Homework 11

This homework is due on Saturday, November 11, 2023 at 11:59PM. Self-grades and HW Resubmissions are due the following Saturday, November 18, 2023 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?)

Solution: Using the fact that S is real and symmetric so $\overline{S} = S = S^\top$, we get

$$\overline{\vec{v}}^\top (S\vec{v}) = \overline{\vec{v}}^\top (\lambda\vec{v}) = \lambda\overline{\vec{v}}^\top\vec{v} = \lambda\|\vec{v}\|^2 \tag{1}$$

$$(\overline{\vec{v}}^\top S)\vec{v} = (S\overline{\vec{v}})^\top\vec{v} = (\overline{S\vec{v}})^\top\vec{v} = (\overline{\lambda\vec{v}})^\top\vec{v} = \overline{\lambda}(\overline{\vec{v}}^\top\vec{v}) = \overline{\lambda}\|\vec{v}\|^2. \tag{2}$$

where $\|\vec{v}\|^2 = \sum_{i=1}^n |v_i|^2$. Since $\vec{v} \neq \vec{0}_n$, we know that $\|\vec{v}\|^2 > 0$, and so $\lambda = \overline{\lambda}$. Thus λ is real.

(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."

[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.

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 $\begin{bmatrix}-1\end{bmatrix}$.)

Solution: Let $S = \begin{bmatrix}s\end{bmatrix}$. Since $\begin{bmatrix}1\end{bmatrix}$ is a real and orthonormal matrix, and $\begin{bmatrix}s\end{bmatrix}$ is diagonal, $S = \begin{bmatrix}1\end{bmatrix}\begin{bmatrix}s\end{bmatrix}\begin{bmatrix}1\end{bmatrix}^\top$ is an orthonormal diagonalization of S. Since $S\vec{x} = s\vec{x}$ for all $\vec{x} \in \mathbb{R}^1$, we see that $\begin{bmatrix}s\end{bmatrix}$ is a matrix of eigenvalues of S, and also that any vector is an eigenvector so an orthonormal matrix of eigenvectors of S is $\begin{bmatrix}1\end{bmatrix}$.

It is also possible to answer with $S = \begin{bmatrix}-1\end{bmatrix}\begin{bmatrix}s\end{bmatrix}\begin{bmatrix}-1\end{bmatrix}^\top$.

(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:**

$$Q^\top S Q = \begin{bmatrix}\lambda & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & S_0\end{bmatrix} \qquad \textbf{where} \qquad S_0 := \widetilde{Q}^\top S \widetilde{Q}. \tag{3}$$

(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.)

Solution: We use block-matrix multiplication:

$$\begin{aligned} Q^\top S Q &= \begin{bmatrix}\vec{q}^\top \\ \widetilde{Q}^\top\end{bmatrix} S \begin{bmatrix}\vec{q} & \widetilde{Q}\end{bmatrix} &(4)\\ &= \begin{bmatrix}\vec{q}^\top \\ \widetilde{Q}^\top\end{bmatrix} \begin{bmatrix}S\vec{q} & S\widetilde{Q}\end{bmatrix} &(5)\\ &= \begin{bmatrix}\vec{q}^\top \\ \widetilde{Q}^\top\end{bmatrix} \begin{bmatrix}\lambda\vec{q} & S\widetilde{Q}\end{bmatrix} &(6)\\ &= \begin{bmatrix}\lambda\vec{q}^\top\vec{q} & \lambda\vec{q}^\top\widetilde{Q} \\ \lambda\widetilde{Q}^\top\vec{q} & \widetilde{Q}^\top S\widetilde{Q}\end{bmatrix}. &(7)\end{aligned}$$

To simplify, we follow the hint, and expand $Q^\top Q = I_n$.

