Spectral methods: ℓ_2 perturbation theory

Cong Ma

University of Chicago, Autumn 2021

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

$$|||AB||| \le |||A||| \cdot \|B\|, \qquad |||AB||| \le |||B||| \cdot \|A\|,$$
$$|||AB||| \ge |||A||| \, \sigma_{\min}(B), \qquad |||AB||| \ge |||B||| \, \sigma_{\min}(A).$$

Exercise: prove the 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*

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

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*

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

eigenvalues of real symmetric matrices are stable against perturbations

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*

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

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

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

Proof of Lemma 2.4

We begin with introducing a useful "dilation" trick:

Definition 2.5 (Symmetric dilation)

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

$$\mathcal{S}(\boldsymbol{A}) = \left[\begin{array}{cc} \boldsymbol{0} & \boldsymbol{A} \\ \boldsymbol{A}^\top & \boldsymbol{0} \end{array} \right].$$

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

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

For $1 \leq i \leq \min\{m, n\}$, $\lambda_i(\mathcal{S}(\boldsymbol{A})) = \sigma_i(\boldsymbol{A})$. In addition, $\|\mathcal{S}(\boldsymbol{E})\| = \|\boldsymbol{E}\|$.

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 define distance?

- $\|\boldsymbol{U} - \boldsymbol{U}^\star\|_{\mathrm{F}}$ and $\|\boldsymbol{U} - \boldsymbol{U}^\star\|$ are not appropriate, since they fall short of accounting for $\underbrace{\text{global orthonormal transformation}}_{\forall \text{ orthonormal } \boldsymbol{R} \in \mathbb{R}^{r\times r},\ \boldsymbol{U} \text{ and } \boldsymbol{U}\boldsymbol{R} \text{ represent same subspace}}$

Valid choices of distance

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

Distance modulo optimal rotation

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

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

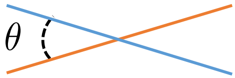
Distance using projection matrices

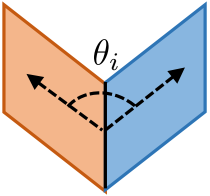
Key observation: projection matrix $\boldsymbol{U}\boldsymbol{U}^\top$ associated with subspace $\mathcal{U}$ is unique

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

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

Distance using principal angles

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

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

Link between projections and principal angles

Lemma 2.6

We have

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

- 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

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

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

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

which further implies

$$\begin{aligned} \|\boldsymbol{U}\boldsymbol{U}^{\top} - \boldsymbol{U}^{\star}\boldsymbol{U}^{\star\top}\| &= \max\left\{\|\boldsymbol{U}^{\top}\boldsymbol{U}_{\perp}^{\star}\|, \|\boldsymbol{U}_{\perp}^{\top}\boldsymbol{U}^{\star}\|\right\}; \\ \|\boldsymbol{U}\boldsymbol{U}^{\top} - \boldsymbol{U}^{\star}\boldsymbol{U}^{\star\top}\|_{\mathrm{F}} &= \left(\|\boldsymbol{U}^{\top}\boldsymbol{U}_{\perp}^{\star}\|_{\mathrm{F}}^{2} + \|\boldsymbol{U}_{\perp}^{\top}\boldsymbol{U}^{\star}\|_{\mathrm{F}}^{2}\right)^{1/2} \end{aligned}$$

Link between optimal rotations and projections

Lemma 2.7

One has

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

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

— *proof left as exercise*

Summary of distance metrics

So far we have discussed

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

Summary of distance metrics

So far we have discussed

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

Our choice of distance:

$$
\begin{aligned}
\mathsf{dist}(\boldsymbol{U}, \boldsymbol{U}^\star) &:= \min_{\boldsymbol{R} \in \mathcal{O}^{r\times r}} \left\| \boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star \right\|; \\
\mathsf{dist}_{\mathrm{F}}(\boldsymbol{U}, \boldsymbol{U}^\star) &:= \min_{\boldsymbol{R} \in \mathcal{O}^{r\times r}} \left\| \boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star \right\|_{\mathrm{F}}
\end{aligned}
$$

