Problem 1 (20 pts)

(a) (5 pts) Define a *unitary* matrix and an *orthogonal* matrix and describe their difference.

Answer: The difference is that the unitarity is for complex-valued matrices and the orthogonality is for real-valued matrices. They are defined as follows.

- Let $A \in \mathbb{C}^{m\times m}$. Then A is a unitary matrix if $A^* A = AA^* = I_{m\times m}$. Alternatively, you can say that the columns of A constitute an orthonormal basis for $\mathbb{C}^m$.
- Let $A \in \mathbb{R}^{m\times m}$. Then A is an orthogonal matrix if $A^T A = AA^T = I_{m\times m}$. Alternatively, you can say that the columns of A constitute an orthonormal basis for $\mathbb{R}^m$.

(b) (7 pts) Let A be a matrix of size $m \times n$. Let P be any unitary matrix of size $m \times m$. Prove that $\|A\|_2 = \|PA\|_2$.

Answer: Consider any $\boldsymbol{x} \in \mathbb{C}^n$. Then

$$\begin{aligned} \|PA\boldsymbol{x}\|_2^2 &= \langle PA\boldsymbol{x}, PA\boldsymbol{x}\rangle \\ &= (PA\boldsymbol{x})^*(PA\boldsymbol{x}) \\ &= \boldsymbol{x}^* A^* P^* PA\boldsymbol{x} \\ &= \boldsymbol{x}^* A^* A\boldsymbol{x} \quad \text{since } P \text{ is unitary;} \\ &= \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle \\ &= \|A\boldsymbol{x}\|_2^2. \end{aligned}$$

Hence,

$$\|PA\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|PA\boldsymbol{x}\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|A\boldsymbol{x}\|_2 = \|A\|_2.$$

(c) (8 pts) Let A be a matrix of size $m \times n$. Let Q be any unitary matrix of size $n \times n$. Prove that $\|A\|_2 = \|AQ\|_2$.

Answer: Consider any $\boldsymbol{x} \in \mathbb{C}^n$. Then let $\boldsymbol{y} = Q\boldsymbol{x} \in \mathbb{C}^n$. Since Q is unitary, $\|\boldsymbol{y}\|_2 = \|Q\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$ with the same argument in Part (a). Now, we have

$$\|AQ\|_2 = \max_{\|\boldsymbol{x}\|_2=1} \|AQ\boldsymbol{x}\|_2 = \max_{\|\boldsymbol{y}\|_2=1} \|Ay\|_2 = \|A\|_2.$$

Problem 2 (20 pts) Consider the following matrix.

$$A = \begin{bmatrix} 1 & 1 & 1 \\ \epsilon & 0 & 0 \\ 0 & \epsilon & 0 \\ 0 & 0 & \epsilon \end{bmatrix},$$

where $\epsilon = 10^{-2}$. Using the 2-digit floating-point arithmetic, compute the reduced QR factorization of A using the *modified* Gram-Schmidt procedure. Note that reduced QR factorization means that Q is of size 4×3 and R is of size 3×3. You can also use the following approximation: $fl(\sqrt{2}) = 1.4$, $fl(1/\sqrt{2}) = 0.71$, $fl(\sqrt{1.5}) = 1.2$, $fl(1/\sqrt{1.5}) = 0.82$, $fl(0.71^2) = 0.50$, $fl(1/0.71) = 1.4$, $fl(1/1.2) = 0.83$.

Answer: Let us proceed step by step. Let $\boldsymbol{a}_j$, $j = 1, 2, 3$ be the column vectors of A.

