CME 307/MS&E 311 Optimization
Assignment 1

April 8, 2023

Due: April 21, 2023 at 5:00PM

Problem 1. Weyl's Inequalities
Recall every symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ has n-real eigenvalues $\lambda_1(\boldsymbol{A}) \geq \lambda_2(\boldsymbol{A}), \ldots, \geq \lambda_n(\boldsymbol{A})$. The eigenvalues of $\boldsymbol{A}$ satisfy a very important property known as the Courant-Fischer minimax principle, which states that

$$\lambda_j(\boldsymbol{A}) = \inf_{\dim(\mathcal{V})=n-j+1} \left(\sup_{v\in\mathcal{V}, \|v\|=1} \langle v, \boldsymbol{A}v\rangle \right),$$

where the infimum is taken over all $(n-j+1)$ dimensional subspaces $\mathcal{V}$ of $\mathbb{R}^n$. Using the Courant-Fischer minimax principle, prove the following special case of Weyl's inequalities:

Proposition 0.1. *Suppose that $\boldsymbol{A}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times n}$ be symmetric matrices. Then for all $1 \leq j \leq n$*

$$\lambda_j(\boldsymbol{A}) + \lambda_n(\boldsymbol{B}) \leq \lambda_j(\boldsymbol{A}+\boldsymbol{B}) \leq \lambda_j(\boldsymbol{A}) + \lambda_1(\boldsymbol{B}).$$

Solultion: We first show the inequality on the right-hand side. By the Courant-Fischer minimax principle, we have

$$\begin{aligned}
\lambda_j(A+B) &= \inf_{\dim(\mathcal{V})=n-j+1} \sup_{v\in\mathcal{V}, \|v\|_2=1} v^\top (A+B)v \\
&\leq \inf_{\dim(\mathcal{V})=n-j+1} \left(\sup_{v\in\mathcal{V}, \|v\|_2=1} v^\top Av + \sup_{v\in\mathcal{V}, \|v\|_2=1} v^\top Bv \right) \\
&\leq \inf_{\dim(\mathcal{V})=n-j+1} \sup_{v\in\mathcal{V}, \|v\|_2=1} v^\top Av + \sup_{\dim(\mathcal{V})=n-j+1} \sup_{v\in\mathcal{V}, \|v\|_2=1} v^\top Bv \\
&= \lambda_j(A) + \sup_{\dim(\mathcal{V})=n-j+1} \sup_{v\in\mathcal{V}, \|v\|_2=1} v^\top Bv \\
&\leq \lambda_j(A) + \lambda_1(B),
\end{aligned}$$

where the first and forth lines come from the Courant-Fischer minimax principle, the second and third lines come from the properties of sup and inf, and the last line comes from the fact that $v^\top Bv \leq \lambda_1(B)$ for any v such that $\|v\|_2 = 1$.

Similarly, we can develop another inequality. Specifically,

$$\begin{aligned}
\lambda_j(A+B) &= \inf_{\dim(\mathcal{V})=n-j+1} \sup_{v\in\mathcal{V},\|v\|_2=1} v^\top (A+B)v \\
&\geq \inf_{\dim(\mathcal{V})=n-j+1} \left(\sup_{v\in\mathcal{V},\|v\|_2=1} v^\top Av + \inf_{v\in\mathcal{V},\|v\|_2=1} v^\top Bv \right) \\
&\geq \inf_{\dim(\mathcal{V})=n-j+1} \sup_{v\in\mathcal{V},\|v\|_2=1} v^\top Av + \inf_{\dim(\mathcal{V})=n-j+1} \inf_{v\in\mathcal{V},\|v\|_2=1} v^\top Bv \\
&= \lambda_j(A) + \inf_{\dim(\mathcal{V})=n-j+1} \inf_{v\in\mathcal{V},\|v\|_2=1} v^\top Bv \\
&\leq \lambda_j(A) + \lambda_n(B),
\end{aligned}$$

where the first and forth lines come from the Courant-Fischer minimax principle, the second and third lines come from the properties of sup and inf, and the last line comes from the fact that $v^\top Bv \geq \lambda_n(B)$ for any v such that $\|v\|_2 = 1$.

