Homework 11

This homework is due on Sunday, November 13, 2022 at 11:59PM. Self-grades and HW Resubmissions are due the following Sunday, November 20, 2022 at 11:59PM.

1. **Spectral Theorem for Real Symmetric Matrices**

   We want to show that every real symmetric matrix can be diagonalized by a matrix of its orthonormal eigenvectors. In other words, a symmetric matrix $S \in \mathbb{R}^{n\times n}$, i.e., a matrix S such that $S = S^\top$, can be written as $S = V\Lambda V^\top$, where $V \in \mathbb{R}^{n\times n}$ is an orthonormal matrix of eigenvectors of S and $\Lambda \in \mathbb{R}^{n\times n}$ is a diagonal matrix of corresponding real eigenvalues of S. This is called the Spectral Theorem for real symmetric matrices.

   To do this, we will use a proof which is similar to the proof of existence of the Schur decomposition. Along the way, we will practice block matrix manipulation and the induction proof technique.

   (a) One part of the spectral theorem can be proved without any further delay. **Prove that the eigenvalues λ of a real, symmetric matrix S are real.**

   (HINT: Let λ be an eigenvalue of S with corresponding nonzero eigenvector $\vec{v}$. Evaluate $\overline{\vec{v}}^\top S\vec{v}$ in two different ways: $\overline{\vec{v}}^\top (S\vec{v})$ and $(\overline{\vec{v}}^\top S)\vec{v}$. What does this show about λ?)

   (b) For the main proof that every real symmetric matrix is diagonalized by a matrix of its orthonormal real eigenvectors, we will proceed by *induction*.

   Recall that an inductive proof trying to prove a statement that depends on n, say P_n[1], is true for all positive integers n, has two steps:

   - A base case – prove that P_1 is true.
   - An inductive step – for every $n \geq 2$, given that P_{n-1} is true, prove that P_n is true.[2]

   By doing these two steps, we show P_n is true for all n.

   In our case, the statement P_n is "every $n \times n$ symmetric matrix S can be diagonalized as $S = V\Lambda V^\top$, where V is the real orthonormal matrix of eigenvectors of S, and Λ is the real diagonal matrix of corresponding eigenvalues of S."

   Show the base case: every 1×1 symmetric matrix S can be written as $S = V\Lambda V^\top$, where V is a real and orthonormal matrix of eigenvectors of S, and Λ is a real and diagonal matrix of corresponding eigenvalues of S.

   (HINT: Every 1×1 matrix is symmetric, and also diagonal, by definition; the only real orthonormal 1×1 matrices are $\begin{bmatrix}1\end{bmatrix}$ and $\begin{bmatrix}-1\end{bmatrix}$.)

[1] Lecture used S_n, but S is already being used for symmetric matrix here.

[2] This is the so-called *weak induction* paradigm; it contrasts with *strong induction*, which you can learn in future classes like CS70.

(c) With the base case done, we are now in the inductive step. Let S be an arbitrary $n \times n$ symmetric matrix; ultimately, we want to show that $S = V\Lambda V^\top$, where V is a real and orthonormal matrix of eigenvectors of S, and Λ is a real and diagonal matrix of corresponding eigenvalues of S.

To start, let λ be an eigenvalue of S, and let $\vec{q}$ be any normalized eigenvector of S corresponding to eigenvalue λ. Let $\widetilde{Q} \in \mathbb{R}^{n\times(n-1)}$ be a set of orthonormal vectors chosen so that $Q := \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix} \in \mathbb{R}^{n\times n}$ is an orthonormal matrix.[3] **Show the following equality:**

$$Q^\top S Q = \begin{bmatrix} \lambda & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & S_0 \end{bmatrix} \qquad \textbf{where} \qquad S_0 := \widetilde{Q}^\top S \widetilde{Q}. \tag{1}$$

(HINT: Expand Q as a block matrix $\begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix}$ and multiply $Q^\top S Q = \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix}^\top S \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix}$.)
(HINT: Since Q is orthonormal, we have $Q^\top Q = I_n$. What does this mean for the values of $\vec{q}^\top \vec{q}$ and $\widetilde{Q}^\top \vec{q}$? Use block matrix multiplication on $Q^\top Q = \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix}^\top \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix}$ again.)

