Homework 6

This homework is due on Sunday, July 31 at 11:59 pm PT.

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."

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

[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.

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:

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

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

$$\begin{aligned} Q^\top Q &= I_n &\quad (8) \\ \begin{bmatrix} \vec{q}^\top \\ \widetilde{Q}^\top \end{bmatrix} \begin{bmatrix} \vec{q} & \widetilde{Q} \end{bmatrix} &= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & I_{n-1} \end{bmatrix} &\quad (9) \end{aligned}$$

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

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

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

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

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

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

$$S_0^\top = (\widetilde{Q}^\top S \widetilde{Q})^\top \tag{12}$$
$$= (\widetilde{Q})^\top (S)^\top (\widetilde{Q}^\top)^\top \tag{13}$$
$$= \widetilde{Q}^\top S^\top \widetilde{Q} \tag{14}$$
$$= \widetilde{Q}^\top S \widetilde{Q} \tag{15}$$
$$= S_0. \tag{16}$$

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 $S_0 = V_0 \Lambda_0 V_0^\top$, 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

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

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

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

$$V^\top V = \left(Q \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \right)^\top \left(Q \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \right) \tag{18}$$
$$= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix}^\top Q^\top Q \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \tag{19}$$
$$= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix}^\top \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \tag{20}$$
$$= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0^\top \end{bmatrix} \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0 \end{bmatrix} \tag{21}$$
$$= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & V_0^\top V_0 \end{bmatrix} \tag{22}$$
$$= \begin{bmatrix} 1 & \vec{0}_{n-1}^\top \\ \vec{0}_{n-1} & I_{n-1} \end{bmatrix} \tag{23}$$

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

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 $\Lambda = V^\top SV$.

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

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

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

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

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

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

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

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

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

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

(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

(a) Consider the matrix

$$A = \begin{bmatrix} -1 & 1 & 5 \\ 3 & 1 & -1 \\ 2 & -1 & 4 \end{bmatrix}.$$

Observe that the columns of matrix A are mutually orthogonal with norms $\sqrt{14}, \sqrt{3}, \sqrt{42}$.

Verify numerically that columns $\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$ **and** $\begin{bmatrix} 5 \\ -1 \\ 4 \end{bmatrix}$ **are orthogonal to each other.**

Solution: Taking the inner product of the two vectors, we have

$$\left\langle \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ -1 \\ 4 \end{bmatrix} \right\rangle = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}^\top \begin{bmatrix} 5 \\ -1 \\ 4 \end{bmatrix} = 5 - 1 - 4 = 0. \tag{34}$$

So the two columns are orthogonal to each other.

(b) **Write $A = BD$, where B is an orthonormal matrix and D is a diagonal matrix.** What is B? What is D?

Solution: We compute the norm for each column and divide each column by its norm to obtain matrix B. Matrix D is formed by placing the norms on the diagonal.

$$B = \begin{bmatrix} -\frac{1}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{5}{\sqrt{42}} \\ \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{42}} \\ \frac{2}{\sqrt{14}} & -\frac{1}{\sqrt{3}} & \frac{4}{\sqrt{42}} \end{bmatrix} \tag{35}$$

$$D = \begin{bmatrix} \sqrt{14} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{42} \end{bmatrix} \tag{36}$$

(c) **Write out a singular value decomposition of $A = U\Sigma V^\top$ using the previous part.** Note the ordering of the singular values in Σ should be from the largest to smallest. *(HINT: There is no need to compute the eigenvalues of anything. Use Theorem 14, Note 16.)*

Solution: Using part b, we can write

$$A = BD = BDI = \begin{bmatrix} -\frac{1}{\sqrt{14}} & \frac{1}{\sqrt{3}} & \frac{5}{\sqrt{42}} \\ \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{42}} \\ \frac{2}{\sqrt{14}} & -\frac{1}{\sqrt{3}} & \frac{4}{\sqrt{42}} \end{bmatrix} \begin{bmatrix} \sqrt{14} & 0 & 0 \\ 0 & \sqrt{3} & 0 \\ 0 & 0 & \sqrt{42} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{37}$$

Reordering the columns and rows of B and I so that the diagonal entries of D are in non-decreasing order, we have

$$\begin{bmatrix} \frac{5}{\sqrt{42}} & -\frac{1}{\sqrt{14}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{42}} & \frac{3}{\sqrt{14}} & \frac{1}{\sqrt{3}} \\ \frac{4}{\sqrt{42}} & \frac{2}{\sqrt{14}} & -\frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{42} & 0 & 0 \\ 0 & \sqrt{14} & 0 \\ 0 & 0 & \sqrt{3} \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \tag{38}$$

Then by Note 16, Theorem 14, this is an SVD of A.

(d) Given the matrix

$$A = \frac{1}{\sqrt{50}} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} + \frac{3}{\sqrt{50}} \begin{bmatrix} -4 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix}, \tag{39}$$

write out a singular value decomposition of matrix A in the form $U\Sigma V^\top$. Note the ordering of the singular values in Σ should be from the largest to smallest. *(HINT: You don't have to compute any eigenvalues for this. Some useful observations are that*

$$\begin{bmatrix} 3, 4 \end{bmatrix} \begin{bmatrix} -4 \\ 3 \end{bmatrix} = 0, \quad \begin{bmatrix} 1, -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 0, \quad \| \begin{bmatrix} 3 \\ 4 \end{bmatrix} \| = \| \begin{bmatrix} -4 \\ 3 \end{bmatrix} \| = 5, \quad \| \begin{bmatrix} 1 \\ -1 \end{bmatrix} \| = \| \begin{bmatrix} 1 \\ 1 \end{bmatrix} \| = \sqrt{2}.$$

)

Solution:

The singular value decomposition can be written in the form

$$A = \sum_{i=1}^{2} \sigma_i \vec{u}_i \vec{v}_i^\top, \tag{40}$$

with unit orthonormal vectors $\{\vec{u}_i\}$ and $\{\vec{v}_i\}$. From the given observations, we can see that the vectors we were provided are orthogonal, so we can just normalize them to get the desired answer. Taking it step by step:

$$\begin{aligned} A &= \frac{1}{\sqrt{50}} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} + \frac{3}{\sqrt{50}} \begin{bmatrix} -4 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} & (41) \\ &= \frac{5}{\sqrt{50}} \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \begin{bmatrix} 1 & -1 \end{bmatrix} + \frac{3 \cdot 5}{\sqrt{50}} \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} & (42) \\ &= \frac{5\sqrt{2}}{\sqrt{50}} \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} + \frac{3 \cdot 5\sqrt{2}}{\sqrt{50}} \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} & (43) \\ &= 1 \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix} + 3 \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}. & (44) \end{aligned}$$

