Spectral methods: ℓ_2 perturbation theory

Cong Ma

University of Chicago, Autumn 2021

Matrix perturbation theory (spectral analysis)

Let $M^\star$ be a "simple" matrix, and E be a perturbation matrix

— *"simple" means spectral structure of $M^\star$ is understood*

Goal of matrix perturbation theory:

Understand how eigenspaces (resp. eigenvalues) / singular subspaces (resp. singular values) of $M^\star + E$ change w.r.t. perturbation E

Outline

- Preliminaries: basic matrix analysis

- Distance between two subspaces

- Eigenspace perturbation theory

- Perturbation bounds for singular subspaces

- Eigenvector perturbation bounds for probability transition matrices

Basic matrix analysis

Unitarily invariant norms

Definition 2.1

A matrix norm $\|\!\|\!\|\cdot\|\!\|\!\|$ on $\mathbb{R}^{m\times n}$ is said to be unitarily invariant if

$$\|\!\|\!\|\boldsymbol{A}\|\!\|\!\| = \|\!\|\!\|\boldsymbol{U}^\top \boldsymbol{A}\boldsymbol{V}\|\!\|\!\|$$

holds for any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and any two square orthonormal matrices $\boldsymbol{U} \in \mathcal{O}^{m\times m}$ and $\boldsymbol{V} \in \mathcal{O}^{n\times n}$.

Examples:

- $\|\boldsymbol{A}\|$: spectral norm (largest singular value of $\boldsymbol{A}$)
- $\|\boldsymbol{A}\|_{\mathrm{F}}$: Frobenius norm ($\|\boldsymbol{A}\|_{\mathrm{F}} = \sqrt{\mathrm{tr}(\boldsymbol{A}^\top \boldsymbol{A})} = \sqrt{\sum_{i,j} A_{i,j}^2}$)

Properties of unitarily invariant norms

Lemma 2.2

For any unitarily invariant norm $\|\cdot\|$, one has

$$\|AB\| \leq \|A\| \cdot \|B\|, \qquad \|AB\| \leq \|B\| \cdot \|A\|;$$
$$\|AB\| \geq \|A\| \sigma_{\min}(B), \qquad \textit{if } B \textit{ is square};$$
$$\|AB\| \geq \|B\| \sigma_{\min}(A), \qquad \textit{if } A \textit{ is square}.$$

Exercise: prove this lemma for special cases $\|\cdot\|$ and $\|\cdot\|_{\mathrm{F}}$

Eigenvalue perturbation bounds

Lemma 2.3 (Weyl's inequality for eigenvalues)

Let $\boldsymbol{A}, \boldsymbol{E} \in \mathbb{R}^{n \times n}$ be two real symmetric matrices. For every $1 \leq i \leq n$, the i-th largest eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{A} + \boldsymbol{E}$ obey

$$|\lambda_i\left(\boldsymbol{A}\right) - \lambda_i\left(\boldsymbol{A} + \boldsymbol{E}\right)| \leq \|\boldsymbol{E}\|.$$

Eigenvalue perturbation bounds

Lemma 2.3 (Weyl's inequality for eigenvalues)

Let $A, E \in \mathbb{R}^{n \times n}$ be two real symmetric matrices. For every $1 \le i \le n$, the i-th largest eigenvalues of A and $A + E$ obey

$$|\lambda_i(A) - \lambda_i(A + E)| \le \|E\|.$$

eigenvalues of real symmetric matrices are stable against perturbations

Eigenvalue perturbation bounds

Lemma 2.3 (Weyl's inequality for eigenvalues)

Let $A, E \in \mathbb{R}^{n \times n}$ be two real symmetric matrices. For every $1 \leq i \leq n$, the i-th largest eigenvalues of A and $A + E$ obey

$$|\lambda_i(A) - \lambda_i(A + E)| \leq \|E\|.$$

— *proof left as exercise*

eigenvalues of real symmetric matrices are stable against perturbations

Singular value perturbation bounds

Lemma 2.4 (Weyl's inequality for singular values)

Let $A, E \in \mathbb{R}^{m \times n}$ be two general matrices. Then for every $1 \leq i \leq \min\{m, n\}$, the i-th largest singular values of A and $A + E$ obey

$$|\sigma_i(A + E) - \sigma_i(A)| \leq \|E\|.$$

singular values are stable against perturbations

Proof of Lemma 2.4

We begin with introducing a useful "dilation" trick:

Definition 2.5 (Symmetric dilation)

For $\boldsymbol{A} \in \mathbb{R}^{n_1 \times n_2}$, define its symmetric dilation $\mathcal{S}(\boldsymbol{A})$ to be

$$\mathcal{S}(\boldsymbol{A}) = \left[\begin{array}{cc} \boldsymbol{0} & \boldsymbol{A} \\ \boldsymbol{A}^\top & \boldsymbol{0} \end{array} \right] \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}.$$

Then one has the following eigendecomposition for $\mathcal{S}(\boldsymbol{A})$:

$$\mathcal{S}(\boldsymbol{A}) = \frac{1}{\sqrt{2}} \left[\begin{array}{cc} \boldsymbol{U} & \boldsymbol{U} \\ \boldsymbol{V} & -\boldsymbol{V} \end{array} \right] \cdot \left[\begin{array}{cc} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{\Sigma} \end{array} \right] \cdot \frac{1}{\sqrt{2}} \left[\begin{array}{cc} \boldsymbol{U} & \boldsymbol{U} \\ \boldsymbol{V} & -\boldsymbol{V} \end{array} \right]^\top.$$

Two observations: for $1 \le i \le \min\{m, n\}$, $\lambda_i(\mathcal{S}(\boldsymbol{A})) = \sigma_i(\boldsymbol{A})$, and $\|\mathcal{S}(\boldsymbol{A})\| = \|\boldsymbol{A}\|$. Apply Lemma 2.3 to finish the proof.