Eigen-space perturbation theory

Setup and notation

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

$$\boldsymbol{M} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^\top \qquad \text{and} \qquad \hat{\boldsymbol{M}} = \sum_{i=1}^{n} \hat{\lambda}_i \hat{\boldsymbol{u}}_i \hat{\boldsymbol{u}}_i^\top$$

where $\lambda_1 \geq \cdots \geq \lambda_n$; $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_n$. For simplicity, write

$$\boldsymbol{M} = [\boldsymbol{U}_0, \boldsymbol{U}_1] \begin{bmatrix} \boldsymbol{\Lambda}_0 & \\ & \boldsymbol{\Lambda}_1 \end{bmatrix} \begin{bmatrix} \boldsymbol{U}_0^\top \\ \boldsymbol{U}_1^\top \end{bmatrix}$$

$$\hat{\boldsymbol{M}} = [\hat{\boldsymbol{U}}_0, \hat{\boldsymbol{U}}_1] \begin{bmatrix} \hat{\boldsymbol{\Lambda}}_0 & \\ & \hat{\boldsymbol{\Lambda}}_1 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{U}}_0^\top \\ \hat{\boldsymbol{U}}_1^\top \end{bmatrix}$$

Here, $\boldsymbol{U}_0 = [\boldsymbol{u}_1, \cdots, \boldsymbol{u}_r]$, $\boldsymbol{\Lambda}_0 = \mathrm{diag}\left([\lambda_1, \cdots, \lambda_r]\right)$, $\cdots$

Setup and notation

$$
\boldsymbol{M} = \Big[\underbrace{\boldsymbol{u}_1 \ \cdots \ \boldsymbol{u}_r}_{\boldsymbol{U}_0}\ \underbrace{\boldsymbol{u}_{r+1} \ \cdots \ \boldsymbol{u}_n}_{\boldsymbol{U}_1}\Big]
\cdot \begin{bmatrix} \underbrace{\begin{matrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{matrix}}_{\boldsymbol{\Lambda}_0} & \\ & \underbrace{\begin{matrix} \lambda_{r+1} & & \\ & \ddots & \\ & & \lambda_n \end{matrix}}_{\boldsymbol{\Lambda}_1} \end{bmatrix}
\begin{bmatrix} \boldsymbol{u}_1^\top \\ \vdots \\ \boldsymbol{u}_r^\top \\ \boldsymbol{u}_{r+1}^\top \\ \vdots \\ \boldsymbol{u}_n^\top \end{bmatrix}
\begin{matrix} \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \boldsymbol{U}_0^\top \\ \left.\begin{matrix} \\ \\ \\ \end{matrix}\right\} \boldsymbol{U}_1^\top \end{matrix}
$$

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

Chandler Davis

William Kahan

Theorem 2.8

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

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

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

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

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

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

Remarks:

- Eigen-gap $\lambda_r(\boldsymbol{M}^\star) = \lambda_r(\boldsymbol{M}^\star) - \lambda_{r+1}(\boldsymbol{M}^\star)$
- Perturbation size $\|\boldsymbol{E}\|$
- Inverse signal-to-noise ratio $\frac{\|\boldsymbol{E}\|}{\lambda_r(\boldsymbol{M}^\star)}$
- Necessity of $\|\boldsymbol{E}\| \lesssim \lambda_r(\boldsymbol{M}^\star)$

What happens when SNR is small?

A toy example

$$\boldsymbol{M}^\star = \begin{bmatrix} 1+\epsilon & 0 \\ 0 & 1-\epsilon \end{bmatrix}, \quad \boldsymbol{E} = \begin{bmatrix} -\epsilon & \epsilon \\ \epsilon & \epsilon \end{bmatrix}, \quad \boldsymbol{M} = \begin{bmatrix} 1 & \epsilon \\ \epsilon & 1 \end{bmatrix},$$

where $0 < \epsilon < 1$ can be arbitrarily small. It is straightforward to check that the leading eigenvectors of $\boldsymbol{M}^\star$ and $\boldsymbol{M}$ are given respectively by

$$\boldsymbol{u}_1^\star = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \text{and} \qquad \boldsymbol{u}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Consequently, we have

$$\left\| \boldsymbol{u}_1 \boldsymbol{u}_1^\top - \boldsymbol{u}_1^\star \boldsymbol{u}_1^{\star\top} \right\| = \frac{1}{\sqrt{2}}, \quad \text{and} \quad \left\| \boldsymbol{u}_1 \boldsymbol{u}_1^\top - \boldsymbol{u}_1^\star \boldsymbol{u}_1^{\star\top} \right\|_{\mathrm{F}} = 1, \qquad (2.6)$$

which are both quite large regardless of the size of ϵ or the size of the perturbation $\|\boldsymbol{E}\|$.

