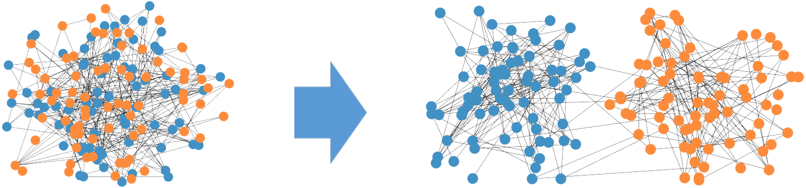
Spectral methods: $\ell_{2,\infty}$ perturbation theory

Cong Ma

University of Chicago, Autumn 2021

Revisit stochastic block model



$\mathcal{G}$

- Community membership vector

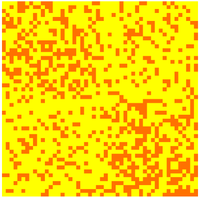
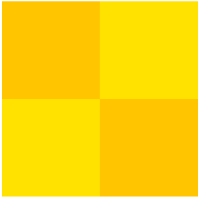
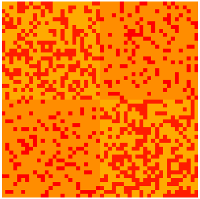
$$x_1^\star = \cdots = x_{n/2}^\star = 1; \; x_{n/2+1}^\star = \cdots = x_n^\star = -1$$

- observe a graph $\mathcal{G}$ (assuming $p > q$)

$$(i,j) \in \mathcal{G} \text{ with prob. } \begin{cases} p, & \text{if } x_i = x_j \\ q, & \text{else} \end{cases}$$

- **Goal:** recover community memberships $\pm \boldsymbol{x}^\star$

Revisit spectral clustering

$$\boldsymbol{A} \quad = \quad \underbrace{\mathbb{E}[\boldsymbol{A}]}_{\text{rank 2}} \quad + \quad \boldsymbol{A} - \mathbb{E}[\boldsymbol{A}]$$

1. computing the leading eigenvector $\boldsymbol{u} = [u_i]_{1 \le i \le n}$ of $\boldsymbol{A} - \frac{p+q}{2}\mathbf{1}\mathbf{1}^\top$
2. rounding: output $x_i = \begin{cases} 1, & \text{if } u_i \ge 0 \\ -1, & \text{if } u_i < 0 \end{cases}$

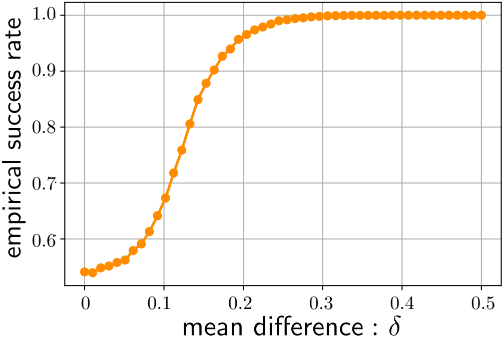
Almost exact recovery

$$\frac{p-q}{\sqrt{p}} \gg \sqrt{\frac{\log n}{n}} \quad \Longrightarrow \quad \text{almost exact recovery}$$

- Almost exact recovery means

$$\min\left\{\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{x_i \neq x_i^\star\}, \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}\{x_i \neq -x_i^\star\}\right\} = o(1)$$

Empirical performance of spectral clustering



ℓ_2 perturbation theory alone cannot explain exact recovery guarantees

— call for fine-grained analysis

Reverse engineering

Spectral clustering uses signs of $\boldsymbol{u}$ to cluster nodes

Reverse engineering

Spectral clustering uses signs of $\boldsymbol{u}$ to cluster nodes

$$\Uparrow$$

It achieves exact recovery iff $u_i u_i^\star > 0$ for all i

Reverse engineering

Spectral clustering uses signs of $\boldsymbol{u}$ to cluster nodes

$$\Uparrow$$

It achieves exact recovery iff $u_i u_i^\star > 0$ for all i

$$\Uparrow$$

A sufficient condition is* $\|\boldsymbol{u} - \boldsymbol{u}^\star\|_\infty < 1/\sqrt{n}$

Reverse engineering

Spectral clustering uses signs of $\boldsymbol{u}$ to cluster nodes

$\Uparrow$

It achieves exact recovery iff $u_i u_i^\star > 0$ for all i

$\Uparrow$

A sufficient condition is* $\|\boldsymbol{u} - \boldsymbol{u}^\star\|_\infty < 1/\sqrt{n}$

$\Uparrow$

Need ℓ_∞ perturbation theory

Outline

- An illustrative example: rank-1 matrix denoising
- General ℓ_∞ perturbation theory: symmetric rank-1 case
- Application: exact recovery in community detection
- General $\ell_{2,\infty}$ perturbation theory: rank-r case
- Application: entrywise error in matrix completion

An illustrative example: rank-1 matrix denoising

Setup and algorithm

- Groundtruth: $\boldsymbol{M}^\star = \lambda^\star \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} \in \mathbb{R}^{n\times n}$, with $\lambda^\star > 0$
- Observation: $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{E}$, where $\boldsymbol{E}$ is symmetric, and its upper triangular part comprises of i.i.d. $\mathcal{N}(0,\sigma^2)$ entries
- Estimate $\boldsymbol{u}^\star$ using $\boldsymbol{u}$, leading eigenvector of $\boldsymbol{M}$
- Goal: characterize entrywise errror

$$\mathsf{dist}_\infty(\boldsymbol{u},\boldsymbol{u}^\star) \coloneqq \min\left\{\|\boldsymbol{u}-\boldsymbol{u}^\star\|_\infty, \|\boldsymbol{u}+\boldsymbol{u}^\star\|_\infty\right\}$$