Distance between two subspaces

Setup and notation

- Two r-dimensional subspaces $\mathcal{U}^\star$ and $\mathcal{U}$ in $\mathbb{R}^n$
- Two orthonormal matrices $\boldsymbol{U}^\star$ and $\boldsymbol{U}$ in $\mathbb{R}^{n \times r}$
- Orthogonal complements: $[\boldsymbol{U}^\star, \boldsymbol{U}_\perp^\star]$, and $[\boldsymbol{U}, \boldsymbol{U}_\perp]$

Question: how to measure distance?

- $\|U - U^\star\|_{\mathrm{F}}$ and $\|U - U^\star\|$ are not appropriate, since they fall short of accounting for $\underbrace{\text{global orthonormal transformation}}$

  $\forall$ orthonormal $R \in \mathbb{R}^{r \times r}$, U and UR represent same subspace

Three valid choices of distance

- Distance modulo *optimal rotation*
- Distance using *projection matrices*
- Geometric construction via *principal/canonical angles*

Distance modulo optimal rotation

Given global rotation ambiguity, it is natural to adjust for rotation before computing distance:

$$\mathsf{dist}_{\|\cdot\|}(\boldsymbol{U}, \boldsymbol{U}^{\star}) := \min_{\boldsymbol{R} \in \mathcal{O}^{r \times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^{\star}\|$$

Distance using projection matrices

Key observation: projection matrix $\boldsymbol{U}\boldsymbol{U}^\top$ associated with subspace $\mathcal{U}$ is unique

$$\mathsf{dist}_{\mathsf{p}, \|\!|\!|\cdot\|\!|\!|}(\boldsymbol{U}, \boldsymbol{U}^\star) \coloneqq \|\!|\!|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\|\!|\!|$$

Principal angles between two eigen-spaces

In addition to "distance", one might also be interested in "angles"
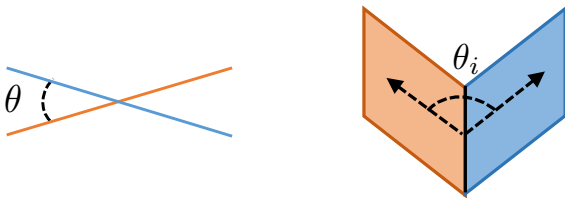
We can quantify the similarity between two lines (represented resp. by unit vectors $\boldsymbol{u}$ and $\boldsymbol{u}^{\star}$) by an angle between them

$$\theta = \arccos\langle \boldsymbol{u}, \boldsymbol{u}^{\star}\rangle$$

Principal angles between two eigen-spaces

More generally, for r-dimensional subspaces, one needs r angles

Specifically, given $\|\boldsymbol{U}^\top \boldsymbol{U}^\star\| \leq 1$, we write the singular value decomposition (SVD) of $\boldsymbol{U}^\top \boldsymbol{U}^\star \in \mathbb{R}^{r \times r}$ as

$$\boldsymbol{U}^\top \boldsymbol{U}^\star = \boldsymbol{X} \underbrace{\begin{bmatrix} \cos\theta_1 & & \\ & \ddots & \\ & & \cos\theta_r \end{bmatrix}}_{=:\cos\boldsymbol{\Theta}} \boldsymbol{Y}^\top =: \boldsymbol{X} \cos\boldsymbol{\Theta}\, \boldsymbol{Y}^\top$$

where $\{\theta_1, \ldots, \theta_r\}$ are called the principal angles between $\boldsymbol{U}$ and $\boldsymbol{U}^\star$

Distance using principal angles

With principal angles in place, we can define $\sin \Theta$ distance between subspaces as

$$\mathsf{dist}_{\mathsf{sin}, \|\!\|\cdot\|\!\|}(\boldsymbol{U}, \boldsymbol{U}^\star) := \|\!\|\sin \Theta\|\!\|$$

where

$$\Theta := \begin{bmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_r \end{bmatrix}, \quad \sin \Theta := \begin{bmatrix} \sin \theta_1 & & \\ & \ddots & \\ & & \sin \theta_r \end{bmatrix}$$

Link between projections and principal angles

Lemma 2.6

The following identities are true:

$$\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\| = \|\sin\boldsymbol{\Theta}\| = \|\boldsymbol{U}_\perp^\top\boldsymbol{U}^\star\| = \|\boldsymbol{U}^\top\boldsymbol{U}_\perp^\star\|;$$
$$\frac{1}{\sqrt{2}}\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\|_{\mathrm{F}} = \|\sin\boldsymbol{\Theta}\|_{\mathrm{F}} = \|\boldsymbol{U}_\perp^\top\boldsymbol{U}^\star\|_{\mathrm{F}} = \|\boldsymbol{U}^\top\boldsymbol{U}_\perp^\star\|_{\mathrm{F}}.$$

