SI231B - Matrix Computations, Spring 2022-23

Homework Set #5

Prof. Ziping Zhao

Acknowledgements:

1) Deadline: **2023-05-06 23:59:59**
2) Please submit your assignments via Gradescope.
3) You can write your homework using latex/word or you can write in handwriting and submit the scanned pdf.

Problem 1. (20 points)

Consider a matrix $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 2 & 1 \end{bmatrix}$.

1) Find the SVD of $\mathbf{A}$. (Both compact SVD and full SVD are correct.) (15 points)
2) Compute the pseudo-inverse of $\mathbf{A}$. (5 points)

Solution:

1) Find the eigenvalues and eigenvectors of $\mathbf{A}\mathbf{A}^T$:

   Eigenvalue: 6, eignevector: $[1, -1, 2]^T$;

   Eigenvalue: 3, eignevector: $[-1, 1, 1]^T$;

   Eigenvalue: 0, eignevector: $[1, 1, 0]^T$.

   Find the square roots of the nonzero eigenvalues:

   $\sigma_1 = \sqrt{6}$;

   $\sigma_2 = \sqrt{3}$.

   The $\mathbf{\Sigma}$ matrix is a zero matrix with σ_i on its diagonal:

$$\mathbf{\Sigma} = \begin{bmatrix} \sqrt{6} & 0 \\ 0 & \sqrt{3} \\ 0 & 0 \end{bmatrix}.$$

   The columns of the matrix $\mathbf{U}$ are the normalized (unit) vectors:

$$\mathbf{U} = \begin{bmatrix} \frac{\sqrt{6}}{6} & -\frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} \\ -\frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{6}}{3} & \frac{\sqrt{3}}{3} & 0 \end{bmatrix}.$$

   Now, $\mathbf{v}_i = \frac{1}{\sigma_i}\mathbf{A}^T\mathbf{u}_i$:

$$\mathbf{v}_1 = \frac{1}{\sigma_1}\mathbf{A}^T\mathbf{u}_1 = [1, 0]^T,$$

$$\mathbf{v}_2 = \frac{1}{\sigma_2}\mathbf{A}^T\mathbf{u}_2 = [0,1]^T.$$

Therefore,

$$\mathbf{V} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

2)

$$\mathbf{U}_1 = \begin{bmatrix} \frac{\sqrt{6}}{6} & -\frac{\sqrt{3}}{3} \\ -\frac{\sqrt{6}}{6} & \frac{\sqrt{3}}{3} \\ \frac{\sqrt{6}}{3} & \frac{\sqrt{3}}{3} \end{bmatrix}, \mathbf{V}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tilde{\mathbf{\Sigma}} = \begin{bmatrix} \sqrt{6} & 0 \\ 0 & \sqrt{3} \end{bmatrix}.$$

The pseudo-inverse of $\mathbf{A}$ is given by

$$\mathbf{A}^\dagger = \mathbf{V}_1\tilde{\mathbf{\Sigma}}^{-1}\mathbf{U}_1^T = \begin{bmatrix} \frac{1}{6} & -\frac{1}{6} & \frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Problem 2. (20 points)

The Frobenius norm of a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is defined as $\|\mathbf{A}\|_F = \sqrt{Tr(\mathbf{A}^T\mathbf{A})}$.

1) Show that

$$\|\mathbf{A}\|_F = \left(\sum_{i,j} |\mathbf{A}_{ij}|^2\right)^{\frac{1}{2}}$$

(5 points)

2) Show that if $\mathbf{U}$ and $\mathbf{V}$ are orthogonal, then $\|\mathbf{UA}\|_F = \|\mathbf{AV}\|_F = \|\mathbf{A}\|_F$. (5 points)
3) Show that $\|\mathbf{A}\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$, where $\sigma_1, \ldots, \sigma_r$ are the singular value of $\mathbf{A}$. (5 points)
4) Assume that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Show that $\sigma_1 \leq \|\mathbf{A}\|_F \leq \sqrt{r}\sigma_1$. (5 points)

Solution:

1)

$$\|\mathbf{A}\|_F = \sqrt{Tr(\mathbf{A}^T\mathbf{A})} = \sum_i \left(\sum_j \mathbf{A}_{ji}^T\mathbf{A}_{ji}\right) = \sum_{i,j} \mathbf{A}_{ij}^2$$

2) First, consider $\|\mathbf{UA}\|_F$, we have

$$\|\mathbf{UA}\| = \sqrt{Tr((\mathbf{UA})^T(\mathbf{UA}))} = \sqrt{Tr(\mathbf{A}^T\mathbf{U}^T\mathbf{UA})} = \sqrt{Tr(\mathbf{A}^T\mathbf{A})} = \|\mathbf{A}\|_F$$

Similarly, we have

$$\|\mathbf{AV}\| = \sqrt{Tr((\mathbf{AV})^T(\mathbf{AV}))} = \sqrt{Tr((\mathbf{AV})(\mathbf{AV}))^T} = \sqrt{Tr(\mathbf{AVV}^T\mathbf{A}^T)} = \sqrt{Tr(\mathbf{AA}^T)} = \sqrt{Tr(\mathbf{A}^T\mathbf{A})} = \|\mathbf{A}\|_F$$

Thus, $\|\mathbf{UA}\|_F = \|\mathbf{AV}\|_F = \|\mathbf{A}\|_F$.

3) Use the singular value decomposition,

$$\|\mathbf{A}\|_F = \|\mathbf{U\Sigma V}^T\|_F = \|\Sigma\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}$$

4) Since $\sigma_i^2 \geq 0$, we have $\|\mathbf{A}\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \geq \sigma_1$. And since $\sigma_2, \ldots, \sigma_r \leq \sigma_1$, we have $\|\mathbf{A}\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_r^2} \leq \sqrt{r}\sigma_1$.

Problem 3. (20 points)

Let $\mathbf{A} \in \mathbb{R}^{n \times n}$.

1) If $\kappa(\mathbf{A}) = 1$. Show that $\mathbf{A}$ is a multiple of an orthogonal matrix. (10 points)
2) If $\mathbf{A} = \gamma \mathbf{U}$, where $\mathbf{U}$ is an orthogonal matrix and $\gamma \in \mathbb{R}$. Show that $\kappa(\mathbf{A}) = 1$. (10 points)

Solution:

1)

$$1 = \kappa(\mathbf{A}) = \|\mathbf{A}\|_2 \|\mathbf{A}^\dagger\|_2 = \sigma_{\max} \cdot \frac{1}{\sigma_{\min}} \quad \Longrightarrow$$

$$\sigma_{\max} = \sigma_{\min}$$

Then, $\mathbf{\Sigma} = \sigma_{\max}\mathbf{I}$, and

$$\mathbf{A} = \mathbf{U\Sigma V}^T = \sigma_{\max}\mathbf{UV}^T \quad \Longrightarrow$$

$$\mathbf{A}^T\mathbf{A} = \mathbf{AA}^T = \sigma_{\max}\mathbf{I}$$

Thus, $\mathbf{A}$ is a scaled orthogonal matrix.

2)

$$\mathbf{A} = \gamma\mathbf{U} = \mathbf{U}\left(|\gamma|\mathbf{I}\right)\mathrm{sgn}(\gamma)\mathbf{I}$$

As we can see, this is a SVD of $\mathbf{A}$, in which $\mathbf{\Sigma} = |\gamma|\mathbf{I}$ and $\mathbf{V} = \mathrm{sgn}(\gamma)\mathbf{I}$. Then we have $\sigma_{\max} = \sigma_{\min} = |\gamma|$. Thus, $\kappa(\mathbf{A}) = 1$.

