STAT 37797: Mathematics of Data Science — Autumn 2021

Homework 1 Solutions

Please do not distribute.

1. Weyl's inequality (*20 points*)

a.(*10 points*) Let $\boldsymbol{A}$ be an $n \times n$ real symmetric matrix, with eigenvalues $\lambda_1(\boldsymbol{A}) \geq \lambda_2(\boldsymbol{A}) \geq \cdots \geq \lambda_n(\boldsymbol{A})$. Then for each $1 \leq i \leq n$, prove the following variational representation of eigenvalues

$$\lambda_i(\boldsymbol{A}) = \sup_{V:\dim(V)=i} \inf_{\boldsymbol{v}\in V:\|\boldsymbol{v}\|_2=1} \boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v}.$$

In the above notation, V is a subspace in $\mathbb{R}^n$, and $\dim(V) = i$ means V is an i-dimensional subspace.

Solution: *Let $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top$ be the eigen-decomposition of $\boldsymbol{A}$, where $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n]$ and $\boldsymbol{\Lambda} = diag(\lambda_1(\boldsymbol{A}), \lambda_2(\boldsymbol{A}), \cdots, \lambda_n(\boldsymbol{A}))$. Pick V to be the subpace spaned by the top-i eigenvectors $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_i\}$. Then every $\boldsymbol{v} \in V$ has the following decomposition*

$$\boldsymbol{v} = \sum_{k=1}^{i} \alpha_k \boldsymbol{u}_k.$$

As a consequence, we have

$$\begin{aligned}\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} &= \boldsymbol{v}^\top \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\top \boldsymbol{v} \\ &= \sum_{k=1}^{i} \alpha_k^2 \lambda_k(\boldsymbol{A}),\end{aligned}$$

which in turn leads to

$$\inf_{\boldsymbol{v}\in V:\|\boldsymbol{v}\|_2=1} \boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} = \inf_{\|\boldsymbol{\alpha}\|_2=1} \sum_{k=1}^{i} \alpha_k^2 \lambda_k(\boldsymbol{A}) = \lambda_i(\boldsymbol{A}).$$

Therefore we obtain

$$\lambda_i(\boldsymbol{A}) \leq \sup_{V:dim(V)=i} \inf_{\boldsymbol{v}\in V:\|\boldsymbol{v}\|_2=1} \boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v}.$$

Now we move on to show the reverse inequality, i.e. for every V with dimension i, we can find some $\boldsymbol{v} \in V, \|\boldsymbol{v}\|_2 = 1$ such that $\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} \leq \lambda_i(\boldsymbol{A})$. Let W be a space spanned by $\{\boldsymbol{u}_i, \boldsymbol{u}_{i+1}, \cdots, \boldsymbol{u}_n\}$ which has dimension $n-i+1$ and codimension $i-1$. In light of this, it must have a nontrivial intersection with V. Let $v \in V \cap W$ and $v = \sum_{k=i}^{n} \alpha_k \boldsymbol{u}_k$. Then one has

$$\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} = \sum_{k=i}^{n} \alpha_k^2 \lambda_k(\boldsymbol{A}) \leq \lambda_i(\boldsymbol{A}).$$

This completes the proof.

b.(*10 points*) Prove that: if $\boldsymbol{A}$ and $\boldsymbol{B}$ are both real and symmetric matrices, then

$$|\lambda_i(\boldsymbol{A}) - \lambda_i(\boldsymbol{B})| \leq \|\boldsymbol{A} - \boldsymbol{B}\|, \qquad \text{for all } 1 \leq i \leq n,$$

where $\|\cdot\|$ denotes the spectral norm.

Solution: *By (a), one sees that*

$$\begin{aligned}\lambda_i(\boldsymbol{A}) &= \sup_{V:dim(V)=i} \inf_{\boldsymbol{v}\in V:\|\boldsymbol{v}\|_2=1} \boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} \\ &= \sup_{V:dim(V)=i} \inf_{\boldsymbol{v}\in V:\|\boldsymbol{v}\|_2=1} \boldsymbol{v}^\top \boldsymbol{B}\boldsymbol{v} + \boldsymbol{v}^\top (\boldsymbol{A}-\boldsymbol{B})\boldsymbol{v}.\end{aligned}$$