- sanity check: if $\boldsymbol{U} = \boldsymbol{U}^\star$, then everything is 0

Proof of Lemma 2.6

We prove the claim for spectral norm; the claim for Frobenius norm
follows similar argument. Note that

$$
\begin{aligned}
\|\boldsymbol{U}^\top \boldsymbol{U}_\perp^\star\| &= \|\boldsymbol{U}^\top \underbrace{\boldsymbol{U}_\perp^\star \boldsymbol{U}_\perp^{\star\top}}_{=\boldsymbol{I}-\boldsymbol{U}^\star \boldsymbol{U}^{\star\top}} \boldsymbol{U}\|^{\frac{1}{2}} \\
&= \|\boldsymbol{U}^\top \boldsymbol{U} - \boldsymbol{U}^\top \boldsymbol{U}^\star \boldsymbol{U}^{\star\top} \boldsymbol{U}\|^{\frac{1}{2}} \\
&= \|\boldsymbol{I} - \boldsymbol{X}\cos^2 \boldsymbol{\Theta}\, \boldsymbol{X}^\top\|^{\frac{1}{2}} \qquad (\text{write } \boldsymbol{U}^\top \boldsymbol{U}^\star = \boldsymbol{X}\cos\boldsymbol{\Theta}\, \boldsymbol{Y}^\top) \\
&= \|\boldsymbol{I} - \cos^2 \boldsymbol{\Theta}\|^{\frac{1}{2}} \\
&= \|\sin^2 \boldsymbol{\Theta}\|^{\frac{1}{2}} \\
&= \|\sin \boldsymbol{\Theta}\|
\end{aligned}
$$

Proof of Lemma 2.6 (cont.)

Given that singular values are unitarily invariant, it suffices to look at the singular values of the following matrix

$$\left[\begin{array}{c} \boldsymbol{U}^\top \\ \boldsymbol{U}_\perp^\top \end{array} \right] (\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top}) [\boldsymbol{U}_\perp^\star, \boldsymbol{U}^\star] = \left[\begin{array}{cc} \boldsymbol{U}^\top \boldsymbol{U}_\perp^\star & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star \end{array} \right]$$

which further implies

$$\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top}\| = \max\left\{\|\boldsymbol{U}^\top \boldsymbol{U}_\perp^\star\|, \|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star\|\right\};$$

$$\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top}\|_{\mathrm{F}} = \left(\|\boldsymbol{U}^\top \boldsymbol{U}_\perp^\star\|_{\mathrm{F}}^2 + \|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star\|_{\mathrm{F}}^2\right)^{1/2}$$

Link between optimal rotations and projections

Lemma 2.7

The following identities are true:

$$\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\| \le \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\| \le \sqrt{2}\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\|;$$

$$\tfrac{1}{\sqrt{2}}\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\|_{\mathrm{F}} \le \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{\mathrm{F}} \le \|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\|_{\mathrm{F}}.$$

— *proof left as exercise*

Summary of distance metrics

So far we have discussed

$$1) \quad \left\| \boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star \boldsymbol{U}^{\star\top} \right\|$$

$$2) \quad \left\| \sin \boldsymbol{\Theta} \right\|$$

$$3) \quad \left\| \boldsymbol{U}_\perp^\top \boldsymbol{U}^\star \right\| = \left\| \boldsymbol{U}^\top \boldsymbol{U}_\perp^\star \right\|$$

$$4) \quad \min_{\boldsymbol{R} \in \mathcal{O}^{r \times r}} \left\| \boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star \right\|$$

Summary of distance metrics

So far we have discussed

$$\begin{align}
&1) \quad \||\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\|| \\
&2) \quad \||\sin\boldsymbol{\Theta}\|| \\
&3) \quad \||\boldsymbol{U}_\perp^\top\boldsymbol{U}^\star\|| = \||\boldsymbol{U}^\top\boldsymbol{U}_\perp^\star\|| \\
&4) \quad \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \||\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\||
\end{align}$$

Our choice of distance:

$$\mathsf{dist}(\boldsymbol{U},\boldsymbol{U}^\star) \coloneqq \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|;$$

$$\mathsf{dist}_{\mathrm{F}}(\boldsymbol{U},\boldsymbol{U}^\star) \coloneqq \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{\mathrm{F}}$$

Eigenspace perturbation theory

Setup and notation

Consider 2 symmetric matrices $M^\star$, $M = M^\star + E \in \mathbb{R}^{n \times n}$ with eigen-decompositions

$$M^\star = \sum_{i=1}^n \lambda_i^\star u_i^\star u_i^{\star\top} = \begin{bmatrix} U^\star & U_\perp^\star \end{bmatrix} \begin{bmatrix} \Lambda^\star & 0 \\ 0 & \Lambda_\perp^\star \end{bmatrix} \begin{bmatrix} U^{\star\top} \\ U_\perp^{\star\top} \end{bmatrix};$$

$$M = \sum_{i=1}^n \lambda_i u_i u_i^\top = \begin{bmatrix} U & U_\perp \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & \Lambda_\perp \end{bmatrix} \begin{bmatrix} U^\top \\ U_\perp^\top \end{bmatrix}$$