Problem 2. Loewner Ordering A fundamental object in matrix analysis is the Loewner ordering on the cone of $n \times n$ symmetric positive semidefinite (psd) matrices $\mathbb{S}_n^+(\mathbb{R})$. Given $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}_n^+(\mathbb{R})$, we write $\boldsymbol{A} \preceq \boldsymbol{B}$ if and only if $\boldsymbol{B} - \boldsymbol{A}$ is a symmetric psd matrix. Prove the following items:

1. Let $\boldsymbol{A} \in \mathbb{S}_n^+(\mathbb{R})$ and $\boldsymbol{M} \in \mathbb{R}^{n\times k}$ be any matrix. Then
$$\boldsymbol{M}^T \boldsymbol{A} \boldsymbol{M} \in \mathbb{S}_k^+(\mathbb{R}).$$
2. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}_n^+(\mathbb{R})$ and $\boldsymbol{A} \preceq \boldsymbol{B}$. Then for each $1 \le j \le n$
$$\lambda_j(\boldsymbol{A}) \le \lambda_j(\boldsymbol{B}).$$
3. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}_n^{++}(\mathbb{R})$ and $\boldsymbol{A} \preceq \boldsymbol{B}$. Then
$$\boldsymbol{B}^{-1} \preceq \boldsymbol{A}^{-1}.$$
4. Let $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{S}_n^{++}(\mathbb{R})$. Then
$$\lambda_n(\boldsymbol{A})\lambda_j(\boldsymbol{B}) \le \lambda_j(\boldsymbol{A}\boldsymbol{B}) \le \lambda_1(\boldsymbol{A})\lambda_j(\boldsymbol{B}).$$

Hint: For item 4. matrix similarity will be helpful.

Solution:

1. We show it by the definition of PSD matrices. For any $x \in \mathbb{R}^k$, since A is PSD,
$$x^\top (M^\top AM)x = (Mx)^\top A(Mx) \ge 0,$$
which implies $M^\top AM$ is also PSD.
2. By the statement of Problem 1, we have
$$\lambda_j(A) + \lambda_n(B - A) \le \lambda_j(B).$$
Then, since $B - A$ is PSD, $\lambda_n(B - A) \ge 0$, which implies
$$\lambda_j(A) \le \lambda_j(B).$$
3. We first show a special case when $B = I$. If $A \preceq I$ for a positive definite matrix A, we have all eigenvalues of A are in $(0, 1]$, which implies all eigenvalues of A^{-1} are larger than 1. Thus, for any $x \in \mathbb{R}^n$,
$$x^T A^{-1}x \ge x^T x = x^T Ix,$$
which is equivalent to $A^{-1} \succeq I$.

   Next, we show the case that $B \ne I$. By definition, $A \preceq B$ implies that $B^{-1/2}AB^{-1/2} \preceq I$ since for any $x \in \mathbb{R}^n$,
$$x^\top B^{-1/2}AB^{-1/2}x = (B^{-1/2}x)^\top A(B^{-1/2}x) \le (B^{-1/2}x)^\top A(B^{-1/2}x) = x^T x.$$
Moreover, by Problem 2.1 and the condition that B is invertible, we have $B^{-1/2}AB^{-1/2}$ is still positive definite. Thus, the previous special case implies $B^{1/2}A^{-1}B^{1/2} \succeq I$. Similarly, it implies
$$A^{-1} \succeq B^{-1}.$$

4. We first show the inequality on the right-hand side that

$$\lambda_j(AB) \le \lambda_1(A)\lambda_j(B).$$

It is equivalent to

$$\lambda_j(B^{1/2}AB^{1/2}) \le \lambda_j(\lambda_1(A)B).$$

since similarity operations do not change eigenvalues. By Problem 2.2, it is sufficient to show that

$$B^{1/2}AB^{1/2} \preceq \lambda_1(A)B,$$

which can be proved by definition. Specifically, for any $x \in \mathbb{R}^n$,

$$x^\top B^{1/2}AB^{1/2}x = (B^{1/2}x)^\top A(B^{1/2}x) \le \lambda_1(A)(B^{1/2}x)^\top(B^{1/2}x) = \lambda_1 x^\top Bx,$$

and thus, $B^{1/2}AB^{1/2} \preceq \lambda_1(A)B$.