In light of the fact that

$$\left|\boldsymbol{v}^\top (\boldsymbol{A}-\boldsymbol{B})\boldsymbol{v}\right| \le \|\boldsymbol{A}-\boldsymbol{B}\|,$$

we have

$$\begin{aligned}\lambda_i(\boldsymbol{A}) &\le \sup_{V:dim(V)=i} \inf_{\boldsymbol{v}\in V:\|\boldsymbol{v}\|_2=1} \boldsymbol{v}^\top \boldsymbol{B}\boldsymbol{v} + \|\boldsymbol{A}-\boldsymbol{B}\| \\ &= \lambda_i(\boldsymbol{B}) + \|\boldsymbol{A}-\boldsymbol{B}\|\end{aligned}$$

and similarly

$$\begin{aligned}\lambda_i(\boldsymbol{A}) &\ge \sup_{V:dim(V)=i} \inf_{\boldsymbol{v}\in V:\|\boldsymbol{v}\|_2=1} \boldsymbol{v}^\top \boldsymbol{B}\boldsymbol{v} - \|\boldsymbol{A}-\boldsymbol{B}\| \\ &= \lambda_i(\boldsymbol{B}) - \|\boldsymbol{A}-\boldsymbol{B}\|.\end{aligned}$$

Combining the above two inequalities yields the desired result.

2. Distance metrics for subspaces (*20 points***)** Consider two orthonormal matrices $\boldsymbol{U}, \boldsymbol{U}^\star \in \mathbb{R}^{n\times r}$, satisfying $\boldsymbol{U}^\top\boldsymbol{U} = \boldsymbol{U}^{\star\top}\boldsymbol{U}^\star = \boldsymbol{I}_r$ with $r < n$. We have discussed extensively the distance using projection matrices

$$\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\|, \quad \text{and} \quad \|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\|_{\mathrm{F}}.$$

Also, our default choice of distance is the one using optimal rotation matrix:

$$\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \left\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\right\|, \quad \text{and} \quad \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{\mathrm{F}}.$$

Here $\mathbb{O}^{r\times r} := \{\boldsymbol{R}\in\mathbb{R}^{r\times r} \mid \boldsymbol{R}\boldsymbol{R}^\top = \boldsymbol{R}^\top\boldsymbol{R} = \boldsymbol{I}_r\}$ is the set of all $r\times r$ orthonormal matrices.

a.(*10 points*) Show that

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

Solution:

As before, suppose that the SVD of $\boldsymbol{U}^\top\boldsymbol{U}^\star$ is given by $\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}^\top$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are $r\times r$ orthonormal matrices whose columns contain the left singular vectors and the right singular vectors of $\boldsymbol{U}^\top\boldsymbol{U}^\star$, respectively, and $\boldsymbol{\Sigma}\in\mathbb{R}^{r\times r} = \cos\boldsymbol{\Theta}$ is a diagonal matrix whose diagonal entries correspond to the singular values of $\boldsymbol{U}^\top\boldsymbol{U}^\star$.

The spectral norm upper bound. *We first observe that*

$$\begin{aligned}\|\boldsymbol{U}\boldsymbol{X}\boldsymbol{Y}^\top - \boldsymbol{U}^\star\|^2 &= \|(\boldsymbol{U}\boldsymbol{X}\boldsymbol{Y}^\top - \boldsymbol{U}^\star)^\top(\boldsymbol{U}\boldsymbol{X}\boldsymbol{Y}^\top - \boldsymbol{U}^\star)\| \\ &= \|2\boldsymbol{I}_r - \boldsymbol{Y}\boldsymbol{X}^\top\boldsymbol{U}^\top\boldsymbol{U}^\star - \boldsymbol{U}^{\star\top}\boldsymbol{U}\boldsymbol{X}\boldsymbol{Y}^\top\| \\ &= \|2\boldsymbol{I}_r - \boldsymbol{Y}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}^\top - \boldsymbol{Y}\boldsymbol{\Sigma}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{Y}^\top\| \\ &= 2\|\boldsymbol{Y}(\boldsymbol{I}_r - \boldsymbol{\Sigma})\boldsymbol{Y}^\top\| = 2\|\boldsymbol{I}_r - \boldsymbol{\Sigma}\|.\end{aligned} \tag{1}$$

Here, the penultimate line relies on the singular value decomposition $\boldsymbol{U}^\top\boldsymbol{U}^\star = \boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}^\top$, while the two identities in the last line result from the orthonormality of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. In addition, note that

$$\begin{aligned}\|\boldsymbol{I}_r - \boldsymbol{\Sigma}\| &= \|\boldsymbol{I}_r - \cos\boldsymbol{\Theta}\| \le \|\boldsymbol{I}_r - \cos^2\boldsymbol{\Theta}\| \\ &= \|\sin^2\boldsymbol{\Theta}\| = \|\sin\boldsymbol{\Theta}\|^2.\end{aligned}$$

This taken together with (1) *leads to*

$$\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \big\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\big\| \le \big\|\boldsymbol{U}\boldsymbol{X}\boldsymbol{Y}^\top - \boldsymbol{U}^\star\big\| \le \sqrt{2}\|\sin\boldsymbol{\Theta}\|,$$

where the first inequality holds since $\boldsymbol{X}$ and $\boldsymbol{Y}$ are both orthonormal matrices and hence $\boldsymbol{X}\boldsymbol{Y}^\top$ is also orthonormal.

