D5: Databases and Information Systems
Data Mining and Matrices, SS 2017
Homework #4: CX, CUR, and ICA
Tutorial: **5 July 2017** at 10:15

You can discuss these problems with other students, but everybody must do and present their own answers. You can use computers etc. to perform the algebraic operations, but you must show the intermediate steps (and "computer said so" is never a valid answer). You are of course free to use material from the Internet, but again, you must present the intermediate steps and you must also be able to explain why the steps are valid and why you chose them. You can mark an answer even if it is not complete or correct, as long as you have made significant progress towards solving it. Note, however, that the TA does the final decision on whether your solution is complete (or correct) enough for a mark. ***We continue to apply more strict evaluation on what constitutes as sufficiently solved problem. You should only mark a problem if you think you have essentially solved it. This is done to give you a better impression on how the exam questions will be graded.***

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

$$\boldsymbol{A\Pi} = \boldsymbol{QR} = \boldsymbol{Q}\begin{pmatrix}\boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ 0 & \boldsymbol{R}_{22}\end{pmatrix} , \tag{1.1}$$

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

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

$$\left\|\boldsymbol{A} - \boldsymbol{CC}^{+}\boldsymbol{A}\right\|_{\xi} = \left\|\boldsymbol{R}_{22}\right\|_{\xi} , \tag{1.2}$$

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

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

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

$$\frac{\sigma_k(\boldsymbol{A})}{p_1(k,m)} \le \sigma_{\min}(\boldsymbol{R}_{11}) \le \sigma_k(\boldsymbol{A}) \tag{2.1}$$

$$\sigma_{k+1}(\boldsymbol{A}) \le \sigma_{\max}(\boldsymbol{R}_{22}) \le p_2(k,m)\sigma_{k+1}(\boldsymbol{A}) . \tag{2.2}$$

Using (1.2) from Problem 1 and the above inequalities, show that

$$\left\|\boldsymbol{A} - \boldsymbol{CC}^{+}\boldsymbol{A}\right\|_2 \le p_2(k,m)\left\|\boldsymbol{A} - \boldsymbol{A}_k\right\|_2 , \tag{2.3}$$

where $\boldsymbol{A}_k = \boldsymbol{U}_k\boldsymbol{\Sigma}_k\boldsymbol{V}_k^T$ is the rank-k truncated SVD of $\boldsymbol{A}$.

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

a) Prove Bob wrong. Construct matrix $\boldsymbol{A}$ such that $\boldsymbol{A}$ is sparse, but in an optimal rank-k CX decomposition matrix $\boldsymbol{C}$ is not sparse. Matrix $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ is sparse if $\mathrm{nnz}(\boldsymbol{A})/(nm) \ll 0.5$ and it is not sparse if $\mathrm{nnz}(\boldsymbol{A})/(nm) \gg 0.5$, where $\mathrm{nnz}(\boldsymbol{A}) = |\{(i,j) : a_{ij} \neq 0\}|$ is the number of non-zero elements in $\boldsymbol{A}$. 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 $\boldsymbol{C}$).

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 $\boldsymbol{C}$ 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.

$\begin{pmatrix} \mathbf{1} & \mathbf{1} & \mathbf{0} \\ \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{1} & \mathbf{1} \end{pmatrix}$

Problem 4 (Generating CUR data). A standard practise 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 $\boldsymbol{W} \in \mathbb{R}_+^{n\times k}$ and $\boldsymbol{H} \in \mathbb{R}_+^{k\times m}$, multiply them to obtain $\boldsymbol{A} = \boldsymbol{W}\boldsymbol{H}$, and add some noise to $\boldsymbol{A}$.

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

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

$$\operatorname{cov}(x_i, x_j) = \mathbb{E}[x_i x_j] \ , \tag{5.1}$$

The *correlation matrix* $\boldsymbol{\Sigma}$ is defined as

$$\boldsymbol{\Sigma} = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^T] = \big(\operatorname{cov}(x_i, x_j)\big)_{i,j} \ . \tag{5.2}$$

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, $\boldsymbol{\Sigma}$ is an identity matrix.

Hint: consider what $\boldsymbol{\Sigma}_{i,i} = \operatorname{cov}(x_i, x_i)$ tells about random variable x_i.

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

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

$$\boldsymbol{U} = \boldsymbol{A}\boldsymbol{Q}\boldsymbol{\Delta}^{-1/2} \ . \tag{6.1}$$

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