ℓ_2 guarantees

We start with characterizing noise size

Lemma 6.1

Assume symmetric Gaussian noise model. With high prob., one has

$$\|\boldsymbol{E}\| \leq 5\sigma\sqrt{n}$$

This in conjunction with Davis-Kahan's $\sin\boldsymbol{\Theta}$ theorem leads to:

$$\mathsf{dist}(\boldsymbol{u}, \boldsymbol{u}^\star) \leq \frac{2\|\boldsymbol{E}\|}{\lambda^\star} \leq \frac{10\sigma\sqrt{n}}{\lambda^\star},$$

as long as $\sigma\sqrt{n} \leq \frac{1-1/\sqrt{2}}{5}\lambda^\star$ so that $\|\boldsymbol{E}\| \leq (1 - 1/\sqrt{2})\lambda^\star$

— *implies* $\mathsf{dist}_\infty\big(\boldsymbol{u}, \boldsymbol{u}^\star\big) \leq \mathsf{dist}\big(\boldsymbol{u}, \boldsymbol{u}^\star\big) \lesssim \frac{\sigma\sqrt{n}}{\lambda}$

Incoherence

Definition 6.2

Fix a unit vector $\boldsymbol{u}^\star \in \mathbb{R}^n$. Define its incoherence to be

$$\mu := n\|\boldsymbol{u}^\star\|_\infty^2$$

- Range of possible values of μ: $1 \le \mu \le n$
- Two extremes: $\boldsymbol{u}^\star = \boldsymbol{e}_1$, and $\boldsymbol{u}^\star = (1/\sqrt{n}) \cdot \mathbf{1}_n$
- Small μ indicates energy of eigenvector is spread across different entries
- Consider SBM and random Gaussian vectors

ℓ_∞ guarantees for matrix denoising

Theorem 6.3

Suppose that $\sigma\sqrt{n} \leq c_0\lambda^\star$ *for some sufficiently small constant* $c_0 > 0$. *Then whp., we have*

$$\mathsf{dist}_\infty(\boldsymbol{u}, \boldsymbol{u}^\star) \lesssim \frac{\sigma(\sqrt{\log n} + \sqrt{\mu})}{\lambda^\star}$$

- When $\mu \lesssim \log n$ (i.e., no entries are significantly larger than average), our bound reads

$$\mathsf{dist}_\infty(\boldsymbol{u}, \boldsymbol{u}^\star) \lesssim \frac{\sigma\sqrt{\log n}}{\lambda^\star}$$

- Much tighter than ℓ_2 bound: $\sqrt{n/\log n}$ times smaller

Technical hurdle: dependency

We would like to understand u_l. Since $\boldsymbol{u}$ is eigenvector of $\boldsymbol{M}$, we have

$$\boldsymbol{M}\boldsymbol{u} = \lambda \boldsymbol{u},$$

which yields

$$u_l = \frac{1}{\lambda}[\boldsymbol{M}]_{l,:}\boldsymbol{u} = \frac{1}{\lambda}[\boldsymbol{M}^\star + \boldsymbol{E}]_{l,:}\boldsymbol{u}$$

$\boldsymbol{u}$ is dependent on $\boldsymbol{E}$; analyzing $[\boldsymbol{M}^\star + \boldsymbol{E}]_{l,:}\boldsymbol{u}$ is challenging

—*how to deal with such dependency*

An independent proxy

Recall our focus is

$$[\boldsymbol{M}^\star + \boldsymbol{E}]_{l,:}\boldsymbol{u}$$

Suppose we have a proxy $\boldsymbol{u}^{(l)}$ which is independent of $[\boldsymbol{E}]_{l,:}$, then

$$[\boldsymbol{M}^\star + \boldsymbol{E}]_{l,:}\boldsymbol{u} = [\boldsymbol{M}^\star + \boldsymbol{E}]_{l,:}\boldsymbol{u}^{(l)} + [\boldsymbol{M}^\star + \boldsymbol{E}]_{l,:}\left(\boldsymbol{u} - \boldsymbol{u}^{(l)}\right)$$

- Independence between $\boldsymbol{u}^{(l)}$ and $[\boldsymbol{E}]_{l,:}$
- Proximity between $\boldsymbol{u}^{(l)}$ and $\boldsymbol{u}$

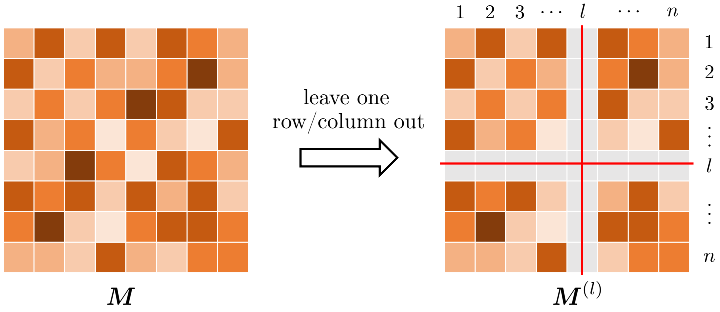
Leave-one-out estimates

For each $1 \leq l \leq n$, construct an auxiliary matrix $\boldsymbol{M}^{(l)}$

$$\boldsymbol{M}^{(l)} := \lambda^\star \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} + \boldsymbol{E}^{(l)},$$

where the noise matrix $\boldsymbol{E}^{(l)}$ is generated according to

$$E_{i,j}^{(l)} := \begin{cases} E_{i,j}, & \text{if } i \neq l \text{ and } j \neq l, \\ 0, & \text{else.} \end{cases}$$



Leave-one-out estimates (cont.)