The spectral norm lower bound. *On the other hand, we make the observation that*

$$\begin{aligned}\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \big\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\big\|^2 &= \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \big\|(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star)^\top(\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star)\big\| \\ &= \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \big\|\boldsymbol{R}^\top\boldsymbol{U}^\top\boldsymbol{U}\boldsymbol{R} + \boldsymbol{U}^{\star\top}\boldsymbol{U}^\star - \boldsymbol{R}^\top\boldsymbol{U}^\top\boldsymbol{U}^\star - \boldsymbol{U}^{\star\top}\boldsymbol{U}\boldsymbol{R}\big\| \\ &= \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \big\|2\boldsymbol{I}_r - \boldsymbol{R}^\top\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}^\top - \boldsymbol{Y}\boldsymbol{\Sigma}\boldsymbol{X}^\top\boldsymbol{R}\big\|, \end{aligned} \tag{2}$$

where the last relation holds since $\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}^\top$ is the SVD of $\boldsymbol{U}^\top\boldsymbol{U}^\star$. Continue the derivation to obtain

$$\begin{aligned}(2) &\overset{\text{(i)}}{=} \min_{\boldsymbol{Q}\in\mathcal{O}^{r\times r}} \big\|2\boldsymbol{I}_r - \boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Y}^\top - \boldsymbol{Y}\boldsymbol{\Sigma}\boldsymbol{Q}^\top\big\| \\ &\overset{\text{(ii)}}{=} \min_{\boldsymbol{Q}\in\mathcal{O}^{r\times r}} \big\|2\boldsymbol{Q}^\top\boldsymbol{Q} - \boldsymbol{Q}^\top\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Y}^\top\boldsymbol{Q} - \boldsymbol{Q}^\top\boldsymbol{Y}\boldsymbol{\Sigma}\boldsymbol{Q}^\top\boldsymbol{Q}\big\| \\ &= \min_{\boldsymbol{Q}\in\mathcal{O}^{r\times r}} \big\|2\boldsymbol{I}_r - \boldsymbol{\Sigma}\boldsymbol{Y}^\top\boldsymbol{Q} - \boldsymbol{Q}^\top\boldsymbol{Y}\boldsymbol{\Sigma}\big\| \\ &\overset{\text{(iii)}}{=} \min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}} \big\|2\boldsymbol{I}_r - \boldsymbol{\Sigma}\boldsymbol{O} - \boldsymbol{O}^\top\boldsymbol{\Sigma}\big\|. \end{aligned} \tag{3}$$

Here, (i) follows by setting $\boldsymbol{Q} = \boldsymbol{R}^\top\boldsymbol{X}$ (since both $\boldsymbol{X}$ and $\boldsymbol{R}$ are orthonormal matrices), (ii) results from the unitary invariance of the spectral norm, whereas (iii) holds by setting $\boldsymbol{O} = \boldsymbol{Y}^\top\boldsymbol{Q}$. Moreover, recognizing that $\|\boldsymbol{\Sigma}\boldsymbol{O}\| \le \|\boldsymbol{\Sigma}\|\cdot\|\boldsymbol{O}\| \le 1$ (and hence $2\boldsymbol{I}_r - \boldsymbol{\Sigma}\boldsymbol{O} - \boldsymbol{O}^\top\boldsymbol{\Sigma} \succeq \boldsymbol{0}$), one can obtain

$$\begin{aligned}\min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}} \big\|2\boldsymbol{I}_r - \boldsymbol{\Sigma}\boldsymbol{O} - \boldsymbol{O}^\top\boldsymbol{\Sigma}\big\| &= \min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}} \lambda_{\max}\big(2\boldsymbol{I}_r - \boldsymbol{\Sigma}\boldsymbol{O} - \boldsymbol{O}^\top\boldsymbol{\Sigma}\big) \\ &= \min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}} \max_{\boldsymbol{u}:\|\boldsymbol{u}\|_2=1} \boldsymbol{u}^\top\big(2\boldsymbol{I}_r - \boldsymbol{\Sigma}\boldsymbol{O} - \boldsymbol{O}^\top\boldsymbol{\Sigma}\big)\boldsymbol{u} \\ &= \min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}} \max_{\boldsymbol{u}:\|\boldsymbol{u}\|_2=1} \big(2 - 2\boldsymbol{u}^\top\boldsymbol{\Sigma}\boldsymbol{O}\boldsymbol{u}\big) \\ &\ge \min_{\boldsymbol{O}\in\mathcal{O}^{r\times r}} \big(2 - 2\boldsymbol{e}_r^\top\boldsymbol{\Sigma}\boldsymbol{O}\boldsymbol{e}_r\big) \\ &= 2 - 2\cos\theta_r \max_{\boldsymbol{O}\in\mathcal{O}^{r\times r}} \boldsymbol{e}_r^\top\boldsymbol{O}\boldsymbol{e}_r \\ &\ge 2 - 2\cos\theta_r = 4\sin^2(\theta_r/2). \end{aligned} \tag{4}$$