From this, we can derive

$$\vec{u}_1 = \begin{bmatrix} -\frac{4}{5} \\ \frac{3}{5} \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix}, \vec{v}_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \tag{45}$$

and corresponding singular values $\sigma_1 = 3, \sigma_2 = 1$ because we need to order them by size in decreasing order. This gives the singular value decomposition

$$A = \begin{bmatrix} -\frac{4}{5} & \frac{3}{5} \\ \frac{3}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^\top \tag{46}$$

(e) Define the matrix

$$A = \begin{bmatrix} -1 & 4 \\ 1 & 4 \end{bmatrix}.$$

Find the SVD of A by following the standard algorithm introduced in Note 16, i.e. by computing the eigendecomposition of $A^\top A$. Also find the eigenvectors and eigenvalues of A. Is there a relationship between the eigenvalues or eigenvectors of A with the SVD of A?

Solution: Since we have a square matrix, we will arbitrarily use $A^\top A$ for our SVD:

$$A^\top A = \begin{bmatrix} -1 & 1 \\ 4 & 4 \end{bmatrix} \begin{bmatrix} -1 & 4 \\ 1 & 4 \end{bmatrix} \tag{47}$$

$$= \begin{bmatrix} 2 & 0 \\ 0 & 32 \end{bmatrix} \tag{48}$$

Next, we find the eigenvalues of the above matrix.

$$\det(A^\top A - \lambda I) = (2 - \lambda)(32 - \lambda) = 0 \tag{49}$$

Hence, the eigenvalues are $\lambda_1 = 32$ and $\lambda_2 = 2$, and the singular values are $\sigma_1 = \sqrt{32} = 4\sqrt{2}$ and $\sigma_2 = \sqrt{2}$.

Next, we find the right singular vectors (i.e. the columns of V). Finding null$(A^\top A - \lambda_1 I)$ and null$(A^\top A - \lambda_2 I)$ will give us $\vec{v}_1$ and $\vec{v}_2$ respectively.

Hence, $\vec{v}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and $\vec{v}_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ (the eigenvectors are already normalized here).

Lastly, we find the right singular vectors (the columns of U)

$$\vec{u}_1 = \frac{1}{\sigma_1} A v_1 \tag{50}$$

$$= \frac{1}{4\sqrt{2}} \begin{bmatrix} -1 & 4 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} \tag{51}$$

$$= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \tag{52}$$

Similarly, we get $\vec{u}_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$.

So the full SVD representation of A is given below

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & \sqrt{2} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{53}$$

Now that we have found the SVD of A, we will find the eigenvalues and eigenvectors of A. Let us start with the eigenvalues:

$$\det(A - \lambda I) = (-1 - \lambda)(4 - \lambda) - 4 = \lambda^2 - 3\lambda - 8 = 0 \tag{54}$$

Using the quadratic formula, the eigenvalues are $\lambda_1 = \frac{3+\sqrt{41}}{2} \approx 4.7$ and $\lambda_2 = \frac{3-\sqrt{41}}{2} \approx -1.7$.

Since we already used $\vec{v}_1, \vec{v}_2$ for the SVD, let us denote the eigenvectors of A as $\vec{r}_1, \vec{r}_2$.

Finding null$(A - \lambda_1 I)$ and null$(A - \lambda_2 I)$ will give us $\vec{r}_1$ and $\vec{r}_2$ respectively.

Hence, the normalized eigenvectors of A are $\vec{r}_1 \approx \begin{bmatrix} -0.98 \\ 0.17 \end{bmatrix}$ and $\vec{r}_2 \approx \begin{bmatrix} -0.57 \\ -0.82 \end{bmatrix}$.

We notice that there is no relationship between the eigenvalues or eigenvectors of A with the SVD of A.

3. The Moore-Penrose pseudoinverse

Say we have a set of linear equations given by $A\vec{x} = \vec{y}$. If A is invertible, then the unique solution for $\vec{x}$ is $\vec{x} = A^{-1}\vec{y}$. However, what if A is not a square matrix, and we still wanted to find an $\vec{x}$ that satisfied $A\vec{x} = \vec{y}$? We know that we could use a linear least-squares approach for "tall" matrices A where it isn't possible to find a solution that exactly matches all the measurements. The linear least-squares solution gives us a reasonable answer that asks for the "best" match in terms of reducing the norm of the error vector.

How about when the matrix A is "wide", i.e. A has more columns than rows? In this case, there are generally going to be lots of possible solutions — so which should we choose? To address this, we introduce the *Moore-Penrose pseudoinverse*, which generalizes the idea of the matrix inverse and can be calculated using the singular value decomposition.

Since the SVD of a matrix always exists, the Moore-Penrose pseudoinverse does as well. Another useful property of the Moore-Penrose pseudoinverse $A^\dagger$ is that the solution it gives, $\hat{\vec{x}} = A^\dagger\vec{y}$, satisfies a minimality property: $\|\hat{\vec{x}}\| \leq \|\vec{z}\|$ for all $\vec{z}$ such that $A\vec{z} = \vec{y}$.

(a) Say we have the matrix

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix}.$$

To find the Moore-Penrose pseudoinverse we start by calculating the SVD of A. That is to say, we find orthonormal matrices U and V, and diagonal matrix Σ, such that $A = U\Sigma V^\top$.