$$Q^\top Q = I_n \tag{8}$$

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

$$\begin{bmatrix} \vec{q}^\top \\ \widetilde{Q}^\top \end{bmatrix} \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix} = \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & I_{n-1} \end{bmatrix} \tag{9}$$

$$\begin{bmatrix} \vec{q}^\top\vec{q} & \vec{q}^\top\widetilde{Q} \\ \widetilde{Q}^\top\vec{q} & \widetilde{Q}^\top\widetilde{Q} \end{bmatrix} = \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & I_{n-1} \end{bmatrix}. \tag{10}$$

Thus we get $\vec{q}^\top\vec{q} = 1$, $\widetilde{Q}^\top\vec{q} = \vec{0}_{n-1}$, $\vec{q}^\top\widetilde{Q} = \vec{0}_{n-1}^\top$, and so we have

$$Q^\top SQ = \begin{bmatrix} \lambda\vec{q}^\top\vec{q} & \vec{q}^\top\widetilde{Q} \\ \lambda\widetilde{Q}^\top\vec{q} & \widetilde{Q}^\top S\widetilde{Q} \end{bmatrix} = \begin{bmatrix} \lambda & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & \widetilde{Q}^\top S\widetilde{Q} \end{bmatrix} \tag{11}$$

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

Solution: We show that $S_0^\top = S_0$.

$$\begin{aligned} S_0^\top &= (\widetilde{Q}^\top S\widetilde{Q})^\top && (12)\\ &= (\widetilde{Q})^\top(S)^\top(\widetilde{Q}^\top)^\top && (13)\\ &= \widetilde{Q}^\top S^\top\widetilde{Q} && (14)\\ &= \widetilde{Q}^\top S\widetilde{Q} && (15)\\ &= S_0. && (16)\end{aligned}$$

where the second-to-last equality is because S is symmetric so $S^\top = S$.

It is not necessary to write in the solution, but to show that S_0 is real, note that $\widetilde{Q}$ is real and S is real, so $S_0 = \widetilde{Q}^\top S\widetilde{Q}$ is real as a matrix product of real matrices.

(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 $S_0 = V_0\Lambda_0V_0^\top$, 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

$$V := Q\begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \qquad \textbf{and} \qquad \Lambda := V^\top SV. \tag{17}$$

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

Solution: We compute $V^\top V$.

$$\begin{aligned} V^\top V &= \left(Q\begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix}\right)^\top\left(Q\begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix}\right) && (18)\\ &= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix}^\top Q^\top Q\begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} && (19)\\ &= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix}^\top\begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} && (20)\\ &= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0^\top \end{bmatrix}\begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} && (21)\\ &= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0^\top V_0 \end{bmatrix} && (22)\end{aligned}$$

$$= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & I_{n-1} \end{bmatrix} \tag{23}$$

$$= I_n. \tag{24}$$

It is not necessary to write in the solution, but to show that V is real, note that Q is real and $\begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix}$ is real, so $V = Q \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix}$ is real as a matrix product of real matrices.

ii. **Show that Λ is diagonal.**

Solution: We compute $\Lambda = V^\top S V$.

$$\Lambda = V^\top S V \tag{25}$$

$$= \left(Q \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \right)^\top S \left(Q \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \right) \tag{26}$$

$$= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix}^\top Q^\top S Q \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \tag{27}$$

$$= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0^\top \end{bmatrix} \begin{bmatrix} \lambda & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & S_0 \end{bmatrix} \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \tag{28}$$

$$= \begin{bmatrix} \lambda & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0^\top S_0 V_0 \end{bmatrix} \tag{29}$$

$$= \begin{bmatrix} \lambda & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & \Lambda_0 \end{bmatrix}. \tag{30}$$

We already know Λ_0 is diagonal so Λ is diagonal.

It is not necessary to write in the solution, but to show that Λ is real, note that λ is real (shown in part (a)) and Λ_0 is real by the induction, so $\Lambda = \begin{bmatrix} \lambda & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & \Lambda_0 \end{bmatrix}$ is real.