Proof of Theorem 2.8

We intend to control $\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star$ by studying their interactions through $\boldsymbol{E}$:

$$\boldsymbol{U}_\perp^\top (\boldsymbol{M} - \boldsymbol{M}^\star)\boldsymbol{U}^\star = \boldsymbol{\Lambda}_\perp \boldsymbol{U}_\perp^\top \boldsymbol{U}^\star - \boldsymbol{U}_\perp^\top \boldsymbol{U}^\star \boldsymbol{\Lambda}^\star, \tag{2.7}$$

which together with triangle inequality implies

$$\begin{aligned} \|\boldsymbol{U}_\perp^\top \boldsymbol{E} \boldsymbol{U}^\star\| &\geq \|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star \boldsymbol{\Lambda}^\star\| - \|\boldsymbol{\Lambda}_\perp \boldsymbol{U}_\perp^\top \boldsymbol{U}^\star\| \\ &\geq \sigma_{\min}(\boldsymbol{\Lambda}^\star)\|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star\| - \|\boldsymbol{\Lambda}_\perp\| \cdot \|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star\| \end{aligned} \tag{2.8}$$

In view of Weyl's lemma, one has $\|\hat{\boldsymbol{\Lambda}}_\perp\| \leq \|\boldsymbol{E}\|$. In addition, we have $\sigma_{\min}(\boldsymbol{\Lambda}^\star) = \lambda_r(\boldsymbol{M}^\star)$. These combined with relation (2.8) give

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

This together with Lemmas 2.6-2.7 completes the proof

Work until here

Davis-Kahan $\sin\Theta$ Theorem: general case

Theorem 2.9 (Davis-Kahan $\sin\Theta$ Theorem)

Suppose $\lambda_r(\boldsymbol{M}) \geq a$ *and* $\lambda_{r+1}(\hat{\boldsymbol{M}}) \leq a - \Delta$ *for some* $\Delta > 0$. *Then*

$$\mathsf{dist}(\hat{\boldsymbol{U}}_0, \boldsymbol{U}_0) \leq \frac{\|\boldsymbol{E}\boldsymbol{U}_0\|}{\Delta} \leq \frac{\|\boldsymbol{E}\|}{\Delta}$$

- immediate consequence: if $\lambda_r(\boldsymbol{M}) > \lambda_{r+1}(\boldsymbol{M}) + \|\boldsymbol{E}\|$, then

$$\mathsf{dist}(\hat{\boldsymbol{U}}_0, \boldsymbol{U}_0) \leq \frac{\|\boldsymbol{E}\|}{\underbrace{\lambda_r(\boldsymbol{M}) - \lambda_{r+1}(\boldsymbol{M})}_{\text{spectral gap}} - \|\boldsymbol{E}\|} \tag{2.9}$$

Perturbation theory for singular subspaces

Singular value decomposition

Consider two matrices $\boldsymbol{M}, \hat{\boldsymbol{M}} = \boldsymbol{M} + \boldsymbol{E} \in \mathbb{R}^{n_1 \times n_2}$ with SVD

$$\boldsymbol{M} = [\boldsymbol{U}_0, \boldsymbol{U}_1] \left[\begin{array}{cc} \boldsymbol{\Sigma}_0 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_1 \\ \boldsymbol{0} & \boldsymbol{0} \end{array} \right] \left[\begin{array}{c} \boldsymbol{V}_0^\top \\ \boldsymbol{V}_1^\top \end{array} \right]$$

$$\hat{\boldsymbol{M}} = \left[\hat{\boldsymbol{U}}_0, \hat{\boldsymbol{U}}_1\right] \left[\begin{array}{cc} \hat{\boldsymbol{\Sigma}}_0 & \boldsymbol{0} \\ \boldsymbol{0} & \hat{\boldsymbol{\Sigma}}_1 \\ \boldsymbol{0} & \boldsymbol{0} \end{array} \right] \left[\begin{array}{c} \hat{\boldsymbol{V}}_0^\top \\ \hat{\boldsymbol{V}}_1^\top \end{array} \right]$$

where $\boldsymbol{U}_0$ (resp. $\hat{\boldsymbol{U}}_0$) and $\boldsymbol{V}_0$ (resp. $\hat{\boldsymbol{V}}_0$) represent the top-r singular subspaces of $\boldsymbol{M}$ (resp. $\hat{\boldsymbol{M}}$)