Here we give you the decomposition of A as:

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{55}$$

where:

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{56}$$

$$\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \tag{57}$$

$$V^\top = \begin{bmatrix} 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{58}$$

It is a good idea to be able to calculate the SVD yourself as you may be asked to solve similar questions on your own in the exam. **You do not have to do any work for this part.**

Solution: Though you did not have to do any work for deriving the SVD the following solutions will walk you through how to solve for the SVD:

$$A = U\Sigma V^\top \tag{59}$$

$$AA^\top = \begin{bmatrix} 3 & -1 \\ -1 & 3 \end{bmatrix}. \tag{60}$$

Which has characteristic polynomial $\lambda^2 - 6\lambda + 8 = 0$, producing eigenvalues 4 and 2. Solving $Av = \lambda_i v$ produces eigenvectors $[\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}]^\top$ and $[\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^\top$ associated with eigenvalues 4 and 2 respectively. The singular values are the square roots of the eigenvalues of $AA^\top$, so

$$\Sigma = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \tag{61}$$

and

$$U = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{62}$$

We can then solve for the $\vec{v}$ vectors using $A^\top \vec{u}_i = \sigma_i \vec{v}_i$, producing $\vec{v}_1 = [0, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^\top$ and $\vec{v}_2 = [1, 0, 0]^\top$. The last $\vec{v}$ must be orthonormal to the other two, so we can pick $[0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^\top$. The SVD is:

$$A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \begin{bmatrix} 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \tag{63}$$

(b) Suppose we have non-zero singular values $\sigma_1, \ldots, \sigma_r$, and that we have written the SVD matrices so that Σ is in the form

$$\Sigma = \underbrace{\begin{bmatrix} \sigma_1 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \sigma_2 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sigma_r & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{bmatrix}}_{\text{Dimension: } m \times n}. \tag{64}$$

Consider the action of Σ on $\vec{v} \in \mathbb{R}^n$, i.e. $\Sigma\vec{v}$. **What is the effect of Σ on each element of $\vec{v}$?**

Let us define the following matrix:

$$\widetilde{\Sigma} = \underbrace{\begin{bmatrix} \frac{1}{\sigma_1} & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \frac{1}{\sigma_2} & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{\sigma_r} & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & 0 & \ldots & 0 \end{bmatrix}}_{\text{Dimension: } n \times m}. \tag{65}$$

What is $\widetilde{\Sigma}\Sigma$? What is the effect of $\widetilde{\Sigma}\Sigma$ on $\vec{v} \in \mathbb{R}^n$?

Solution: As described in the problem, we can represent $\vec{v}$ as $\vec{v} = [v_1, v_2, ..., v_n]^T$. Then, we have

$$\Sigma\vec{v} = [\sigma_1 v_1, \sigma_2 v_2, ..., \sigma_r v_r, 0, 0, ..., 0]. \tag{66}$$

Therefore, the effect of Σ is to scale the i-th element of $\vec{v}$: v_i by σ_i, when $i \leq r$. When $i > r$, Σ wipes out the original value with a 0.

Following the matrix multiplication, we have

$$\widetilde{\Sigma}\Sigma = \underbrace{\begin{bmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{bmatrix}}_{\text{Dimension: } n \times n}, \tag{67}$$

where $\widetilde{\Sigma}\Sigma$ is a diagonal matrix with the first r diagonal values equal to **1**, and **0** for the rest of the elements. Therefore, $\widetilde{\Sigma}\Sigma\vec{v} = [v_1, v_2, ..., v_r, 0, ..., 0]^T$. This operation keeps the first r values of $\vec{v}$, and turns the rest to **0**.

(c) Consider when $A = U\Sigma V^\top$ acts on $\vec{x}$ to give the result $\vec{y}$, i.e.

$$A\vec{x} = U\Sigma V^\top \vec{x} = \vec{y}. \tag{68}$$

Observe that $V^\top \vec{x}$ rotates $\vec{x}$ without changing its length, and U rotates $\Sigma V^\top \vec{x}$ again. The *Moore-Penrose pseudoinverse* $A^\dagger$ is given as

$$A^\dagger = V\widetilde{\Sigma}U^\top, \tag{69}$$

where $\widetilde{\Sigma}$ is given in (c). Consider if we apply the Moore-Penrose pseudoinverse to find a candidate solution $A^\dagger y$:

$$\vec{y} = U\Sigma V^\top \vec{x} \tag{70}$$

$$A^\dagger y = (V\widetilde{\Sigma}U^\top)(U\Sigma V^\top)\vec{x}. \tag{71}$$

Qualitatively, what are the effects of the matrices V, $\widetilde{\Sigma}$, and $U^\top$ in the Moore-Penrose pseudoinverse when finding a solution?

Solution: Given

$$A^\dagger y = V\widetilde{\Sigma}U^\top(U\Sigma V^\top)\vec{x}, \tag{72}$$

and $U^\top U = I$, we know that $U^\top$ "undoes" the rotation effect of U and brings the vector $U\Sigma V^\top \vec{x}$ back to $\Sigma V^\top \vec{x}$.

Now we have

$$A^\dagger y = V\widetilde{\Sigma}U^\top(U\Sigma V^\top)\vec{x} \tag{73}$$

$$= V\widetilde{\Sigma}\Sigma V^\top \vec{x}. \tag{74}$$

$\widetilde{\Sigma}$ scales the first r values of $\Sigma V^\top \vec{x}$ by $\frac{1}{\sigma_i}$, for $i = 1, 2, 3, ..., r$, and returns $\mathbf{0}$ for the rest. Meanwhile, as shown in part (b), we know that $\widetilde{\Sigma}\Sigma$ keeps the first r values of $V^\top \vec{x}$.

Finally, since $VV^\top = I$, matrix V again "undoes" the rotation effect of $V^\top$ without changing its length.

(d) **What does the Moore-Penrose pseudoinverse give as a solution $\vec{x}$ in the following system of equations?**

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \vec{x} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}.$$

Confirm that your solution indeed satisfies the system of equations.

Solution: From the above, we have the solution given by:

$$\vec{x} = A^\dagger \vec{y} = V\widetilde{\Sigma}U^\top \vec{y} \tag{75}$$

$$= \begin{bmatrix} 0 & 1 & 0 \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{\sqrt{2}} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} \tag{76}$$

$$= \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & -\frac{1}{4} \end{bmatrix} \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 3 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} \tag{77}$$

Therefore, a reasonable solution to the system of equations is:

$$\vec{x} = \begin{bmatrix} 3 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} \tag{78}$$

Confirming that the solution works, we have

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}. \tag{79}$$

4. Rank 1 Decomposition

In this problem, we will decompose a few images into linear combinations of rank 1 matrices. Remember that outer product of two vectors $\vec{s}\vec{g}^\top$ gives a rank 1 matrix. It has rank 1 because the column span is one-dimensional — multiples of $\vec{s}$ only — and the row span is also one dimensional — multiples of $\vec{g}^\top$ only.

