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 $1 \leq j \leq n$*

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

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 $1 \leq j \leq n$
$$\lambda_j(\boldsymbol{A}) \leq \lambda_j(\boldsymbol{B}).$$
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
$$\lambda_n(\boldsymbol{A})\lambda_j(\boldsymbol{B}) \leq \lambda_j(\boldsymbol{A}\boldsymbol{B}) \leq \lambda_1(\boldsymbol{A})\lambda_j(\boldsymbol{B}).$$

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

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:

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