Solutions to Graded Homework 3
CS-526 Learning Theory

Problem 1: a small application of Jennrich's theorem

1. (a) $M = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}$.

   (b) Using that $RR^T = I_2$:

$$M = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \otimes \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} R \cdot R^T \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \vec{a} \otimes \vec{b} + \vec{c} \otimes \vec{d}$$

   where

$$\begin{bmatrix} \vec{a} & \vec{c} \end{bmatrix} = \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} R = \begin{pmatrix} \cos\theta + \sin\theta & \cos\theta - \sin\theta \\ 2\cos\theta & -2\sin\theta \end{pmatrix} ;$$
$$\begin{bmatrix} \vec{b} & \vec{d} \end{bmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \cos\theta + \sin\theta & \cos\theta - \sin\theta \end{pmatrix} .$$

2. We have $T = \vec{a}_1 \otimes \vec{b}_1 \otimes \vec{c}_1 + \vec{a}_2 \otimes \vec{b}_2 \otimes \vec{c}_2 + \vec{a}_3 \otimes \vec{b}_3 \otimes \vec{c}_3$ where

$$\begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \vec{a}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 3 \\ 1 & 2 & 5 \end{bmatrix} \text{ has pairwise independent columns;}$$
$$\begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \vec{b}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ has linearly independent columns;}$$
$$\text{and } \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \vec{c}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix} \text{ has linearly independent columns.}$$

   By Jennrich's theorem the decomposition is therefore unique and the rank of T is 3.

3. (a) We have $T = \vec{a}_1 \otimes \vec{b}_1 \otimes \vec{c}_1 + \vec{a}_2 \otimes \vec{b}_2 \otimes \vec{c}_2$ where $\vec{c}_1 = \vec{c}_2 = \vec{c}$. We cannot invoke Jennrich's theorem because the vectors $\vec{c}_1, \vec{c}_2$ are not pairwise independent.

   (b) The tensor rank is obviously less than or equal to 2. We will prove by contradiction that it cannot be equal to 1.
   Assume the rank is one. Then there exist vectors $\vec{e}, \vec{f}, \vec{g}$ such that $T = \vec{e} \otimes \vec{f} \otimes \vec{g}$. Pick any vector $\vec{x}$ that is not orthogonal to $\vec{c}$. We have:

$$(\vec{e} \otimes \vec{f})(\vec{g}^T \vec{x}) = (\vec{a}_1 \otimes \vec{b}_1 + \vec{a}_2 \otimes \vec{b}_2)(\vec{c}^T \vec{x})$$

   The matrix $(\vec{e} \otimes \vec{f})(\vec{g}^T \vec{x})$ has rank 0 or 1 while the matrix $(\vec{a}_1 \otimes \vec{b}_1 + \vec{a}_2 \otimes \vec{b}_2)(\vec{c}^T \vec{x})$ has rank 2 because $\vec{a}_1 \otimes \vec{b}_1 + \vec{a}_2 \otimes \vec{b}_2$ has rank 2 and $\vec{c}^T \vec{x} \neq 0$. This is a contradiction.

Problem 2: Kronecker and Khatri-Rao products

1) To show that $A \odot_{\mathrm{KhR}} B$ is full column rank, we have to prove that the kernel of the linear application $\underline{x} \mapsto (A \odot_{\mathrm{KhR}} B)\underline{x}$ is $\{0\}$. Let $\underline{x} \in \mathbb{R}^R$ with components $(x^1, x^2, \cdots, x^R)$ be such that $(A \odot_{KhR} B)\underline{x} = 0$. Then, $\forall \alpha \in [I_1]$:

$$\sum_{r=1}^{R} a_r^\alpha x^r \underline{b}_r = 0\,.$$

Because B is full column rank, $\sum_{r=1}^{R} a_r^\alpha x^r \underline{b}_r = 0 \Rightarrow \forall r \in [R] : a_r^\alpha x^r = 0$. Hence, $\forall r \in [R] : x_r \underline{a}_r = 0$. A is full column rank so none of its columns can be the all-zero vector. It follows that x_r must be zero for all $r \in [R]$, i.e., $\underline{x} = 0$. $A \odot_{KhR} B$ is full column rank.

