Problem 1 (CX and RRQR). Recall that an RRQR decomposition of a matrix $A \in \mathbb{R}^{n \times m}$ is of form

$$A \Pi = QR = Q \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix},$$

(1.1)

where $\Pi \in \{0,1\}^{m \times m}$ is a permutation matrix, $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, $R_{11} \in \mathbb{R}^{k \times k}$ is upper-triangular with positive values in diagonal, and $R_{12} \in \mathbb{R}^{k \times (m-k)}$ and $R_{22} \in \mathbb{R}^{(n-k) \times (m-k)}$ are arbitrary.

Let $\Pi_k \{0,1\}^{n \times k}$ be the first $k$ columns of $\Pi$ and set $C = A \Pi_k \in \mathbb{R}^{n \times k}$. Show that

$$\|A - CC^+ A\|_\xi = \|R_{22}\|_\xi,$$

(1.2)

where $\xi$ is either $F$ or $2$ (i.e. we compute either the Frobenius or spectral norm).

**Hint:** Use the fact that $R_{11}$ is guaranteed to be invertible and that both of the studied norms are orthogonally invariant.

Problem 2 (CX and RRQR again). Let $A \Pi = QR$ be the RRQR factorization of $A$ as above. Assume the factorization admits the following inequalities for some polynomials $p_1$ and $p_2$ over $k$ and $m$:

$$\frac{\sigma_k(A)}{p_1(k,m)} \leq \sigma_{\text{min}}(R_{11}) \leq \sigma_k(A),$$

(2.1)

$$\sigma_{k+1}(A) \leq \sigma_{\text{max}}(R_{22}) \leq p_2(k,m) \sigma_{k+1}(A).$$

(2.2)

Using $\mathbf{1.2}$ from Problem 1 and the above inequalities, show that

$$\|A - CC^+ A\|_2 \leq p_2(k,m) \|A - A_k\|_2,$$

(2.3)

where $A_k = U_k \Sigma_k V_k^T$ is the rank-$k$ truncated SVD of $A$.

Problem 3 (CX and sparse decompositions). Bob is a big proponent of CX decomposition, and he claims that if matrix $A$ is sparse and you do a normal CX decomposition to it, the column matrix $C$ must also be sparse.

a) Prove Bob wrong. Construct matrix $A$ such that $A$ is sparse, but in an optimal rank-$k$ CX decomposition matrix $C$ is not sparse. Matrix $A \in \mathbb{R}^{n \times m}$ is sparse if $\text{nnz}(A)/(nm) \ll 0.5$ and it is not sparse if $\text{nnz}(A)/(nm) \gg 0.5$, where $\text{nnz}(A) = \{(i,j): a_{ij} \neq 0\}$ is the number of non-zero elements in $A$. Your matrix can be of any size, you can choose any rank $k > 0$ and the non-sparse optimal decomposition does not have to be unique (i.e. there can be other decompositions that yield equal reconstruction error, but have sparse $C$).

b) Bob insists that even if CX doesn’t yield sparse decompositions, NNCX will. Prove Bob wrong again by constructing sparse nonnegative $A$ that has an NNCX decomposition where $C$ is not sparse. The rules are as above, but you must construct a new example even if your previous example was already an NNCX decomposition.
**Problem 4** (Generating CUR data). A standard practice when validating that a proposed matrix factorization algorithm works in practice is to generate random data that has the kind of structure the factorization aims at finding, add some random, structure-less noise, and use the resulting matrix as an input for the algorithm. For example, for NMF, we would first choose some \( n \), \( m \), and \( k \), then we would generate random matrices \( W \in \mathbb{R}^{n \times k} \) and \( H \in \mathbb{R}^{k \times m} \), multiply them to obtain \( A = WH \), and add some noise to \( A \).

Design a method that creates random synthetic matrices for CUR decomposition. That is, explain how to generate matrices \( C \in \mathbb{R}^{n \times k} \), \( U \in \mathbb{R}^{k \times k} \), and \( R \in \mathbb{R}^{k \times m} \) (\( k < n, m \)) such that matrix \( A = CUR \) has \( k \) columns that are exactly the columns of \( C \) and \( k \) rows that are exactly the rows of \( R \). The factor matrices cannot be completely random, but try to have as much randomness as possible.

**Problem 5** (Correlation matrix). Let \( x = (x_i)_{i=1}^n \) be a (column) vector of \( n \) zero-centered random variables. The covariance \( \text{cov}(x_i, x_j) \) is defined as

\[
\text{cov}(x_i, x_j) = \mathbb{E}[x_i x_j],
\]

(5.1)

The correlation matrix \( \Sigma \) is defined as

\[
\Sigma = \mathbb{E}[xx^T] = (\text{cov}(x_i, x_j))_{i,j}.
\]

(5.2)

What are the requirements for random variables \( x_i \) that ensure that the covariance matrix is an identity matrix? Give the requirements, and prove that if all \( x_i \) satisfy them, \( \Sigma \) is an identity matrix.

**Hint:** consider what \( \Sigma_{i,i} = \text{cov}(x_i, x_i) \) tells about random variable \( x_i \).

**Problem 6** (Whitening). Most textbooks (and Wikipedia) explain the whitening process as follows: Given data matrix \( A \) (where rows are observations and columns variables), compute the correlation matrix \( C = A^T A \). Then, compute the eigendecomposition of \( C \), \( C = Q \Delta Q^T \), where \( Q \) is an orthogonal matrix and \( \Delta \) is diagonal matrix with non-negative entries. To whiten \( A \), we multiply \( A \) from right with \( Q \Delta^{1/2} \), where \( (\Delta^{1/2})_{ii} = 1/\sqrt{(\Delta)_{ii}} \) if \( (\Delta)_{ii} \neq 0 \) and \( (\Delta^{-1/2})_{ii} = 0 \) otherwise.

In the lectures it was claimed that if \( U \Sigma V^T \) is the SVD of \( A \), then the whitened \( A \) is \( U \). Prove that these two processes yield the same solution, that is

\[
U = AQ \Delta^{-1/2}.
\]

(6.1)

**Hint:** eigendecomposition is unique, that is, if \( C = Q \Delta Q^T \) for some orthogonal \( Q \) and diagonal \( \Delta \) with nonnegative entries, then \( Q \Delta Q^T \) is the eigendecomposition of \( C \). Use the SVD of \( A \) to express \( C \) and find a definition of \( Q \) and \( \Delta \) in terms of SVD of \( A \).