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

Solution: We have

$$\Lambda = V^\top S V \tag{31}$$

$$\implies V\Lambda = SV \tag{32}$$

$$\implies V\Lambda V^\top = S. \tag{33}$$

(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!

2. SVD and the fundamental subspaces

Consider a matrix $A \in \mathbb{R}^{m\times n}$ with $\text{rank}(A) = r$. The compact SVD of A is given by $A = U_r \Sigma_r V_r^\top$ where

$$U_r = \begin{bmatrix} \vec{u}_1 \cdots \vec{u}_r \end{bmatrix} \in \mathbb{R}^{m\times r}, \quad \Sigma_r = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_r \end{bmatrix} \in \mathbb{R}^{r\times r}, \quad V_r = \begin{bmatrix} \vec{v}_1 \cdots \vec{v}_r \end{bmatrix} \in \mathbb{R}^{n\times r}$$

with $\sigma_1 \geq \cdots \geq \sigma_r > 0$ being the singular values of A.

(a) **Which one of the following sets is always guaranteed to form an orthonormal basis for** $\text{Col}(A)$**?** *(Please fill in one of the circles for the options below. You will only be graded on your final answer.)*

i. $\{\vec{u}_1, \cdots, \vec{u}_r\}$
ii. $\{\sigma_1\vec{u}_1, \ldots, \sigma_r\vec{u}_r\}$
iii. $\{\vec{v}_1, \ldots, \vec{v}_r\}$
iv. $\{\sigma_1\vec{v}_1, \ldots, \sigma_r\vec{v}_r\}$

Option	i	ii	iii	iv
Answer	◯	◯	◯	◯

Solution: Only options i. and ii. form bases for $\text{Col}(A)$. Since the question is asking for an orthonormal basis the correct answer is i.

(b) **Which one of the following sets is always guaranteed to form an orthonormal basis for** $\text{Col}(A^\top)$**?** *(Please fill in one of the circles for the options below. You will only be graded on your final answer.)*

i. $\{\vec{u}_1, \cdots, \vec{u}_r\}$
ii. $\{\sigma_1\vec{u}_1, \ldots, \sigma_r\vec{u}_r\}$
iii. $\{\vec{v}_1, \ldots, \vec{v}_r\}$
iv. $\{\sigma_1\vec{v}_1, \ldots, \sigma_r\vec{v}_r\}$

Option	i	ii	iii	iv
Answer	◯	◯	◯	◯

Solution: Only options iii. and iv. form bases for $\text{Col}(A^\top)$. Since the question is asking for an orthonormal basis the correct answer is iii.

Now suppose that the considered A matrix has the following compact SVD components:

$$U_r = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} \\ 1 & 0 \\ 0 & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \Sigma_r = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, \quad V_r = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

(c) **Using the given compact SVD, state α, where α is the tightest upper bound $\|A\vec{x}\| \leq \alpha$ for any $\vec{x}$ such that $\|\vec{x}\| \leq 1$.**

Solution: The largest amplification factor of a matrix is given by its largest singular value. Thus for A this is 2.

(d) **Given the compact SVD, which of the following provides a valid full SVD for $A = U\Sigma V^\top$?**
(Please fill in one of the circles for the options below. You will only be graded on your final answer.)

i. $U = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$, $\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

ii. $U = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & 0 \\ 1 & 0 & -1 \\ 0 & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}$, $\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

iii. $U = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$, $\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $V = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

iv. $U = \begin{bmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}$, $\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$, $V = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

Option	i	ii	iii	iv
Answer	◯	◯	◯	◯

Solution: Only option iv. forms a valid SVD as it has orthonormal U and V matrices and has strictly positive singular values ordered from largest to smallest in the Σ matrix.