For each $1 \leq l \leq n$, construct an auxiliary matrix $\boldsymbol{M}^{(l)}$

$$\boldsymbol{M}^{(l)} := \lambda^{\star} \boldsymbol{u}^{\star} \boldsymbol{u}^{\star\top} + \boldsymbol{E}^{(l)},$$

where the noise matrix $\boldsymbol{E}^{(l)}$ is generated according to

$$E_{i,j}^{(l)} := \begin{cases} E_{i,j}, & \text{if } i \neq l \text{ and } j \neq l, \\ 0, & \text{else.} \end{cases}$$

Let $\lambda^{(l)}$ and $\boldsymbol{u}^{(l)}$ denote respectively leading eigenvalue and leading eigenvector of $\boldsymbol{M}^{(l)}$

—$\boldsymbol{u}^{(l)}$ *is independent of* $[\boldsymbol{E}]_{l,:}$

Intuition

- Since $\boldsymbol{u}^{(l)}$ is obtained by dropping only a tiny fraction of data, we expect $\boldsymbol{u}^{(l)}$ to be extremely close to $\boldsymbol{u}$, i.e., $\boldsymbol{u} \approx \pm \boldsymbol{u}^{(l)}$
- By construction,

$$\begin{aligned} u_l^{(l)} &= \frac{1}{\lambda^{(l)}} \boldsymbol{M}_{l,\cdot}^{(l)} \boldsymbol{u}^{(l)} = \frac{1}{\lambda^{(l)}} \boldsymbol{M}_{l,\cdot}^{\star} \boldsymbol{u}^{(l)} = \frac{\lambda^{\star}}{\lambda^{(l)}} u_l^{\star} \boldsymbol{u}^{\star\top} \boldsymbol{u}^{(l)} \\ &\approx \pm u_l^{\star}. \end{aligned}$$

Proof of Theorem 6.3

What we have learned from ℓ_2 analysis

$$
\begin{aligned}
\|\boldsymbol{E}\| &\le 5\sigma\sqrt{n} & \qquad \|\boldsymbol{E}^{(l)}\| \le \|\boldsymbol{E}\| &\le 5\sigma\sqrt{n} \\
\mathsf{dist}(\boldsymbol{u}, \boldsymbol{u}^\star) &\le \frac{10\sigma\sqrt{n}}{\lambda^\star} & \qquad \mathsf{dist}(\boldsymbol{u}^{(l)}, \boldsymbol{u}^\star) &\le \frac{10\sigma\sqrt{n}}{\lambda^\star} \\
|\lambda - \lambda^\star| &\le 5\sigma\sqrt{n} & \qquad |\lambda^{(l)} - \lambda^\star| &\le 5\sigma\sqrt{n} \\
\max_{j: j\ge 2} |\lambda_j(\boldsymbol{M})| &\le 5\sigma\sqrt{n} & \qquad \max_{j: j\ge 2} |\lambda_j(\boldsymbol{M}^{(l)})| &\le 5\sigma\sqrt{n}
\end{aligned}
$$

Addressing ambiguity

Assume WLOG,

$$\|\boldsymbol{u} - \boldsymbol{u}^\star\|_2 = \mathsf{dist}(\boldsymbol{u}, \boldsymbol{u}^\star),$$
$$\big\|\boldsymbol{u}^{(l)} - \boldsymbol{u}^\star\big\|_2 = \mathsf{dist}(\boldsymbol{u}^{(l)}, \boldsymbol{u}^\star), \quad 1 \le l \le n$$

A useful byproduct: if $20\sigma\sqrt{n} < \lambda^\star$, then one necessarily has

$$\big\|\boldsymbol{u} - \boldsymbol{u}^{(l)}\big\|_2 = \mathsf{dist}\big(\boldsymbol{u}, \boldsymbol{u}^{(l)}\big), \qquad 1 \le l \le n$$

—*check this*

Bounding $\|\boldsymbol{u} - \boldsymbol{u}^{(l)}\|_2$

Key: view $\boldsymbol{M}$ as perturbation of $\boldsymbol{M}^{(l)}$; apply "sharper" version of Davis-Kahan

$$\left\|\boldsymbol{u} - \boldsymbol{u}^{(l)}\right\|_2 \le \frac{2\|(\boldsymbol{M} - \boldsymbol{M}^{(l)})\boldsymbol{u}^{(l)}\|_2}{\lambda^{(l)} - \max_{j\ge 2}\left|\lambda_j(\boldsymbol{M}^{(l)})\right|} \le \frac{4\|(\boldsymbol{M} - \boldsymbol{M}^{(l)})\boldsymbol{u}^{(l)}\|_2}{\lambda^\star}$$

as long as

$$\|\boldsymbol{M} - \boldsymbol{M}^{(l)}\| \le (1 - 1/\sqrt{2})\Big(\lambda^{(l)} - \max_{j\ge 2}\left|\lambda_j(\boldsymbol{M}^{(l)})\right|\Big),$$

$$\lambda^{(l)} - \max_{j\ge 2}\left|\lambda_j(\boldsymbol{M}^{(l)})\right| \ge \lambda^\star/2$$

Bounding $\|\big(\boldsymbol{M}-\boldsymbol{M}^{(l)}\big)\boldsymbol{u}^{(l)}\|_2$

By design,

$$(\boldsymbol{M}-\boldsymbol{M}^{(l)})\boldsymbol{u}^{(l)} = \boldsymbol{e}_l \boldsymbol{E}_{l,\cdot}\boldsymbol{u}^{(l)} + u_l^{(l)}(\boldsymbol{E}_{\cdot,l} - E_{l,l}\boldsymbol{e}_l),$$

which together with triangle inequality yields

$$\begin{aligned}
\|\big(\boldsymbol{M}-\boldsymbol{M}^{(l)}\big)\boldsymbol{u}^{(l)}\|_2 &\leq \big|\boldsymbol{E}_{l,\cdot}\boldsymbol{u}^{(l)}\big| + \big\|\boldsymbol{E}_{\cdot,l}\big\|_2 \cdot \big|u_l^{(l)}\big| \\
&\leq 5\sigma\sqrt{\log n} + \big\|\boldsymbol{E}_{\cdot,l}\big\|_2\big(|u_l| + \big\|\boldsymbol{u}-\boldsymbol{u}^{(l)}\big\|_\infty\big) \\
&\leq 5\sigma\sqrt{\log n} + 5\sigma\sqrt{n}\|\boldsymbol{u}\|_\infty + 5\sigma\sqrt{n}\big\|\boldsymbol{u}-\boldsymbol{u}^{(l)}\big\|_2
\end{aligned}$$

Bounding $\|\boldsymbol{u} - \boldsymbol{u}^{(l)}\|_2$ (cont.)