These decompositions are useful for understanding the outer product form of the SVD.

(a) Consider a standard 8×8 chessboard shown in Figure 1. Assume that black colors represent -1 and that white colors represent 1.

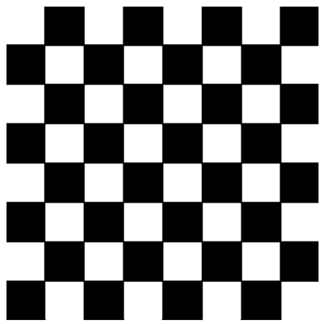


Figure 1: 8×8 chessboard.

In particular, the chessboard is given by the following 8×8 matrix C_1:

$$C_1 = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{bmatrix} \tag{80}$$

Express C_1 as a linear combination of outer products.

(HINT: In order to determine how many rank 1 matrices you need to combine to represent the matrix, find the rank of the matrix you are trying to represent.)

Solution: The matrix C_1 only has rank 1, since column vectors 1, 3, 5, and 7 are the same, and column vectors 2, 4, 6, and 8 are multiples of the other columns. This means that we can express C_1 by multiplying the first column vector $\begin{bmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \end{bmatrix}^\top$ by the multiples required to generate the other columns, which are $1, -1, 1, \ldots, -1$. As a result, we get the following outer product form:

$$C_1 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}^\top \tag{81}$$

(b) For the same chessboard shown in Figure 1, now assume that black colors represent 0 and that white colors represent 1.

In particular, the chessboard is given by the following 8×8 matrix C_2:

$$C_2 = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \tag{82}$$

Express C_2 as a linear combination of outer products.

(HINT: If you find yourself getting stuck, one way to proceed is to note that $\text{rank}(C_2) = 2$*, so you will need to sum two outer products. Try to have one outer product that fills in the odd columns and is* 0 *elsewhere, and one outer product that fills in the even columns and is* 0 *elsewhere.)*

Solution: There are multiple valid solutions. This is just one of them. Give yourself full credit for any valid solution.

The chessboard is now a rank 2 image, so we need to decompose it.

To fill in the odd columns, we can use the following outer product:

$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}^\top = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{83}$$

To fill in the even columns, we can use the following outer product:

$$\begin{bmatrix}0\\1\\0\\1\\0\\1\\0\\1\end{bmatrix}\begin{bmatrix}0\\1\\0\\1\\0\\1\\0\\1\end{bmatrix}^{\top} = \begin{bmatrix}0&0&0&0&0&0&0&0\\0&1&0&1&0&1&0&1\\0&0&0&0&0&0&0&0\\0&1&0&1&0&1&0&1\\0&0&0&0&0&0&0&0\\0&1&0&1&0&1&0&1\\0&0&0&0&0&0&0&0\\0&1&0&1&0&1&0&1\end{bmatrix}. \tag{84}$$

Then the final chessboard is the sum of these odd and even columns:

$$C_2 = \begin{bmatrix}1\\0\\1\\0\\1\\0\\1\\0\end{bmatrix}\begin{bmatrix}1\\0\\1\\0\\1\\0\\1\\0\end{bmatrix}^{\top} + \begin{bmatrix}0\\1\\0\\1\\0\\1\\0\\1\end{bmatrix}\begin{bmatrix}0\\1\\0\\1\\0\\1\\0\\1\end{bmatrix}^{\top} = \begin{bmatrix}1&0&1&0&1&0&1&0\\0&1&0&1&0&1&0&1\\1&0&1&0&1&0&1&0\\0&1&0&1&0&1&0&1\\1&0&1&0&1&0&1&0\\0&1&0&1&0&1&0&1\\1&0&1&0&1&0&1&0\\0&1&0&1&0&1&0&1\end{bmatrix}. \tag{85}$$

(c) Now consider the Swiss flag shown in Figure 2. Assume that red colors represent 0 and that white colors represent 1.

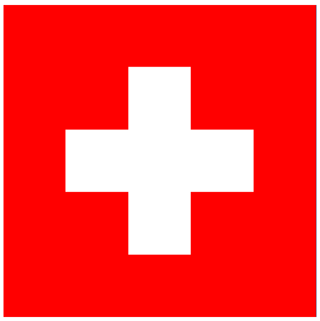


Figure 2: Swiss flag.

Assume that the Swiss flag is given by the following 5×5 matrix S:

$$S = \begin{bmatrix}0&0&0&0&0\\0&0&1&0&0\\0&1&1&1&0\\0&0&1&0&0\\0&0&0&0&0\end{bmatrix} \tag{86}$$

Furthermore, we know that the Swiss flag can be viewed as a superposition of the following pairs of images:

Figure 3: Pairs of images - Option 1

Figure 4: Pairs of images - Option 2

Express the S in two different ways:

i. **as a linear combination of outer products inspired by the Option 1 images.**
Solution:

$$S = \begin{bmatrix}0\\1\\1\\1\\0\end{bmatrix}\begin{bmatrix}0\\1\\1\\1\\0\end{bmatrix}^\top - \begin{bmatrix}0\\1\\0\\1\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\\1\\0\end{bmatrix}^\top \tag{87}$$

Note here that the there does not necessarily exist a unique decomposition for an image.

ii. **as a linear combination of outer products inspired by the Option 2 images.**
Solution:

$$S = \begin{bmatrix}0\\1\\1\\1\\0\end{bmatrix}\begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix}^\top + \begin{bmatrix}0\\0\\1\\0\\0\end{bmatrix}\begin{bmatrix}0\\1\\0\\1\\0\end{bmatrix}^\top \tag{88}$$

Note here that the there does not necessarily exist a unique decomposition for an image.

5. 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

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |A_{ij}|^2}. \tag{89}$$

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**:**

$$\|A\|_F = \sqrt{\operatorname{tr}(A^\top A)}. \tag{90}$$

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

$$\operatorname{tr}(A) = \sum_{i=1}^{\min(n,m)} A_{ii} \tag{91}$$

Think about how/whether this expression eq. (90) 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.

$$\operatorname{tr}\left(A^\top A\right) = \sum_{i=1}^{n} (A^\top A)_{ii} \tag{92}$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} (A^\top)_{ij} A_{ji} \right) \tag{93}$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{m} A_{ji} A_{ji} \right) \tag{94}$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} (A_{ji}^2) \tag{95}$$

$$= \|A\|_F^2 \tag{96}$$

In the above solution, step eq. (92) writes out the trace definition, step eq. (93) 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. (94) applies the definition of matrix transpose, and the last two steps collects the result into the definition of Frobenius norm.

