Problem 1 (Maximum rank). It was stated in the lectures that the rank of a tensor $T \in \mathbb{R}^{I \times J \times K}$ is never more than $\min\{IJ, IK, JK\}$.

Let $I, J$, and $K$ be that $JK = \min\{IJ, IK, JK\}$ and let $T \in \mathbb{R}^{I \times J \times K}$ be arbitrary. Your task is to construct $A \in \mathbb{R}^{I \times JK}$, $B \in \mathbb{R}^{J \times JK}$, and $C \in \mathbb{R}^{K \times JK}$ such that $T_{(1)} = A(C \odot B)^T$.

Hint: Construct $B$ from identity matrices.

Problem 2 (Nonnegative INDSCAL). Present an algorithm for nonnegative 3-way INDSCAL. That is, given a nonnegative 3-way tensor $T \in \mathbb{R}_{\geq 0}^{I \times J \times K}$ and an integer $R$, find matrices $A \in \mathbb{R}_{\geq 0}^{I \times R}$, $B \in \mathbb{R}_{\geq 0}^{J \times R}$, and $C \in \mathbb{R}_{\geq 0}^{K \times R}$ that aim at minimizing $\|T - [A, B, C]\|$.

Problem 3 (CP-APR for KL-divergence). In CP-APR, we need to find a matrix $A$ that minimizes

$$L(A) = A(C \odot B)^T - T_{(1)} \ast \log(A(C \odot B)^T) .$$

This is a type of a KL divergence. In nonnegative matrix factorization (NMF), we are given a nonnegative matrix $A \in \mathbb{R}_{\geq 0}^{I \times J}$ and an integer $K$ and we have to find nonnegative matrices $W \in \mathbb{R}_{\geq 0}^{I \times K}$ and $H \in \mathbb{R}_{\geq 0}^{K \times J}$ such that $A \approx WH$.

The standard NMF algorithm for KL divergence has the following update rule:

$$W_{ik} \leftarrow W_{ik} \frac{\sum_{j=1}^{m} (A_{ij} / (WH)_{ij}) H_{kj}}{\sum_{j=1}^{m} H_{kj}} .$$

Adapt this update rule for the factor matrix $A$ in the CP decomposition. How does it relate to the update rule

$$ A \leftarrow A \ast (T_{(1)} \odot (A(C \odot B)^T))(C \odot B)^T ,$$

presented in the lecture? (To recap, $\odot$ is the element-wise division.)
**Problem 4 (PARAFAC2).** The PARAFAC2 decomposition is another variant of the CP decomposition, defined slice-wise as follows. Given \( K \) matrices \( X_k \in \mathbb{R}^{I_k \times J} \) and rank \( R \), find \( K \) matrices \( U_k \in \mathbb{R}^{I_k \times R} \), diagonal matrices \( S_k \in \mathbb{R}^{R \times R} \), and a matrix \( V \in \mathbb{R}^{J \times R} \) such that

\[
\sum_{k=1}^{K} \left\| X_k - U_k S_k V^T \right\|_F
\]

is minimized.

a) PARAFAC2 is related to CP, but how? Under which conditions is PARAFAC2 the same as the CP decomposition?

b) Consider following kind of health records data: We have longitudinal health records data over \( K \) patients and \( J \) attributes, such as diagnoses and medication. For each patient, we have collected these attributes over different time span and at different times, and each patient \( k \) is represented by an \( I_k \)-by-\( J \) matrix \( X_k \), where \( I_k \) is the number of observations for this patient, and \( (X_k)_{ij} \) is the value of variable \( j \) that observation point \( i \). Notice that the observation points do not align between the users, that is, they correspond to different points in time. Assume we do rank-\( R \) PARAFAC2 to the collection of such matrices \( \{X_k\}_{k=1}^{K} \) and obtain \( \{U_k, S_k\}_{k=1}^{K} \), and \( V \).

We can assume that the columns of the \( J \)-by-\( R \) matrix \( V \) corresponds to some latent *phenotypes*, that is, they encode which diagnoses and medication “go together.” How would you interpret the other factors?