Combining previous bounds, we arrive at

$$\begin{aligned}\left\|\boldsymbol{u} - \boldsymbol{u}^{(l)}\right\|_2 &\leq \frac{20\sigma\sqrt{\log n} + 20\sigma\sqrt{n}\|\boldsymbol{u}\|_\infty + 20\sigma\sqrt{n}\left\|\boldsymbol{u} - \boldsymbol{u}^{(l)}\right\|_2}{\lambda^\star} \\ &\leq \frac{20\sigma\sqrt{\log n} + 20\sigma\sqrt{n}\|\boldsymbol{u}\|_\infty}{\lambda^\star} + \frac{1}{2}\left\|\boldsymbol{u} - \boldsymbol{u}^{(l)}\right\|_2,\end{aligned}$$

provided that $40\sigma\sqrt{n} \leq \lambda^\star$
Rearranging terms and taking the union bound, we demonstrate that whp.,

$$\left\|\boldsymbol{u} - \boldsymbol{u}^{(l)}\right\|_2 \leq \frac{40\sigma\sqrt{\log n} + 40\sigma\sqrt{n}\|\boldsymbol{u}\|_\infty}{\lambda^\star} \qquad 1 \leq l \leq n$$

Analyzing leave-one-out iterates

Recall that

$$u_l^{(l)} = \frac{1}{\lambda^{(l)}} \boldsymbol{M}_{l,\cdot}^{(l)} \boldsymbol{u}^{(l)} = \frac{1}{\lambda^{(l)}} \boldsymbol{M}_{l,\cdot}^{\star} \boldsymbol{u}^{(l)} = \frac{\lambda^\star}{\lambda^{(l)}} u_l^\star \boldsymbol{u}^{\star\top} \boldsymbol{u}^{(l)}$$

This implies

$$\begin{aligned} u_l^{(l)} - u_l^\star &= u_l^\star \Big(\frac{\lambda^\star}{\lambda^{(l)}} \boldsymbol{u}^{\star\top} \boldsymbol{u}^{(l)} - \boldsymbol{u}^{\star\top} \boldsymbol{u}^\star \Big) \\ &= u_l^\star \Big(\frac{\lambda^\star - \lambda^{(l)}}{\lambda^{(l)}} \boldsymbol{u}^{\star\top} \boldsymbol{u}^{(l)} \Big) + u_l^\star \boldsymbol{u}^{\star\top} \big(\boldsymbol{u}^{(l)} - \boldsymbol{u}^\star \big) \end{aligned}$$

Analyzing leave-one-out iterates (cont.)

Triangle inequality gives

$$
\begin{aligned}
\left|u_l^{(l)} - u_l^\star\right| &\leq \left|u_l^\star\right| \cdot \frac{\left|\lambda^\star - \lambda^{(l)}\right|}{\left|\lambda^{(l)}\right|} \cdot \|\boldsymbol{u}^\star\|_2 \cdot \|\boldsymbol{u}^{(l)}\|_2 \\
&\quad + \left|u_l^\star\right| \cdot \|\boldsymbol{u}^\star\|_2 \cdot \left\|\boldsymbol{u}^{(l)} - \boldsymbol{u}^\star\right\|_2 \\
&\leq \left|u_l^\star\right| \cdot \frac{10\sigma\sqrt{n}}{\lambda^\star} + \left|u_l^\star\right| \cdot \frac{10\sigma\sqrt{n}}{\lambda^\star} \\
&\leq \frac{20\sigma\sqrt{n}}{\lambda^\star} \|\boldsymbol{u}^\star\|_\infty
\end{aligned}
$$

Putting pieces together

Now we come to conclude that

$$\left\|\boldsymbol{u}-\boldsymbol{u}^{\star}\right\|_{\infty}=\max_{l}\left|u_{l}-u_{l}^{\star}\right| \leq \max_{l}\left\{\left|u_{l}^{(l)}-u_{l}^{\star}\right|+\left\|\boldsymbol{u}-\boldsymbol{u}^{(l)}\right\|_{2}\right\}$$

$$\leq \frac{20\sigma\sqrt{n}}{\lambda^{\star}}\left\|\boldsymbol{u}^{\star}\right\|_{\infty}+\frac{40\sigma\sqrt{\log n}+40\sigma\sqrt{n}\|\boldsymbol{u}\|_{\infty}}{\lambda^{\star}}$$

One more triangle inequality gives

$$\left\|\boldsymbol{u}-\boldsymbol{u}^{\star}\right\|_{\infty} \leq \frac{40\sigma\sqrt{\log n}+60\sigma\sqrt{n}\left\|\boldsymbol{u}^{\star}\right\|_{\infty}}{\lambda^{\star}}+\frac{1}{2}\left\|\boldsymbol{u}-\boldsymbol{u}^{\star}\right\|_{\infty},$$

provided that $80\sigma\sqrt{n} \leq \lambda^{\star}$. Rearranging terms yields

$$\left\|\boldsymbol{u}-\boldsymbol{u}^{\star}\right\|_{\infty} \leq \frac{80\sigma\sqrt{\log n}+120\sigma\sqrt{n}\left\|\boldsymbol{u}^{\star}\right\|_{\infty}}{\lambda^{\star}}=\frac{80\sigma\sqrt{\log n}+120\sigma\sqrt{\mu}}{\lambda^{\star}},$$

where the last identity results from the definition of μ

General ℓ_∞ perturbation theory

—symmetric rank-1 case

Setup and notation

Groundtruth: consider a rank-1 psd matrix $\boldsymbol{M}^\star = \lambda^\star \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} \in \mathbb{R}^{n\times n}$

Incoherence:

$$\mu := n\|\boldsymbol{u}^\star\|_\infty^2 \qquad (1 \le \mu \le n)$$

Observations:

$$\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{E} \in \mathbb{R}^{n\times n}$$

with $\boldsymbol{E}$ a symmetric noise matrix

Spectral method: return $\boldsymbol{u}$ leading eigenvector of $\boldsymbol{M}$

Noise assumptions

The entries in the lower triangular part of $\boldsymbol{E} = [E_{i,j}]_{1 \le i,j \le n}$ are independently generated obeying

$$\mathbb{E}[E_{i,j}] = 0, \quad \mathbb{E}[E_{i,j}^2] \le \sigma^2, \quad |E_{i,j}| \le B, \quad \text{for all } i \ge j$$