(d) **Show that the matrix S_0 is a real symmetric matrix.**

(e) Since S_0 is a real symmetric $(n-1)\times(n-1)$ matrix, by our inductive assumption, S_0 can be orthonormally diagonalized as $S_0 = V_0 \Lambda_0 V_0^\top$, where Λ_0 is a real diagonal matrix of eigenvalues of S_0 and $V_0 \in \mathbb{R}^{(n-1)\times(n-1)}$ is a real orthonormal matrix of corresponding eigenvectors of S_0.

Define

$$V := Q \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \qquad \textbf{and} \qquad \Lambda := V^\top S V. \tag{2}$$

i. **Show that V is orthonormal.**

ii. **Show that Λ is diagonal.**

iii. **Show that $S = V\Lambda V^\top$.**

(HINT: Use block matrix multiplication again.)

Thus, we have found a real orthonormal V and real diagonal Λ such that $S = V\Lambda V^\top = V\Lambda V^{-1}$. We have seen in a previous homework that if $A = V\Lambda V^{-1}$, then Λ are the eigenvalues of A, and V are the corresponding eigenvectors. Thus, given P_{n-1} – the fact that we can orthonormally diagonalize $(n-1)\times(n-1)$ real symmetric matrices – we have proven P_n – the fact that we can orthonormally diagonalize $n \times n$ real symmetric matrices. Thus, we've proved the Spectral Theorem for real symmetric matrices by induction!

[3] This matrix $\widetilde{Q}$ can be generated via Gram-Schmidt, for example.

2. QR System ID Revisited

Recall from your previous homework that, if $D \in \mathbb{R}^{m\times n}$ where $m < n$ and $\text{rank}(D) = m$, then we can write the QR decomposition of its transpose as

$$D^\top = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \begin{bmatrix} R_1 \\ 0_{(n-m)\times m} \end{bmatrix} \tag{3}$$

The previous homework problem focused on solving a system ID problem, namely solving for $\vec{p}$ in

$$D\vec{p} = \vec{s} \tag{4}$$

where $\vec{s} \in \mathbb{R}^m$. Since this is an underdetermined system, we can have multipled choices for $\vec{p}$. As in the previous homework, we want to find the unique solution that minimizes $\|\vec{p}\|$. To do this, we said that we want to set $Q_2^\top \vec{p} = \vec{0}$. In this problem, we will examine why this minimizes $\|\vec{p}\|$.

(a) **First, show that $\|\vec{p}\| = \|U\vec{p}\|$ where U is a matrix with orthonormal columns.** *Warning: a matrix with orthonormal columns is not necessarily an orthonormal matrix. (HINT: Consider squaring both sides of the equation.) (HINT: Recall that $\|\vec{v}\|^2 = \vec{v}^\top \vec{v}$.) (HINT: It may be useful to note that $\left(U^\top U\right)_{ij}$ (the (i,j)th entry of $U^\top U$) is $\vec{u}_i^\top \vec{u}_j$.)*

(b) **Next, show that $\|\vec{v} + \vec{u}\|^2 = \|\vec{v}\|^2 + \|\vec{u}\|^2$ for nonzero $\vec{u}, \vec{v} \in \mathbb{R}^n$ if and only if $\vec{u}$ and $\vec{v}$ are orthogonal.**

(c) Recall from the previous homework that we determined that the value of $Q_2^\top \vec{p}$ does not matter. More explicitly, as long as $Q_1^\top \vec{p} = \left(R_1^\top\right)^{-1} \vec{s}$, we can choose $\vec{p}$ such that $Q_2^\top \vec{p}$ can be any value, and it will not invalidate our solution. **First, show that we can write $\vec{p}$ as**

$$\vec{p} = Q_1 Q_1^\top \vec{p} + Q_2 Q_2^\top \vec{p} \tag{5}$$

Then, using this representation as well as the previous two parts, conclude that we should set $Q_2^\top \vec{p} = \vec{0}$.