Here, the inequality follows by taking $\boldsymbol{u}$ to be $\boldsymbol{e}_r$ (recall that by construction, $\sigma_r = \cos\theta_r \ge 0$ is the smallest singular value of $\boldsymbol{\Sigma}$), and the penultimate line holds by combining the facts $|\boldsymbol{e}_r^\top\boldsymbol{O}\boldsymbol{e}_r| \le \|\boldsymbol{O}\| = 1$ and $\boldsymbol{e}_r^\top\boldsymbol{e}_r = 1$. Putting (4) and (3) together yields

$$\begin{aligned}\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}} \big\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\big\| &\ge \sqrt{4\sin^2(\theta_r/2)} = 2\sin(\theta_r/2) = \|2\sin(\boldsymbol{\Theta}/2)\| \\ &\ge \|\sin\boldsymbol{\Theta}\|,\end{aligned}$$

where we again use the inequality $2\sin(\theta/2) \geq \sin\theta$ *for all* $\theta \in [0, \pi/2]$.

Finally, invoking the relation $\|\sin\boldsymbol{\Theta}\| = \|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\|$ *establishes the claimed spectral norm bounds.*

b.(*10 points*) Show that

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

Solution:

The Frobenius norm upper bound. *Regarding the Frobenius norm upper bound, one sees that*

$$\begin{aligned}\big\|\boldsymbol{U}\boldsymbol{X}\boldsymbol{Y}^\top - \boldsymbol{U}^\star\big\|_{\mathrm{F}}^2 &= \|\boldsymbol{U}\|_{\mathrm{F}}^2 + \big\|\boldsymbol{U}^\star\big\|_{\mathrm{F}}^2 - 2\mathsf{Tr}\big(\boldsymbol{Y}\boldsymbol{X}^\top\boldsymbol{U}^\top\boldsymbol{U}^\star\big)\\ &\overset{\text{(i)}}{=} r + r - 2\mathsf{Tr}\big(\boldsymbol{Y}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}^\top\big) \overset{\text{(ii)}}{=} 2r - 2\mathsf{Tr}\left(\boldsymbol{\Sigma}\right), \end{aligned} \tag{5}$$

where (i) holds since $\boldsymbol{U}$ *and* $\boldsymbol{U}^\star$ *are both* $n \times r$ *matrices with orthonormal columns, and (ii) follows since* $\boldsymbol{X}^\top\boldsymbol{X} = \boldsymbol{Y}^\top\boldsymbol{Y} = \boldsymbol{I}$ *(and hence* $\mathsf{Tr}(\boldsymbol{Y}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}^\top) = \mathsf{Tr}(\boldsymbol{Y}^\top\boldsymbol{Y}\boldsymbol{X}^\top\boldsymbol{X}\boldsymbol{\Sigma}) = \mathsf{Tr}(\boldsymbol{\Sigma})$*). Furthermore,*

$$\begin{aligned}2r - 2\mathsf{Tr}\left(\boldsymbol{\Sigma}\right) &\overset{\text{(iii)}}{=} 2\sum\nolimits_i(1-\cos\theta_i) \leq 2\sum\nolimits_i(1-\cos^2\theta_i)\\ &= 2\left\|\sin\boldsymbol{\Theta}\right\|_{\mathrm{F}}^2 = \big\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\big\|_{\mathrm{F}}^2,\end{aligned}$$

where (iii) holds by construction, and the last identity results from the lemma in class. This taken collectively with (5) *reveals that*

$$\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\|_{\mathrm{F}}^2 \leq \big\|\boldsymbol{U}\boldsymbol{X}\boldsymbol{Y}^\top - \boldsymbol{U}^\star\big\|_{\mathrm{F}}^2 \leq \big\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\big\|_{\mathrm{F}}^2,$$

where the first inequality holds since $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *are both orthonormal matrices and hence* $\boldsymbol{X}\boldsymbol{Y}^\top$ *is also orthonormal.*