Further, assume that

$$c_{\mathsf{b}} := \frac{B}{\sigma\sqrt{n/(\mu \log n)}} = O(1)$$

ℓ_∞ perturbation theory

Theorem 6.4

With high prob, there exists $z \in \{1, -1\}$ such that

$$\left\|z\boldsymbol{u} - \boldsymbol{u}^\star\right\|_\infty \lesssim \frac{\sigma\sqrt{\mu} + \sigma\sqrt{\log n}}{\lambda^\star}, \tag{6.3a}$$

$$\left\|z\boldsymbol{u} - \frac{1}{\lambda^\star}\boldsymbol{M}\boldsymbol{u}^\star\right\|_\infty \lesssim \frac{\sigma\sqrt{\mu}}{\lambda^\star} + \frac{\sigma^2\sqrt{n\log n} + \sigma B\sqrt{\mu\log^3 n}}{(\lambda^\star)^2} \tag{6.3b}$$

provided that $\sigma\sqrt{n\log n} \leq c_\sigma\lambda^\star$ for some sufficiently small constant $c_\sigma > 0$.

- Delocalization of error

First-order expansion

Chain of approximation

$$\boldsymbol{u} = \frac{\boldsymbol{M}\boldsymbol{u}}{\lambda} \approx \frac{\boldsymbol{M}\boldsymbol{u}^\star}{\lambda^\star} \approx \frac{\boldsymbol{M}^\star\boldsymbol{u}^\star}{\lambda^\star} = \boldsymbol{u}^\star$$

- first approximation is much tighter than the second one
- important in certain applications such as SBM

Application: exact recovery in community detection

Exact recovery using spectral method

We consider the case when (why?)

$$p = \frac{\alpha \log n}{n}, \quad \text{and} \quad q = \frac{\beta \log n}{n}$$

Theorem 6.5

Fix any constant $\varepsilon > 0$. Suppose $\alpha > \beta > 0$ are sufficiently large, and*

$$\left(\sqrt{\alpha} - \sqrt{\beta}\right)^2 \geq 2\left(1 + \varepsilon\right).$$

With probability $1 - o(1)$, spectral method achieves exact recovery.

Optimality of spectral method

It turns out that when

$$\left(\sqrt{\alpha} - \sqrt{\beta}\right)^2 \leq 2\left(1 - \varepsilon\right),$$

no method whatsoever can achieve exact recovery

—*what's special about* $\left(\sqrt{\alpha} - \sqrt{\beta}\right)^2$ *or* $\left(\sqrt{p} - \sqrt{q}\right)^2$*?*

Squared Hellinger distance

Definition 6.6

Consider two distributions P and Q over a finite alphabet $\mathcal{Y}$. The squared Hellinger distance $\mathsf{H}^2(P \,\|\, Q)$ between P and Q is defined as

$$\mathsf{H}^2(P \,\|\, Q) := \frac{1}{2} \sum_{y \in \mathcal{Y}} \left(\sqrt{P(y)} - \sqrt{Q(y)} \right)^2. \tag{6.4}$$

Consider squared Hellinger distance between $\mathsf{Bern}(p)$ and $\mathsf{Bern}(q)$:

$$\begin{aligned} \mathsf{H}^2(\mathsf{Bern}(p), \mathsf{Bern}(q)) &:= \frac{1}{2}(\sqrt{p} - \sqrt{q})^2 + \frac{1}{2}(\sqrt{1-p} - \sqrt{1-q})^2 \\ &= (1 + o(1)) \frac{1}{2}(\sqrt{p} - \sqrt{q})^2, \end{aligned}$$

when $p = o(1)$ and $q = o(1)$

Optimality of spectral method (cont.)

The phase transition phenomenon can then be described as

$$\text{spectral method works} \quad \text{if } \mathsf{H}^2\big(\mathsf{Bern}(p), \mathsf{Bern}(q)\big) \geq (1+\varepsilon)\frac{\log n}{n}$$

$$\text{no algorithm works} \quad \text{if } \mathsf{H}^2\big(\mathsf{Bern}(p), \mathsf{Bern}(q)\big) \leq (1-\varepsilon)\frac{\log n}{n}$$

for an arbitrary small constant $\varepsilon > 0$

Fine-grained analysis of spectral clustering

Consider "ground-truth" matrix

$$\boldsymbol{M}^\star := \mathbb{E}[\boldsymbol{A}] - \frac{p+q}{2}\mathbf{1}\mathbf{1}^\top = \frac{p-q}{2}\left[\begin{array}{c} \mathbf{1} \\ -\mathbf{1} \end{array}\right]\left[\begin{array}{cc} \mathbf{1}^\top & -\mathbf{1}^\top \end{array}\right],$$

which obeys

$$\lambda_1(\boldsymbol{M}^\star) := \frac{(p-q)n}{2}, \quad \text{and} \quad \boldsymbol{u}^\star := \frac{1}{\sqrt{n}}\left[\begin{array}{c} \mathbf{1}_{n/2} \\ -\mathbf{1}_{n/2} \end{array}\right].$$

These imply

$$\begin{aligned} &\lambda^\star = \tfrac{n(p-q)}{2}, && \mu = 1, \\ &B = 1, && \sigma^2 \le \max\{p,q\} = p \end{aligned}$$

Invoke ℓ_∞ perturbation theory

ℓ_∞ perturbation bound (6.3b) yields

$$\begin{aligned}\left\|z\lambda^\star \boldsymbol{u} - \boldsymbol{M}\boldsymbol{u}^\star\right\|_\infty &\lesssim \sigma + \frac{\sigma^2\sqrt{n\log n}}{\lambda^\star} + \frac{\sigma B\ \log^{3/2} n}{\lambda^\star} \\ &\leq C\Big(\sqrt{p} + \frac{p\sqrt{\log n}}{\sqrt{n}(p-q)} + \frac{\sqrt{p}\log^{3/2} n}{n(p-q)}\Big) =: \Delta\end{aligned}$$

for some constant $C > 0$

it boils down to characterizing the entrywise behavior of $\boldsymbol{M}\boldsymbol{u}^\star$