Step 1: Normalize $\boldsymbol{a}_1$ to get $\boldsymbol{q}_1$, i.e.,

$$r_{11} = \|\boldsymbol{a}_1\|_2 = fl(\sqrt{1+\epsilon^2}) = 1.$$

$$\boldsymbol{q}_1 = \boldsymbol{a}_1 / r_{11} = \begin{bmatrix} 1 \\ \epsilon \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.010 \\ 0 \\ 0 \end{bmatrix}.$$

Step 2: Remove the $\boldsymbol{q}_1$ component from $\boldsymbol{a}_2$ and $\boldsymbol{a}_3$ immediately. Thus, we have:

$$r_{12} = \langle \boldsymbol{q}_1, \boldsymbol{a}_2 \rangle = \begin{bmatrix} 1 & \epsilon & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \epsilon \\ 0 \end{bmatrix} = 1. \quad \widetilde{\boldsymbol{a}}_2 = \boldsymbol{a}_2 - r_{12}\boldsymbol{q}_1 = \begin{bmatrix} 1 \\ 0 \\ \epsilon \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ \epsilon \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -\epsilon \\ \epsilon \\ 0 \end{bmatrix}.$$

$$r_{13} = \langle \boldsymbol{q}_1, \boldsymbol{a}_3 \rangle = \begin{bmatrix} 1 & \epsilon & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ \epsilon \end{bmatrix} = 1. \quad \widetilde{\boldsymbol{a}}_3 = \boldsymbol{a}_3 - r_{13}\boldsymbol{q}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \epsilon \end{bmatrix} - \begin{bmatrix} 1 \\ \epsilon \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -0.010 \\ 0 \\ 0.010 \end{bmatrix}.$$

Step 3: Normalize $\widetilde{\boldsymbol{a}}_2$ to get $\boldsymbol{q}_2$. Then immediately subtract the $\boldsymbol{q}_2$ component from $\widetilde{\boldsymbol{a}}_3$.

$$r_{22} = \|\widetilde{\boldsymbol{a}}_2\|_2 = fl(\epsilon \cdot \sqrt{2}) = 0.014. \quad \boldsymbol{q}_2 = \widetilde{\boldsymbol{a}}_2 / r_{22} = fl(1/0.014) \cdot \begin{bmatrix} 0 \\ -\epsilon \\ \epsilon \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -0.71 \\ 0.71 \\ 0 \end{bmatrix}.$$

$$r_{23} = \langle \boldsymbol{q}_2, \widetilde{\boldsymbol{a}}_3 \rangle = \begin{bmatrix} 0 & -0.71 & 0.71 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ -0.010 \\ 0 \\ 0.010 \end{bmatrix} = 0.0071.$$

$$\tilde{\boldsymbol{a}}_3 \leftarrow \tilde{\boldsymbol{a}}_3 - r_{23}\boldsymbol{q}_2 = fl\left(\begin{bmatrix} 0 \\ -0.010 \\ 0 \\ 0.010 \end{bmatrix} - 0.0071 \cdot \begin{bmatrix} 0 \\ -0.71 \\ 0.71 \\ 0 \end{bmatrix}\right) = 0.010 \cdot \begin{bmatrix} 0 \\ fl(-1 + \cdot 0.71^2) \\ fl(-0.71^2) \\ 1 \end{bmatrix} = 0.010 \begin{bmatrix} 0 \\ -0.50 \\ -0.50 \\ 1 \end{bmatrix}.$$

Step 4: Normalize $\tilde{\boldsymbol{a}}_3$ to get $\boldsymbol{q}_3$.

$$r_{33} = \|\tilde{\boldsymbol{a}}_3\|_2 = fl(0.010 \cdot \sqrt{0.5^2 + 0.5^2 + 1}) = 0.010 \cdot fl(\sqrt{1.5}) = 0.012.$$

$$\boldsymbol{q}_3 = \tilde{\boldsymbol{a}}_3 / r_{33} = fl\left(0.010/0.012 \cdot \begin{bmatrix} 0 \\ -0.50 \\ -0.50 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0 \\ -0.42 \\ -0.42 \\ 0.83 \end{bmatrix}.$$

Step 5: Finally, we can form the reduced QR factorization as

$$\boxed{A = \hat{Q}\hat{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0.010 & -0.71 & -0.42 \\ 0 & 0.71 & -0.42 \\ 0 & 0 & 0.83 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0.014 & 0.0071 \\ 0 & 0 & 0.012 \end{bmatrix}.}$$

Problem 3 (20 pts) Let $P = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$.

(a) (5 pts) Show P is a projector, but not an orthogonal projector.