Setup and notation

$$M = \Big[\underbrace{\boldsymbol{u}_1 \ \cdots \ \boldsymbol{u}_r}_{\boldsymbol{U}} \ \underbrace{\boldsymbol{u}_{r+1} \ \cdots \ \boldsymbol{u}_n}_{\boldsymbol{U}_\perp} \Big]$$

$$\cdot \left[\begin{array}{cccccc} \lambda_1 & & & & & \\ & \ddots & & & & \\ & & \underbrace{\lambda_r}_{\boldsymbol{\Lambda}} & & & \\ & & & \lambda_{r+1} & & \\ & & & & \ddots & \\ & & & & & \underbrace{\lambda_n}_{\boldsymbol{\Lambda}_\perp} \end{array} \right] \left. \left[\begin{array}{c} \boldsymbol{u}_1^\top \\ \vdots \\ \boldsymbol{u}_r^\top \\ \boldsymbol{u}_{r+1}^\top \\ \vdots \\ \boldsymbol{u}_n^\top \end{array} \right] \begin{array}{l} \Big\} \ \boldsymbol{U}^\top \\ \\ \Big\} \ \boldsymbol{U}_\perp^\top \end{array} \right.$$

Davis-Kahan's $\sin\Theta$ theorem: a simple case

Chandler Davis

William Kahan

Theorem 2.8 (Davis-Kahan's $\sin\Theta$ theorem: simple version)

Suppose $\boldsymbol{M}^\star \succeq \boldsymbol{0}$ and is rank-r. If $\|\boldsymbol{E}\| < (1 - 1/\sqrt{2})\lambda_r(\boldsymbol{M}^\star)$, then

$$\mathsf{dist}(\boldsymbol{U}, \boldsymbol{U}^\star) \le \sqrt{2}\, \|\sin\boldsymbol{\Theta}\| \le \frac{2\|\boldsymbol{E}\boldsymbol{U}^\star\|}{\lambda_r(\boldsymbol{M}^\star)} \le \frac{2\|\boldsymbol{E}\|}{\lambda_r(\boldsymbol{M}^\star)};$$

$$\mathsf{dist}_{\mathrm{F}}(\boldsymbol{U}, \boldsymbol{U}^\star) \le \sqrt{2}\, \|\sin\boldsymbol{\Theta}\|_{\mathrm{F}} \le \frac{2\|\boldsymbol{E}\boldsymbol{U}^\star\|_{\mathrm{F}}}{\lambda_r(\boldsymbol{M}^\star)} \le \frac{2\sqrt{r}\|\boldsymbol{E}\|}{\lambda_r(\boldsymbol{M}^\star)}.$$

Interpretations of Davis-Kahan's $\sin\Theta$ theorem

Suppose $M^\star \succeq 0$ and is rank-r. If $\|E\| < (1 - 1/\sqrt{2})\lambda_r(M^\star)$, then

$$\mathsf{dist}(U, U^\star) \le \sqrt{2}\,\|\sin\Theta\| \le \frac{2\|EU^\star\|}{\lambda_r(M^\star)} \le \frac{2\|E\|}{\lambda_r(M^\star)}.$$

Remarks:

- Eigen-gap $\lambda_r(M^\star) = \lambda_r(M^\star) - \lambda_{r+1}(M^\star)$

- Perturbation size $\|E\|$

- Signal-to-noise (SNR) ratio $\frac{\lambda_r(M^\star)}{\|E\|}$

- $\|EU^\star\|$ is sometimes useful; we will see benefit later

- Necessity of $\|E\| \lesssim \lambda_r(M^\star)$

What happens when SNR is small?

A toy example (with $0 < \epsilon < 1$)

$$M^\star = \left[\begin{array}{cc} 1 + \epsilon & 0 \\ 0 & 1 - \epsilon \end{array} \right], \quad E = \left[\begin{array}{cc} -\epsilon & \epsilon \\ \epsilon & \epsilon \end{array} \right], \quad M = \left[\begin{array}{cc} 1 & \epsilon \\ \epsilon & 1 \end{array} \right]$$

Leading eigenvectors of $M^\star$ and M are given respectively by

$$u_1^\star = \left[\begin{array}{c} 1 \\ 0 \end{array} \right], \qquad \text{and} \qquad u_1 = \frac{1}{\sqrt{2}} \left[\begin{array}{c} 1 \\ 1 \end{array} \right]$$

Consequently, we have

$$\|u_1 u_1^\top - u_1^\star u_1^{\star\top}\| = \frac{1}{\sqrt{2}}, \quad \text{and} \quad \|u_1 u_1^\top - u_1^\star u_1^{\star\top}\|_{\mathrm{F}} = 1$$

— *large regardless of size of ϵ or size of the perturbation $\|E\|$*

Proof of Theorem 2.8

We intend to control $U_\perp^\top U^\star$ by studying their interactions through E:

$$U_\perp^\top E U^\star = U_\perp^\top (M - M^\star) U^\star = \Lambda_\perp U_\perp^\top U^\star - U_\perp^\top U^\star \Lambda^\star,$$

which together with triangle inequality implies

$$\||U_\perp^\top E U^\star\|| \geq \||U_\perp^\top U^\star \Lambda^\star\|| - \||\Lambda_\perp U_\perp^\top U^\star\||$$
$$\geq \sigma_{\min}(\Lambda^\star)\||U_\perp^\top U^\star\|| - \|\Lambda_\perp\| \cdot \||U_\perp^\top U^\star\|| \qquad (2.6)$$

In view of Weyl's inequality, one has $\|\Lambda_\perp\| \leq \|E\|$. In addition, we have $\sigma_{\min}(\Lambda^\star) = \lambda_r(M^\star)$. These combined with relation (2.6) give

$$\||U_\perp^\top U^\star\|| \leq \frac{\||U_\perp^\top E U^\star\||}{\lambda_r(M^\star) - \|E\|} \leq \frac{\sqrt{2}\|U_\perp\| \cdot \||E U^\star\||}{\lambda_r(M^\star)} = \frac{\sqrt{2}\||E U^\star\||}{\lambda_r(M^\star)}$$