Bounding entries in $\boldsymbol{M}\boldsymbol{u}^\star$

Lemma 6.7

Suppose that

$$\left(\sqrt{\alpha} - \sqrt{\beta}\right)^2 \geq 2\,(1+\varepsilon)$$

for some quantity $\varepsilon > 0$. *Then with probability exceeding* $1 - o(1)$, *one has*

$$\boldsymbol{M}_{l,\cdot}\boldsymbol{u}^\star \geq \frac{\eta \log n}{\sqrt{n}} \;\; \textit{for all } l \leq \frac{n}{2} \textit{ and } \boldsymbol{M}_{l,\cdot}\boldsymbol{u}^\star \leq -\frac{\eta \log n}{\sqrt{n}} \;\; \textit{for all } l > \frac{n}{2},$$

where $\eta > 0$ *obeys* $(\sqrt{\alpha} - \sqrt{\beta})^2 - \eta \log(\alpha/\beta) > 2$.

Key message: entries in $\boldsymbol{M}\boldsymbol{u}^\star$ are bounded away from 0 with correct sign

Completing the picture

On one hand

$$\boldsymbol{M}_{l,\cdot}\boldsymbol{u}^\star \geq \frac{\eta \log n}{\sqrt{n}} \text{ for all } l \leq \frac{n}{2} \text{ and } \boldsymbol{M}_{l,\cdot}\boldsymbol{u}^\star \leq -\frac{\eta \log n}{\sqrt{n}} \text{ for all } l > \frac{n}{2}$$

On the other hand

$$\left\| z\lambda^\star \boldsymbol{u} - \boldsymbol{M}\boldsymbol{u}^\star \right\|_\infty \leq \Delta$$

In sum, if one can show

$$\frac{\eta \log n}{\sqrt{n}} > \Delta \tag{6.5}$$

then it follows that

$$z u_l u_l^\star > 0 \quad \text{for all } 1 \leq l \leq n \quad \Longrightarrow \quad \text{exact recovery}$$

Proof of relation (6.5)

Our goal is to show

$$\frac{\eta \log n}{\sqrt{n}} \geq C\Big(\sqrt{p} + \frac{p\sqrt{\log n}}{\sqrt{n}(p-q)} + \frac{\sqrt{p}\log^{3/2} n}{n(p-q)}\Big)$$

- 1st term: $\sqrt{p} \asymp \sqrt{\frac{\log n}{n}} \ll \frac{\eta \log n}{\sqrt{n}}$
- 2nd term: $\frac{p\sqrt{\log n}}{\sqrt{n}(p-q)} \asymp \sqrt{\frac{\log n}{n}} \ll \frac{\eta \log n}{\sqrt{n}}$
- 3rd term: divide discussion into two cases $\alpha/\beta \leq 2$, and $\alpha/\beta \geq 2$

Compare two sets of Bernoullis

Lemma 6.8

Suppose $\alpha > \beta$, $\{W_i\}_{1\leq i\leq n/2}$ *are i.i.d.* $\mathrm{Bern}(\frac{\alpha\log n}{n})$, *and* $\{Z_i\}_{1\leq i\leq n/2}$ *are i.i.d.* $\mathrm{Bern}(\frac{\beta\log n}{n})$, *which are independent of* W_i. *For any* $t > 0$, *one has*

$$\mathbb{P}\left(\sum_{i=1}^{n/2} W_i - \sum_{i=1}^{n/2} Z_i \leq t\log n\right) \leq n^{-(\sqrt{a}-\sqrt{b})^2/2+t\log(a/b)/2}.$$

Proof of Lemma 6.7

Note that $\boldsymbol{M}\boldsymbol{u}^\star = (\boldsymbol{A} - \frac{p+q}{2}\mathbf{1}\mathbf{1}^\top)\boldsymbol{u}^\star = \boldsymbol{A}\boldsymbol{u}^\star$. Hence

$$\boldsymbol{M}_{1,:}\boldsymbol{u}^\star = \boldsymbol{A}_{1,:}\boldsymbol{u}^\star = \frac{1}{\sqrt{n}}\left(\sum_{j=1}^{n/2} A_{1,j} - \sum_{j=n/2+1}^{n} A_{1,j}\right)$$

Apply Lemma 6.8 to obtain with probability at least $1 - n^{-(\sqrt{a}-\sqrt{b})^2/2+\eta\log(a/b)/2} = 1 - o(n^{-1})$

$$\boldsymbol{M}_{1,:}\boldsymbol{u}^\star \geq \frac{\eta \log n}{\sqrt{n}}$$

Invoke union bound to complete proof

Proof of Lemma 6.8

We apply the Laplace transform method: for any $\lambda < 0$

$$\begin{aligned}
&\mathbb{P}\left(\sum_{i=1}^{n/2} W_i - \sum_{i=1}^{n/2} Z_i \le t \log n\right) \\
&\quad = \mathbb{P}\left(\exp\left(\lambda\left(\sum_{i=1}^{n/2} W_i - \sum_{i=1}^{n/2} Z_i\right)\right) \ge \exp\left(\lambda t \log n\right)\right) \\
&\quad \le \frac{\mathbb{E}\left[\exp\left(\lambda\left(\sum_{i=1}^{n/2} W_i - \sum_{i=1}^{n/2} Z_i\right)\right)\right]}{\exp\left(\lambda t \log n\right)}
\end{aligned}$$

By independence, one has

$$\mathbb{E}\left[\exp\left(\lambda\left(\sum_{i=1}^{n/2} W_i - \sum_{i=1}^{n/2} Z_i\right)\right)\right] = \prod_{i=1}^{n/2} \mathbb{E}\left[\exp\left(\lambda W_i\right)\right] \mathbb{E}\left[\exp\left(-\lambda Z_i\right)\right]$$

Proof of Lemma 6.8 (cont.)

