

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

$$\min\{IJ, IK, JK\}.$$

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

$$\mathbf{T}_{(1)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T.$$

Hint: Construct \mathbf{B} from identity matrices.

Solution. Let

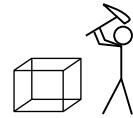
$$\begin{aligned} \mathbf{A} &= \mathbf{T}_{(1)} \\ \mathbf{B} &= \underbrace{[\mathbf{I}_J \mathbf{I}_J \cdots \mathbf{I}_J]}_{K \text{ times}} \\ \mathbf{C} &= \begin{pmatrix} \mathbf{j}_J^T & 0 & \cdots & 0 \\ 0 & \mathbf{j}_J^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{j}_J^T \end{pmatrix} (K \text{ rows}), \end{aligned}$$

where \mathbf{I}_J is J -by- J identity matrix and \mathbf{j}_J^T is J -dimensional row vector of all 1s. Now

$$\begin{aligned} \mathbf{C} \odot \mathbf{B} &= [\mathbf{c}_1 \otimes \mathbf{b}_1 \quad \mathbf{c}_2 \otimes \mathbf{b}_2 \quad \cdots \quad \mathbf{c}_{JK} \otimes \mathbf{b}_{JK}] \\ &= \begin{pmatrix} \mathbf{I}_J & 0 & \cdots & 0 \\ 0 & \mathbf{I}_J & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{I}_J \end{pmatrix} \\ &= \mathbf{I}_{JK} = (\mathbf{C} \odot \mathbf{B})^T. \end{aligned}$$

Hence, $\mathbf{A}(\mathbf{C} \odot \mathbf{B})^T = \mathbf{A}\mathbf{I}_{JK} = \mathbf{A} = \mathbf{T}_{(1)}$.

One can also show that this construct admits $\mathbf{T}_{(2)} = \mathbf{B}(\mathbf{C} \odot \mathbf{A})^T$ and $\mathbf{T}_{(3)} = \mathbf{C}(\mathbf{B} \odot \mathbf{A})^T$, though those proofs require much more complex subscripting.



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

$$\|\mathcal{T} - \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket\| .$$

Solution. The problem statement is wrong. The real problem should be: Given $\mathcal{T} \in \mathbb{R}_{\geq 0}^{I \times I \times K}$, find $\mathbf{A} \in \mathbb{R}_{\geq 0}^{I \times R}$ and $\mathbf{C} \in \mathbb{R}_{\geq 0}^{K \times R}$ that minimize $\|\mathcal{T} - \llbracket \mathbf{A}, \mathbf{A}, \mathbf{C} \rrbracket\|$.

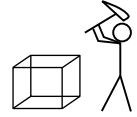
One option is to use the multiplicative update rules for NCP. If we let $\mathcal{Q} = \llbracket \mathbf{A}, \mathbf{A}, \mathbf{C} \rrbracket$, we have

$$a_{ir} \leftarrow a_{ir} \frac{\sum_{j,k} a_{jr} c_{kr} (t_{ijk}/q_{ijk})}{\sum_{j,k} a_{jr} c_{kr}}$$

$$c_{kr} \leftarrow c_{kr} \frac{\sum_{i,j} a_{ir} a_{jr} (t_{ijk}/q_{ijk})}{\sum_{i,j} a_{ir} a_{jr}} .$$

The initialization of the factor matrices requires some attention. They should naturally be nonnegative, and we should aim to have them in a correct scale. For multiplicative update rules we also cannot have zero entries. For example, we can sample from uniform distribution over $(0, u)$, where we set u so that the expected value of the CP product of the random matrices, $\mathbb{E}[\llbracket \mathbf{A}, \mathbf{A}, \mathbf{C} \rrbracket_{ijk}]$, is equal to the average value in the tensor, $\frac{1}{IJK} \sum_{i,j,k} t_{ijk}$. We can obtain this by setting

$$u = \frac{2}{R\sqrt[3]{IJK}} \sqrt[3]{\sum_{i,j,k} t_{ijk}} .$$



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

$$L(\mathbf{A}) = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T - \mathbf{T}_{(1)} * \log(\mathbf{A}(\mathbf{C} \odot \mathbf{B})^T).$$