This together with Lemmas 2.6-2.7 completes the proof

Davis-Kahan's $\sin\Theta$ theorem: general case

Theorem 2.9 (Davis-Kahan's sinΘ theorem: general version)

Assume that

$$\text{eigenvalues}(\boldsymbol{\Lambda}^\star) \subseteq (-\infty, \alpha - \Delta] \cup [\beta + \Delta, \infty); \qquad (2.7a)$$

$$\text{eigenvalues}(\boldsymbol{\Lambda}_\perp) \subseteq [\alpha, \beta]. \qquad (2.7b)$$

for some quantities $\alpha, \beta \in \mathbb{R}$ and eigengap $\Delta > 0$. Then one has

$$\text{dist}(\boldsymbol{U}, \boldsymbol{U}^\star) \leq \sqrt{2}\|\sin\boldsymbol{\Theta}\| \leq \frac{\sqrt{2}\|\boldsymbol{E}\boldsymbol{U}^\star\|}{\Delta} \leq \frac{\sqrt{2}\|\boldsymbol{E}\|}{\Delta};$$

$$\text{dist}_{\mathrm{F}}(\boldsymbol{U}, \boldsymbol{U}^\star) \leq \sqrt{2}\|\sin\boldsymbol{\Theta}\|_{\mathrm{F}} \leq \frac{\sqrt{2}\|\boldsymbol{E}\boldsymbol{U}^\star\|_{\mathrm{F}}}{\Delta} \leq \frac{\sqrt{2r}\|\boldsymbol{E}\|}{\Delta}.$$

— conclusion remains valid if Assumption (2.7) *is reversed*

Perturbation theory for singular subspaces

Singular value decomposition

Let $\boldsymbol{M}^\star$ and $\boldsymbol{M} = \boldsymbol{M}^\star + \boldsymbol{E}$ be two matrices in $\mathbb{R}^{n_1 \times n_2}$ (WLOG, we assume $n_1 \leq n_2$), whose SVDs are given respectively by

$$\boldsymbol{M}^\star = \sum_{i=1}^{n_1} \sigma_i^\star \boldsymbol{u}_i^\star \boldsymbol{v}_i^{\star\top} = \left[\begin{array}{cc} \boldsymbol{U}^\star & \boldsymbol{U}_\perp^\star \end{array} \right] \left[\begin{array}{ccc} \boldsymbol{\Sigma}^\star & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_\perp^\star & \boldsymbol{0} \end{array} \right] \left[\begin{array}{c} \boldsymbol{V}^{\star\top} \\ \boldsymbol{V}_\perp^{\star\top} \end{array} \right]$$

$$\boldsymbol{M} = \sum_{i=1}^{n_1} \sigma_i \boldsymbol{u}_i \boldsymbol{v}_i^\top \ = \left[\begin{array}{cc} \boldsymbol{U} & \boldsymbol{U}_\perp \end{array} \right] \left[\begin{array}{ccc} \boldsymbol{\Sigma} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_\perp & \boldsymbol{0} \end{array} \right] \left[\begin{array}{c} \boldsymbol{V}^\top \\ \boldsymbol{V}_\perp^\top \end{array} \right]$$

- $\sigma_1 \geq \cdots \geq \sigma_{n_1}$ (resp. $\sigma_1^\star \geq \cdots \geq \sigma_{n_1}^\star$) stand for the singular values of $\boldsymbol{M}$ (resp. $\boldsymbol{M}^\star$) arranged in descending order
- $\boldsymbol{U}, \boldsymbol{U}^\star \in \mathbb{R}^{n_1 \times r}$ have orthonormal columns

Wedin's $\sin\Theta$ theorem

Davis-Kahan's theorem generalizes to singular subspace perturbation:

Theorem 2.10 (Wedin's sinΘ theorem)

If $\|\boldsymbol{E}\| < \sigma_r^\star - \sigma_{r+1}^\star$, then one has

$$\max\left\{\mathsf{dist}(\boldsymbol{U}, \boldsymbol{U}^\star), \mathsf{dist}(\boldsymbol{V}, \boldsymbol{V}^\star)\right\} \leq \frac{\sqrt{2}\max\left\{\|\boldsymbol{E}^\top\boldsymbol{U}^\star\|, \|\boldsymbol{E}\boldsymbol{V}^\star\|\right\}}{\sigma_r^\star - \sigma_{r+1}^\star - \|\boldsymbol{E}\|};$$

$$\max\left\{\mathsf{dist}_\mathrm{F}(\boldsymbol{U}, \boldsymbol{U}^\star), \mathsf{dist}_\mathrm{F}(\boldsymbol{V}, \boldsymbol{V}^\star)\right\} \leq \frac{\sqrt{2}\max\left\{\|\boldsymbol{E}^\top\boldsymbol{U}^\star\|_\mathrm{F}, \|\boldsymbol{E}\boldsymbol{V}^\star\|_\mathrm{F}\right\}}{\sigma_r^\star - \sigma_{r+1}^\star - \|\boldsymbol{E}\|}$$

— *can be simplified if* $\|E\| < (1 - 1/\sqrt{2})(\sigma_r^\star - \sigma_{r+1}^\star)$

