# Spectral methods: $\ell_2$ perturbation theory

Cong Ma

University of Chicago, Autumn 2021

# Matrix perturbation theory (spectral analysis)

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

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

**Goal of matrix perturbation theory:**

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

# 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 3.1**

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 3.2**

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

$$
\begin{aligned}
&|||\boldsymbol{AB}||| \le |||\boldsymbol{A}||| \cdot \|\boldsymbol{B}\|\,, && |||\boldsymbol{AB}||| \le |||\boldsymbol{B}||| \cdot \|\boldsymbol{A}\|\,; \\
&|||\boldsymbol{AB}||| \ge |||\boldsymbol{A}|||\,\sigma_{\min}(\boldsymbol{B})\,, && \textit{if } \boldsymbol{B} \textit{ is square}; \\
&|||\boldsymbol{AB}||| \ge |||\boldsymbol{B}|||\,\sigma_{\min}(\boldsymbol{A})\,, && \textit{if } \boldsymbol{A} \textit{ is square}.
\end{aligned}
$$

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

# Eigenvalue perturbation bounds

**Lemma 3.3 (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*

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

# Eigenvalue perturbation bounds

**Lemma 3.3 (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*

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

eigenvalues of real symmetric matrices are stable against perturbations

# Eigenvalue perturbation bounds

**Lemma 3.3 (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 3.4 (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*

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

singular values are stable against perturbations

# Proof of Lemma 3.4

We begin with introducing a useful "dilation" trick:

**Definition 3.5 (Symmetric dilation)**

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

$$\mathcal{S}(\boldsymbol{A}) = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{A} \\ \boldsymbol{A}^\top & \boldsymbol{0} \end{bmatrix} \in \mathbb{R}^{(n_1+n_2)\times(n_1+n_2)}.$$

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

$$\mathcal{S}(\boldsymbol{A}) = \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{U} & \boldsymbol{U} \\ \boldsymbol{V} & -\boldsymbol{V} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{0} \\ \boldsymbol{0} & -\boldsymbol{\Sigma} \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} \boldsymbol{U} & \boldsymbol{U} \\ \boldsymbol{V} & -\boldsymbol{V} \end{bmatrix}^\top .$$

Two observations: for $1 \leq i \leq \min\{m, n\}$, $\lambda_i(\mathcal{S}(\boldsymbol{A})) = \sigma_i(\boldsymbol{A})$, and $\|\mathcal{S}(\boldsymbol{A})\| = \|\boldsymbol{A}\|$. Apply Lemma 3.3 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: $[\boldsymbol{U}^\star, \boldsymbol{U}^\star_\perp]$, and $[\boldsymbol{U}, \boldsymbol{U}_\perp]$

# Question: how to measure distance?

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

# 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:

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

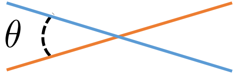
# Distance using projection matrices

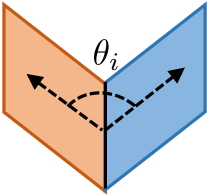
Key observation: projection matrix $\boldsymbol{U}\boldsymbol{U}^\top$ associated with subspace $\mathcal{U}$ is unique

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

# 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 $\sin \boldsymbol{\Theta}$ distance between subspaces as

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

where

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

# Link between projections and principal angles

**Lemma 3.6**

*The following identities are true:*

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

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

# Proof of Lemma 3.6

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

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

# Proof of Lemma 3.6 (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 3.7**

*The following identities are true:*

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

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

— *proof left as exercise*

# Summary of distance metrics

So far we have discussed

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

# Summary of distance metrics

So far we have discussed

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

Our choice of distance:

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

# Eigenspace perturbation theory

# Setup and notation

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

$$\boldsymbol{M}^\star = \sum_{i=1}^{n} \lambda_i^\star \boldsymbol{u}_i^\star \boldsymbol{u}_i^{\star\top} = \begin{bmatrix} \boldsymbol{U}^\star & \boldsymbol{U}_\perp^\star \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}^\star & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_\perp^\star \end{bmatrix} \begin{bmatrix} \boldsymbol{U}^{\star\top} \\ \boldsymbol{U}_\perp^{\star\top} \end{bmatrix};$$

$$\boldsymbol{M} = \sum_{i=1}^{n} \lambda_i \boldsymbol{u}_i \boldsymbol{u}_i^{\top} = \begin{bmatrix} \boldsymbol{U} & \boldsymbol{U}_\perp \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Lambda}_\perp \end{bmatrix} \begin{bmatrix} \boldsymbol{U}^{\top} \\ \boldsymbol{U}_\perp^{\top} \end{bmatrix}$$