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

$$\|UA\|_F = \|AV\|_F = \|A\|_F. \tag{97}$$

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

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

$$\|UA\|_F = \sqrt{\operatorname{tr}((UA)^\top (UA))} = \sqrt{\operatorname{tr}(A^\top U^\top UA)} = \sqrt{\operatorname{tr}(A^\top A)} = \|A\|_F \tag{98}$$

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:

$$\|AV\|_F = \left\|(AV)^\top\right\|_F = \left\|V^\top A^\top\right\|_F \tag{99}$$

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

$$\left\|V^\top A^\top\right\|_F = \left\|A^\top\right\|_F = \|A\|_F \tag{100}$$

(c) **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:

$$\|A\|_F = \left\|U\Sigma V^\top\right\|_F = \left\|\Sigma V^\top\right\|_F = \|\Sigma\|_F \tag{101}$$

$$= \sqrt{\sum_{i=1}^n \sigma_i^2} \tag{102}$$

6. Movie Ratings and PCA

Recall from the lecture on PCA that we can think of movie ratings as a structured set of data. For every person i and movie j, we have that person's rating $R_{i,j}$ (thought of as a real number).

Suppose that there are m movies and n people. Let's think about arranging this data into a big $n \times m$ matrix R with people corresponding to rows and movies corresponding to columns. So the entry in the i-th row and j-th column should be $R_{i,j}$ above. Note that this is organized differently from how it was in lecture. Each row corresponds to a unique person and each column to a unique movie.

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{22} & \cdots & R_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ R_{n1} & R_{n2} & \cdots & R_{nm} \end{bmatrix} \tag{103}$$

(a) Suppose we believe that there is actually an underlying pattern to this rating data and that a separate study has revealed that every movie is characterized by a set of characteristics: say action and comedy. This means that every movie j has a pair of numbers a_j (for action) and c_j (for comedy) that define it. At the same time, every person i has a pair of sensitivities f_i and g_i that defines that person's preferences for action vs. comedy movies respectively. A person i will rate the movie j as $R_{i,j} = f_i a_j + g_i c_j$.

If we arrange the sensitivities into a pair of n-dimensional vectors $\vec{f}, \vec{g}$ for our group of n people, and the movie characteristics into a pair of m-dimensional vectors $\vec{a}, \vec{c}$ for our group of m movies, **use outer products to express the rating matrix R in terms of these vectors $\vec{f}, \vec{g}, \vec{a}, \vec{c}$.**

Solution:

$$R = \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ R_{21} & R_{22} & \cdots & R_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ R_{n1} & R_{n2} & \cdots & R_{nm} \end{bmatrix} \tag{104}$$

$$= \begin{bmatrix} f_1 a_1 + g_1 c_1 & f_1 a_2 + g_1 c_2 & \cdots & f_1 a_m + g_1 c_m \\ f_2 a_1 + g_2 c_1 & f_2 a_2 + g_2 c_2 & \cdots & f_2 a_m + g_2 c_m \\ \vdots & \vdots & \cdots & \vdots \\ f_n a_1 + g_n c_1 & f_n a_2 + g_n c_2 & \cdots & f_n a_m + g_n c_m \end{bmatrix} \tag{105}$$

$$= \begin{bmatrix} f_1 a_1 & f_1 a_2 & \cdots & f_1 a_m \\ f_2 a_1 & f_2 a_2 & \cdots & f_2 a_m \\ \vdots & \vdots & \cdots & \vdots \\ f_n a_1 & f_n a_2 & \cdots & f_n a_m \end{bmatrix} + \begin{bmatrix} g_1 c_1 & g_1 c_2 & \cdots & g_1 c_m \\ g_2 c_1 & g_2 c_2 & \cdots & g_2 c_m \\ \vdots & \vdots & \cdots & \vdots \\ g_n c_1 & g_n c_2 & \cdots & g_n c_m \end{bmatrix} \tag{106}$$

$$= \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} + \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix} \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} \tag{107}$$

$$= \vec{f}\vec{a}^\top + \vec{g}\vec{c}^\top \tag{108}$$

(b) Now suppose that we want to discover the underlying nature of movies from the data R itself.

Suppose for this part, that you have four observed rating data vectors (corresponding to four different movies being rated by six individuals).

All of the movie data vectors just happened to be multiples of the following 6-dimensional vector $\vec{w} = \begin{bmatrix} 2 \\ -2 \\ 3 \\ -4 \\ 4 \\ 0 \end{bmatrix}$. (For your convenience, note that $\|\vec{w}\| = 7$.)

You arrange the movie data vectors as the columns of a matrix R given by:

$$R = \begin{bmatrix} | & | & | & | \\ -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} \\ | & | & | & | \end{bmatrix} \tag{109}$$

You want to perform PCA (for movies) using the SVD of the matrix R to better understand the pattern in your data.

The first "principal component vector" is a unit vector that tells which direction we would want to project the columns of R onto to get the best rank-1 approximation for R.

Find this first principal component vector of the columns of R to explain the nature of your movie data points.

(HINT: You don't need to actually compute any SVDs in this simple case. Also, be sure to think about what size vector you want as the answer. Don't forget that you want a unit vector!)

Solution: Principal component analysis is in general about understanding how best to approximate our (potentially) high-dimensional data (like recordings from a microphone, or in this case, a movie's ratings by lots of people) with its lower-dimensional essence. The first principal component is about seeing which one-dimensional line best approximates the data points — i.e. which is the line for which projecting the data points onto it results in "estimates" that are as close as possible to the data points.

In the case of this problem, every point is explicitly given as a multiple of a single vector $\vec{w}$ and so the data already lies on such a straight line going through the origin. So, the first principal component is just along the direction of $\vec{w}$. Because a principal component represents a direction, it is conventional to normalize the vector to have unit length. In this case, we are told that the vector $\vec{w}$ has length 7, and so the answer is $\frac{\vec{w}}{7}$.

(Because the line is all that matters, you could also have used the negative of this $-\frac{\vec{w}}{7}$.)