Then, we can follow a similar proof to show the inequality on the left-hand side. Correspondingly, it is sufficient to show

$$B^{1/2}AB^{1/2} \succeq \lambda_n(A)B.$$

To show this, for any $x \in \mathbb{R}^n$,

$$x^\top B^{1/2}AB^{1/2}x = (B^{1/2}x)^\top A(B^{1/2}x) \ge \lambda_n(A)(B^{1/2}x)^\top(B^{1/2}x) = \lambda_n x^\top Bx.$$

As a result, $B^{1/2}AB^{1/2} \succeq \lambda_n(A)B$, and we finish the proof.

Problem 3. Nyström Approximation Let $\boldsymbol{A} \in \mathbb{S}_n^+(\mathbb{R})$ and $\boldsymbol{X} \in \mathbb{R}^{n\times k}$. Then the Nyström approximation of $\boldsymbol{A}$ with respect to $\boldsymbol{X}$ is given by

$$\hat{\boldsymbol{A}} = (\boldsymbol{A}\boldsymbol{X})(\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X})^\dagger(\boldsymbol{A}\boldsymbol{X})^T.$$

In the world of randomized numerical linear algebra, $\boldsymbol{X}$ is referred to as the *test matrix*. Observe as $\boldsymbol{X} \in \mathbb{R}^{n\times k}$ that $\hat{\boldsymbol{A}}$ has rank at most k, hence $\hat{\boldsymbol{A}}$ yields a low-rank approximation to $\boldsymbol{A}$. It turns out a near optimal low-rank approximation to $\boldsymbol{A}$ can be obtained by choosing $\boldsymbol{X}$ to be a suitable random matrix. Canonical choices for $\boldsymbol{X}$ include standard normal random matrices and column selection matrices (matrices that select a subset of the columns of $\boldsymbol{A}$). Thanks to its ability to produce high-quality low-rank approximations, the randomized Nyström approximation has come to play a fundamental role in modern machine learning, with applications including large-scale kernel learning, linear system solving, and low-rank semidefinite programming.

In this problem, you will establish some of the randomized Nyström approximation's key properties.

1. Show that $\boldsymbol{M}(\boldsymbol{M}^T\boldsymbol{M})^\dagger\boldsymbol{M}^T = \boldsymbol{\Pi}_{\text{range}(\boldsymbol{M})}$, where $\boldsymbol{\Pi}_{\text{range}(\boldsymbol{M})}$ is the orthogonal projector onto range$(\boldsymbol{M})$.
2. Show that $\hat{\boldsymbol{A}} \succeq 0$ and $\hat{\boldsymbol{A}} \preceq \boldsymbol{A}$. (**Hint:** Try and rewrite $\hat{\boldsymbol{A}}$ in a way so that you can exploit item 1.)
3. Show that $\hat{\boldsymbol{A}}$ and $\boldsymbol{A}$ agree on the subspace $\boldsymbol{X}$, that is show the equality

$$\hat{\boldsymbol{A}}\boldsymbol{X} = \boldsymbol{A}\boldsymbol{X}.$$

Remark 0.2. The notation $(\boldsymbol{M}^T\boldsymbol{M})^\dagger$ denotes the pseudo-inverse of the matrix $\boldsymbol{M}^T\boldsymbol{M}$. If $\boldsymbol{M}^T\boldsymbol{M}$ is invertible, the pseudo-inverse corresponds with the inverse. In general, given a matrix $\boldsymbol{B}$, the pseudo-inverse of $\boldsymbol{B}$ is defined using its singular value decomposition (SVD) as follows:

Definition 0.3. Let $\boldsymbol{B} \in \mathbb{R}^{m\times n}$ and consider its SVD $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T$. Then the *pseudo-inverse* $\boldsymbol{B}^\dagger \in \mathbb{R}^{n\times m}$ of $\boldsymbol{B}$, is defined to be the matrix

$$\boldsymbol{B}^\dagger = \boldsymbol{V}\boldsymbol{\Sigma}^+\boldsymbol{U}^T.$$

Here $\boldsymbol{\Sigma}^+ \in \mathbb{R}^{n\times m}$ is a diagonal matrix whose entries are computed as follows:

$$\begin{cases} (\boldsymbol{\Sigma}^+)_{ii} = (\boldsymbol{\Sigma}_{ii})^{-1}, & \boldsymbol{\Sigma}_{ii} \neq 0, \\ (\boldsymbol{\Sigma}^+)_{ii} = 0, & \text{otherwise.} \end{cases}$$

Solution:

1. Denote the SVD of M as $M = U\Sigma V^\top$. Without loss of generality, we assume there exists a positive integer $k \in [1, n]$ such that $\Sigma_{ii} \neq 0$ for $i \leq k$, and $\Sigma_{ii} = 0$ for all $i > k$.

