Spectral methods: ℓ_∞ perturbation theory
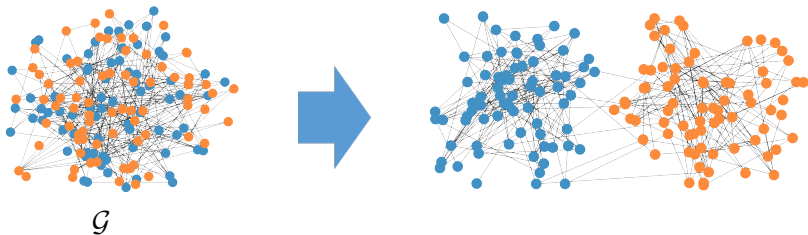
Cong Ma

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 $\boldsymbol{u} = [u_i]_{1 \leq i \leq n}$ of $\boldsymbol{A} - \frac{p+q}{2}\mathbf{1}\mathbf{1}^\top$

2. rounding: output $x_i = \begin{cases} 1, & \text{if } u_i \geq 0 \\ -1, & \text{if } u_i < 0 \end{cases}$

Almost exact recovery

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

- 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
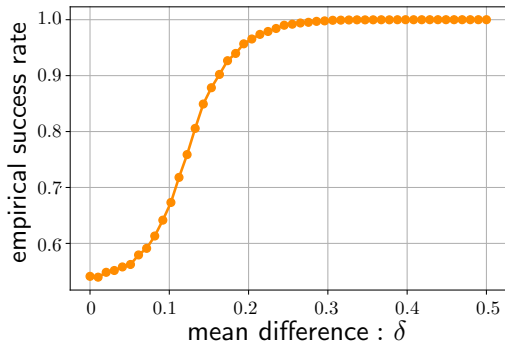
ℓ_2 perturbation theory alone cannot explain exact recovery guarantees

— call for fine-grained analysis

Reverse engineering

Spectral clustering uses signs of 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

⇧

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

⇧

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

⇑

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

⇑

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

⇑

Need ℓ_∞ perturbation theory

Outline

- An illustrative example: rank-1 matrix denoising

- General ℓ_∞ perturbation theory

Setup and algorithm

- Groundtruth: $M^\star = \lambda^\star u^\star u^{\star\top} \in \mathbb{R}^{n \times n}$, with $\lambda^\star > 0$

- Observation: $M = M^\star + E$, where E is symmetric, and its upper triangular part comprises of i.i.d. $\mathcal{N}(0, \sigma^2)$ entries

- Estimate $u^\star$ using u, leading eigenvector of M

- Goal: characterize entrywise errror

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

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 $\|\boldsymbol{E}\| \leq (1 - 1/\sqrt{2})\lambda^\star$

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

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: $1 \leq \mu \leq n$
- 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
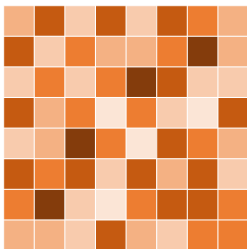
$$E_{i,j}^{(l)} := \begin{cases} E_{i,j}, & \text{if } i \neq l \text{ and } j \neq l, \\ 0, & \text{else.} \end{cases}$$

Leave-one-out estimates (cont.)

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

$$\boldsymbol{M}^{(l)} := \lambda^\star \boldsymbol{u}^\star \boldsymbol{u}^{\star\top} + \boldsymbol{E}^{(l)},$$
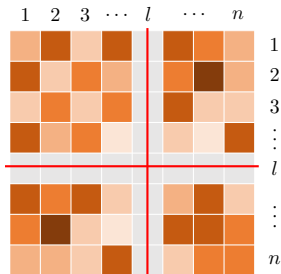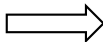
where the noise matrix $\boldsymbol{E}^{(l)}$ is generated according to

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

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,

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

Proof of Theorem 6.3

$$\|\boldsymbol{E}\| \leq 5\sigma\sqrt{n} \qquad\qquad \|\boldsymbol{E}^{(l)}\| \leq \|\boldsymbol{E}\| \leq 5\sigma\sqrt{n}$$

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

$$|\lambda - \lambda^{\star}| \leq 5\sigma\sqrt{n} \qquad\qquad |\lambda^{(l)} - \lambda^{\star}| \leq 5\sigma\sqrt{n}$$

$$\max_{j:j\geq 2} |\lambda_j(\boldsymbol{M})| \leq 5\sigma\sqrt{n} \qquad\qquad \max_{j:j\geq 2} |\lambda_j(\boldsymbol{M}^{(l)})| \leq 5\sigma\sqrt{n}$$

Assume WLOG,

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

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

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

—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

$$\|\boldsymbol{u} - \boldsymbol{u}^{(l)}\|_2 \leq \frac{2\|(\boldsymbol{M} - \boldsymbol{M}^{(l)})\boldsymbol{u}^{(l)}\|_2}{\lambda^{(l)} - \max_{j \geq 2} |\lambda_j(\boldsymbol{M}^{(l)})|} \leq \frac{4\|(\boldsymbol{M} - \boldsymbol{M}^{(l)})\boldsymbol{u}^{(l)}\|_2}{\lambda^\star}$$

as long as

$$\|\boldsymbol{M} - \boldsymbol{M}^{(l)}\| \leq (1 - 1/\sqrt{2})\Big(\lambda^{(l)} - \max_{j \geq 2} |\lambda_j(\boldsymbol{M}^{(l)})|\Big),$$

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

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

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

Combining previous bounds, we arrive at

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

provided that $40\sigma\sqrt{n} \leq \lambda^\star$

Rearranging terms and taking the union bound, we demonstrate that whp.,

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

Recall that

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

This implies

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

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

$$\left\| \boldsymbol{u} - \boldsymbol{u}^\star \right\|_\infty = \max_l |u_l - u_l^\star| \leq \max_l \left\{ |u_l^{(l)} - u_l^\star| + \left\| \boldsymbol{u} - \boldsymbol{u}^{(l)} \right\|_2 \right\}$$
$$\leq \frac{20\sigma\sqrt{n}}{\lambda^\star} \|\boldsymbol{u}^\star\|_\infty + \frac{40\sigma\sqrt{\log n} + 40\sigma\sqrt{n}\|\boldsymbol{u}\|_\infty}{\lambda^\star}$$

One more triangle inequality gives

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

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

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

where the last identity results from the definition of μ

General ℓ_∞ perturbation theory

—rank-1 case