Lecture 12

Last time

- Distances continued.
- Davis-Kahan $\sin\Theta$ Theorem

Today

- Proof of Davis-Kahan $\sin\Theta$.
- Wedin's theorem
- Community detection

Proof of Davis-Kahan: We will prove a silightly more general bound on

$$|||U_\perp^T U^*||| \leftarrow \text{Any rotationally invariant norm.}$$

The final bound will follow from Lemma ⚡ (Lecture 10). Assume condition 1) holds true. WLOG we can assume that $a = -b \le 0$. Otherwise, we can modify

$$M^* \leftarrow M^* - \frac{a+b}{2} I_n \text{ and } M \leftarrow M - \frac{a+b}{2} I_n.$$

You can check that this doesn't change the eigenvectors, it only shifts eigenvalues. Thus, we get

$$\underbrace{\|\Lambda^*\|_{op}}_{\max_{i\in[r]} |\lambda_i^*|} \le b \quad \text{and} \quad \underbrace{\sigma_{\min}(\Lambda_\perp)}_{\min_{i\in[n]\setminus[r]} |\lambda_i|} \ge b + \Delta.$$

We will use these to control $|||U_\perp^T U^*|||$.
The next decomposition will be useful:

$$U_\perp^T (M - M^*) U^* = \Lambda_\perp U_\perp^T U^* - U_\perp^T U^* \Lambda^*.$$

Thus, by noting $EU^* = (M-M^*)U^*$, yields

$$\begin{aligned}
|||U_\perp^T E U^*||| &\overset{\text{Reverse triangle ineq.}}{\geq} |||\Lambda_\perp U_\perp^T U^*||| - |||U_\perp^T U^* \Lambda^*||| \\
&\underset{\text{Fact } \star \text{ in Lecture 10}}{\geq} \sigma_{\min}(\Lambda_\perp) \cdot |||U_\perp^T U^*||| - |||U_\perp^T U^*||| \cdot \|\Lambda^*\|_{op} \\
&\geq (b + \Delta - b) |||U_\perp^T U^*||| \\
&= \Delta \, |||U_\perp^T U^*|||.
\end{aligned}$$

Consequently,

$$|||U_\perp^T U^*||| \leq \frac{|||U_\perp^T E U^*|||}{\Delta} \underset{\text{Fact } \star}{\leq} \|U_\perp^T\|_{op} \frac{|||EU^*|||}{\Delta} \overset{\|U_\perp^T\|_{op} = 1}{=} \frac{|||EU^*|||}{\Delta}$$

When $|||\cdot||| = \|\cdot\|_{op}$, we can apply Fact $\star$ once more to derive $\|\sin\Theta\|_{op} \leq \frac{\|E\|_{op}}{\Delta}$.
When $|||\cdot||| = \|\cdot\|_F$, we do the same and note $\|U^*\|_F = \sqrt{r}$, which gives $\|\sin\Theta\|_F \leq \frac{\sqrt{r}\,\|E\|_{op}}{\Delta}$.
The proof with condition 2) is similar and we leave it as an exercise.

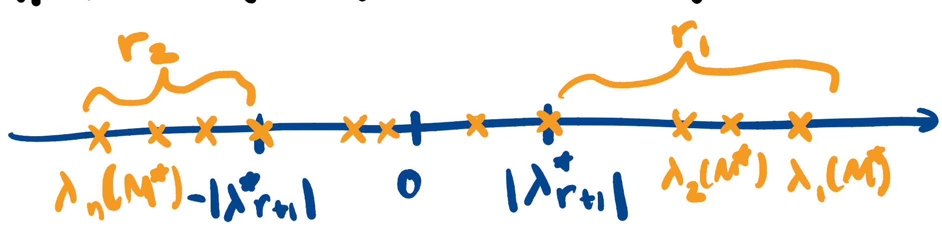
$\square$

Proof of Corollary (☺): We shall prove that the the eigenvalues satisfies condition 2). Let $\lambda_i(M^*)$ (resp. $\lambda_i(M)$) be the ith largest eigenvalue of M^* (resp. M) sorted by value (as opposed to absolute value.

By Weyl's ineq.

(W) $$|\lambda_i(M) - \lambda_i(M^*)| \leq \|E\|_{op} \quad \forall\, i \in [n].$$

Let $r_1 = \#\{i \mid \lambda_i(M) \geq |\lambda^*_{r+1}|\}$ and $r_2 = \#\{i \mid \lambda_i(M) \leq -|\lambda^*_{r+1}|\}$. That is



By construction, $r = r_1 + r_2$. Using (W) together with triangle ineq we get that for i s.t. $i \in [r_1]$ or $i > n - r_2$ we have

$$\begin{aligned}
|\lambda_i(M)| &\geq |\lambda_i(M^*)| - \|E\|_{op} \\
&\geq |\lambda^*_r| - (1 - 1/\sqrt{2})(|\lambda^*_r| - |\lambda^*_{r+1}|) \\
&= 1/\sqrt{2}\, |\lambda^*_r| + (1 - 1/\sqrt{2})|\lambda^*_{r+1}| \\
&\geq 1/\sqrt{2}\, |\lambda^*_{r+1}| + (1 - 1/\sqrt{2})|\lambda^*_r| \\
&= |\lambda^*_{r+1}| + (1 - 1/\sqrt{2})(|\lambda^*_r| - |\lambda^*_{r+1}|) \\
&\geq |\lambda^*_{r+1}| + \|E\|_{op}.
\end{aligned}$$

Convex combination with more weight on the smallest value.