Proof of Theorem 2.10

Similar to proof of Davis-Kahan theorem, we concentrate on $U_\perp^\top U^\star$

$$
\begin{aligned}
U_\perp^\top U^\star &= U_\perp^\top (U^\star \Sigma^\star V^{\star\top}) V^\star \Sigma^{\star-1} \\
&= U_\perp^\top \left(M - E - U_\perp^\star \Sigma_\perp^\star V_\perp^{\star\top} \right) V^\star \Sigma^{\star-1} \\
&= U_\perp^\top \left(U\Sigma V^\top + U_\perp \Sigma_\perp V_\perp^\top - E - U_\perp^\star \Sigma_\perp^\star V_\perp^{\star\top} \right) V^\star \Sigma^{\star-1} \\
&= \Sigma_\perp V_\perp^\top V^\star \Sigma^{\star-1} - U_\perp^\top E V^\star \Sigma^{\star-1}. \tag{2.9}
\end{aligned}
$$

Applying triangle inequality and Lemma 2.2 to identity (2.9) yields

$$
\begin{aligned}
\|U_\perp^\top U^\star\| &\leq \|\Sigma_\perp\| \cdot \|V_\perp^\top V^\star\| \cdot \|\Sigma^{\star-1}\| + \|U_\perp^\top\| \cdot \|E V^\star\| \cdot \|\Sigma^{\star-1}\| \\
&= \sigma_{r+1} \cdot \|V_\perp^\top V^\star\| \cdot \frac{1}{\sigma_r^\star} + \|E V^\star\| \cdot \frac{1}{\sigma_r^\star} \\
&\leq \frac{\sigma_{r+1}^\star + \|E\|}{\sigma_r^\star} \|V_\perp^\top V^\star\| + \frac{\|E V^\star\|}{\sigma_r^\star} \tag{2.10}
\end{aligned}
$$

Repeating the same argument yields

$$\|\!|\boldsymbol{V}_\perp^\top \boldsymbol{V}^\star|\!\| \leq \frac{\|\!|\boldsymbol{E}^\top \boldsymbol{U}^\star|\!\|}{\sigma_r^\star} + \frac{\sigma_{r+1}^\star + \|\boldsymbol{E}\|}{\sigma_r^\star}\|\!|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star|\!\| \qquad (2.11)$$

To finish up, combine inequalities (2.10) and (2.11) to obtain

$$\begin{aligned}
\max\left\{\|\!|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star|\!\|, \|\!|\boldsymbol{V}_\perp^\top \boldsymbol{V}^\star|\!\|\right\} &\leq \frac{\max\left\{\|\!|\boldsymbol{E}^\top \boldsymbol{U}^\star|\!\|, \|\!|\boldsymbol{E}\boldsymbol{V}^\star|\!\|\right\}}{\sigma_r^\star} \\
&+ \frac{\sigma_{r+1}^\star + \|\boldsymbol{E}\|}{\sigma_r^\star}\max\left\{\|\!|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star|\!\|, \|\!|\boldsymbol{V}_\perp^\top \boldsymbol{V}^\star|\!\|\right\}.
\end{aligned}$$

When $\|\boldsymbol{E}\| < \sigma_r^\star - \sigma_{r+1}^\star$, we can rearrange terms to obtain desired results

Extensions of Wedin's theorem

- Single rotation matrix: Wedin shows us existence of two unitary matrices $\boldsymbol{R}_U, \boldsymbol{R}_V$ such that

$$\max \left\{ \|\boldsymbol{U}\boldsymbol{R}_U - \boldsymbol{U}^\star\|_{\mathrm{F}}, \|\boldsymbol{V}\boldsymbol{R}_V - \boldsymbol{V}^\star\|_{\mathrm{F}} \right\} \quad \text{is small}$$

  ○ Can actually take same unitary matrix (exercise; hint "dilation")

- Separate bounds for left and right singular vectors:
  ○ Can treat $\boldsymbol{U}$ and $\boldsymbol{V}$ differently and obtain sharper bounds
  ○ Useful when n_1 and n_2 are drastically different

Eigenvector perturbation for probability transition matrices
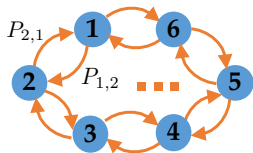
Eigen-decomposition for asymmetric matrices

Eigen-decomposition for asymmetric matrices is trickier:

1. both eigenvalues and eigenvectors might be complex-valued

2. eigenvectors might not be orthogonal to each other

This lecture focuses on a special case: **probability transition matrices**

Probability transition matrices

Consider a Markov chain $\{X_t\}_{t \geq 0}$

- n states
- transition probability $\mathbb{P}\{X_{t+1} = j \mid X_t = i\} = P_{i,j}$
- transition matrix $\boldsymbol{P} = [P_{i,j}]_{1 \leq i,j \leq n}$

Stationary distribution

Recall P is probability transition matrix

- $\pi = [\pi_i]_{1 \le i \le n}$ is stationary distribution of P if

$$\pi \ge 0, \qquad \mathbf{1}^\top \pi = 1, \qquad \text{and} \qquad \pi^\top P = \pi^\top$$

- π is in fact left eigenvector of P with eigenvalue 1
- 1 is largest eigenvalue of P in absolute sense: $|\lambda_i(P)| \le 1$ by Gershgorin circle theorem