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

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

$$\mathbf{W}_{ik} \leftarrow \mathbf{W}_{ik} \frac{\sum_{j=1}^m (\mathbf{A}_{ij} / (\mathbf{W}\mathbf{H})_{ij}) \mathbf{H}_{kj}}{\sum_{j=1}^m \mathbf{H}_{kj}}.$$

Adapt this update rule for the factor matrix \mathbf{A} in the CP decomposition. How does it relate to the update rule

$$\mathbf{A} \leftarrow \mathbf{A} * (\mathbf{T}_{(1)} \oslash (\mathbf{A}(\mathbf{C} \odot \mathbf{B})^T)) (\mathbf{C} \odot \mathbf{B})^T,$$

presented in the lecture? (To recap, \oslash is the element-wise division.)

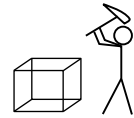
Solution. The NMF KL update rule adapted to matrix \mathbf{A} in NCP is

$$a_{ir} \leftarrow a_{ir} \frac{\sum_{j=1}^J \sum_{k=1}^K (t_{ijk} / (\mathbf{A}(\mathbf{C} \odot \mathbf{B})^T)_{ijk}) (\mathbf{C} \odot \mathbf{B})_{r,(jk)}}{\sum_{j=1}^J \sum_{k=1}^K (\mathbf{C} \odot \mathbf{B})_{r,(jk)}}$$

We can write this in a matrix format:

$$\mathbf{A} \leftarrow \mathbf{A} * \left((\mathbf{T}_{(1)} \oslash (\mathbf{A}(\mathbf{C} \odot \mathbf{B})^T)) (\mathbf{C} \odot \mathbf{B})^T \text{diag}((\mathbf{C} \odot \mathbf{B})^T \mathbf{1}_{JK})^{-1} \right).$$

Compared to the update rule for CP-APR, this has a normalization factor $\text{diag}((\mathbf{C} \odot \mathbf{B})^T \mathbf{1}_{JK})^{-1}$.



Problem 4 (PARAFAC2). The PARAFAC2 decomposition is another variant of the CP decomposition, defined slice-wise as follows. Given K matrices $\mathbf{X}_k \in \mathbb{R}^{I_k \times J}$ and rank R , find K matrices $\mathbf{U}_k \in \mathbb{R}^{I_k \times R}$, diagonal matrices $\mathbf{S}_k \in \mathbb{R}^{R \times R}$, and a matrix $\mathbf{V} \in \mathbb{R}^{J \times R}$ such that

$$\sum_{k=1}^K \left\| \mathbf{X}_k - \mathbf{U}_k \mathbf{S}_k \mathbf{V}^T \right\|_F$$

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 a I_k -by- J matrix \mathbf{X}_k , where I_k is the number of observations for this patient, and $(\mathbf{X}_k)_{ij}$ 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 $\{\mathbf{X}_k\}_{k=1}^K$ and obtain $\{\mathbf{U}_k, \mathbf{S}_k\}_{k=1}^K$, and \mathbf{V} .

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

Solution.

- a) For PARAFAC2 to be equal to CP, it has to be that (1) $I_1 = I_2 = \dots = I_K = I$, (2) the rows of \mathbf{X}_k correspond to each other, and (3) $\mathbf{U}_1 = \mathbf{U}_2 = \dots = \mathbf{U}_k = \mathbf{U}$. Then we can take the matrices \mathbf{X}_k as the frontal slices of tensor \mathcal{X} , set $\mathbf{A} = \mathbf{U}$, $\mathbf{B} = \mathbf{V}$, and arrange the values in the diagonal of \mathbf{S}_k as the k th row of \mathbf{C} . In this case $\mathbf{X}_k \approx \mathbf{U}_k \mathbf{S}_k \mathbf{V}^T$ for all k is equivalent to $\mathcal{X} \approx \llbracket \mathbf{A}, \mathbf{B}, \mathbf{C} \rrbracket$.
- b) The diagonal matrices \mathbf{S}_k indicate the *importance* or *strength* of each of the R phenotypes in the k th subject. The most relevant phenotype is the one with the highest value. Each column of \mathbf{U}_k provides a *temporal signature* for each of the phenotypes in patient k , that is, they indicate when the phenotype has been observed and at which level.