2) Suppose we are given a tensor (the weights λ_r that usually appear in the sum are absorbed in the vectors $\underline{a}_r$)

$$\mathcal{X} = \sum_{r=1}^{R} \underline{a}_r \otimes \underline{b}_r \otimes \underline{c}_r\ , \tag{1}$$

where $A = [\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_R] \in \mathbb{R}^{I_1 \times R}$, $B = [\underline{b}_1, \underline{b}_2, \ldots, \underline{b}_R] \in \mathbb{R}^{I_2 \times R}$ and $C = [\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_R] \in \mathbb{R}^{I_3 \times R}$ are full column rank. By Jennrich's algorithm, the decomposition (**??**) is unique up to trivial rank permutation and feature scaling and Jennrich's algorithm is a way to recover this decomposition. At the end of the step (5) of the algorithm, we have computed A, B and it remains to recover C. We now show how the result in question **1)** allows to recover C uniquely. For each $\gamma \in [I_3]$, define the slice $\mathcal{X}_\gamma$ as the $I_1 \times I_2$ matrix with entries $(\mathcal{X}_\gamma)^{\alpha\beta} = \mathcal{X}^{\alpha\beta\gamma}$ and denote $F(\mathcal{X}_\gamma)$ the $I_1 I_2$ column vector with entries $F(\mathcal{X}_\gamma)^{\beta + I_2(\alpha-1)} = \mathcal{X}^{\alpha\beta\gamma}$. We have:

$$\forall (\alpha, \beta) \in [I_1] \times [I_2] : F(\mathcal{X}_\gamma)^{\beta + I_2(\alpha-1)} = \sum_{r=1}^{R} a_r^\alpha b_r^\beta c_r^\gamma = \sum_{r=1}^{R} (A \odot_{\mathrm{KhR}} B)^{\beta + I_2(\alpha-1), r} c_r^\gamma\ .$$

Therefore, the $I_1 I_2 \times I_3$ matrix $F(\mathcal{X}) = [F(\mathcal{X}_1), F(\mathcal{X}_2), \ldots, F(\mathcal{X}_{I_3})]$ satisfies:

$$F(\mathcal{X}) = (A \odot_{\mathrm{KhR}} B) C^T\ .$$

Because $A \odot_{\mathrm{KhR}} B$ is full column rank, we can invert the system with the Moore-Penrose pseudoinverse: $C^T = (A \odot_{\mathrm{KhR}} B)^\dagger F(\mathcal{X})$.

Problem 3: Jennrich's type algorithm for order 4 tensors

1) To apply Jennrich's algorithm we need to prove that the matrix $E = [\underline{c}_1 \otimes_{\mathrm{Kro}} \underline{d}_1, \ldots, \underline{c}_R \otimes_{\mathrm{Kro}} \underline{d}_R]$ is full column rank (A, B are full column rank by assumption). Note that the same proof as the one in Problem 4 question 1 applies. Nevertheless we repeat the argument here.
Let $\underline{v} \in \mathbb{R}^R$ a column vector in the kernel of E, i.e., $E\underline{v} = 0$. Then:

$$\forall \gamma \in [I_3] : \sum_{r=1}^{R} (c_r^\gamma v^r) \underline{d}_r = 0 \ \Rightarrow\ \forall \gamma \in [I_3], \forall r \in [R] : c_r^\gamma v^r = 0 \ \Rightarrow\ C\underline{v} = 0 \ \Rightarrow\ \underline{v} = 0\ .$$

The first implication follows from D being full column rank and the third one from C being full column rank. We conclude that the kernel of E is $\{0\}$: E is full column rank.
We can therefore apply Jennrich's algorithm.

2) We recover the rank R as well as A, B and E by applying Jennrich's algorithm to $\widetilde{T}$. From E we can then determine C and D. Fix $r \in [R]$. Since C is full column rank, there exists $\alpha_r \in [I_3]$

such that $c_r^{\alpha_r} \neq 0$. As $c_r^{\alpha_r} \neq 0$, the I_4-dimensional column vector $\underline{\tilde{d}}_r = c_r^{\alpha} \underline{d}_r$ contained in the r^{th} column of E recovers $\underline{d}_r$ up to some feature scaling. Doing this for every $r \in [R]$ we build the matrix $\widetilde{D} = \begin{bmatrix} \underline{\tilde{d}}_1 & \underline{\tilde{d}}_2 & \dots & \underline{\tilde{d}}_R \end{bmatrix}$ that recovers D up to some feature scaling and is full column rank (because D is). Finally, for every $r \in R$, pick $\beta_r \in [I_4]$ such that $\tilde{d}_r^{\beta_r} \neq 0$ (such β_r exists because $\widetilde{D}$ is full column rank) and use the entries of E corresponding to $c_r^{\alpha} d_r^{\beta_r}$, $\alpha \in [I_3]$, to build the vector $\underline{\widetilde{c}}_r = \frac{d_r^{\beta_r}}{\tilde{d}_r^{\beta_r}} \underline{c}_r$. The matrix $\widetilde{C} = \begin{bmatrix} \underline{\tilde{c}}_1 & \underline{\tilde{c}}_2 & \dots & \underline{\tilde{c}}_R \end{bmatrix}$ recovers C up to some feature scaling.