Answer: First of all, P is a square matrix. Now we need to show $P^2 = P$. But this is easy:

$$P^2 = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} = P.$$

Therefore, P is a projector. However, P is not an orthogonal projector because it is clear that $P^T \neq P$ in this case.

(b) (5 pts) Show that $\mathbb{R}^2 = \mathcal{R}(P) \oplus \mathcal{N}(P)$. Moreover, show that $\mathcal{R}(P)$ is not orthogonal to $\mathcal{N}(P)$. [Hint: Obtain the spanning sets of $\mathcal{R}(P)$ and $\mathcal{N}(P)$.]

Answer: Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ be any vector in $\mathbb{R}^2$. Then,

$$P\boldsymbol{x} = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - x_2 \\ 0 \end{bmatrix} = (x_1 - x_2) \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

where x_1, x_2 are arbitrary. Therefore, clearly, $\mathcal{R}(P) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$.

Now take any $\boldsymbol{x} \in \mathcal{N}(P)$. Then from $P\boldsymbol{x} = \mathbf{0}$, we immediately see that $x_1 = x_2$. In other words, $\mathcal{N}(P) = \operatorname{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$.

Since $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ are linearly independent, but not mutually orthogonal, we clearly have:

$$\mathbb{R}^2 = \mathcal{R}(P) \oplus \mathcal{N}(P),$$

but $\mathcal{R}(P)$ and $\mathcal{N}(P)$ are not orthogonal.

(c) (5 pts) Show that $\mathbb{R}^2 = \mathcal{R}(P^T) \overset{\perp}{\oplus} \mathcal{N}(P)$. Note that you need to show that $\mathcal{R}(P^T)$ is orthogonal to $\mathcal{N}(P)$. [Hint: Obtain the spanning sets of $\mathcal{R}(P^T)$ and $\mathcal{N}(P)$.]

Answer: We do similarly with Part (b). Let $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ be any vector in $\mathbb{R}^2$.

$$P^T \boldsymbol{x} = \begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ -x_1 \end{bmatrix} = x_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

So, $\mathcal{R}(P^T) = \text{span}\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$. From Part (b), we already know that $\mathcal{N}(P) = \text{span}\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$. Because $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ are linearly independent, and mutually orthogonal, we clearly have:

$$\mathbb{R}^2 = \mathcal{R}(P^T) \overset{\perp}{\oplus} \mathcal{N}(P).$$

(d) (5 pts) Show that $\mathbb{R}^2 = \mathcal{R}(P) \overset{\perp}{\oplus} \mathcal{N}(P^T)$. Note that you need to show that $\mathcal{R}(P)$ is orthogonal to $\mathcal{N}(P^T)$. [Hint: Obtain the spanning sets of $\mathcal{R}(P)$ and $\mathcal{N}(P^T)$.]

Answer: We still need to compute $\mathcal{N}(P^T)$. For any $\boldsymbol{x} \in \mathcal{N}(P^T)$, we have $P^T \boldsymbol{x} = x_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \mathbf{0}$. Thus, we must have $x_1 = 0$. But x_2 is arbitrary. So, $\mathcal{N}(P^T) = \text{span}\left\{ \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$. We already know that $\mathcal{R}(P) = \text{span}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$ from Part (b). Because $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ are linearly independent, and mutually orthogonal, we clearly have:

$$\mathbb{R}^2 = \mathcal{R}(P) \overset{\perp}{\oplus} \mathcal{N}(P^T).$$

Problem 4 (20 pts) Consider the following matrix:

$$A = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}.$$

(a) (10 pts) Compute the *full* SVD of A.