On the other hand, if $r_1 < i \le n - r_2$ we have

$$|\lambda_i(M)| \le |\lambda_i(M^*)| + |\lambda_i(M) - \lambda_i(M^*)|$$
$$\le |\lambda^*_{r+1}| + \|E\|_{op}.$$

Thus, setting $b = |\lambda^*_{r+1}| + \|E\|_{op}$, gives

$$\{\lambda_{r+1}, \dots, \lambda_n\} \subseteq [-b, b]$$

Using that $|\lambda^*_1| \ge \dots \ge |\lambda^*_n|$ yields

$$\{\lambda^*_1, \dots, \lambda^*_r\} \subseteq (-\infty, -|\lambda^*_r|] \cup [|\lambda^*_r|, \infty).$$

We can set $a = -b$ and

$$\Delta = |\lambda^*_r| - |\lambda^*_{r+1}| - \|E\|_{op} > (|\lambda^*_r| - |\lambda^*_{r+1}|)/\sqrt{2}.$$

Invoking Davis-Kahan with condition 2) gives the result.

$\square$

Wedin's $\sin\Theta$ theorem

Wedin (1972) developed a parallel perturbation result for singular vectors. Assume $M = M^* + E \in \mathbb{R}^{n \times m}$

WLOG assume $n < m$

(♥)
$$M^* = \sum_{i=1}^{n} \lambda_i^* u_i^* v_i^{*T} = [U^* \; U_\perp^*] \begin{bmatrix} \Sigma^* & 0 & 0 \\ 0 & \Sigma_\perp^* & 0 \end{bmatrix} \begin{bmatrix} V^{*T} \\ V_\perp^{*T} \end{bmatrix},$$

$$M = \sum_{i=1}^{n} \lambda_i u_i v_i^T = [U \; U_\perp] \begin{bmatrix} \Sigma & 0 & 0 \\ 0 & \Sigma_\perp & 0 \end{bmatrix} \begin{bmatrix} U^T \\ V_\perp^T \end{bmatrix},$$

where $U^* = [u_1^*, \dots, u_r^*]$, $U_\perp^* = [u_{r+1}^*, \dots, u_n^*]$, $V^* = [v_1^*, \dots, v_r^*]$, $V_\perp^* = [v_{r+1}^*, \dots, v_n^*]$, $\Sigma^* = \mathrm{diag}(\sigma_1^*, \dots, \sigma_r^*)$, and $\Sigma_\perp^* = \mathrm{diag}(\sigma_{r+1}^*, \dots, \sigma_n^*)$. The matrices $U, U_\perp, V, V_\perp, \Sigma$ and $\Sigma_\perp$ are defined analogously.

Theorem (Wedin's sin Θ): Consider the setting in (♥). Assume $\|E\|_{op} \le \sigma_r^* - \sigma_{r+1}^*/\sqrt{2}$. Then

$$\mathrm{dist}_{op}(U, U^*) \vee \mathrm{dist}_{op}(V, V^*) \le 2 \frac{\|E^T U^*\|_{op} \vee \|E V^*\|_{op}}{\sigma_r^* - \sigma_{r+1}^*}.$$

$$\mathrm{dist}_F(U, U^*) \vee \mathrm{dist}_F(V, V^*) \le 2 \frac{\|E^T U^*\|_F \vee \|E V^*\|_F}{\sigma_r^* - \sigma_{r+1}^*}.$$

⊣

The proof is similar in spirit to that of Davis-Kahan sin Θ. If you are interested it appears in Spectral Methods for Data Science by Chen, Chi, Fan & Ma

Section 2.4.3.

Community detection in networks

We will apply these perturbation results to analyze an spectral method for community detection.

We consider the so-called stochastic block model where we observe a random graph $G(n, p, q)$ with n nodes that can be split into two disjoint comminities of size $n/2$ s.t.

$$\mathbb{P}((v_1, v_2) \text{ is an edge}) = \begin{cases} p & v_1, v_2 \text{ are in the same community,} \\ q & \text{otherwise.} \end{cases}$$

The goal is to identify the two communities from one draw of $G(n, p, q)$.

Consider the adjacency matrix of G given by

$$A_{ij} = \begin{cases} 1 & \text{if } (i,j) \text{ is an edge} \\ 0 & \text{otherwise.} \end{cases}$$

In expectation this matrix has a block structured. Assuming the first community is $[n/2]$, we have

$$\mathbb{E}A = \left[\begin{array}{ccc|ccc} p & \cdots & p & q & \cdots & q \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ p & \cdots & p & q & \cdots & q \\ \hline q & \cdots & q & p & \cdots & p \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ q & \cdots & q & p & \cdots & p \end{array}\right]$$

Check that this matrix is rank 2 with $\lambda_1 = \frac{p+q}{2}$, $\lambda_2 = \frac{p-q}{2}$,

$$u_1 = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \text{and} \quad u_2 = \begin{bmatrix} 1 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ -1 \end{bmatrix} \begin{matrix} \left.\vphantom{\begin{matrix}1\\ \vdots \\ 1\end{matrix}}\right\} n/2 \\ \vphantom{\begin{matrix}-1\\ \vdots \\ -1\end{matrix}} \end{matrix}$$

Key Insight: Thus, if A is close to $\mathbb{E}A$ we could use its second eigenvector to identify the communities.