Fast Matrix Algorithms for Data Analytics: Problem Set 2 Solutions

1. Let $\mathbf{A} \in \mathbb{R}^{m\times n}$, J be an $n \times 1$ permutation vector, and $J_s = J(1:k)$ for some $k \in \mathbb{N}$. Now suppose

$$\underset{m\times n}{\mathbf{A}} = \underset{m\times k}{\mathbf{E}} \; \underset{k\times n}{\mathbf{F}}, \tag{1}$$

and suppose

$$\mathbf{F}(:,J) = \mathbf{F}(:,J_s)\begin{bmatrix}\mathbf{I} & \mathbf{T}\end{bmatrix}.$$

Prove that

$$\mathbf{A}(:,J) = \mathbf{A}(:,J_s)\begin{bmatrix}\mathbf{I} & \mathbf{T}\end{bmatrix}.$$

Solution: Start with the ID of $\mathbf{F}$ and multiply by $\mathbf{E}$:

$$\mathbf{EF}(:,J) = \mathbf{EF}(:,J_s)[\mathbf{I} \quad \mathbf{T}].$$

Then $\mathbf{EF}(:,J) = \mathbf{A}(:,J)$ by (1). Similarly, $\mathbf{EF}(:,J_s) = \mathbf{A}(:,J_s)$, so

$$\mathbf{A}(:,J) = \mathbf{EF}(:,J) = \mathbf{EF}(:,J_s)[\mathbf{I}\,\mathbf{T}] = \mathbf{A}(:,J_s)[\mathbf{I} \quad \mathbf{T}].$$

2. Suppose that $\mathbf{A}$ is an $m \times n$ matrix of precise rank k. Moreover, suppose that you have available a double-sided ID of $\mathbf{A}$ of the form

$$\underset{m\times n}{\mathbf{A}} = \underset{m\times k}{\mathbf{X}} \; \underset{k\times k}{\mathbf{A}_{\rm s}} \; \underset{k\times n}{\mathbf{Z}}, \tag{2}$$

where $\mathbf{A}_{\rm s} = \mathbf{A}(I_{\rm s}, J_{\rm s})$ for some index vectors $I_{\rm s}$ and $J_{\rm s}$.

   (a) Prove that $\mathbf{A}_{\rm s}$ must be of full rank.

   Solution: Since the rows of $\mathbf{I}_k$ are a subset of the rows of $\mathbf{X}$ and the columns of $\mathbf{I}_k$ are a subset of the columns of $\mathbf{Z}$, we know that $\mathbf{X}$ and $\mathbf{Z}$ have full column and row rank, respectively. Therefore, we have

$$k = \text{rank}(\mathbf{A}) = \text{rank}(\mathbf{X}\mathbf{A}_s\mathbf{Z}) = \text{rank}(\mathbf{A}_s\mathbf{Z}) = \text{rank}(\mathbf{A}_s).$$

   Thus $\mathbf{A}_s$ has full rank.

   (b) Set $\mathbf{U} = \mathbf{A}_{\rm s}^{-1}$, so that (2) can be written

$$\underset{m\times n}{\mathbf{A}} = \underset{m\times k}{\mathbf{X}\mathbf{A}_{\rm s}} \; \underset{k\times k}{\mathbf{U}} \; \underset{k\times n}{\mathbf{A}_{\rm s}\mathbf{Z}}. \tag{3}$$

Prove that $\mathbf{X}\mathbf{A}_\mathrm{s} = \mathbf{A}(:, J_\mathrm{s})$ and $\mathbf{A}_\mathrm{s}\mathbf{Z} = \mathbf{A}(I_\mathrm{s}, :)$ so that (3) is a CUR factorization.

Solution: We have $\mathbf{A}(:, J_s) = \mathbf{X}\mathbf{A}_s\mathbf{Z}(:, J_s)$. Since, in the double-sided ID, we have

$$\mathbf{A} = \mathbf{A}(:, J_s)\mathbf{Z} \implies \mathbf{A}(:, J_s) = \mathbf{A}(:, J_s) = \mathbf{A}(:, J_s)\mathbf{Z}(:, J_s) \implies \mathbf{Z}(:, J_s) = \mathbf{I}_k,$$

it follows that

$$\mathbf{A}(:, J_s) = \mathbf{X}\mathbf{A}_s.$$

A similar argument shows that

$$\mathbf{A}(I_s, :) = \mathbf{X}(I_s, :)\mathbf{A}_s\mathbf{Z}(:, J_s) = \mathbf{A}_s\mathbf{Z}(:, J_s).$$