Problem 4. (20 points)

Consider the problem of partitioning the vertex set $\mathcal{V}$ of a directed graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ into two subsets $\mathcal{S}_1$ and $\mathcal{S}_2$ (i.e., finding a cut for a directed graph) so that the number of edges from $\mathcal{S}_1$ to $\mathcal{S}_2$ is maximized. This problem can be modeled in the following way. For each vertex i with $i = 1, \ldots, n$, we associate it with an indicator variable x_i, which equals to 1 if $i \in \mathcal{S}_1$ and 0 if $i \in \mathcal{S}_2$. The number of edges from $\mathcal{S}_1$ to $\mathcal{S}_2$ is given by

$$\sum_{i,j} a_{ij} x_i (1 - x_j) = \mathbf{x}^T \mathbf{A} (\mathbf{1} - \mathbf{x})$$

where $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a 0-1 matrix called adjacency matrix with $a_{ij} = 1$ if there is an edge from i to j and $a_{ij} = 0$ otherwise. Then the problem is to obtain a 0-1 vector $\mathbf{x} \in \mathbb{R}^n$ by solving the problem

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \mathbf{x}^T \mathbf{A} (\mathbf{1} - \mathbf{x}) \\ \text{subject to} \quad & x_i \in \{0, 1\}, \ \forall i. \end{aligned}$$

The problem is NP-hard. In practice an approximation solution is preferred, which is from the following problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{maximize}} \quad & \mathbf{x}^T \mathbf{A}_k (\mathbf{1} - \mathbf{x}) \\ \text{subject to} \quad & x_i \in \{0, 1\}, \ \forall i, \end{aligned}$$

where $\mathbf{A}_k = \sum_{i=1}^{k} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$ denotes the truncated SVD for $\mathbf{A}$.

Prove that for any 0-1 vector $\mathbf{x}$ the following approximation bound holds:

$$\|\mathbf{x}^T \mathbf{A} (\mathbf{1} - \mathbf{x}) - \mathbf{x}^T \mathbf{A}_k (\mathbf{1} - \mathbf{x})\|_2 \leq \frac{n^2}{\sqrt{k+1}}.$$

Solution:

Since $\mathbf{x}$ and $\mathbf{1} - \mathbf{x}$ are both 0-1 vectors, we have $\|\mathbf{x}\|_2 \leq \sqrt{n}$ and $\|\mathbf{1} - \mathbf{x}\|_2 \leq \sqrt{n}$. Then,

$$\begin{aligned} \|\mathbf{x}^T \mathbf{A} (\mathbf{1} - \mathbf{x}) - \mathbf{x}^T \mathbf{A}_k (\mathbf{1} - \mathbf{x})\|_2 =& \|\mathbf{x}^T (\mathbf{A} - \mathbf{A}_k)(\mathbf{1} - \mathbf{x})\|_2 \leq \|\mathbf{x}\|_2 \|(\mathbf{A} - \mathbf{A}_k)(\mathbf{1} - \mathbf{x})\|_2 \\ \leq& \|\mathbf{x}\|_2 \|\mathbf{A} - \mathbf{A}_k\|_2 \|\mathbf{1} - \mathbf{x}\|_2 \leq n \|\mathbf{A} - \mathbf{A}_k\|_2 = n\sigma_{k+1}(\mathbf{A}). \end{aligned}$$

Based on the basic matrix norm properties and the fact that $\mathbf{A}$ is a 0-1 matrix, we have

$$(k+1)\sigma_{k+1}^2(\mathbf{A}) \leq \sum_{i=1}^{k+1} \sigma_i^2(\mathbf{A}) \leq \sum_{i=1}^{n} \sigma_i^2(\mathbf{A}) = \|\mathbf{A}\|_F^2 = \sum_{i,j=1}^{n} a_{ij}^2 \leq n^2,$$

and then it easily follows that

$$\|\mathbf{x}^T \mathbf{A} (\mathbf{1} - \mathbf{x}) - \mathbf{x}^T \mathbf{A}_k (\mathbf{1} - \mathbf{x})\|_2 \leq n\sigma_{k+1}(\mathbf{A}) \leq n \frac{n}{\sqrt{k+1}} = \frac{n^2}{\sqrt{k+1}}.$$

Problem 5. (20 points)

Show that $\mathbf{A}\mathbf{A}^\dagger$ is the orthogonal projection onto the range space of $\mathbf{A}$, and $\mathbf{A}^\dagger\mathbf{A}$ is the orthogonal projection on the orthogonal complement of $\mathcal{N}(\mathbf{A})$.