The Frobenius norm lower bound. *With regards to the Frobenius norm lower bound, it is seen that*

$$\begin{aligned}\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\big\|\boldsymbol{U}\boldsymbol{R} - \boldsymbol{U}^\star\big\|_{\mathrm{F}}^2 &= \min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\Big\{\|\boldsymbol{U}\boldsymbol{R}\|_{\mathrm{F}}^2 + \|\boldsymbol{U}^\star\|_{\mathrm{F}}^2 - 2\big\langle\boldsymbol{U}\boldsymbol{R}, \boldsymbol{U}^\star\big\rangle\Big\}\\ &\overset{\text{(i)}}{=} 2\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\Big\{r - \big\langle\boldsymbol{R}, \boldsymbol{U}^\top\boldsymbol{U}^\star\big\rangle\Big\}\\ &\overset{\text{(ii)}}{=} 2\min_{\boldsymbol{R}\in\mathcal{O}^{r\times r}}\Big\{r - \big\langle\boldsymbol{R}, \boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}^\top\big\rangle\Big\},\end{aligned} \tag{6}$$

where (i) holds since $\|\boldsymbol{U}\|_{\mathrm{F}} = \|\boldsymbol{U}^\star\|_{\mathrm{F}} = \sqrt{r}$*, and (ii) relies on the SVD* $\boldsymbol{X}\boldsymbol{\Sigma}\boldsymbol{Y}^\top$ *of* $\boldsymbol{U}^\top\boldsymbol{U}^\star$*. Continue the derivation to obtain*

$$\begin{aligned}(6) &\overset{\text{(iii)}}{=} 2\min_{\boldsymbol{Q}\in\mathcal{O}^{r\times r}}\Big\{r - \big\langle\boldsymbol{Q}, \cos\boldsymbol{\Theta}\big\rangle\Big\} \overset{\text{(iv)}}{\geq} 2\min_{\boldsymbol{Q}\in\mathcal{O}^{r\times r}}\Big\{r - \|\boldsymbol{Q}\|\,\|\cos\boldsymbol{\Theta}\|_*\Big\}\\ &= 2\big(r - \sum\nolimits_i\cos\theta_i\big).\end{aligned} \tag{7}$$

Here, (iii) sets $\boldsymbol{Q} = \boldsymbol{X}^\top\boldsymbol{R}\boldsymbol{Y}$ *and identifies* $\boldsymbol{\Sigma}$ *as* $\cos\boldsymbol{\Theta}$*, (iv) comes from the elementary inequality* $\langle\boldsymbol{A}, \boldsymbol{B}\rangle \leq \|\boldsymbol{A}\|\,\|\boldsymbol{B}\|_*$*, whereas the last line follows since* $\cos\theta_i \geq 0$*. Additionally, it is easily seen that*

$$\begin{aligned}(7) &= 2\sum\nolimits_i(1-\cos\theta_i) = 4\sum\nolimits_i\sin^2(\theta_i/2)\\ &\geq \sum\nolimits_i\sin^2\theta_i = \frac{1}{2}\big\|\boldsymbol{U}\boldsymbol{U}^\top - \boldsymbol{U}^\star\boldsymbol{U}^{\star\top}\big\|_{\mathrm{F}}^2,\end{aligned} \tag{8}$$

where the penultimate relation follows from the elementary inequality $2\sin(\theta/2) \geq \sin\theta$ (which holds for any $0 \leq \theta \leq \pi/2$). Combining the inequalities (7) and (8), we establish the claimed lower bound.

3. Variant of Wedin's theorem (*10 points*) Wedin's $\sin\boldsymbol{\Theta}$ theorem tells us that if $\|\boldsymbol{E}\| < \sigma_r^\star - \sigma_{r+1}^\star$, then there exist two orthonormal matrices $\boldsymbol{R_U}, \boldsymbol{R_U} \in \mathbb{R}^{r\times r}$ such that

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

However, in some cases, we hope for a single rotation matrix that could align both $(\boldsymbol{U}, \boldsymbol{U}^\star)$ and $(\boldsymbol{V}, \boldsymbol{V}^\star)$. It turns out that this is achievable. Show that if $\|\boldsymbol{E}\| < \sigma_r^\star - \sigma_{r+1}^\star$, there exists a single orthonormal matrix $\boldsymbol{R} \in \mathcal{O}^{r\times r}$ such that

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

You are allowed to invoke the general Davis-Kahan $\sin\boldsymbol{\Theta}$ theorem given in the class.

Solution: *Apply Davis-Kahan to the symmetric dilation of $\boldsymbol{M}$ and $\boldsymbol{M}^\star$.*

4. Quadratic systems of equations (*10 points*) Suppose that our goal is to estimate an unknown vector $\boldsymbol{x}^\star \in \mathbb{R}^n$ (obeying $\|\boldsymbol{x}\|_2 = 1$) based on m i.i.d. samples of the form

$$y_i = (\boldsymbol{a}_i^\top \boldsymbol{x}^\star)^2, \qquad i = 1, \ldots, m,$$

where $\boldsymbol{a}_i \in \mathbb{R}^n$ are independent vectors (known *a priori*) obeying $\boldsymbol{a}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$.

Suggest a spectral method for estimating $\boldsymbol{x}^\star$ that is consistent with either $\boldsymbol{x}^\star$ or $-\boldsymbol{x}^\star$ in the limit of infinite data, i.e., as m goes to infinity.