3. SVD of a matrix with orthogonal columns

Let $A = \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \in \mathbb{R}^{m \times n}$ where $\vec{a}_i^\top \vec{a}_j = 0$ for all $1 \leq i, j \leq n$ such that $i \neq j$, and $\vec{a}_i^\top \vec{a}_i \neq 0$ for all $i = 1, \ldots, n$. **What is the set of singular values of A *for all such matrices* A?**

(Please fill in one of the circles for the options below. You will only be graded on your final answer.)

(a) $\{0\}$ (all zero)

(b) $\{\sqrt{\|\vec{a}_1\|}, \ldots, \sqrt{\|\vec{a}_n\|}\}$

(c) $\{\|\vec{a}_1\|, \ldots, \|\vec{a}_n\|\}$

(d) $\{\|\vec{a}_1\|^2, \ldots, \|\vec{a}_n\|^2\}$

(e) $\{1\}$ (all one)

Option	a	b	c	d	e
Answer	◯	◯	◯	◯	◯

Solution: (c). The singular values of A are the square roots of the eigenvalues of $A^\top A$. Evaluating this, we have

$$A^\top A = \begin{bmatrix} \vec{a}_1^\top \\ \vdots \\ \vec{a}_n^\top \end{bmatrix} \begin{bmatrix} \vec{a}_1 & \cdots & \vec{a}_n \end{bmatrix} \tag{34}$$

$$= \begin{bmatrix} \vec{a}_1^\top \vec{a}_1 & \cdots & \vec{a}_1^\top \vec{a}_n \\ \vdots & \ddots & \vdots \\ \vec{a}_n^\top \vec{a}_1 & \cdots & \vec{a}_n^\top \vec{a}_n \end{bmatrix} \tag{35}$$

$$= \begin{bmatrix} \|\vec{a}_1\|^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|\vec{a}_n\|^2 \end{bmatrix} \tag{36}$$

Since this is a diagonal matrix, we can read off its eigenvalues as $\|\vec{a}_1\|^2, \ldots, \|\vec{a}_n\|^2$. Then the singular values of A are $\{\|\vec{a}_1\|, \ldots, \|\vec{a}_n\|\}$ (note that we ask for the *set* of singular values in the question because singular values are specified in order of decreasing size).

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

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2}. \tag{37}$$

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**:**

$$\|A\|_F = \sqrt{\text{tr}(A^\top A)}. \tag{38}$$

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

$$\text{tr}(A) = \sum_{i=1}^{\min(n,m)} A_{ii} \tag{39}$$

Think about how/whether this expression eq. (38) generalizes to general $m \times n$ matrices.

Solution: This proof is for the general case of $m \times n$ matrices. You should give yourself full credit if you did this calculation only on the 2×2 case.

$$\text{tr}\left(A^\top A\right) = \sum_{i=1}^{n} (A^\top A)_{ii} \tag{40}$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} (A^\top)_{ij} A_{ji} \right) \tag{41}$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} A_{ji} A_{ji} \right) \tag{42}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} (A_{ji}^2) \tag{43}$$

$$= \|A\|_F^2 \tag{44}$$

In the above solution, step eq. (40) writes out the trace definition, step eq. (41) expands the matrix multiplication on the diagonal indices (i.e. index (i,i) is the real inner product of row i and column i), step eq. (42) applies the definition of matrix transpose, and the last two steps collects the result into the definition of Frobenius norm.

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

$$\|A\|_F = \left\|A^\top\right\|_F \tag{45}$$

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

(HINT: For the mathematical solution, use the trace interpretation from eq. (37).)

Solution: Written Solution: Intuitively, we know that since the Frobenius norm sums the squares of the elements of a matrix and that transposes change the orientation rather than the contents of the matrix, the norms should be equivalent.

Mathematical Solution: Assume without loss of generality that we sum row by row when calculating the Frobenius norm.

