Homework 10
CS-526 Learning Theory

**Note:** The tensor product is denoted by $\otimes$. In other words, for vectors $\vec{a}, \vec{b}, \vec{c}$ we have that $\vec{a} \otimes \vec{b}$ is the square array $a^\alpha b^\beta$ where the superscript denotes the components, and $\vec{a} \otimes \vec{b} \otimes \vec{c}$ is the cubic array $a^\alpha b^\beta c^\gamma$. We denote components by superscripts because we need the lower index to label vectors themselves.

### Problem 1: Comparison of tensor rank and multilinear rank

Recall that the "tensor rank" (usually called "rank") is the smallest $R$ such that the multi-array $T^{\alpha\beta\gamma}$ can be decomposed as a sum of rank one terms in the form

$$T^{\alpha\beta\gamma} = \sum_{j=1}^{R} a_j^\alpha b_j^\beta c_j^\gamma \qquad \text{or equivalently} \qquad T = \sum_{j=1}^{R} \vec{a}_j \otimes \vec{b}_j \otimes \vec{c}_j \ .$$

This is often denoted $\mathrm{rank}_\otimes(T) = R$. On the other hand, the multilinear rank is the tuple $\mathrm{rank}_\boxplus(T) = (R_1, R_2, R_3)$ where $R_1$, $R_2$, $R_3$ are the ranks of the three matricizations $T_{(1)}$, $T_{(2)}$, $T_{(3)}$ defined in class.

1. Show that $\max \mathrm{rank}_\boxplus(T) \leq \mathrm{rank}_\otimes(T)$.

### Problem 2: Non-unicity of the Tucker decomposition

Let $T = (T^{\alpha\beta\gamma})$, $\alpha = 1, \dots I_1$, $\beta = 1, \dots I_2$, $\gamma = 1, \dots I_3$ an order-three tensor. Suppose that its multilinear rank is $\mathrm{rank}_\boxplus(T) = (R_1, R_2, R_3)$ which means that $R_1$, $R_2$, $R_3$ are the ranks of the three matricizations $T_{(1)}$, $T_{(2)}$, $T_{(3)}$ defined in class. We have seen in class that any such tensor has a so-called *Tucker decomposition* (also called higher order singular value decomposition):

$$T = \sum_{p,q,r=1}^{R_1,R_2,R_3} G^{pqr}\, \vec{u}_p \otimes \vec{v}_q \otimes \vec{w}_r$$

where each of the matrices $[\vec{u}_1, \cdots, \vec{u}_{R_1}]$, $[\vec{v}_1, \cdots, \vec{v}_{R_2}]$, $[\vec{w}_1, \cdots, \vec{w}_{R_3}]$ are made of orthogonal unit vectors. $G = (G^{pqr})$ is called the core tensor (and is not diagonal in general). In this problem you will prove that this decomposition is not unique and, in fact, that there exist an infinity of such decompositions related by orthogonal transformations.

Let $M^{(u)} = (M^{(u)}_{pp'})$, $M^{(v)} = (M^{(v)}_{qq'})$ and $M^{(w)} = (M^{(w)}_{rr'})$ be three orthogonal matrices of dimensions $R_1 \times R_1$, $R_2 \times R_2$ and $R_3 \times R_3$. Define the vectors:

$$\vec{x}_{p'} = \sum_{p=1}^{R_1} M^{(u)}_{p'p} \vec{u}_p, \quad \vec{y}_{q'} = \sum_{p=1}^{R_2} M^{(v)}_{q'q} \vec{v}_q, \quad \vec{z}_{r'} = \sum_{p=1}^{R_3} M^{(w)}_{r'r} \vec{w}_r \ .$$

Show that there exist a core tensor $H = (H^{pqr})$ of dimension $R_1 \times R_2 \times R_3$ such that

$$T = \sum_{p,q,r=1}^{R_1,R_2,R_3} H^{pqr}\, \vec{x}_p \otimes \vec{y}_q \otimes \vec{z}_r \ .$$

**Problem 3: Whitening of a tensor**

Consider the tensor

$$T = \sum_{i=1}^{K} \lambda_i\, \vec{\mu}_i \otimes \vec{\mu}_i \otimes \vec{\mu}_i$$

where $\vec{\mu}_i \in \mathbb{R}^D$ are linearly independent (so $K \leq D$) and $\lambda_i$ are strictly positive. Consider the matrix

$$M = \sum_{i=1}^{K} \lambda_i\, \vec{\mu}_i \otimes \vec{\mu}_i = \sum_{i=1}^{K} \lambda_i\, \vec{\mu}_i \vec{\mu}_i^T \ .$$

Note that this is a rank-$K$ symmetric positive semi-definite matrix (there are $D - K$ zero eigenvalues). Denote $d_1 \geq d_2 \geq \cdots \geq d_K$ the strictly positive eigenvalues of $M$ and $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_K$ the corresponding eigenvectors. Hence $M = U \mathrm{Diag}(d_1, \ldots, d_K) U^T$ where $U = \begin{bmatrix} \vec{u}_1 & \vec{u}_2 & \cdots & \vec{u}_K \end{bmatrix}$. Define the $D \times K$ matrix:

$$W = U \mathrm{Diag}(d_1^{-1/2}, d_2^{-1/2}, \cdots, d_K^{-1/2}) \ .$$

The whitening of $T$ is defined as the new tensor obtained by the multilinear transform

$$T(W,W,W) := \sum_{i=1}^{K} \lambda_i\, (W^T \vec{\mu}_i) \otimes (W^T \vec{\mu}_i) \otimes (W^T \vec{\mu}_i) = \sum_{i=1}^{K} \nu_i\, \vec{v}_i \otimes \vec{v}_i \otimes \vec{v}_i$$

where $\nu_i = \lambda_i^{-1/2}$ and $\vec{v}_i = \sqrt{\lambda_i}\, W^T \vec{\mu}_i$.

1. Show that $W^T M W = I$ where $I$ is the $K \times K$ identity matrix. Deduce that the $\vec{v}_i$'s are orthonormal, i.e., $V^T V = I$ where $V = \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_K \end{bmatrix}$.

2. Suppose we are given a tensor $T$ of the form $T = \sum_{i=1}^{K} \lambda_i\, \vec{\mu}_i \otimes \vec{\mu}_i \otimes \vec{\mu}_i$ and a matrix $M = \sum_{i=1}^{K} \lambda_i\, \vec{\mu}_i \vec{\mu}_i^T$ where $\vec{\mu}_i \in \mathbb{R}^D$ are linearly independent and $\lambda_i > 0$.
   Explain how applying the tensor power method to the whitened tensor $T(W,W,W)$ helps you recover the $\lambda_i$'s and $\mu_i$'s, and give a closed-form formula for the matrix $\mu = \begin{bmatrix} \vec{\mu}_1 & \cdots & \vec{\mu}_K \end{bmatrix}$ that uses $V$, $\mathrm{Diag}(\nu_1, \ldots, \nu_K)$ and $W$.