Numerical Questions on Matrix Decompositions, SVD, and LSI

1. SVD Decomposition

Given the following 3×2 term-document matrix C:

$$C = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 4 \end{pmatrix}$$

a) Perform the **Singular Value Decomposition (SVD)** of matrix C. Find matrices U, Σ, and V^T.

b) Verify that the decomposition $C = U\Sigma V^T$ holds true by multiplying the decomposed matrices.

2. Low-Rank Approximation

Consider the same term-document matrix C from Question 1:

$$C = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 3 & 4 \end{pmatrix}$$

a) Compute a **rank-1 approximation** of C using the **largest singular value** from the SVD. Use $C_1 = U_1\Sigma_1 V_1^T$, where U_1, Σ_1, and V_1^T retain only the top singular value.

b) Calculate the **Frobenius norm** of the error between the original matrix C and its rank-1 approximation C_1. The Frobenius norm is given by:

$$\|C - C_1\|_F = \sqrt{\sum (C_{ij} - (C_1)_{ij})^2}$$

3. Eigenvalues and Eigenvectors

For the matrix:

$$A = \begin{pmatrix} 4 & 1 \\ 2 & 3 \end{pmatrix}$$

a) Find the **eigenvalues** and **eigenvectors** of matrix A.

b) Use these eigenvalues and eigenvectors to verify the **matrix diagonalization theorem**: $A = U\Lambda U^{-1}$.

4. Term-Document Similarity in LSI

Given the following term-document matrix C:

$$C = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

a) Perform the **SVD** of the matrix C to find the reduced representation C_2, retaining only the top 2 singular values.

b) Suppose you have a query vector $q = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix}$. Transform the query vector into the **LSI space** using the reduced matrices U, Σ_2, and V_2^T.

c) Calculate the **cosine similarity** between the transformed query vector and the documents in the reduced space.

5. Low-Rank Approximation Error

Consider a 3×3 term-document matrix:

$$C = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 3 & 1 \\ 0 & 1 & 3 \end{pmatrix}$$

a) Perform the **SVD** of matrix C and find the top 2 singular values.

b) Construct the **rank-2 approximation** C_2.

c) Compute the **Frobenius norm** of the error between the original matrix C and the rank-2 approximation C_2.

6. Query-Document Similarity in LSI

Given the term-document matrix C:

$$C = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 0 & 1 \\ 2 & 1 & 0 \end{pmatrix}$$

a) Perform the **SVD** of matrix C and keep the top 2 singular values to create a reduced representation C_2.

b) Suppose you have a query vector $q = \begin{pmatrix} 1 & 0 & 1 \end{pmatrix}$. Transform this query vector into the reduced LSI space.

c) Calculate the **cosine similarity** between the transformed query and each document in the reduced space.

7. Dimensionality Reduction

Given a term-document matrix:

$$C = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

a) Perform **SVD** and retain only the top singular value to form a rank-1 approximation.

b) Compute the approximation matrix C_1 and compare it with the original matrix.