Answer: If $A = U\Sigma V^T$, then $A^T A = V\Sigma^T U^T U \Sigma V^T = V\Sigma^T \Sigma V^T$. From this, we have $A^T AV = V\Sigma^T\Sigma$, or $A^T A\mathbf{v}_j = \sigma_j^2 \mathbf{v}_j$, $j = 1,2$ in this case. So, we need to solve the eigenvalue problem for $A^T A$.

$$A^T A = \begin{bmatrix} -1 & 10 & \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

Hence,

$$\det(A^T A - \lambda I) = \det\left(\begin{bmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{bmatrix}\right) = (2-\lambda)^2 - (-1)^2 = \lambda^2 - 4\lambda + 3 = (\lambda-3)(\lambda-1) = 0 \Longrightarrow \lambda = 3, 1.$$

Since all the singular values must be nonnegative, we have $\sigma_1 = \sqrt{3}$ and $\sigma_2 = 1$. Now let us compute $\mathbf{v}_1 = \begin{bmatrix} x \\ y \end{bmatrix}$.

$$(A^T A - 3I)\mathbf{v}_1 = \mathbf{0} \Longleftrightarrow \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} -x-y \\ -x-y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Longleftrightarrow y = -x.$$

In other words, $\mathbf{v}_1 = x \begin{bmatrix} 1 \\ -1 \end{bmatrix}$. But we must have $\|bv_1\|_2 = 1$. So, $\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Now, let us compute $\mathbf{v}_2 = \begin{bmatrix} x \\ y \end{bmatrix}$.

$$(A^T A - I)\mathbf{v}_1 = \mathbf{0} \Longleftrightarrow \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x-y \\ -x+y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Longleftrightarrow y = x.$$

In other words, we have $\mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.

From $A = U\Sigma V^T$, we have $AV = U\Sigma$. In other words, $A\mathbf{v}_j = \sigma_j \mathbf{u}_j$, $j = 1,2$. So,

$$\sqrt{3}\mathbf{u}_1 = A = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{bmatrix} \Longleftrightarrow \mathbf{u}_1 = \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix}.$$

$$\mathbf{u}_2 = A = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

We still need to find $\mathbf{u}_3$ that is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$ and that has a unit length. Let $\mathbf{u}_3 = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$.

$$\langle \mathbf{u}_1, \mathbf{u}_3 \rangle = \frac{1}{\sqrt{6}}(-x+2y-z) = 0; \quad \langle \mathbf{u}_2, \mathbf{u}_3 \rangle = \frac{1}{\sqrt{2}}(-x+z) = 0; \quad \Longleftrightarrow x = y = z.$$

Thus, we have $\mathbf{u}_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. Finally, we have the following full SVD of A:

$$\boxed{A = \begin{bmatrix} -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}.}$$

(b) (5 pts) Compute the rank 1 approximation A_1 of A.

Answer: By definition of the rank 1 approximation of A, we have

$$A_1 = \sigma_1 \mathbf{u}_1 \mathbf{v}_1^T = \sqrt{3} \cdot \frac{1}{\sqrt{6}} \begin{bmatrix} -1 \\ 2 \\ -1 \end{bmatrix} \cdot \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -1 & 1 \\ 2 & -2 \\ -1 & 1 \end{bmatrix} = \boxed{\begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ 1 & -1 \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}}.$$

(c) (5 pts) Compute the error of the rank 1 approximation using the Frobenius norm, i.e., $\|A - A_1\|_F$.

Answer: We know that for a general A of size $m \times n$

$$\|A - A_1\|_F = \sqrt{\sigma_2^2 + \cdots \sigma_n^2}.$$

But in our particular A, $n = 2$. So,

$$\boxed{\|A - A_1\|_F = \sqrt{\sigma_2^2} = 1.}$$

Alternative Solution:

$$A - A_1 = \begin{bmatrix} -1 & 0 \\ 1 & -1 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} \\ 1 & -1 \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

Hence,

$$\|A - A_1\|_F = \sqrt{\left(-\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = 1.$$

Problem 5 (20 pts) Consider the following documents consisting of titles of some actual mathematical papers.

D1: Neural network and regression spline value function approximations for stochastic dynamic programming

D2: Hardy inequalities and dynamic instability of singular Yamabe metrics

D3: Boundedness of weak solutions of nondiagonal singular parabolic system equations

D4: On one nonlinear analogue of the mean value property and its application to the investigation of the nonlinear Goursat problem

D5: Atomic decompositions of weak Hardy spaces of B-valued martingales

D6: δ-sequence approach to a two-point boundary value problem using Daubechies wavelets

D7: Decomposition strategies for large-scale continuous location-allocation problems

Now, let us consider the following terms:

T1: decomposition, decompositions

T2: singular

T3: value, values, valued

(a) (5 pts) Construct the *Term-by-Document Matrix* from the terms and documents above. Note that the terms are *not* case sensitive.