Wedin's $\sin \boldsymbol{\Theta}$ theorem

The Davis-Kahan 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*

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

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

Proof of Theorem 2.10

$$
\begin{aligned}
\boldsymbol{U}_{\perp}^{\top}\boldsymbol{U}^{\star} &= \boldsymbol{U}_{\perp}^{\top}(\boldsymbol{U}^{\star}\boldsymbol{\Sigma}^{\star}\boldsymbol{V}^{\star\top})\boldsymbol{V}^{\star}\boldsymbol{\Sigma}^{\star-1} \\
&= \boldsymbol{U}_{\perp}^{\top}\left(\boldsymbol{M} - \boldsymbol{E} - \boldsymbol{U}_{\perp}^{\star}\boldsymbol{\Sigma}_{\perp}^{\star}\boldsymbol{V}_{\perp}^{\star\top}\right)\boldsymbol{V}^{\star}\boldsymbol{\Sigma}^{\star-1} \\
&= \boldsymbol{U}_{\perp}^{\top}\left(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\top} + \boldsymbol{U}_{\perp}\boldsymbol{\Sigma}_{\perp}\boldsymbol{V}_{\perp}^{\top} - \boldsymbol{E} - \boldsymbol{U}_{\perp}^{\star}\boldsymbol{\Sigma}_{\perp}^{\star}\boldsymbol{V}_{\perp}^{\star\top}\right)\boldsymbol{V}^{\star}\boldsymbol{\Sigma}^{\star-1} \\
&= \boldsymbol{\Sigma}_{\perp}\boldsymbol{V}_{\perp}^{\top}\boldsymbol{V}^{\star}\boldsymbol{\Sigma}^{\star-1} - \boldsymbol{U}_{\perp}^{\top}\boldsymbol{E}\boldsymbol{V}^{\star}\boldsymbol{\Sigma}^{\star-1}. \qquad (2.10)
\end{aligned}
$$

Applying the triangle inequality and Lemma 2.2 in Section **??** to the identity (2.10) yields

$$
\begin{aligned}
|||\boldsymbol{U}_{\perp}^{\top}\boldsymbol{U}^{\star}||| &\le \|\boldsymbol{\Sigma}_{\perp}\| \cdot |||\boldsymbol{V}_{\perp}^{\top}\boldsymbol{V}^{\star}||| \cdot \|\boldsymbol{\Sigma}^{\star-1}\| + \|\boldsymbol{U}_{\perp}^{\top}\| \cdot |||\boldsymbol{E}\boldsymbol{V}^{\star}||| \cdot \|\boldsymbol{\Sigma}^{\star-1}\| \\
&= \sigma_{r+1} \cdot |||\boldsymbol{V}_{\perp}^{\top}\boldsymbol{V}^{\star}||| \cdot \frac{1}{\sigma_r^{\star}} + |||\boldsymbol{E}\boldsymbol{V}^{\star}||| \cdot \frac{1}{\sigma_r^{\star}} \\
&\le \frac{\sigma_{r+1}^{\star} + \|\boldsymbol{E}\|}{\sigma_r^{\star}} |||\boldsymbol{V}_{\perp}^{\top}\boldsymbol{V}^{\star}||| + \frac{|||\boldsymbol{E}\boldsymbol{V}^{\star}|||}{\sigma_r^{\star}}. \qquad (2.11)
\end{aligned}
$$

Proof of Theorem 2.10 (cont.)