# Setup and notation

$$
M = \Big[ \underbrace{u_1 \ \cdots \ u_r}_{U} \ \underbrace{u_{r+1} \ \cdots \ u_n}_{U_\perp} \Big]
\cdot \begin{bmatrix} \underbrace{\begin{matrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{matrix}}_{\Lambda} & \\ & \underbrace{\begin{matrix} \lambda_{r+1} & & \\ & \ddots & \\ & & \lambda_n \end{matrix}}_{\Lambda_\perp} \end{bmatrix}
\begin{bmatrix} u_1^\top \\ \vdots \\ u_r^\top \\ u_{r+1}^\top \\ \vdots \\ u_n^\top \end{bmatrix}
\begin{matrix} \left.\vphantom{\begin{matrix} u_1^\top \\ \vdots \\ u_r^\top \end{matrix}}\right\} U^\top \\ \left.\vphantom{\begin{matrix} u_{r+1}^\top \\ \vdots \\ u_n^\top \end{matrix}}\right\} U_\perp^\top \end{matrix}
$$

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

Chandler Davis

William Kahan

**Theorem 3.8 (Davis-Kahan's $\sin\Theta$ theorem: simple version)**

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

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

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

# 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

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

Remarks:

- Eigen-gap $\lambda_r(\boldsymbol{M}^\star) = \lambda_r(\boldsymbol{M}^\star) - \lambda_{r+1}(\boldsymbol{M}^\star)$
- Perturbation size $\|\boldsymbol{E}\|$
- Signal-to-noise (SNR) ratio $\frac{\lambda_r(\boldsymbol{M}^\star)}{\|\boldsymbol{E}\|}$
- $\|\boldsymbol{E}\boldsymbol{U}^\star\|$ is sometimes useful; we will see benefit later
- Necessity of $\|\boldsymbol{E}\| \lesssim \lambda_r(\boldsymbol{M}^\star)$

# What happens when SNR is small?

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

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

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

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

— *large regardless of size of $\epsilon$ or size of the perturbation $\|\boldsymbol{E}\|$*

# Proof of Theorem 3.8

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

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

which together with triangle inequality implies

$$\begin{aligned} \|\!|\!|\boldsymbol{U}_\perp^\top \boldsymbol{E}\boldsymbol{U}^\star|\!|\!\| &\geq \|\!|\!|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star \boldsymbol{\Lambda}^\star|\!|\!\| - \|\!|\!|\boldsymbol{\Lambda}_\perp \boldsymbol{U}_\perp^\top \boldsymbol{U}^\star|\!|\!\| \\ &\geq \sigma_{\min}(\boldsymbol{\Lambda}^\star)\|\!|\!|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star|\!|\!\| - \|\boldsymbol{\Lambda}_\perp\| \cdot \|\!|\!|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star|\!|\!\| \end{aligned} \tag{3.6}$$

In view of Weyl's inequality, one has $\|\boldsymbol{\Lambda}_\perp\| \leq \|\boldsymbol{E}\|$. In addition, we have $\sigma_{\min}(\boldsymbol{\Lambda}^\star) = \lambda_r(\boldsymbol{M}^\star)$. These combined with relation (3.6) 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 3.6-3.7 completes the proof

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

— eigenvalues($\boldsymbol{A}$): *set of eigenvalues of* $\boldsymbol{A}$

**Theorem 3.9 (Davis-Kahan's sinΘ theorem: general version)**

*Assume that*

$$\text{eigenvalues}(\boldsymbol{\Lambda}^\star) \subseteq (-\infty, \alpha - \Delta] \cup [\beta + \Delta, \infty); \tag{3.7a}$$

$$\text{eigenvalues}(\boldsymbol{\Lambda}_\perp) \subseteq [\alpha, \beta]. \tag{3.7b}$$