3. Suppose that $\mathbf{A}$ is an $m \times n$ matrix of approximate rank k, and that we have identified two index sets I_s and J_s such that the matrices

$$\mathbf{C} = \mathbf{A}(:, J_s), \quad \mathbf{R} = \mathbf{A}(I_s, :)$$

hold k columns/rows that span the column/row space of $\mathbf{A}$. Then

$$\mathbf{A} \approx \mathbf{C}\mathbf{C}^\dagger\mathbf{A}\mathbf{R}^\dagger\mathbf{R},$$

and the optimal choice for the "U" factor in the CUR decomposition is

$$\mathbf{U} = \mathbf{C}^\dagger\mathbf{A}\mathbf{R}^\dagger.$$

Set $\mathbf{X} = \mathbf{C}\mathbf{C}^\dagger$.

(a) Suppose that $\mathbf{C}$ has the SVD $\mathbf{C} = \mathbf{W}\mathbf{D}\mathbf{V}^*$. Prove that $\mathbf{X} = \mathbf{W}\mathbf{W}^*$.

Solution: To start with, we make the reasonable assumption that the columns of $\mathbf{C}$ form a linearly independent set (if $\dim(\mathrm{Col}(\mathbf{C})) = p < k$ but $\mathrm{Col}(\mathbf{C}) \approx \mathrm{Col}(\mathbf{A})$, then the problem statement would have said that $\mathbf{A}$ has approximate rank p). Therefore, assuming without loss of generality that the given SVD of $\mathbf{C}$ is the "economic" version (to ensure that $\mathbf{\Sigma}^{-1}$ exists and therefore $\mathbf{\Sigma}^\dagger = \mathbf{\Sigma}^{-1}$), we have by the definition of the pseudoinverse that

$$\mathbf{C}^\dagger = \mathbf{V}\mathbf{\Sigma}^\dagger\mathbf{W}^* = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}^*.$$

Therefore, using the fact that $\mathbf{V}$ is orthonormal implies $\mathbf{V}^*\mathbf{V} = \mathbf{I}$, we have

$$\mathbf{X} = \mathbf{C}\mathbf{C}^\dagger = \mathbf{W}\mathbf{\Sigma}\mathbf{V}^*\mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{W}^* = \mathbf{W}\mathbf{\Sigma}\mathbf{\Sigma}^{-1}\mathbf{W}^* = \mathbf{W}\mathbf{W}^*.$$

(note that the distribution of the † is legal because $\mathbf{W}$ has orthonormal columns and $\mathbf{V}^*$ has orthonormal rows.)

(b) Suppose that $\mathbf{C}$ has the QR factorization $\mathbf{CP} = \mathbf{QS}$. Prove that $\mathbf{X} = \mathbf{QQ}^*$.

Solution: First, we mention that we will again use the assumption that $\mathbf{C}$ has full column rank. This, along with the assumption that the QR factorization given is the economic factorization so that $\mathbf{S} \in \mathbb{R}^{\min(m,n)\times n}$, guarantees that, in fact, $\mathbf{S} \in \mathbb{R}^{n\times n}$ and $\mathbf{S}$ is invertible. Therefore, since $\mathbf{CP} = \mathbf{QS} \implies \mathbf{C} = \mathbf{QSP}^*$, we have that

$$\mathbf{X} = \mathbf{CC}^\dagger = \mathbf{QSP}^*(\mathbf{QSP}^*)^\dagger = \mathbf{QSP}^*\mathbf{PS}^\dagger\mathbf{Q}^\dagger = \mathbf{QSS}^{-1}\mathbf{Q}^* = \mathbf{QQ}^*.$$

(note that the distribution of the † is again legal because $\mathbf{Q}$ has orthonormal columns and $\mathbf{P}^*$ has orthonormal rows.)

(c) Prove that $\mathbf{X}$ is the orthogonal projection onto $\mathrm{Col}(\mathbf{C})$.

Solution: Assume as in part a) that $\mathbf{C}$ has the SVD $\mathbf{C} = \mathbf{WDV}^*$. Then the result of part a) gives us that $\mathbf{X} = \mathbf{WW}^*$. Therefore,

$$\mathbf{X}^2 = \mathbf{W}(\mathbf{W}^*\mathbf{W})\mathbf{W}^* = \mathbf{WW}^* = \mathbf{X},$$

so $\mathbf{X}$ is a projection. Furthermore,

$$\mathbf{X}^* = (\mathbf{WW}^*)^* = \mathbf{WW}^* = \mathbf{X},$$

so $\mathbf{X}$ is self-adjoint and thus an orthogonal projection. Now we only need to prove that $\mathrm{Col}(\mathbf{C}) = \mathrm{Col}(\mathbf{X})$.