Solution: *Construct a surrogate matrix*

$$\boldsymbol{Y} = \frac{1}{m}\sum_{i=1}^m y_i \boldsymbol{a}_i\boldsymbol{a}_i^\top = \frac{1}{m}\sum_{i=1}^m (\boldsymbol{a}_i^\top\boldsymbol{x})^2\boldsymbol{a}_i\boldsymbol{a}_i^\top.$$

Then compute the leading eigenvalue $\boldsymbol{u}$ of $\boldsymbol{Y}$. When $m \to \infty$,

$$\boldsymbol{Y} \to \mathbb{E}[\boldsymbol{Y}] = \|\boldsymbol{x}\|_2^2\boldsymbol{I} + 2\boldsymbol{x}\boldsymbol{x}^\top = \boldsymbol{I} + 2\boldsymbol{x}\boldsymbol{x}^\top,$$

whose leading eigenvector is exactly $\pm\boldsymbol{x}$.

5. Matrix completion (*20 points*) Suppose the ground-truth matrix is

$$\boldsymbol{M}^\star = \boldsymbol{u}^\star\boldsymbol{v}^{\star\top} \in \mathbb{R}^{n\times n},$$

where $\boldsymbol{u}^\star = \tilde{\boldsymbol{u}}/\|\tilde{\boldsymbol{u}}\|_2$ and $\boldsymbol{v}^\star = \tilde{\boldsymbol{v}}/\|\tilde{\boldsymbol{v}}\|_2$, with $\tilde{\boldsymbol{u}}, \tilde{\boldsymbol{v}} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_n)$ generated independently. Each entry of $\boldsymbol{M}^\star = [M_{i,j}^\star]_{1\leq i,j\leq n}$ is observed independently with probability p. In the lectures, we have constructed a matrix $\boldsymbol{M} = [M_{i,j}]_{1\leq i,j\leq n}$, where

$$M_{i,j} = \begin{cases} \frac{1}{p}M_{i,j}^\star, & \text{if } M_{i,j}^\star \text{ is observed;} \\ 0, & \text{else.} \end{cases}$$

We have shown in class that with high probability, the leading left singular vector $\boldsymbol{u}$ of $\boldsymbol{M}$ is a reliable estimate of $\boldsymbol{u}^\star$, provided that $p \gg \frac{\log^3 n}{n}$.

Now, consider a new matrix $\boldsymbol{M}^{(1)} = [M_{i,j}^{(1)}]_{1\le i,j\le n}$ obtained by zeroing out the 1st column and 1st row of $\boldsymbol{M}$. More precisely, for any $1 \le i, j \le n$,

$$M_{i,j}^{(1)} = \begin{cases} M_{i,j}, & \text{if } i \neq 1 \text{ and } j \neq 1; \\ 0, & \text{else.} \end{cases}$$

Let $\boldsymbol{u}^{(1)}$ (resp. $\boldsymbol{v}^{(1)}$) be the leading left (resp. right) singular vector of $\boldsymbol{M}^{(1)}$.

a.**(***10 points***)** Recall that Wedin's $\sin\boldsymbol{\Theta}$ Theorem states that: for any two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, their leading left singular vectors (denoted by $\boldsymbol{u}_A$ and $\boldsymbol{u}_B$ respectively) satisfy

$$\mathsf{dist}(\boldsymbol{u}_A, \boldsymbol{u}_B) \le \frac{\|\boldsymbol{A}-\boldsymbol{B}\|}{\sigma_1(\boldsymbol{A}) - \sigma_2(\boldsymbol{A}) - \|\boldsymbol{A}-\boldsymbol{B}\|}.$$

Use it to derive an upper bound on $\mathsf{dist}(\boldsymbol{u}^{(1)}, \boldsymbol{u})$ in terms of n and p.

Solution:

To begin with, using Matlab notation we have

$$\begin{aligned} \sigma_1(\boldsymbol{M}^{(1)}) &\ge \sigma_1(\boldsymbol{M}^\star) - \|\boldsymbol{M}^\star - \boldsymbol{M}^{(1)}\| \\ &\ge 1 - \|\boldsymbol{M}^\star_{2:n,2:n} - \boldsymbol{M}_{2:n,2:n}\| - \|\boldsymbol{M}^\star_{1:n,1}\|_2 - \|\boldsymbol{M}^\star_{1,1:n}\|_2 \\ &\ge 1 - o(1), \end{aligned}$$

with high probability. Here, the last inequality follows since

- *it has been shown in the lecture notes that* $\|\boldsymbol{M}^\star_{2:n,2:n} - \boldsymbol{M}_{2:n,2:n}\| \le \|\boldsymbol{M}^\star - \boldsymbol{M}\| \ll 1$ *if* $p \gg \frac{\log^3 n}{n}$;
- $\|\boldsymbol{M}^\star_{1:n,1}\|_2 = |v_1^\star| \cdot \|\boldsymbol{u}^\star\|_2 = |v_1^\star| = \frac{|\tilde{v}_1|}{\|\tilde{\boldsymbol{v}}\|_2} \lesssim \sqrt{\frac{\log n}{n}}$ *with high probability (because* $|\tilde{v}_1| \lesssim \sqrt{\log n}$ *and* $\|\tilde{\boldsymbol{v}}\|_2 = (1-o(1))\sqrt{n}$ *with high probability);*
- *Similarly,* $\|\boldsymbol{M}^\star_{1,1:n}\|_2 \lesssim \sqrt{\frac{\log n}{n}}$ *with high probability.*