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

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

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

— *conclusion remains valid if Assumption* (3.7) *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 $n_1 \le n_2$), whose SVDs are given respectively by

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

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

- $\sigma_1 \ge \cdots \ge \sigma_{n_1}$ (resp. $\sigma_1^\star \ge \cdots \ge \sigma_{n_1}^\star$) 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 $\sin \boldsymbol{\Theta}$ theorem

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

**Theorem 3.10 (Wedin's sinΘ theorem)**

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

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

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

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

# Proof of Theorem 3.10

Similar to proof of Davis-Kahan theorem, we concentrate on $\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star$

$$
\begin{aligned}
\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star &= \boldsymbol{U}_\perp^\top (\boldsymbol{U}^\star \boldsymbol{\Sigma}^\star \boldsymbol{V}^{\star\top}) \boldsymbol{V}^\star \boldsymbol{\Sigma}^{\star -1} \\
&= \boldsymbol{U}_\perp^\top \left( \boldsymbol{M} - \boldsymbol{E} - \boldsymbol{U}_\perp^\star \boldsymbol{\Sigma}_\perp^\star \boldsymbol{V}_\perp^{\star\top} \right) \boldsymbol{V}^\star \boldsymbol{\Sigma}^{\star -1} \\
&= \boldsymbol{U}_\perp^\top \left( \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{V}^\top + \boldsymbol{U}_\perp \boldsymbol{\Sigma}_\perp \boldsymbol{V}_\perp^\top - \boldsymbol{E} - \boldsymbol{U}_\perp^\star \boldsymbol{\Sigma}_\perp^\star \boldsymbol{V}_\perp^{\star\top} \right) \boldsymbol{V}^\star \boldsymbol{\Sigma}^{\star -1} \\
&= \boldsymbol{\Sigma}_\perp \boldsymbol{V}_\perp^\top \boldsymbol{V}^\star \boldsymbol{\Sigma}^{\star -1} - \boldsymbol{U}_\perp^\top \boldsymbol{E} \boldsymbol{V}^\star \boldsymbol{\Sigma}^{\star -1}. \qquad (3.9)
\end{aligned}
$$

Applying triangle inequality and Lemma 3.2 to identity (3.9) yields

$$
\begin{aligned}
\left\|\left\| \boldsymbol{U}_\perp^\top \boldsymbol{U}^\star \right\|\right\| &\leq \|\boldsymbol{\Sigma}_\perp\| \cdot \left\|\left\| \boldsymbol{V}_\perp^\top \boldsymbol{V}^\star \right\|\right\| \cdot \|\boldsymbol{\Sigma}^{\star -1}\| + \|\boldsymbol{U}_\perp^\top\| \cdot \left\|\left\| \boldsymbol{E} \boldsymbol{V}^\star \right\|\right\| \cdot \|\boldsymbol{\Sigma}^{\star -1}\| \\
&= \sigma_{r+1} \cdot \left\|\left\| \boldsymbol{V}_\perp^\top \boldsymbol{V}^\star \right\|\right\| \cdot \frac{1}{\sigma_r^\star} + \left\|\left\| \boldsymbol{E} \boldsymbol{V}^\star \right\|\right\| \cdot \frac{1}{\sigma_r^\star} \\
&\leq \frac{\sigma_{r+1}^\star + \|\boldsymbol{E}\|}{\sigma_r^\star} \left\|\left\| \boldsymbol{V}_\perp^\top \boldsymbol{V}^\star \right\|\right\| + \frac{\left\|\left\| \boldsymbol{E} \boldsymbol{V}^\star \right\|\right\|}{\sigma_r^\star} \qquad (3.10)
\end{aligned}
$$

# Proof of Theorem 3.10 (cont.)

Repeating the same argument yields

$$\|\boldsymbol{V}_\perp^\top \boldsymbol{V}^\star\| \le \frac{\|\boldsymbol{E}^\top \boldsymbol{U}^\star\|}{\sigma_r^\star} + \frac{\sigma_{r+1}^\star + \|\boldsymbol{E}\|}{\sigma_r^\star} \|\boldsymbol{U}_\perp^\top \boldsymbol{U}^\star\| \tag{3.11}$$