Reversible Markov chains

- Markov chain $\{X_t\}_{t \geq 0}$ with transition matrix $\boldsymbol{P}$ and stationary distribution $\boldsymbol{\pi}$ is said to be reversible if

$$\pi_i P_{i,j} = \pi_j P_{j,i} \quad \text{for all } i, j$$

— *detailed balance condition*

- Nice consequence: if $\boldsymbol{P}$ is reversible, all eigenvalues are real

— *will see proof later*

Setup

- Probability transition matrix $P^\star$ of reversible Markov chain
- Perturbed transition matrix $P = P^\star + E$
- $\pi^\star$, π are leading left eigenvectors of $P^\star$, P, respectively
- Question: how does E affect perturbation $\pi - \pi^\star$

New norms

Fix a strictly positive probability vector $\boldsymbol{\pi} = [\pi_i]_{1 \leq i \leq n}$, define

- Vector norm: $\|\boldsymbol{x}\|_{\boldsymbol{\pi}} := \sqrt{\sum_i \pi_i x_i^2}$ with $\boldsymbol{x} = [x_i]_{1 \leq i \leq n}$
- Matrix norm: $\|\boldsymbol{A}\|_{\boldsymbol{\pi}} := \sup_{\|\boldsymbol{x}\|_{\boldsymbol{\pi}}=1} \|\boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{\pi}}$ with $\boldsymbol{A} = [A_{i,j}]_{1 \leq i,j \leq n}$

Eigenvector perturbation for transition matrices

Theorem 2.11 (Chen, Fan, Ma, Wang '17)

Suppose that $\boldsymbol{P}^\star$ represents a reversible Markov chain, whose stationary distribution vector $\boldsymbol{\pi}^\star$ is strictly positive. Assume that

$$\|\boldsymbol{E}\|_{\boldsymbol{\pi}^\star} < 1 - \max\left\{\lambda_2(\boldsymbol{P}^\star), -\lambda_n(\boldsymbol{P}^\star)\right\}.$$

Then one has

$$\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_{\boldsymbol{\pi}^\star} \leq \frac{\|\boldsymbol{\pi}^{\star\top}\boldsymbol{E}\|_{\boldsymbol{\pi}^\star}}{1 - \max\left\{\lambda_2(\boldsymbol{P}^\star), -\lambda_n(\boldsymbol{P}^\star)\right\} - \|\boldsymbol{E}\|_{\boldsymbol{\pi}^\star}}.$$

- Similar to Davis-Kahan
- Eigengap: $1 - \max\left\{\lambda_2(\boldsymbol{P}^\star), -\lambda_n(\boldsymbol{P}^\star)\right\}$ since $1 = \lambda_1(\boldsymbol{P})$
- Noise size: $\|\boldsymbol{\pi}^{\star\top}\boldsymbol{E}\|_{\boldsymbol{\pi}^\star}$

Proof of Theorem 2.11

By definitions of $\boldsymbol{\pi}^\star$ and $\boldsymbol{\pi}$, we have

$$\boldsymbol{\pi}^{\star\top} \boldsymbol{P}^\star = \boldsymbol{\pi}^{\star\top}, \qquad \text{and} \qquad \boldsymbol{\pi}^\top \boldsymbol{P} = \boldsymbol{\pi}^\top,$$

which imply the following decomposition of $\boldsymbol{\pi} - \boldsymbol{\pi}^\star$

$$\begin{aligned}
\boldsymbol{\pi}^\top - \boldsymbol{\pi}^{\star\top} &= \boldsymbol{\pi}^\top \boldsymbol{P} - \boldsymbol{\pi}^{\star\top} \boldsymbol{P}^\star = (\boldsymbol{\pi} - \boldsymbol{\pi}^\star)^\top \boldsymbol{P} + \boldsymbol{\pi}^{\star\top} (\boldsymbol{P} - \boldsymbol{P}^\star) \\
&= (\boldsymbol{\pi} - \boldsymbol{\pi}^\star)^\top (\boldsymbol{P} - \boldsymbol{P}^\star) + (\boldsymbol{\pi} - \boldsymbol{\pi}^\star)^\top \boldsymbol{P}^\star + \boldsymbol{\pi}^{\star\top} (\boldsymbol{P} - \boldsymbol{P}^\star) \\
&= (\boldsymbol{\pi} - \boldsymbol{\pi}^\star)^\top (\boldsymbol{P} - \boldsymbol{P}^\star) + (\boldsymbol{\pi} - \boldsymbol{\pi}^\star)^\top (\boldsymbol{P}^\star - \mathbf{1}\boldsymbol{\pi}^{\star\top}) + \boldsymbol{\pi}^{\star\top} (\boldsymbol{P} - \boldsymbol{P}^\star)
\end{aligned}$$

In last step, we use $(\boldsymbol{\pi} - \boldsymbol{\pi}^\star)^\top \mathbf{1} = 1 - 1 = 0$

Proof of Theorem 2.11 (cont.)