Similarly, with high probability one has

$$\sigma_2(\boldsymbol{M}^{(1)}) \le \sigma_2(\boldsymbol{M}^\star) + \|\boldsymbol{M}^\star - \boldsymbol{M}^{(1)}\| \le o(1).$$

The above two bounds taken collectively give

$$\sigma_1(\boldsymbol{M}^{(1)}) - \sigma_2(\boldsymbol{M}^{(1)}) \ge 1 - o(1).$$

As a result, applying Wedin's $\sin\boldsymbol{\Theta}$ *Theorem gives*

$$\mathsf{dist}\big(\boldsymbol{u}, \boldsymbol{u}^{(1)}\big) \lesssim \frac{\|\boldsymbol{M} - \boldsymbol{M}^{(1)}\|}{\sigma_1(\boldsymbol{M}^{(1)}) - \sigma_2(\boldsymbol{M}^{(1)})} \lesssim \|\boldsymbol{M} - \boldsymbol{M}^{(1)}\|. \tag{9}$$

In addition,

$$\begin{aligned} \|\boldsymbol{M} - \boldsymbol{M}^{(1)}\| &\le \|\boldsymbol{M}_{1:n,1}\|_2 + \|\boldsymbol{M}_{1,1:n}\|_2 \\ &\le \|\boldsymbol{M}_{1:n,1}\|_\infty \sqrt{\|\boldsymbol{M}_{1:n,1}\|_0} + \|\boldsymbol{M}_{1,1:n}\|_\infty \sqrt{\|\boldsymbol{M}_{1,1:n}\|_0} \\ &\lesssim \frac{1}{p}\|\boldsymbol{u}^\star\|_\infty \|\boldsymbol{v}^\star\|_\infty \sqrt{np} \\ &\lesssim \frac{1}{p} \cdot \frac{\log n}{n} \cdot \sqrt{np} \asymp \frac{\log n}{\sqrt{np}}, \end{aligned}$$

where we have used the fact that $\|\boldsymbol{M}_{1:n,1}\|_0 \asymp np$ *as long as* $p \gg \frac{\log n}{n}$ *(Chernoff bound). Substitution into (9) gives*

$$\mathsf{dist}\big(\boldsymbol{u}, \boldsymbol{u}^{(1)}\big) \lesssim \frac{\log n}{\sqrt{np}}. \tag{10}$$

b.(*10 points*) Recall that a more refined version of Wedin's $\sin \boldsymbol{\Theta}$ Theorem states that: for any two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, their leading left singular vectors (denoted by $\boldsymbol{u}_A$ and $\boldsymbol{u}_B$ respectively) satisfy

$$\mathsf{dist}(\boldsymbol{u}_A, \boldsymbol{u}_B) \leq \frac{\max\left\{\big\|(\boldsymbol{A}-\boldsymbol{B})\boldsymbol{v}_A\big\|, \big\|(\boldsymbol{A}-\boldsymbol{B})^\top \boldsymbol{u}_A\big\|\right\}}{\sigma_1\big(\boldsymbol{A}\big) - \sigma_2(\boldsymbol{A}) - \|\boldsymbol{A}-\boldsymbol{B}\|}$$

where $\boldsymbol{v}_A$ is the leading right singular vector of $\boldsymbol{A}$. Can you use this refined version to derive a sharper upper bound on $\mathsf{dist}(\boldsymbol{u}^{(1)}, \boldsymbol{u})$? Here, you can assume without proof that $\|\boldsymbol{u}\|_\infty, \|\boldsymbol{u}^{(1)}\|_\infty, \|\boldsymbol{v}\|_\infty, \|\boldsymbol{v}^{(1)}\|_\infty \lesssim \sqrt{\frac{\log n}{n}}$ with high probability.