   Then, we have $M^\top M = V\Sigma^\top\Sigma V$, where $\Sigma^\top\Sigma$ is a diagonal matrix in $\mathbb{R}^{m\times m}$,

$$M(M^\top M)^\dagger M^\top = U\Sigma V^\top V(\Sigma^\top\Sigma)^+V^\top V\Sigma U^\top = U\Sigma(\Sigma^\top\Sigma)^+\Sigma U^\top.$$

   By the definition of $(\cdot)^+$ of a diagonal matrix, we have that $\tilde{\Sigma} := \Sigma(\Sigma^\top\Sigma)^+\Sigma$ is a diagonal matrix satisfying

$$(\tilde{\Sigma})_{ii} = \begin{cases} 1, \text{if } i \leq k, \\ 0, \text{if } i > k. \end{cases} \tag{1}$$

   Denote the column vectors of U as $u_1, ..., u_n$, and define $\tilde{U} = (u_1, ..., u_k)$. Based on inequality (1), we have

$$M(M^\top M)^\dagger M^\top = \tilde{U}\tilde{U}^T,$$

   which is the orthogonal projector onto the span of $u_1, ..., u_k$. Finally, we finish the proof by noticing that $\text{span}(u_1, ..., u_k) = \text{Range}(M)$.

2. We first show that $\hat{A} \succeq 0$. By the definition of the pseudo-inverse, we have that A is PSD if and only if $A^\dagger$ is PSD. By Problem 2.1, we have that $X^T AX$ is PSD, which implies $(X^T AX)^\dagger$ is PSD. Then, applying Problem 2.1 again, we have $\hat{A} = (AX)(X^T AX)^\dagger (AX)^\top$ is PSD.

   Next, we show that $A \succeq \hat{A}$. Let $M = A^{1/2}X$ in Problem 3.1, we have

   $$\hat{A} = A^{1/2} M (M^\top M)^\dagger M A^{1/2} = A^{1/2} \Pi_{\mathrm{Range}(A^{1/2}X)} A^{1/2}. \tag{2}$$

   For any $x \in \mathbb{R}^d$,

   $$\begin{aligned} x^\top A x &= (A^{1/2}x)^\top (A^{1/2}x) \\ &= (A^{1/2}x)^\top (\Pi_{\mathrm{Range}(A^{1/2}X)} + (I - \Pi_{\mathrm{Range}(A^{1/2}X)}))(A^{1/2}x) \\ &= (A^{1/2}x)^\top \Pi_{\mathrm{Range}(A^{1/2}X)} (A^{1/2}x) + (A^{1/2}x)^\top (I - \Pi_{\mathrm{Range}(A^{1/2}X)})(A^{1/2}x) \\ &\geq (A^{1/2}x)^\top \Pi_{\mathrm{Range}(A^{1/2}X)} (A^{1/2}x) \\ &\geq x^\top \hat{A} x, \end{aligned}$$

   where the first and second lines come from the direct calculation, the third line comes from the property of any orthogonal projector, the forth line comes from the fact that $(I - \Pi_{\mathrm{Range}(A^{1/2}X)})$ is also PSD for any orthogonal projector, and the last line comes from (2).

3. We direct show it as follows:

   $$\begin{aligned} \hat{A}X &= A^{1/2} \Pi_{\mathrm{Range}(A^{1/2}X)} A^{1/2} X \\ &= A^{1/2} A^{1/2} X \\ &= AX, \end{aligned}$$

   where the first line comes from (2), the second line comes from the fact that $A^{1/2}X$ is invariant under the orthogonal projection on $\mathrm{Range}(A^{1/2}X)$, and the last line comes from matrix multiplication.