We can also do this using the SVD.

The singular value decomposition of a matrix R is a way of decomposing R into a sum of rank 1 matrices. In this sum the i^{th} rank 1 matrix is formed from taking the outer product of normalized column vectors $\vec{u}_i$ and normalized row vectors $\vec{w}_i^\top$, scaled by their respective singular values σ_i.

Looking at our given R, we can see that the matrix itself is rank 1 as the columns are all multiples of the same vector: $\vec{w}$. Seeing this we realize we can rewrite the matrix R as the following outer product:

$$R = \begin{bmatrix} | \\ \vec{w} \\ | \end{bmatrix} \begin{bmatrix} -1 & -2 & 2 & 4 \end{bmatrix} \tag{110}$$

However the SVD requires we normalize the vectors $\vec{u_1}$ and $\vec{v_1}^{\top}$. In order to reconstruct A properly we must scale back with the norms that we divided out to normalize.

$\|[-1,-2,2,4]\| = \sqrt{25} = 5$ and $\|\vec{w}\| = 7$. Consequently, when we pull that out, we get $\sigma_1 = 35$ as the singular value that corresponds to the first (and only) principal component.
Thus we can write the SVD of R as:

$$R = \begin{bmatrix} | \\ \frac{\vec{w}}{7} \\ | \end{bmatrix} 35 \begin{bmatrix} \frac{-1}{5} & \frac{-2}{5} & \frac{2}{5} & \frac{4}{5} \end{bmatrix} \tag{111}$$

Now we just have to pick which normalized vector to deem the principal component. Since our data (the movie ratings) are collected as columns we choose $\frac{\vec{w}}{7}$ as the principal component.

Alternate Solution:

To find the first principal component along the columns, we can use Σ and U. This is because our data is stored in the columns of R. We know that

$$R = U\Sigma V^T \tag{112}$$

$$\implies RR^T = U\Sigma\Sigma^T U^T \tag{113}$$

where $\Sigma\Sigma^T$ represents the square diagonal matrix with $min(m,n)$ singular values squared on the diagonal. Plugging in for R gives

$$RR^T = \begin{bmatrix} -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} \end{bmatrix} \begin{bmatrix} -\vec{w} \\ -2\vec{w} \\ 2\vec{w} \\ 4\vec{w} \end{bmatrix} \tag{114}$$

$$RR^T = \vec{w}\vec{w}^T + 4\vec{w}\vec{w}^T + 4\vec{w}\vec{w}^T + 16\vec{w}\vec{w}^T \tag{115}$$

$$RR^T = 25\vec{w}\vec{w}^T \tag{116}$$

By the Spectral Theorem, we know that the eigenvectors of RR^T correspond to the columns of U and the eigenvalues of RR^T correspond to the singular values squared. By inspection, we can see that the eigenvector is $\vec{w}$. Solving for the eigenvalue and singular value:

$$RR^T\vec{w} = 25\vec{w}\vec{w}^T\vec{w} \tag{117}$$

$$RR^T\vec{w} = (25\vec{w}^T\vec{w})\vec{w} \tag{118}$$
$$\implies \lambda = 25\vec{w}^T\vec{w} = 5^2 7^2 \tag{119}$$
$$\implies \sigma = \sqrt{\lambda} = 35 \tag{120}$$

For PCA, we require normalized vectors, so for that reason our first principal component is $\frac{\vec{w}}{\|\vec{w}\|} = \frac{\vec{w}}{7}$ with a corresponding singular value of $\sigma = 35$.

(c) Suppose that now, we have two more data points (corresponding to two more movies being rated by the same set of six people, i.e. we added two columns to our matrix) that are multiples of a different vector $\vec{p}$ where:

$\vec{p} = \begin{bmatrix} 6 \\ 3 \\ -2 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. (For your convenience, note that $\|\vec{p}\| = 7$ and that $\vec{p}^\top\vec{w} = 0$.)

We augment our ratings data matrix with these two new data points to get:

$$R = \begin{bmatrix} | & | & | & | & | & | \\ -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} & -3\vec{p} & 3\vec{p} \\ | & | & | & | & | & | \end{bmatrix} \tag{121}$$

Find the first two principal components corresponding to the nonzero singular values of R. This is what we would use to best project the movie data points onto a two-dimensional subspace. **What is the first principal component vector? What is the second principal component vector? Justify your answer.** *(HINT: Think about the inner product of $\vec{w}$ and $\vec{p}$ and what that implies for being able to appropriately decompose R. Again, very little computation is required here.)*

Solution: The solution to the previous part tells you what we need to do. We need to find the best two-dimensional subspace that best represents our data.

We start by taking the SVD of R.

The columns of R are all multiples of two vectors: $\vec{w}$ and $\vec{p}$. Each of these can be used to create a rank 1 matrix, and these can be summed together to form R.

Since $\vec{w}$ and $\vec{p}$ are orthogonal to one another, our life is easier. This problem's R matrix is made especially nice by seeing that a data point is either purely in the $\vec{w}$ direction, or the $\vec{p}$ direction.

Using this knowledge we rewrite R as:

$$R = \begin{bmatrix} | \\ \vec{w} \\ | \end{bmatrix} \begin{bmatrix} -1 & -2 & 2 & 4 & 0 & 0 \end{bmatrix} + \begin{bmatrix} | \\ \vec{p} \\ | \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0 & -3 & 3 \end{bmatrix}. \tag{122}$$

The orthogonality relationships demanded by the SVD are clearly satisfied since the row-vectors involved above have disjoint support (i.e. when one is nonzero, the other is zero) and the columns are orthogonal since we've been told so.

However for the SVD the vectors: $\vec{u}_1, \vec{v}_1^\top, \vec{u}_2$ and $\vec{v}_2^\top$ must be normalized and each rank 1 matrix must be scaled by the appropriate σ_i to allow the sum to properly reconstruct R. We also need to figure out which σ_i is bigger so we can order them properly. In the previous part, we have already done the calculations for $\vec{w}$'s part in this story. So what remains is the $\vec{p}$ part. Clearly the norm of the relevant row is $3\sqrt{2}$ which the norm of the relevant column is 7. So the singular value in question is $21\sqrt{2}$.