Repeating the same argument yields

$$\|\boldsymbol{V}_\perp^\top \boldsymbol{V}^\star\| \le \frac{\|\boldsymbol{E}^\top \boldsymbol{U}^\star\|}{\sigma_r^\star} + \frac{\sigma_{r+1}^\star + \|\boldsymbol{E}\|}{\sigma_r^\star}\|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star\|. \qquad (2.12)$$

To finish up, combine the inequalities (2.11) and (2.12) to obtain

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

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

$$\max\left\{\|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star\|, \|\boldsymbol{V}_\perp^\top \boldsymbol{V}^\star\|\right\} \le \frac{\max\left\{\|\boldsymbol{E}^\top \boldsymbol{U}^\star\|, \|\boldsymbol{E}\boldsymbol{V}^\star\|\right\}}{\sigma_r^\star - \sigma_{r+1}^\star - \|\boldsymbol{E}\|}.$$

The proof is then completed by invoking Lemmas **??** and 2.7.

Extensions of Wedin's theorem

- Single rotation matrix
- Separate bounds for left and right singular vectors

Eigenvector perturbation for probability transition matrices

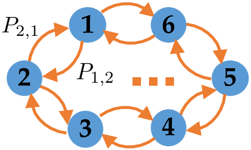
Eigen-decomposition for asymmetric matrices

Eigen-decomposition for asymmetric matrices is much more tricky:

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 $\underbrace{\boldsymbol{\pi} = [\pi_1, \cdots, \pi_n]}_{\pi_1+\cdots+\pi_n=1}$ is 1st eigenvector of $\boldsymbol{P}$

$$\boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{\pi}$$

- $\{X_t\}_{t\geq 0}$ is said to be reversible if $\pi_i P_{i,j} = \pi_j P_{j,i}$ for all i, j

Eigenvector perturbation for transition matrices

Define $\|\boldsymbol{a}\|_{\boldsymbol{\pi}} := \sqrt{\pi_1 a_1^2 + \cdots + \pi_n a_n^2}$

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

Suppose $\boldsymbol{P}$, $\hat{\boldsymbol{P}}$ *are transition matrices with stationary distributions* $\boldsymbol{\pi}$, $\hat{\boldsymbol{\pi}}$, *respectively. Assume* $\boldsymbol{P}$ *induces a reversible Markov chain. If* $1 > \max\left\{\lambda_2(\boldsymbol{P}), -\lambda_n(\boldsymbol{P})\right\} + \|\hat{\boldsymbol{P}} - \boldsymbol{P}\|_{\boldsymbol{\pi}}$, *then*

$$\|\hat{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_{\boldsymbol{\pi}} \leq \frac{\|\boldsymbol{\pi}(\hat{\boldsymbol{P}} - \boldsymbol{P})\|_{\boldsymbol{\pi}}}{\underbrace{1 - \max\left\{\lambda_2(\boldsymbol{P}), -\lambda_n(\boldsymbol{P})\right\}}_{\text{spectral gap}} - \underbrace{\|\hat{\boldsymbol{P}} - \boldsymbol{P}\|_{\boldsymbol{\pi}}}_{\text{perturbation}}}$$

- $\hat{\boldsymbol{P}}$ does not need to induce a reversible Markov chain

Reference

- "*The rotation of eigenvectors by a perturbation*," C. Davis, W. Kahan, *SIAM Journal on Numerical Analysis*, 1970.
- "*Perturbation bounds in connection with singular value decomposition*," P. Wedin, *BIT Numerical Mathematics*, 1972.
- "*Inference, estimation, and information processing, EE 378B lecture notes*," A. Montanari, Stanford University.
- "*COMS 4772 lecture notes*," D. Hsu, Columbia University.
- "*High-dimensional statistics: a non-asymptotic viewpoint*," M. Wainwright, *Cambridge University Press*, 2019.
- "*Community detection and stochastic block models*," E. Abbe, *Foundations and Trends in Communications and Information Theory*, 2018.

Reference

- "*Consistency thresholds for the planted bisection model*," E. Mossel, J. Neeman, A. Sly, *ACM Symposium on Theory of Computing*, 2015.
- "*Matrix completion from a few entries*," R. Keshavan, A. Montanari, S. Oh, *IEEE Transactions on Information Theory*, 2010.
- "*The PageRank citation ranking: bringing order to the web*," L. Page, S. Brin, R. Motwani, T. Winograd, 1999.
- "*Rank centrality: ranking from pairwise comparisons*," S. Negahban, S. Oh, D. Shah, *Operations Research*, 2016.
- "*Spectral method and regularized MLE are both optimal for top-K ranking*," Y. Chen, J. Fan, C. Ma, K. Wang, *Annals of Statistics*, 2019.