Solutions to Homework 11: 1 June 2021
CS-526 Learning Theory

Problem 1: Whitening of a tensor

1. We have $M = U\text{Diag}(d_1, \ldots, d_K)U^T$ and, by definition, $W := U\text{Diag}(d_1^{-1/2}, \ldots, d_K^{-1/2})$. A direct computation gives:

$$\begin{aligned} W^T M W &= \text{Diag}(d_1^{-1/2}, \ldots, d_K^{-1/2})(U^T U)\text{Diag}(d_1, \ldots, d_K)(U^T U)\text{Diag}(d_1^{-1/2}, \ldots, d_K^{-1/2}) \\ &= \text{Diag}(d_1^{-1/2}, \ldots, d_K^{-1/2})\text{Diag}(d_1, \ldots, d_K)\text{Diag}(d_1^{-1/2}, \ldots, d_K^{-1/2}) \\ &= I \; . \end{aligned}$$

We used that the columns of U are orthogonal unit vectors: $U^T U = I$. By definition of $\vec{v}_i$, we have $V := \begin{bmatrix} \vec{v}_1 & \cdots & \vec{v}_K \end{bmatrix} = W^T \mu \, \text{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_K})$ where $\mu := \begin{bmatrix} \vec{\mu}_1 & \cdots & \vec{\mu}_K \end{bmatrix}$. It also follows from the definition of M that $M = \mu \, \text{Diag}(\lambda_1, \ldots, \lambda_K)\mu^T$. Hence:

$$\begin{aligned} V V^T &= W^T \mu \, \text{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_K})\text{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_K})\mu^T W \\ &= W^T M W \\ &= I \; . \end{aligned}$$

The matrix V is square and satisfies $VV^T = I$, thefore $V^T V = I$ meaning that the vector $\vec{v}_i$ are orthonormal.

2. Because M is known we can compute the matrix W and use it to obtained the whitened tensor $T(W, W, W) = \sum_{i=1}^K \nu_i \, \vec{v}_i \otimes \vec{v}_i \otimes \vec{v}_i$ where $\nu_i = \lambda_i^{-1/2}$ and $\vec{v}_i = \sqrt{\lambda_i} \, W^T \vec{\mu}_i$. We have shown in the previous question that $\vec{v}_1, \ldots, \vec{v}_K$ are orthogonal unit vectors. Thus, we can use the tensor power method to recover each of the pair $\pm(\nu_i, \vec{v}_i)$ for $i \in [K]$. Because $\nu_i > 0$ we can disambiguate the sign and determine $(\nu_i, \vec{v}_i)$ from $\pm(\nu_i, \vec{v}_i)$.
Now that all the $(\nu_i, \vec{v}_i)$ are known, we need to show that the whitening transformation can be inverted to give back $(\lambda_i, \vec{\mu}_i)$. The relation between λ_i and ν_i is easy to invert: $\lambda_i = 1/\nu_i^2$. To recover $\mu = \begin{bmatrix} \vec{\mu}_1 & \cdots & \vec{\mu}_K \end{bmatrix}$, we need to invert the system of equations

$$V = W^T \mu \, \text{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_K}) \Leftrightarrow V\text{Diag}(\nu_1, \ldots, \nu_K) = W^T \mu \; . \tag{1}$$

The matrix $W^T = \text{Diag}(d_1^{-1/2}, \ldots, d_K^{-1/2})U^T$ has full row rank and its Moore-Penrose pseudo-inverse reads $(W^T)^\dagger = U\text{Diag}(\sqrt{d_1}, \ldots, \sqrt{d_K})$. Multiplying both sides of (1) by $(W^T)^\dagger$ yields:

$$(W^T)^\dagger V\text{Diag}(\nu_1, \ldots, \nu_K) = UU^T \mu \; . \tag{2}$$

At this point we might be tempted to say that $UU^T = I$, yielding $\mu = (W^T)^\dagger V\text{Diag}(\nu_1, \ldots, \nu_K)$. However, U is in general not a square matrix and we cannot conclude $UU^T = I$ from $U^T U = I$. This is only a minor setback. Note that (the left-hand side is the definition of M, the right-hand side is its diagonalization):

$$\mu\text{Diag}(\lambda_1, \ldots, \lambda_K)\mu^T = U\text{Diag}(d_1, \ldots, d_K)U^T \; ,$$

where μ, U are $D \times K$ full column rank matrices. It follows that $\text{span}(\mu) = \text{span}(U)$ and there exists a $K \times K$ matrix P such that $\mu = UP$. Hence, $UU^T\mu = U(U^TU)P = UP = \mu$ and (2) reads:

$$\mu = (W^T)^\dagger V \text{Diag}(\nu_1, \dots, \nu_K) = U \text{Diag}(\sqrt{d_1}, \dots, \sqrt{d_K}) V \text{Diag}(\nu_1, \dots, \nu_K) \ .$$

We are thus able to recover μ from the knowledge of W, V and $\text{Diag}(\nu_1, \dots, \nu_K)$.