Since $\mathbf{X} = \mathbf{WW}^\dagger$, every column of $\mathbf{X}$ is a linear combination of the columns of $\mathbf{W}$, so $\mathrm{Col}(\mathbf{X}) \subseteq \mathrm{Col}(\mathbf{W})$. Finally, since $\mathbf{W}^*$ has rank k and $\mathbf{W}$ does as well, since the columns of $\mathbf{W}$ form an orthonormal basis for $\mathrm{Col}(\mathbf{C})$, and since $\mathrm{Col}(\mathbf{X}) \subseteq \mathrm{Col}(\mathbf{C})$, we get $\mathrm{Col}(\mathbf{X}) = \mathrm{Col}(\mathbf{WW}^*) = \mathrm{Col}(\mathbf{W}) = \mathrm{Col}(\mathbf{C})$.

(d) Suppose that $\mathbf{A}$ has precisely rank k and that $\mathbf{C}$ and $\mathbf{R}$ are both of rank k. Prove that then $\mathbf{C}^\dagger\mathbf{AR}^\dagger = (\mathbf{A}(I_s, J_s))^{-1}$.

Solution: As we showed in class, since $\mathbf{A}$ is rank k and $\mathbf{C} = \mathbf{A}(:, J_s)$ and $\mathbf{R} = \mathbf{A}(I_s, :)$ are rank k as well, we have that

$$\mathbf{A} = \mathbf{C}(\mathbf{A}(I_s, J_s)^{-1})\mathbf{R}.$$

Therefore, we have that

$$\mathbf{C}^\dagger\mathbf{AR}^\dagger = \mathbf{C}^\dagger\mathbf{C}(\mathbf{A}(I_s, J_s))^{-1}\mathbf{RR}^\dagger.$$

Let $\mathbf{C} = \mathbf{U}_C \mathbf{D}_C \mathbf{V}_C^*$ be the SVD of $\mathbf{C}$. Then since $\mathbf{C} \in \mathbb{C}^{m \times k}$, $\mathbf{V_C} \in \mathbb{C}^{k \times k}$, and so $\mathbf{V}_C \mathbf{V}_C^* = \mathbf{I}$. Therefore, since $\mathbf{C}$ has full rank implies that $\mathbf{D}_C$ is invertible, we have that

$$\mathbf{C}^\dagger \mathbf{C} = \mathbf{V}_C \mathbf{D}_C^{-1} \mathbf{U}_C^* \mathbf{U}_C \mathbf{D}_C \mathbf{V}_C^* = \mathbf{V}_C \mathbf{V}_C^* = I.$$

Similarly, if $\mathbf{R} = \mathbf{U}_R \mathbf{D}_R \mathbf{V}_R^*$ is the SVD of $\mathbf{R}$, we have that

$$\mathbf{R}\mathbf{R}^\dagger = \mathbf{U}_R \mathbf{D}_R \mathbf{V}_R^* \mathbf{V}_R \mathbf{D}_R^{-1} \mathbf{U}_R^* = \mathbf{U}_R \mathbf{U}_R^* = \mathbf{I},$$

where $\mathbf{U}_R \mathbf{U}_R^* = \mathbf{I}$ because $\mathbf{U} \in \mathbb{C}^{k \times k}$ and is orthonormal, and $\mathbf{D}_R^{-1}$ exists because $\mathbf{R}$ is full rank. Therefore, we have shown that

$$\mathbf{C}^\dagger \mathbf{A} \mathbf{R}^\dagger = \mathbf{C}^\dagger \mathbf{C} (\mathbf{A}(I_s, J_s))^{-1} \mathbf{R}\mathbf{R}^\dagger = \mathbf{I} (\mathbf{A}(I_s, J_s))^{-1} \mathbf{I} = (\mathbf{A}(I_s, J_s))^{-1}.$$

4. Let $\mathbf{A} \in \mathbb{R}^{m \times n}$ have rank exactly k. In this problem, we will prove that $\mathbf{A}$ admits a factorization $\mathbf{A} = \mathbf{A}(:, J_s)\mathbf{Z}$, where $\mathbf{A}(:, J_s) \in \mathbb{R}^{m \times k}$ and $\mathbf{Z} \in \mathbb{R}^{k \times n}$ such that $\mathbf{Z}(:, J_s) = \mathbf{I}_k$ and $\max_{i,j} |\mathbf{Z}(i,j)| \leq 1$.