To finish up, combine inequalities (3.10) and (3.11) to obtain

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

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

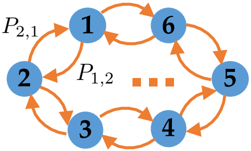
# Eigen-decomposition for asymmetric matrices

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 $\boldsymbol{P}$ is probability transition matrix

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

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

- $\boldsymbol{\pi}$ is in fact left eigenvector of $\boldsymbol{P}$ with eigenvalue $1$
- 1 is largest eigenvalue of $\boldsymbol{P}$ in absolute sense: $|\lambda_i(\boldsymbol{P})| \le 1$ 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 $\boldsymbol{P}^\star$ of reversible Markov chain
- Perturbed transition matrix $\boldsymbol{P} = \boldsymbol{P}^\star + \boldsymbol{E}$
- $\boldsymbol{\pi}^\star$, $\boldsymbol{\pi}$ are leading left eigenvectors of $\boldsymbol{P}^\star$, $\boldsymbol{P}$, respectively
- Question: how does $\boldsymbol{E}$ affect perturbation $\boldsymbol{\pi} - \boldsymbol{\pi}^\star$

# New norms

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

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

# Eigenvector perturbation for transition matrices

**Theorem 3.11 (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*

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

- 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: $\left\|\boldsymbol{\pi}^{\star\top}\boldsymbol{E}\right\|_{\boldsymbol{\pi}^\star}$

# Proof of Theorem 3.11

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 $\boldsymbol{\pi} - \boldsymbol{\pi}^\star$

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

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

# Proof of Theorem 3.11 (cont.)

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

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

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

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

Proof will be complete if one can show

$$
\left\|\boldsymbol{P}^\star-\boldsymbol{1}\boldsymbol{\pi}^{\star\top}\right\|_{\boldsymbol{\pi}^\star}=\max\left\{\lambda_2(\boldsymbol{P}^\star),-\lambda_n(\boldsymbol{P}^\star)\right\} \qquad (3.12)
$$

# **Proof of identity** (3.12)

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

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

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

As a consequence, one has

$$\begin{aligned}\|\boldsymbol{P}^\star - \boldsymbol{1}\boldsymbol{\pi}^{\star\top}\|_{\boldsymbol{\pi}^\star} &= \|\left(\mathbf{\Pi}^\star\right)^{1/2}\left(\boldsymbol{P}^\star - \boldsymbol{1}\boldsymbol{\pi}^{\star\top}\right)\left(\mathbf{\Pi}^\star\right)^{-1/2}\| \\ &= \|\boldsymbol{S}^\star - \boldsymbol{\pi}^\star_{1/2}(\boldsymbol{\pi}^\star_{1/2})^\top\|\end{aligned}$$

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

# Proof of identity (3.12) (cont.)

Several properties of $\boldsymbol{S}^\star$:

- Symmetric: all eigenvalues are real

  — *check detailed balance*

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

$$\boldsymbol{S}^\star \boldsymbol{\pi}^\star_{1/2} = \boldsymbol{\pi}^\star_{1/2}$$

- Eigenvalues of $\boldsymbol{S}^\star - \boldsymbol{\pi}^\star_{1/2}(\boldsymbol{\pi}^\star_{1/2})^\top$ are $0, \lambda_2(\boldsymbol{S}^\star), \dots, \lambda_n(\boldsymbol{S}^\star)$

Combine all to see

$$\begin{aligned}
\big\|\boldsymbol{S}^\star - \boldsymbol{\pi}^\star_{1/2}(\boldsymbol{\pi}^\star_{1/2})^\top\big\| &\overset{\text{(i)}}{=} \max\big\{|\lambda_2(\boldsymbol{S}^\star)|, |\lambda_n(\boldsymbol{S}^\star)|\big\} \\
&= \max\big\{\lambda_2(\boldsymbol{S}^\star), -\lambda_n(\boldsymbol{S}^\star)\big\} \overset{\text{(ii)}}{=} \max\big\{\lambda_2(\boldsymbol{P}^\star), -\lambda_n(\boldsymbol{P}^\star)\big\}.
\end{aligned}$$