By definition and using $1 + x \le e^x$, one has

$$\begin{aligned}\mathbb{E}\left[\exp\left(\lambda W_i\right)\right] &= \frac{\alpha \log n}{n} \exp\left(\lambda\right) + \left(1 - \frac{\alpha \log n}{n}\right) \\ &\le \exp\left(\frac{\alpha \log n}{n} \exp\left(\lambda\right) - \frac{\alpha \log n}{n}\right)\end{aligned}$$

Similarly for Z_i, one has

$$\mathbb{E}\left[\exp\left(-\lambda W_i\right)\right] \le \exp\left(\frac{\beta \log n}{n} \exp\left(-\lambda\right) - \frac{\beta \log n}{n}\right)$$

Combine these two to see that

$$\begin{aligned}&\mathbb{E}\left[\exp\left(\lambda W_i\right)\right] \mathbb{E}\left[\exp\left(-\lambda Z_i\right)\right] \\ &\quad \le \exp\left(\frac{\log n}{n}\left(\alpha \exp\left(\lambda\right) + \beta \exp\left(-\lambda\right) - \alpha - \beta\right)\right)\end{aligned}$$

Proof of Lemma 6.8 (cont.)

Combine previous two pages to see

$$\log \mathbb{P}\left(\sum_{i=1}^{n/2} W_i - \sum_{i=1}^{n/2} Z_i \le t \log n\right)$$
$$\le -\lambda t \log n + \frac{n}{2}\frac{\log n}{n}\left(\alpha \exp\left(\lambda\right) + \beta \exp\left(-\lambda\right) - \alpha - \beta\right)$$

Set $\lambda = -\log\left(\alpha/\beta\right)/2$ to obtain

$$\alpha \exp\left(\lambda\right) + \beta \exp\left(-\lambda\right) - \alpha - \beta = \alpha\sqrt{\frac{\beta}{\alpha}} + \beta\sqrt{\frac{\alpha}{\beta}} - \alpha - \beta = -\left(\sqrt{\alpha} - \sqrt{\beta}\right)^2$$

and proof is finished

General $\ell_{2,\infty}$ perturbation theory

—rank-r case

Setup and notation

Groundtruth: consider a rank-r matrix $\boldsymbol{M}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{V}^{\star\top} \in \mathbb{R}^{n_1 \times n_2}$, with singular values $\sigma_1^\star \geq \sigma_2^\star \geq \cdots \geq \sigma_r^\star > 0$ (assume $n_1 \leq n_2$)

Two convenient notation:

$$\kappa := \frac{\sigma_1^\star}{\sigma_r^\star}, \qquad n := n_1 + n_2$$

Observations:

$$\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{E} \in \mathbb{R}^{n_1 \times n_2}$$

with $\boldsymbol{E}$ a noise matrix

Spectral method: return $\boldsymbol{U}, \boldsymbol{V}$ where $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top + \boldsymbol{U}_\perp \boldsymbol{\Sigma}_\perp \boldsymbol{V}_\perp^\top$

Noise assumptions

The entries in $\boldsymbol{E} = [E_{i,j}]_{1\leq i\leq n_1, 1\leq j\leq n_2}$ are independently generated obeying

$$\mathbb{E}[E_{i,j}] = 0, \quad \mathbb{E}[E_{i,j}^2] \leq \sigma^2, \quad |E_{i,j}| \leq B, \quad \text{for all } i,j$$

Further, assume that

$$c_{\mathsf{b}} := \frac{B}{\sigma\sqrt{n_1/(\mu \log n)}} = O(1)$$

$\ell_{2,\infty}$ distance between U and $U^\star$

Need to take into account rotation ambiguity

—*which rotation matrix to use?*

Definition 6.9

For any square matrix Z with SVD $Z = U_Z \Sigma_Z V_Z^\top$, define

$$\mathsf{sgn}(Z) := U_Z V_Z^\top \tag{6.6}$$

to be the matrix sign function of Z.

Use $\mathsf{sgn}(U^\top U^\star)$—solution to procrustes problem, which yields

$$\|U \mathsf{sgn}(U^\top U^\star) - U^\star\|_{2,\infty}$$

Incoherence of subspace

Definition 6.10

Fix an orthonormal matrix $\boldsymbol{U}^\star \in \mathbb{R}^{n\times r}$. Define its incoherence to be

$$\mu(\boldsymbol{U}^\star) := \frac{n\|\boldsymbol{U}^\star\|_{2,\infty}^2}{r}$$

—*recover incoherence of eigenvector when* $r = 1$

- For $\boldsymbol{M}^\star = \boldsymbol{U}^\star\boldsymbol{\Sigma}^\star\boldsymbol{V}^{\star\top}$, define $\mu(\boldsymbol{M}^\star) := \max\{\mu(\boldsymbol{U}^\star), \mu(\boldsymbol{V}^\star)\}$

$\ell_{2,\infty}$ perturbation theory

Define $\boldsymbol{H_U} := \boldsymbol{U}^\top \boldsymbol{U}^\star$ and $\boldsymbol{H_V} := \boldsymbol{V}^\top \boldsymbol{V}^\star$

Theorem 6.11

With probability at least $1 - O(n^{-5})$, one has

$$\max\left\{\|\boldsymbol{U}\mathsf{sgn}(\boldsymbol{H_U}) - \boldsymbol{U}^\star\|_{2,\infty}, \|\boldsymbol{V}\mathsf{sgn}(\boldsymbol{H_V}) - \boldsymbol{V}^\star\|_{2,\infty}\right\} \lesssim \frac{\sigma\sqrt{r}\left(\kappa\sqrt{\frac{n_2}{n_1}\mu} + \sqrt{\log n}\right)}{\sigma_r^\star},$$

provided that $\sigma\sqrt{n\log n} \le c_1\sigma_r^\star$ for some sufficiently small constant $c_1 > 0$.