Using this we can rewrite R as:

$$= \begin{bmatrix} | \\ \frac{\vec{w}}{7} \\ | \end{bmatrix} 35 \begin{bmatrix} \frac{-1}{5} & \frac{-2}{5} & \frac{2}{5} & \frac{4}{5} & 0 & 0 \end{bmatrix} + \begin{bmatrix} | \\ \frac{\vec{p}}{7} \\ | \end{bmatrix} 21\sqrt{2} \begin{bmatrix} 0 & 0 & 0 & 0 & \frac{-3}{3\sqrt{2}} & \frac{3}{3\sqrt{2}} \end{bmatrix}. \tag{123}$$

Since $35 > 21\sqrt{2}$, we arrange them in descending order such that our singular values are $\sigma_1 = 35$ and $\sigma_2 = 21\sqrt{2}$. Thus $\frac{\vec{w}}{7}$ which corresponds to σ_1 is still the first principal component vector and $\frac{\vec{p}}{7}$ which corresponds to σ_2 is the second principal component vector.

Alternate Solution:

Again, to find the first principal component along the columns, we can use Σ and U. This is because our data is stored in the columns of R. We know that

$$R = U\Sigma V^T \tag{124}$$

$$\implies RR^T = U\Sigma\Sigma^T U^T \tag{125}$$

where $\Sigma\Sigma^T$ represents the square diagonal matrix with $min(m,n)$ singular values squared on the diagonal. Plugging in for R gives

$$RR^T = \begin{bmatrix} -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} & -3\vec{p} & 3\vec{p} \end{bmatrix} \begin{bmatrix} -\vec{w} \\ -2\vec{w} \\ 2\vec{w} \\ 4\vec{w} \\ -3\vec{p} \\ 3\vec{p} \end{bmatrix} \tag{126}$$

$$RR^T = \vec{w}\vec{w}^T + 4\vec{w}\vec{w}^T + 4\vec{w}\vec{w}^T + 16\vec{w}\vec{w}^T + 9\vec{p}\vec{p}^T + 9\vec{p}\vec{p}^T \tag{127}$$

$$RR^T = 25\vec{w}\vec{w}^T + 18\vec{p}\vec{p}^T \tag{128}$$

By the spectral decomposition, we know that the eigenvectors of RR^T correspond to the columns of U and the eigenvalues of RR^T correspond to the singular values squared. By inspection, we can see that the first eigenvector is $\vec{w}$ and the second one is $\vec{p}$. Starting with the first:

$$RR^T\vec{w} = 25\vec{w}\vec{w}^T\vec{w} + 18\vec{p}(\vec{p}^T\vec{w}) \tag{129}$$

$$RR^T\vec{w} = (25\vec{w}^T\vec{w})\vec{w} + 0 \tag{130}$$

$$\implies \lambda_1 = 25\vec{w}^T\vec{w} = 5^2 7^2 \tag{131}$$

$$\implies \sigma_1 = \sqrt{\lambda_1} = 35 \tag{132}$$

When computing the first eigenvalue, we used the fact that $\vec{p}^T\vec{w} = 0$. We can repeat the same process for the second eigenvalue:

$$RR^T\vec{p} = 25\vec{w}(\vec{w}^T\vec{p}) + 18\vec{p}(\vec{p}^T\vec{w}) \tag{133}$$

$$RR^T\vec{p} = 0 + 18(\vec{p}^T\vec{p})\vec{p} \tag{134}$$
$$\implies \lambda_2 = 18\vec{p}^T\vec{p} = 3^2 7^2 2 \tag{135}$$
$$\implies \sigma_2 = \sqrt{\lambda_2} = 21\sqrt{2} \tag{136}$$

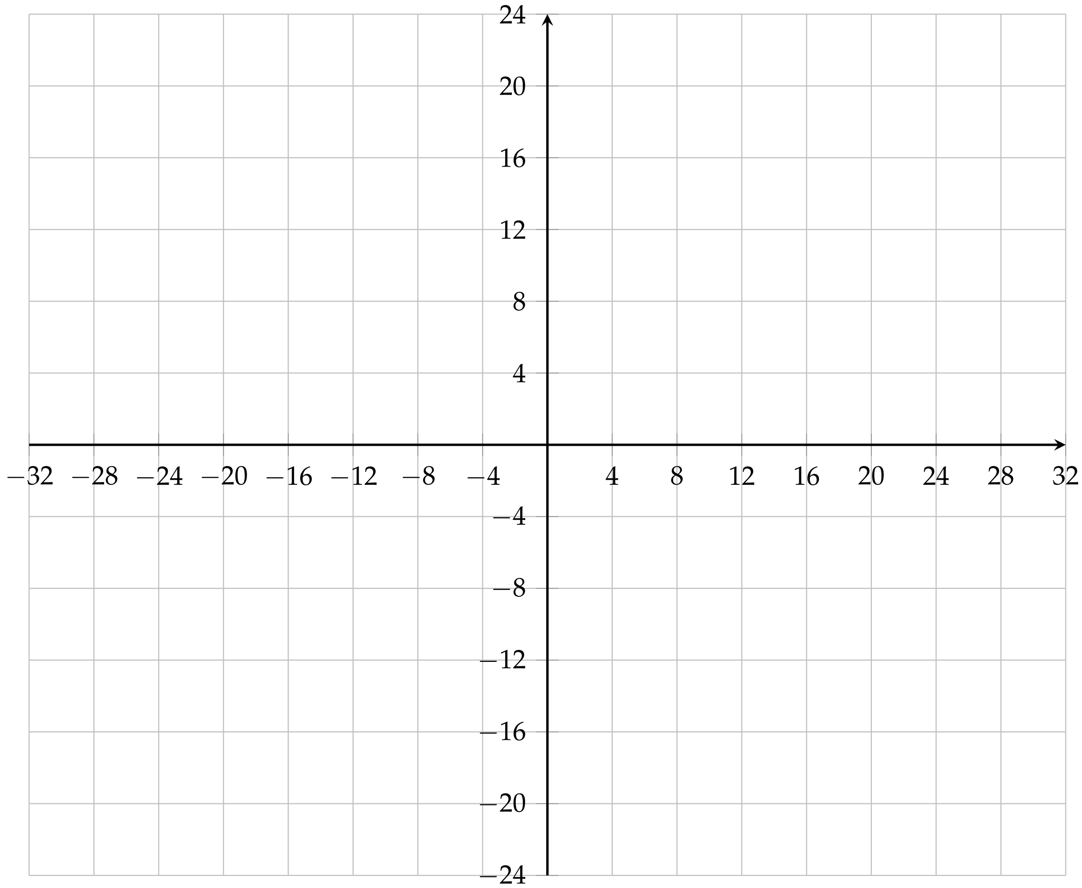
For PCA, we require normalized vectors, so for that reason our first principal component is $\frac{\vec{w}}{\|\vec{w}\|} = \frac{\vec{w}}{7}$ with a corresponding singular value of $\sigma_1 = 35$. Our second principal component is $\frac{\vec{p}}{\|\vec{p}\|} = \frac{\vec{p}}{7}$ with a corresponding singular value of $\sigma_2 = 21\sqrt{2}$. As mentioned in the solution above, we choose $\sigma_1 = 35$, $\sigma_2 = 21\sqrt{2}$ in this order because $35 > 21\sqrt{2}$.