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|^2} \tag{46}$$

$$= \sqrt{\sum_{j=1}^{n}\sum_{i=1}^{m}|A_{ij}|^2} \tag{47}$$

$$= \sqrt{\sum_{j=1}^{n}\sum_{i=1}^{m}|A^{\top}_{ji}|^2} \tag{48}$$

$$= \left\|A^{\top}\right\|_F \tag{49}$$

The first equality is given. The second equality stems from swapping from summing by each row to summing by each column (equivalent since contents don't change). The third equality stems from the fact that 2D transposed matrices have the indices of their corresponding elements swapped. Finally, since we defined the Frobenius norm as a row-wise sum of the matrix, we know this last equation to be the Frobenius norm of the transposed matrix.

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

$$\|UA\|_F = \|AV\|_F = \|A\|_F\,. \tag{50}$$

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

Solution: The direct path is just to compute using the trace formula:

$$\|UA\|_F = \sqrt{\operatorname{tr}\left((UA)^{\top}(UA)\right)} = \sqrt{\operatorname{tr}\left(A^{\top}U^{\top}UA\right)} = \sqrt{\operatorname{tr}\left(A^{\top}A\right)} = \|A\|_F \tag{51}$$

Another path is to note that the Frobenius norm squared of a matrix is the sum of squared Euclidean norms of the columns of the matrix. Matrix multiplication UA proceeds to act on each column of A independently. None of those norms change since U is orthonormal, and so the Frobenius norm also doesn't change.

To show the second equality, we must first note that $\|A^{\top}\|_F = \|A\|_F$, because we are just summing over the same numbers, just in a different order. Hence:

$$\|AV\|_F = \left\|(AV)^{\top}\right\|_F = \left\|V^{\top}A^{\top}\right\|_F \tag{52}$$

But the transpose of a square orthonormal matrix is also orthonormal, hence this case reduces to the previous case, implying

$$\left\|V^{\top}A^{\top}\right\|_F = \left\|A^{\top}\right\|_F = \|A\|_F \tag{53}$$

(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.)

Solution:

$$\|A\|_F = \left\|U\Sigma V^\top\right\|_F = \left\|\Sigma V^\top\right\|_F = \|\Sigma\|_F \tag{54}$$

$$= \sqrt{\sum_{i=1}^{n}\sigma_i^2} \tag{55}$$

(e) **Show that for any matrix $A \in \mathbb{R}^{m\times n}$ and any vector $\vec{x} \in \mathbb{R}^n$:**

$$\|A\vec{x}\|^2 \leq \|A\|_F^2\|\vec{x}\|^2 \tag{56}$$

(HINT: Use the summation form of matrix multiplication to find an expression for each element of $A\vec{x}$ and use this to find the expression for $\|A\vec{x}\|^2$. Then, use the fact that $|\sum ab|^2 \leq \left(\sum |a|^2\right)\left(|\sum |b|^2\right)$ (called the Cauchy-Schwarz inequality).)

Solution: To start out, we can write each element of $A\vec{x}$ as the dot product or inner product of the rows of A and the column vector $\vec{x}$, which in summation notation is equivalent to:

$$(A\vec{x})_i = \sum_{j=1}^{n} a_{ij}x_j \tag{57}$$

The value of $\|A\vec{x}\|^2$ can be written as a sum of the magnitude squared of each of the elements:

$$\|A\vec{x}\|^2 = \sum_{i=1}^{m}|(A\vec{x})_i|^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}x_j|^2 \tag{58}$$

Then, we can use the Cauchy-Schwarz inequality provided by the hint to relate this to $\|A\|_F^2\|\vec{x}\|^2$:

$$\|A\vec{x}\|^2 = \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}x_j|^2 \tag{59}$$

$$\leq \sum_{i=1}^{m}\left(\sum_{j=1}^{n}|a_{ij}|^2\right)\left(\sum_{k=1}^{n}|x_k|^2\right) \tag{60}$$

$$\leq \left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)\left(\sum_{k=1}^{n}|x_k|^2\right) \tag{61}$$

$$\leq \|A\|_F^2\|\vec{x}\|^2 \tag{62}$$

Contributors:

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