   (a) **case 1:** $m = k$.

      i. Pick a permutation vector J_s such that $|\det(\mathbf{A}(:, J_s))|$ is maximized, and let J_r denote the remaining indices so that $[J_s \quad J_r]$ is some permutation of the vector $[1 \ 2 \quad \cdots \quad n]$. Then we have that

$$\mathbf{A}(:, [J_s \quad J_r]) = [\mathbf{A}(:, J_s) \quad \mathbf{A}(:, J_r)]$$

      can be written as $\mathbf{AP}$ for some permutation matrix $\mathbf{P}$. Find an interpolative decomposition $\mathbf{A} = \mathbf{CZ}$ of $\mathbf{A}$, where the columns of $\mathbf{C}$ are some of the columns of $\mathbf{A}$. $\mathbf{C}$ and $\mathbf{Z}$ should be in terms of $\mathbf{A}(:, J_s), \mathbf{A}(:, J_r), \mathbf{P}$, and the identity matrix $\mathbf{I}$.
      Solution: Since we have $\mathbf{AP} = [\mathbf{A}(:, J_s) \quad \mathbf{A}(:, J_r)]$, we can write

$$\mathbf{A} = \mathbf{A}(:, J_s)[\mathbf{I}_k \quad \mathbf{A}(:, J_s)^{-1}\mathbf{A}(:, J_r)]\mathbf{P}^*.$$

      Thus, setting $\mathbf{C} = \mathbf{A}(:, J_s)$ and $\mathbf{Z} = [\mathbf{I}_k \quad \mathbf{A}(:, J_s)^{-1}\mathbf{A}(:, J_r)]\mathbf{P}^*$, we have our interpolative decomposition.

      ii. Consider the matrix $\mathbf{T} = \mathbf{A}(:, J_s)^{-1}\mathbf{A}(:, J_r)$. If we can show that

$$\max_{i,j} |\mathbf{T}(i,j)| \leq 1, \tag{4}$$

then we will be done with the case $m = k$ (why?). Find a way to show (4) by applying Cramer's Rule to our definition of $\mathbf{T}$.

Cramer's Rule: Consider the linear system $\mathbf{Ax} = \mathbf{b}$. The i-th entry of the solution $\mathbf{x}$ is given by

$$x_i = \frac{\det(\mathbf{A}_i)}{\det(\mathbf{A})},$$

where $\mathbf{A}_i$ is matrix formed by replacing the i-th column of $\mathbf{A}$ with $\mathbf{b}$.
Solution: We have $\mathbf{T} = \mathbf{A}(:, J_s)^{-1}\mathbf{A}(:, J_r)$. We can also write this as $\mathbf{A}(:, J_s)\mathbf{T} = \mathbf{A}(:, J_r)$, where $\mathbf{T}$ is the solution to this equation. By Cramer's Rule, we have that

$$\mathbf{T}(1, 1) = \frac{\det(\mathbf{A}_1)}{\det(\mathbf{A}(:, J_s))},$$

where $\mathbf{A}_1$ is $\mathbf{A}(:, J_s)$ with the first column from $\mathbf{A}(:, J_r)$ replacing the first column of $\mathbf{A}(:, J_s)$. Because of our criterion in choosing J_s, we know that $|\det(\mathbf{A}(:, J_s))| \geq |\det(\mathbf{A}_1)|$, so we have

$$|\mathbf{T}(1, 1)| = \left|\frac{\det(\mathbf{A}_1)}{\det(\mathbf{A}(:, J_s))}\right| \leq 1.$$

Similarly, $|\mathbf{T}(i, j)| \leq 1$ for all i, j.

(b) **case 2:** $m \geq k$.
Then $\mathbf{A}$ admits a factorization $\mathbf{A} = \mathbf{EF}$, where $\mathbf{E}$ is $m \times k$ and $\mathbf{F}$ is $k \times n$. Apply case 1 to $\mathbf{F}$ to show the result for this case.

Solution: We showed that $\mathbf{F}$ admits a factorization $\mathbf{F} = \mathbf{F}(:, J_s)\mathbf{Z}$ for a $\mathbf{Z}$ that satisfies all the criteria. Then by problem 1,

$$\mathbf{A} = \mathbf{A}(:, J_s)\mathbf{Z}.$$