Answer: We just need to count the occurrences of each term in each document. Thus, we have

	D1	D2	D3	D4	D5	D6	D7
T1	0	0	0	0	1	0	1
T2	0	1	1	0	0	0	0
T3	1	0	0	1	1	1	0

(b) (5 pts) Let $\boldsymbol{q} = (1,1,1)^T$ be your query vector. Compute $\cos\theta_j$, $j = 1,\ldots,7$ where θ_j is an angle between $\boldsymbol{q}$ and $\boldsymbol{d}_j$, i.e., the jth column vector of the term-by-document matrix you obtained in Part (a). Then, find the best matching document to your query.

Answer: We can compute $\cos\theta_j$ as

$$\cos\theta_j = \frac{\langle \boldsymbol{q}, \boldsymbol{d}_j \rangle}{\|\boldsymbol{q}\|_2 \|\boldsymbol{d}_j\|_2}, \quad j = 1,\ldots,7,$$

where $\boldsymbol{d}_j$ is the jth column vector of the term-by-document matrix A computed in Part (a). Now, $\|\boldsymbol{q}\|_2 = \sqrt{3}$ and $\|\boldsymbol{d}_j\|_2$ is $1,1,1,1,\sqrt{2},1,1$ for $j = 1,\ldots,7$. Hence

$$(\cos\theta_j)_{j=1}^{7} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \sqrt{\frac{2}{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right).$$

Therefore, the best match to this query is **D5**.

(c) (10 pts) Describe why SVD is helpful for searching documents containing a given set of terms for a large term-by-document matrix. (Note: No need to compute the SVD of the term-by-document matrix of Part (a).)

Answer: There are at least two advantages to use SVD, in particular, the low rank approximation of a large term-by-document matrix A.

(1) A contains a lot of noise (variation and ambiguity in the use of vocabulary, etc.). Thus, the low rank approximation via SVD can filter out such noise.

(2) The computation of the match via $\cos\theta_j$ can be faster thanks to the low rank approximation. In fact, let $A_k = \sum_{j=1}^{k} \sigma_j \mathbf{u}_j \mathbf{v}_j^T$ be the rank k approximation of A where $1 < k \geq \text{rank}(A)$. Let us write

$$A_k = U_k \Sigma_k V_k^T = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_k \end{bmatrix} \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_k \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_k^T \end{bmatrix}.$$

Then, the better (and computationally faster) version of the measure of match is:

$$\cos\phi_j = \frac{\langle \boldsymbol{q}, A_k \boldsymbol{e}_j \rangle}{\|\boldsymbol{q}\|_2 \|A_k \boldsymbol{e}_j\|_2}.$$

This version is faster than the original version because

$$\begin{aligned} \|A_k \boldsymbol{e}_j\|_2 &= \|U_k \Sigma_k V_k^T \boldsymbol{e}_j\|_2 \\ &= \|U_k S_k \boldsymbol{e}_j\|_2 \quad \text{by setting } S_k = \Sigma_k V_k^T; \\ &= \|U_k \boldsymbol{s}_j\|_2 \quad \boldsymbol{s}_j \text{ is the } j\text{th column of } S_k; \\ &= \|\boldsymbol{s}_j\|_2 \quad \text{since } U_k\text{'s columns are orthonormal,} \end{aligned}$$

and consequently we have

$$\cos\phi_j = \frac{\langle \boldsymbol{q}, U_k \boldsymbol{s}_j \rangle}{\|\boldsymbol{q}\|_2 \|\boldsymbol{s}_j\|_2}.$$

Now, we only need U_k, Σ_k, V_k separately without explicitly forming $A_k = U_k \Sigma_k V_k^T$, and moreover, $\boldsymbol{s}_j$'s do not depend on a query vector $\boldsymbol{q}$ so that they can be precomputed for a given A.