Solution:

Since $\mathbf{A}\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger = \mathbf{A}\mathbf{A}^\dagger$, $\mathbf{A}^\dagger\mathbf{A}\mathbf{A}^\dagger\mathbf{A} = \mathbf{A}^\dagger\mathbf{A}$, $\left(\mathbf{A}\mathbf{A}^\dagger\right)^T = \mathbf{A}\mathbf{A}^\dagger$, and $\left(\mathbf{A}^\dagger\mathbf{A}\right)^T = \mathbf{A}^\dagger\mathbf{A}$, we can conclude that $\mathbf{A}\mathbf{A}^\dagger$ and $\mathbf{A}^\dagger\mathbf{A}$ are orthogonal projections.

Observe that

$$\mathbf{y} \in \mathcal{R}\left(\mathbf{A}\mathbf{A}^\dagger\right) \Rightarrow \mathbf{y} = \mathbf{A}\mathbf{A}^\dagger\mathbf{x} \text{ for some } \mathbf{x} \Rightarrow \mathbf{y} \in \mathcal{R}(\mathbf{A}) \Rightarrow \mathcal{R}\left(\mathbf{A}\mathbf{A}^\dagger\right) \subseteq \mathcal{R}(\mathbf{A})$$

and

$$\mathbf{y} \in \mathcal{R}(\mathbf{A}) \Rightarrow \mathbf{y} = \mathbf{A}\mathbf{x} \text{ for some } \mathbf{x} \Rightarrow \mathbf{A}\mathbf{A}^\dagger\mathbf{y} = \mathbf{A}\mathbf{A}^\dagger\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x} = \mathbf{y} \Rightarrow \mathbf{y} \in \mathcal{R}\left(\mathbf{A}\mathbf{A}^\dagger\right),$$

we can get $\mathcal{R}\left(\mathbf{A}\mathbf{A}^\dagger\right) = \mathcal{R}(\mathbf{A})$.

Similarly, observe that

$$\mathbf{x} \in \mathcal{N}(\mathbf{A}) \Rightarrow \mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{A}^\dagger\mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x} \in \mathcal{N}\left(\mathbf{A}^\dagger\mathbf{A}\right) \Rightarrow \mathcal{N}(\mathbf{A}) \subseteq \mathcal{N}\left(\mathbf{A}^\dagger\mathbf{A}\right),$$

and

$$\mathbf{x} \in \mathcal{N}\left(\mathbf{A}^\dagger\mathbf{A}\right) \Rightarrow \mathbf{A}^\dagger\mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{A}\mathbf{A}^\dagger\mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{A}\mathbf{x} = \mathbf{0} \Rightarrow \mathbf{x} \in \mathcal{N}(\mathbf{A}) \Rightarrow \mathcal{N}\left(\mathbf{A}^\dagger\mathbf{A}\right) \subseteq \mathcal{N}(\mathbf{A}),$$

we can get

$$\mathcal{N}\left(\left(\mathbf{A}^\dagger\mathbf{A}\right)^T\right) = \mathcal{N}\left(\mathbf{A}^\dagger\mathbf{A}\right) = \mathcal{N}(\mathbf{A}) \Rightarrow \mathcal{R}\left(\mathbf{A}^\dagger\mathbf{A}\right) = \mathcal{N}(\mathbf{A})^\perp.$$

To sum up, we have

$$\mathcal{R}\left(\mathbf{A}\mathbf{A}^\dagger\right) = \mathcal{R}(\mathbf{A}), \quad \mathcal{R}\left(\mathbf{A}^\dagger\mathbf{A}\right) = \mathcal{N}(\mathbf{A})^\perp.$$

Therefore, $\mathbf{A}\mathbf{A}^\dagger$ is the orthogonal projection onto the range space of $\mathbf{A}$, and $\mathbf{A}^\dagger\mathbf{A}$ is the orthogonal projection on the orthogonal complement of $\mathcal{N}(\mathbf{A})$.