Apply triangle inequality w.r.t. norm $\|\cdot\|_{\pi^\star}$ to obtain

$$\begin{aligned}
\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_{\pi^\star} &\leq \left\| (\boldsymbol{\pi} - \boldsymbol{\pi}^\star)^\top (\boldsymbol{P} - \boldsymbol{P}^\star) \right\|_{\pi^\star} + \left\| (\boldsymbol{\pi} - \boldsymbol{\pi}^\star)^\top (\boldsymbol{P}^\star - \mathbf{1}\boldsymbol{\pi}^{\star\top}) \right\|_{\pi^\star} \\
&\quad + \left\| \boldsymbol{\pi}^{\star\top} (\boldsymbol{P} - \boldsymbol{P}^\star) \right\|_{\pi^\star} \\
&\leq \left(\|\boldsymbol{P} - \boldsymbol{P}^\star\|_{\pi^\star} + \|\boldsymbol{P}^\star - \mathbf{1}\boldsymbol{\pi}^{\star\top}\|_{\pi^\star} \right) \|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_{\pi^\star} \\
&\quad + \left\| \boldsymbol{\pi}^{\star\top} (\boldsymbol{P} - \boldsymbol{P}^\star) \right\|_{\pi^\star}
\end{aligned}$$

Assuming $\|\boldsymbol{P} - \boldsymbol{P}^\star\|_{\pi^\star} + \|\boldsymbol{P}^\star - \mathbf{1}\boldsymbol{\pi}^{\star\top}\|_{\pi^\star} < 1$, rearrangement gives

$$\|\boldsymbol{\pi} - \boldsymbol{\pi}^\star\|_{\pi^\star} \leq \frac{\left\| \boldsymbol{\pi}^{\star\top} (\boldsymbol{P} - \boldsymbol{P}^\star) \right\|_{\pi^\star}}{1 - \|\boldsymbol{P} - \boldsymbol{P}^\star\|_{\pi^\star} - \|\boldsymbol{P}^\star - \mathbf{1}\boldsymbol{\pi}^{\star\top}\|_{\pi^\star}}$$

Proof will be complete if one can show

$$\left\| \boldsymbol{P}^\star - \mathbf{1}\boldsymbol{\pi}^{\star\top} \right\|_{\pi^\star} = \max \left\{ \lambda_2(\boldsymbol{P}^\star), -\lambda_n(\boldsymbol{P}^\star) \right\} \qquad (2.12)$$

Proof of identity (2.12)

Define diagonal matrix $\boldsymbol{\Pi}^\star = \mathsf{diag}([\pi_1^\star, \cdots, \pi_n^\star]) \in \mathbb{R}^{n \times n}$. Observe

$$\|\boldsymbol{A}\|_{\boldsymbol{\pi}^\star} = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_{\boldsymbol{\pi}^\star}}{\|\boldsymbol{x}\|_{\boldsymbol{\pi}^\star}} = \sup_{\boldsymbol{x} \neq \boldsymbol{0}} \frac{\left\|(\boldsymbol{\Pi}^\star)^{1/2} \boldsymbol{A} (\boldsymbol{\Pi}^\star)^{-1/2} (\boldsymbol{\Pi}^\star)^{1/2} \boldsymbol{x}\right\|_2}{\left\|(\boldsymbol{\Pi}^\star)^{1/2} \boldsymbol{x}\right\|_2}$$

$$= \sup_{\boldsymbol{v} \neq \boldsymbol{0}} \frac{\left\|(\boldsymbol{\Pi}^\star)^{1/2} \boldsymbol{A} (\boldsymbol{\Pi}^\star)^{-1/2} \boldsymbol{v}\right\|_2}{\|\boldsymbol{v}\|_2} = \|(\boldsymbol{\Pi}^\star)^{1/2} \boldsymbol{A} (\boldsymbol{\Pi}^\star)^{-1/2}\|$$

As a consequence, one has

$$\|\boldsymbol{P}^\star - \boldsymbol{1}\boldsymbol{\pi}^{\star\top}\|_{\boldsymbol{\pi}^\star} = \|(\boldsymbol{\Pi}^\star)^{1/2} \left(\boldsymbol{P}^\star - \boldsymbol{1}\boldsymbol{\pi}^{\star\top}\right) (\boldsymbol{\Pi}^\star)^{-1/2}\|$$

$$= \|\boldsymbol{S}^\star - \boldsymbol{\pi}_{1/2}^\star (\boldsymbol{\pi}_{1/2}^\star)^\top\|$$

with $\boldsymbol{S}^\star = (\boldsymbol{\Pi}^\star)^{1/2} \boldsymbol{P}^\star (\boldsymbol{\Pi}^\star)^{-1/2}$ and $\boldsymbol{\pi}_{1/2}^\star = [(\pi_j^\star)^{1/2}]_{1 \leq j \leq n}$

Proof of identity (2.12) (cont.)

Several properties of $S^\star$:

- Symmetric: all eigenvalues are real

  — *check detailed balance*

- Similar to $P^\star$: $S^\star$ have same eigenvalues as $P^\star$, and

$$S^\star \pi_{1/2}^\star = \pi_{1/2}^\star$$

- Eigenvalues of $S^\star - \pi_{1/2}^\star (\pi_{1/2}^\star)^\top$ are $0, \lambda_2(S^\star), \ldots, \lambda_n(S^\star)$

Combine all to see

$$\begin{aligned}
\big\| S^\star - \pi_{1/2}^\star (\pi_{1/2}^\star)^\top \big\| &\overset{(\mathrm{i})}{=} \max\left\{ |\lambda_2(S^\star)|, |\lambda_n(S^\star)| \right\} \\
&= \max\left\{ \lambda_2(S^\star), -\lambda_n(S^\star) \right\} \overset{(\mathrm{ii})}{=} \max\left\{ \lambda_2(P^\star), -\lambda_n(P^\star) \right\}.
\end{aligned}$$