Problem 6 (20 pts) Let $A \in \mathbb{C}^{m\times n}$ and its SVD be $A = U\Sigma V^*$. Let us assume that $\text{rank}(A) = r$. Let us now define the *pseudoinverse* of A as:

$$A^\dagger = V\Sigma^\dagger U^*, \quad \Sigma^\dagger = \begin{bmatrix} D^{-1}_{r\times r} & O_{r\times(m-r)} \\ O_{(n-r)\times r} & O_{(n-r)\times(m-r)} \end{bmatrix}, \quad D^{-1}_{r\times r} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{1}{\sigma_r} \end{bmatrix},$$

where $O_{k\times \ell}$ denotes $k \times \ell$ matrix whose entries are all zeros.

(a) (10 pts) Suppose $r = n \le m$. Then show that $A^\dagger = (A^* A)^{-1} A^*$.

Answer: Because $r = n \le m$, the Σ part of the SVD of A is of the form

$$\Sigma = \left[\begin{array}{ccc} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n \\ \hline & O_{(m-n)\times n} & \end{array}\right] \in \mathbb{R}^{m\times n}.$$

Therefore,

$$A^* A = V\Sigma^* U^* U\Sigma V^* = V\Sigma^*\Sigma V^* = V \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{bmatrix} V^*.$$

Thus,

$$(A^* A)^{-1} = V \begin{bmatrix} \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix} V^*.$$

Finally, we have

$$\begin{aligned}
(A^*A)^{-1}A^* &= V \begin{bmatrix} \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix} V^*V\Sigma^*U^* \\
&= V \begin{bmatrix} \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix} \left[\begin{array}{ccc|c} \sigma_1 & \cdots & 0 & \\ \vdots & \ddots & \vdots & O_{n\times(m-n)} \\ 0 & \cdots & \sigma_n & \end{array}\right] U^* \\
&= V \left[\begin{array}{ccc|c} \frac{1}{\sigma_1} & \cdots & 0 & \\ \vdots & \ddots & \vdots & O_{n\times(m-n)} \\ 0 & \cdots & \frac{1}{\sigma_n} & \end{array}\right] U^* \\
&= V\Sigma^\dagger U^* \\
&= A^\dagger.
\end{aligned}$$

(b) (10 pts) Prove $AA^\dagger$ is an orthogonal projector onto $\mathcal{R}(A)$. Hint: You can use that $A^\dagger$ satisfies the following *Moore-Penrose* condition:

$$\begin{aligned} AA^\dagger A &= A & \qquad A^\dagger AA^\dagger &= A^\dagger \\ (AA^\dagger)^* &= AA^\dagger & \qquad (A^\dagger A)^* &= A^\dagger A. \end{aligned}$$

Answer: To show $AA^\dagger$ is an orthogonal projector onto $\mathcal{R}(A)$, we need to show the following four items.

(1) $AA^\dagger$ is a square matrix of size $m \times m$:
This is obvious because $A \in \mathbb{C}^{m\times n}$ and $A^\dagger \in \mathbb{C}^{n\times m}$ from its definition.

(2) $(AA^\dagger)^2 = AA^\dagger$:
This can be easily show as follows.

$$(AA^\dagger)^2 = AA^\dagger AA^\dagger = (AA^\dagger A)A^\dagger = AA^\dagger,$$

using the first of the Moore-Penrose condition.

(3) $(AA^\dagger)^* = AA^\dagger$:
This is simply the third of the Moore-Penrose condition listed above.

(4) For any $\boldsymbol{x} \in \mathcal{R}(A)$, $AA^\dagger \boldsymbol{x} = \boldsymbol{x}$:
Because $\boldsymbol{x} \in \mathcal{R}(A)$, there exists $\boldsymbol{y} \in \mathbb{C}^n$ such that $\boldsymbol{x} = A\boldsymbol{y}$. Now,

$$AA^\dagger \boldsymbol{x} = AA^\dagger A\boldsymbol{y} = A\boldsymbol{y} = \boldsymbol{x},$$

where the first of the Moore-Penrose condition was used again.

Therefore, $AA^\dagger$ is an orthogonal projector onto $\mathcal{R}(A)$.