Solution: *Applying the refined version of Wedin's* $\sin \boldsymbol{\Theta}$ *Theorem gives*

$$\begin{aligned} \mathsf{dist}\big(\boldsymbol{u}, \boldsymbol{u}^{(1)}\big) &\lesssim \frac{\max\left\{\|(\boldsymbol{M}-\boldsymbol{M}^{(1)})\boldsymbol{v}^{(1)}\|_2, \|(\boldsymbol{M}-\boldsymbol{M}^{(1)})^\top \boldsymbol{u}^{(1)}\|_2\right\}}{\sigma_1(\boldsymbol{M}^{(1)}) - \sigma_2(\boldsymbol{M}^{(1)})} & (11) \\ &\lesssim \max\left\{\|(\boldsymbol{M}-\boldsymbol{M}^{(1)})\boldsymbol{v}^{(1)}\|_2, \|(\boldsymbol{M}-\boldsymbol{M}^{(1)})^\top \boldsymbol{u}^{(1)}\|_2\right\}. & (12) \end{aligned}$$

To bound $\|(\boldsymbol{M}-\boldsymbol{M}^{(1)})\boldsymbol{v}^{(1)}\|_2$, *we have*

$$\|(\boldsymbol{M}-\boldsymbol{M}^{(1)})\boldsymbol{v}^{(1)}\|_2 \leq \big|\boldsymbol{M}_{1,1:n}\boldsymbol{v}^{(1)}\big| + \|\boldsymbol{M}_{1:n,1}\|_2 \big|v_1^{(1)}\big|.$$

It has been shown above that $\|\boldsymbol{M}_{1:n,1}\|_2 \lesssim \frac{\log n}{\sqrt{np}}$, *which together with the assumption* $\|\boldsymbol{v}^{(1)}\|_\infty \lesssim \sqrt{\frac{\log n}{n}}$ *gives*

$$\|\boldsymbol{M}_{1:n,1}\|_2 \big|v_1^{(1)}\big| \lesssim \sqrt{\frac{\log^3 n}{n^2 p}}.$$

In addition, given that $\boldsymbol{M}_{1,1:n}$ *and* $\boldsymbol{v}^{(1)}$ *are statistically independent, we have*

$$\begin{aligned} \mathsf{Var}\left(\big|\boldsymbol{M}_{1,1:n}\boldsymbol{v}^{(1)}\big|\right) &\leq \mathbb{E}\left[\big|\boldsymbol{M}_{1,1:n}\boldsymbol{v}^{(1)}\big|^2\right] \\ &= \sum_{i=1}^n \mathbb{E}\left[M_{1,i}^2 (v_i^{(1)})^2\right] \\ &= \sum_{i=1}^n \mathbb{E}\left[M_{1,i}^2\right] \mathbb{E}\left[(v_i^{(1)})^2\right] \\ &= \sum_{i=1}^n \frac{1}{p} M_{1,i}^{\star 2} \mathbb{E}\left[(v_i^{(1)})^2\right] \\ &\leq \frac{1}{p}\|\boldsymbol{M}_{1,1:n}^\star\|_\infty^2 \mathbb{E}\left[\sum_{i=1}^n (v_i^{(1)})^2\right] \\ &= \frac{1}{p}\|\boldsymbol{M}_{1,1:n}^\star\|_\infty^2 \\ &\lesssim \frac{1}{p}\frac{\log^2 n}{n^2}, \end{aligned}$$

and hence by Chebyshev's inequality,

$$\left|\boldsymbol{M}_{1,1:n}\boldsymbol{v}^{(1)}\right| \lesssim \sqrt{\mathsf{Var}\left(\left|\boldsymbol{M}_{1,1:n}\boldsymbol{v}^{(1)}\right|\right)\log n} \lesssim \sqrt{\frac{\log^3 n}{n^2 p}}$$

with high probability. In summary,

$$\|(\boldsymbol{M}-\boldsymbol{M}^{(1)})\boldsymbol{v}^{(1)}\|_2 \lesssim \sqrt{\frac{\log^3 n}{n^2 p}}.$$

Similarly,

$$\|(\boldsymbol{M}-\boldsymbol{M}^{(1)})^\top \boldsymbol{u}^{(1)}\|_2 \lesssim \sqrt{\frac{\log^3 n}{n^2 p}}.$$

Putting the above bounds together, we obtain

$$\mathsf{dist}\big(\boldsymbol{u}, \boldsymbol{u}^{(1)}\big) \le \sqrt{\frac{\log^3 n}{n^2 p}}. \tag{13}$$

This bound is significantly tighter than the one obtain in Part (a).

6. Community detection experiments (*20 points*) Consider the SBM model discussed in class. Fix the number n of nodes in a graph to be 100. Set $p = \frac{1+\varepsilon}{2}$ and $q = \frac{1-\varepsilon}{2}$ for some quantity $\varepsilon \in [0, 1/2]$. Generate a random graph and then use the spectral method to cluster the nodes. Please plot the mis-clustering rate vs. the probability gap ε. At the minimum, you should take 50 different values of ε (with linear spacing) in $[0, 1/2]$. For each value of ε, you need to run the experiment with at least 200 Monte-Carlo trials.