(HINT: For the first part of this problem, there are many ways to approach it, but one way would be to say that $\vec{p} = \underbrace{QQ^\top}_{I} \vec{p}$, write $Q := \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$, and then use block matrix multiplication. Another way to approach this part would be to consider projecting $\vec{p}$ onto the column space of Q_1 and Q_2 separately (and then show why the summation of these two projections will equal $\vec{p}$).)

3. SVD System ID

Previously, we saw instances for how to solve system ID problems when $D \in \mathbb{R}^{m\times n}$ is full rank (separately, for $m > n$ and $n > m$). Now, let us consider more generally the following problem of estimating $\vec{p}$ in

$$D\vec{p} = \vec{s} \tag{6}$$

where $\vec{p} \in \mathbb{R}^n$, $\vec{s} \in \mathbb{R}^m$, and $D \in \mathbb{R}^{m\times n}$. We assume that $\text{rank}(D) = r < \min(m, n)$, and we do not make any further assumptions on the relationship between m and n. Let's assume that D has an SVD given by

$$D = U\Sigma V^\top \tag{7}$$

where $U \in \mathbb{R}^{m\times m}$ and $V \in \mathbb{R}^{n\times n}$ are orthonormal matrices. $\Sigma \in \mathbb{R}^{m\times n}$ has the following form:

$$\Sigma = \begin{bmatrix} \Sigma_r & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix} \tag{8}$$

where $\Sigma_r = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$ is a $r \times r$ diagonal matrix with nonzero elements along its diagonal.

Using this problem setup, we can rewrite our original system ID problem as

$$U\Sigma V^\top \vec{p} = \vec{s} \tag{9}$$

Our goal is to find $\vec{p}$ with smallest norm that best estimates $\vec{s}$.

For notational convenience, denote $U := \begin{bmatrix} U_r & U_{m-r} \end{bmatrix}$ where $U_r \in \mathbb{R}^{m\times r}$ is a matrix with the first r columns of U and $U_{m-r} \in \mathbb{R}^{m\times(m-r)}$ is a matrix with the last $m - r$ columns of U. Also, denote $V := \begin{bmatrix} V_r & V_{n-r} \end{bmatrix}$ where $V_r \in \mathbb{R}^{n\times r}$ is a matrix that has the first r columns of V, and $V_{n-r} \in \mathbb{R}^{n\times(n-r)}$ is a matrix that has the last $n - r$ columns of V. From SVD properties, we know that the columns of U_r form an orthonormal basis for $\text{Col}(D)$ and that the columns of V_{n-r} form an orthonormal basis for $\text{Null}(D)$.

(a) **Using the fact that U is orthonormal, show that $\Sigma V^\top \vec{p} = U^\top \vec{s}$.**

(b) **Show that we can write $\vec{p} = \begin{bmatrix} V_r & V_{n-r} \end{bmatrix} \begin{bmatrix} \vec{\alpha} \\ \vec{\beta} \end{bmatrix}$ for some vectors $\vec{\alpha}$ and $\vec{\beta}$ (i.e., find $\vec{\alpha} \in \mathbb{R}^r$ and $\vec{\beta} \in \mathbb{R}^{n-r}$). Show that changing $\vec{\beta}$ will not affect the result of $D\vec{p}$ and that we should set $\vec{\beta} = \vec{0}$ if we want to minimize $\|p\|$.** This result justifies that we are achieving a $\vec{p}$ with smallest norm. *(HINT: For the second part of this question, consider using block matrix multiplication on $\begin{bmatrix} V_r & V_{n-r} \end{bmatrix} \begin{bmatrix} \vec{\alpha} \\ \vec{\beta} \end{bmatrix}$ (don't substitute for $\vec{\alpha}$ and $\vec{\beta}$) and leverage the result from the QR decomposition problem on this homework.)*

