimates* $\boldsymbol{w}$ *returned by both methods obey (up to global scaling)*

$$\frac{\|\boldsymbol{w} - \boldsymbol{w}^*\|_\infty}{\|\boldsymbol{w}^*\|_\infty} < \frac{1}{2}\Delta_K$$

Key ingredient: leave-one-out analysis

For each $1 \leq m \leq n$, introduce leave-one-out estimate $\boldsymbol{w}^{(m)}$



$$\Longrightarrow \ \boldsymbol{w}^{(m)}$$

$$\boldsymbol{y} = [y_{i,j}]_{1 \leq i,j \leq n}$$

Key ingredient: leave-one-out analysis

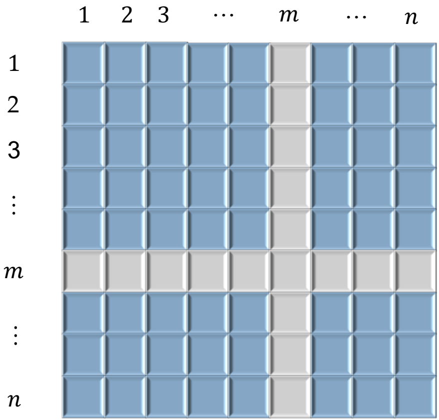
For each $1 \leq m \leq n$, introduce leave-one-out estimate $\boldsymbol{w}^{(m)}$

$$|w_m - w_m^*| \leq \underbrace{\left|w_m^{(m)} - w_m^*\right|}_{\text{Leave-one-out estimation error}} + \underbrace{\|\boldsymbol{w} - \boldsymbol{w}^{(m)}\|_2}_{\text{Leave-one-out perturbation}}$$

$\Downarrow$ statistical independence

$\Downarrow$ stability

Exploit statistical independence



$$\Longrightarrow \ \boldsymbol{w}^{(m)}$$

$$\boldsymbol{y} = [y_{i,j}]_{1 \le i,j \le n}$$

leave-one-out estimate $\boldsymbol{w}^{(m)}$ $\perp\!\!\!\perp$ all data related to mth item

Leave-one-out stability

leave-one-out estimate $\boldsymbol{w}^{(m)} \;\approx\;$ true estimate $\boldsymbol{w}$

Leave-one-out stability

leave-one-out estimate $\boldsymbol{w}^{(m)} \approx$ true estimate $\boldsymbol{w}$

- Spectral method: eigenvector perturbation bound

$$\|\boldsymbol{\pi} - \widehat{\boldsymbol{\pi}}\|_{\boldsymbol{\pi}^*} \lesssim \frac{\|\boldsymbol{\pi}(\boldsymbol{P} - \widehat{\boldsymbol{P}})\|_{\boldsymbol{\pi}^*}}{\text{spectral-gap}}$$

  - new Davis-Kahan bound for $\underbrace{\text{probability transition matrices}}_{\textit{asymmetric}}$

A small sample of related works

- **Parametric models**
  - Ford '57
  - Hunter '04
  - Negahban, Oh, Shah '12
  - Rajkumar, Agarwal '14
  - Hajek, Oh, Xu '14
  - Chen, Suh '15
  - Rajkumar, Agarwal '16
  - Jang, Kim, Suh, Oh '16
  - Suh, Tan, Zhao '17
- **Non-parametric models**
  - Shah, Wainwright '15
  - Shah, Balakrishnan, Guntuboyina, Wainwright '16
  - Chen, Gopi, Mao, Schneider '17
- **Leave-one-out analysis**
  - El Karoui, Bean, Bickel, Lim, Yu '13
  - Zhong, Boumal '17
  - Abbe, Fan, Wang, Zhong '17
  - Ma, Wang, Chi, Chen '17
  - Chen, Chi, Fan, Ma '18
  - Chen, Chi, Fan, Ma, Yan '19
  - Chen, Fan, Ma, Yan '19

Summary for this part

	Optimal sample complexity	Linear-time computational complexity
Spectral method	✔	✔
Regularized MLE	✔	✔

Novel entrywise perturbation analysis for spectral method and convex optimization

Paper: "*Spectral method and regularized MLE are both optimal for top-K ranking*", Y. Chen, J. Fan, C. Ma, K. Wang, *Annals of Statistics*, vol. 47, 2019