(d) In the previous part, you had

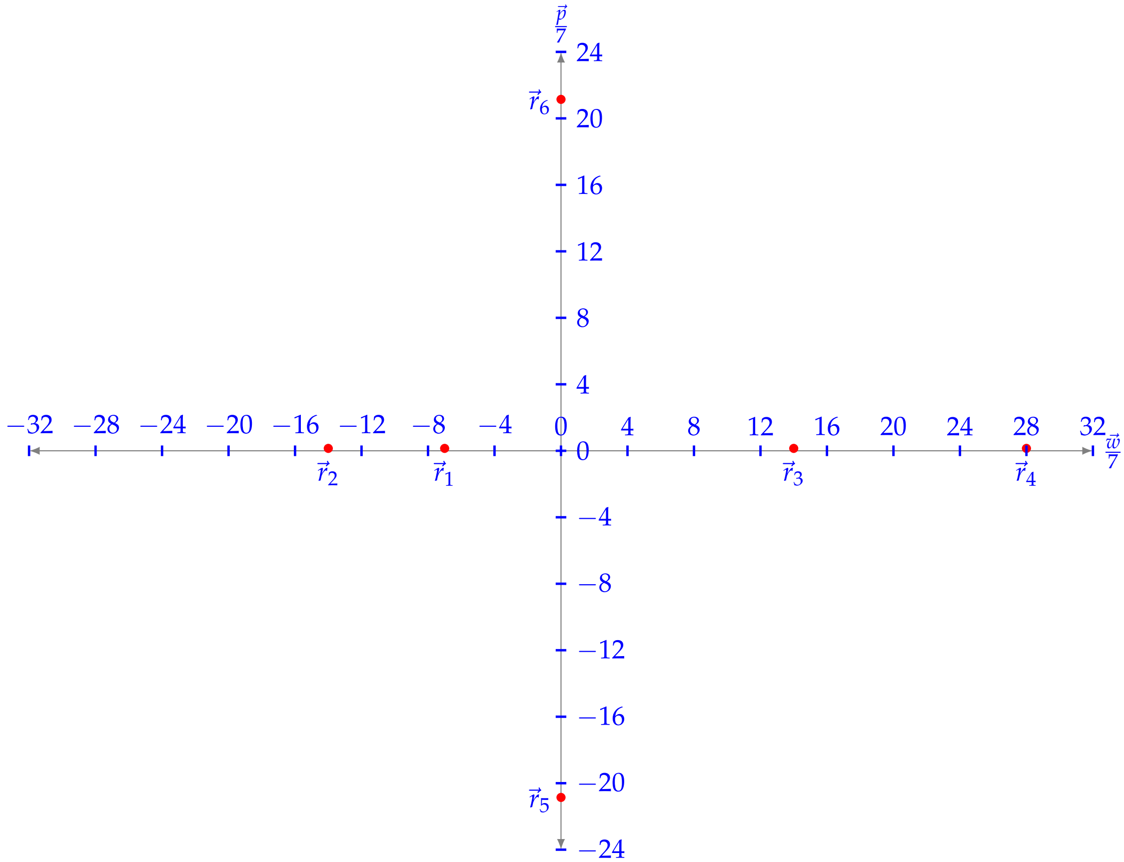
$$R = \begin{bmatrix} | & | & | & | & | & | \\ -\vec{w} & -2\vec{w} & 2\vec{w} & 4\vec{w} & -3\vec{p} & 3\vec{p} \\ | & | & | & | & | & | \end{bmatrix}$$

with $\|\vec{w}\| = 7$ and $\|\vec{p}\| = 7$, satisfying $\vec{p}^\top\vec{w} = 0$.

If we use $\vec{r}_i$ to denote the i-th column of R, **plot the movie data points $\vec{r}_i$ (for all i) projected onto the first and second principal component vectors along the columns of R.** The coordinate along the first principal component should be represented by horizontal axis and the coordinate along the second principal component should be the vertical axis. **Label each point, and the axes. Remember that principal component vectors are normalized**.



Solution: Once we know what the principal components are, we know that the first four data points are just multiples of the first principal component and the last two data points are just multiples of the second principal component. What multiples? For the first four, the multiples are clearly $-7, -14, 14, 28$ since the norm of $\vec{w}$ is 7. For the final two, the multiples are clearly $-21, +21$ since the norm of $\vec{p}$ is also 7. Plotting:



7. Homework Process, Study Group, and Course Weekly Survey

Citing sources and collaborators are an important part of life, including being a student!

We also want to understand what resources you find helpful and how much time homework is taking, so we can change things in the future if possible.

At the same time, we want to check-in weekly regarding Discussions, Lectures, Lab, and Office Hours and see how effective they have all been for you as students.

Please fill out this survey link. For your submission, please attach a screenshot indicating that you have completed the survey this week.

Contributors:
- Anirudh Rengarajan.
- Siddharth Iyer.
- Yu-Yun Dai.
- Sanjit Batra.
- Anant Sahai.
- Sidney Buchbinder.
- Gaoyue Zhou.
- Druv Pai.
- Elena Jia.
- Tanmay Gautam.
- Justin Yim.
- Sally Hui.
- Titan Yuan.
- Aditya Arun.
- Kourosh Hakhamaneshi.
- Antroy Roy Chowdhury.
- Nikhil Shinde.
- Nathan Lambert.
- Daniel Abraham.
- Wahid Rahman.