(c) From the previous part, we can rewrite $\vec{p} = V \begin{bmatrix} \vec{\alpha} \\ \vec{0} \end{bmatrix}$. This simplifies our system ID problem as follows:

$$\Sigma V^\top V \begin{bmatrix} \vec{\alpha} \\ \vec{0} \end{bmatrix} = U^\top \vec{s} \tag{10}$$

$$\Sigma \begin{bmatrix} \vec{\alpha} \\ \vec{0} \end{bmatrix} = U^\top \vec{s} \tag{11}$$

Simplify the left hand side of eq. (11) using eq. (8). Rewrite $U^\top \vec{s}$ **as** $\begin{bmatrix} U_r^\top \vec{s} \\ U_{m-r}^\top \vec{s} \end{bmatrix}$ **and find an expression for** $\vec{\alpha}$**.** *(HINT: Block matrix multiplication will work like normal matrix-vector multiplication here since* $\Sigma_r \in \mathbb{R}^{r \times r}$ *and* $\vec{\alpha} \in \mathbb{R}^r$*.)*

(d) **Use the previous part to come up with a solution for** $\vec{p}$**.**

(e) From the concept of projections, we know that the optimal solution for $\vec{p}$ satisfies the property that the projection error, namely $\vec{s} - D\vec{p}$, is orthogonal to the projection itself, namely $D\vec{p}$. Write $\vec{s} := \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} \vec{w} \\ \vec{z} \end{bmatrix}$ for some vectors $\vec{w} \in \mathbb{R}^r$ and $\vec{z} \in \mathbb{R}^{m-r}$. **Find** $\vec{w}$ **and** $\vec{z}$**. Using this, show that our solution for** $\vec{p}$ **is optimal.**

4. Frobenius Norm

In this problem we will investigate the basic properties of the Frobenius norm.

Similar to how the norm of vector $\vec{x} \in \mathbb{R}^n$ is defined as $\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}$, the Frobenius norm of a matrix $A \in \mathbb{R}^{m \times n}$ is defined as

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2}. \tag{12}$$

A_{ij} is the entry in the i^{th} row and the j^{th} column. This is basically the norm that comes from treating a matrix like a big vector filled with numbers.

(a) With the above definitions, **show that for a** 2×2 **matrix** A**:**

$$\|A\|_F = \sqrt{\operatorname{tr}\left(A^\top A\right)}. \tag{13}$$

Note: The trace of a matrix is the sum of its diagonal entries. For example, let $A \in \mathbb{R}^{m \times n}$, then,

$$\operatorname{tr}(A) = \sum_{i=1}^{\min(n,m)} A_{ii} \tag{14}$$

Think about how/whether this expression eq. (13) generalizes to general $m \times n$ matrices.

(b) **Show for any matrix** $A \in \mathbb{R}^{m \times n}$**:**

$$\|A\|_F = \left\|A^\top\right\|_F \tag{15}$$

A purely written or mathematical solution will be sufficient for this problem.

(HINT: For the mathematical solution, use the trace interpretation from eq. (12).)

(c) **Show that if** U **and** V **are square orthonormal matrices, then**

$$\|UA\|_F = \|AV\|_F = \|A\|_F. \tag{16}$$

(HINT: Use the trace interpretation from part (a) and the equation from part (b).)

(d) **Use the SVD decomposition to show that** $\|A\|_F = \sqrt{\sum_{i=1}^{n} \sigma_i^2}$**, where** $\sigma_1, \ldots, \sigma_n$ **are the singular values of** A**.**

(HINT: The previous part might be quite useful.)

Contributors:

- Siddharth Iyer.
- Yu-Yun Dai.
- Sanjit Batra.
- Anant Sahai.
- Sidney Buchbinder.
- Gaoyue Zhou.
- Druv Pai.
- Anish Muthali.
- Anirudh Rengarajan.
- Aditya Arun.
- Kourosh Hakhamaneshi.
- Antroy Roy Chowdhury.