Entrywise reconstruction error

Recall $\boldsymbol{M} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top + \boldsymbol{U}_\perp \boldsymbol{\Sigma}_\perp \boldsymbol{V}_\perp^\top$

Corollary 6.12

In addition, if $\sigma\kappa\sqrt{n \log n} \leq c_2 \sigma_r^\star$ for some small enough constant $c_2 > 0$, then the following holds with probability at least $1 - O(n^{-5})$:

$$\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top - \boldsymbol{M}^\star\|_\infty \lesssim \sigma\kappa^2 \mu r \sqrt{\frac{(n_2/n_1)\log n}{n_1}}$$

De-localization of estimation error

For simplicity, let us consider the case where $\mu, \kappa, n_2/n_1 = O(1)$. Davis-Kahan theorem results in the following ℓ_2 estimation guarantees

$$\mathsf{dist}_{\mathrm{F}}(\boldsymbol{U}, \boldsymbol{U}^\star) \leq \sqrt{r}\, \mathsf{dist}(\boldsymbol{U}, \boldsymbol{U}^\star) \lesssim \frac{\sigma\sqrt{nr}}{\sigma_r^\star}$$

In comparison, the $\ell_{2,\infty}$ bound derived in Theorem 6.11 simplifies to

$$\min_{\boldsymbol{R} \in \mathcal{O}^{r\times r}} \left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\right\|_{2,\infty} \leq \left\|\boldsymbol{U}\mathsf{sgn}(\boldsymbol{H}) - \boldsymbol{U}^\star\right\|_{2,\infty} \lesssim \frac{\sigma\sqrt{r\log n}}{\sigma_r^\star}$$

De-localization of estimation error (cont.)

For the matrix reconstruction error, one has

$$\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top - \boldsymbol{M}^\star\| \le 2\|\boldsymbol{M} - \boldsymbol{M}^\star\| \lesssim \sigma\sqrt{n},$$

which implies $\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top - \boldsymbol{M}^\star\|_{\mathsf{F}} \lesssim \sigma\sqrt{nr}$

In comparison, one has

$$\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top - \boldsymbol{M}^\star\|_\infty \lesssim \sigma r\sqrt{\frac{\log n}{n}}$$

Application: entrywise error in matrix completion

Low-rank matrix completion

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

figure credit: Candès

- consider a low-rank matrix $\boldsymbol{M}^\star = \boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{V}^{\star\top}$
- each entry $M^\star_{i,j}$ is observed independently with prob. p
- **intermediate goal:** estimate $\boldsymbol{U}^\star, \boldsymbol{V}^\star$

Spectral method for matrix completion

1. identify the key matrix $\boldsymbol{M}^\star$
2. construct surrogate matrix $\boldsymbol{M} \in \mathbb{R}^{n\times n}$ as

$$M_{i,j} = \begin{cases} \frac{1}{p} M^\star_{i,j}, & \text{if } M^\star_{i,j} \text{ is observed} \\ 0, & \text{else} \end{cases}$$

   - **rationale for rescaling:** ensures $\mathbb{E}[\boldsymbol{M}] = \boldsymbol{M}^\star$

3. compute the rank-r SVD $\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top$ of $\boldsymbol{M}$, and return $(\boldsymbol{U}, \boldsymbol{\Sigma}, \boldsymbol{V})$

ℓ_2 guarantees for matrix completion

Theorem 6.13

Suppose that $n_1 p \geq C_1 \kappa^2 \mu r \log n_2$ *for some sufficiently large constant* $C_1 > 0$*. Then with probability exceeding* $1 - O(n_2^{-10})$,

$$\max\left\{\mathsf{dist}\left(\boldsymbol{U}, \boldsymbol{U}^\star\right), \mathsf{dist}\left(\boldsymbol{V}, \boldsymbol{V}^\star\right)\right\} \lesssim \kappa\sqrt{\frac{\mu r \log n_2}{n_1 p}}.$$

- Key: bound $\|\boldsymbol{M} - \boldsymbol{M}^\star\|$ by $\sqrt{\frac{\mu r \log n_2}{n_1 p}}\|\boldsymbol{M}^\star\|$ (homework)

$\ell_{2,\infty}$ guarantees for matrix completion

Theorem 6.14

Suppose that $n_1 \leq n_2$ *and* $n_1 p \geq C\kappa^4 \mu^2 r^2 \log n$ *for some sufficiently large constant* $C > 0$. *Then with high prob., we have*

$$\max\{\|\boldsymbol{U}\mathsf{sgn}(\boldsymbol{H}_{\boldsymbol{U}}) - \boldsymbol{U}^\star\|_{2,\infty}, \|\boldsymbol{V}\mathsf{sgn}(\boldsymbol{H}_{\boldsymbol{V}}) - \boldsymbol{V}^\star\|_{2,\infty}\} \leq \kappa^2 \sqrt{\frac{\mu^3 r^3 \log n}{n_1^2 p}};$$

$$\|\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^\top - \boldsymbol{M}^\star\|_\infty \lesssim \kappa^2 \mu^2 r^2 \sqrt{\frac{\log n}{n_1^3 p}} \|\boldsymbol{M}^\star\|$$

Proof of Theorem 6.14

Recall our notation $\boldsymbol{E} = \boldsymbol{M} - \boldsymbol{M}^\star = p^{-1}\mathcal{P}_\Omega(\boldsymbol{M}^\star) - \boldsymbol{M}^\star$. It is straightforward to check that $\boldsymbol{E}$ satisfies noise assumptions with

$$\sigma^2 := \frac{\|\boldsymbol{M}^\star\|_\infty^2}{p}, \qquad \text{and} \qquad B := \frac{\|\boldsymbol{M}^\star\|_\infty}{p}$$

In addition, from the relation $B = c_{\mathsf{b}}\sigma\sqrt{n_1/(\mu \log n)}$, it is seen that $c_{\mathsf{b}} = O(1)$ holds as long as $n_1 p \gtrsim \mu \log n$.

With these preparations in place, the claims in Theorem 6.14 follow directly from Theorem 6.11 and

$$\|\boldsymbol{M}^\star\|_\infty \le \mu r \|\boldsymbol{M}^\star\| / \sqrt{n_1 n_2}$$

What we have not discussed so far

- More applications of spectral methods
- Uncertainty quantification for spectral estimators
- Precise asymptotic analysis of spectral estimators
- Variants of spectral methods with certain advantages
- ℓ_p analysis of spectral methods