Learning Theory - Homework 4

Alexandru Mocanu - SCIPER 295172

May 2019

1 Exercise 1

1) For the first moment, we have:

$$\mathbb{E}[\underline{x}] = \int\limits_{\underline{x}} \underline{x} p(\underline{x}) = \sum_{i=1}^{K} w_i (2\pi\sigma^2)^{-D/2} \int\limits_{\underline{x}} \exp\left\{-\frac{\|\underline{x} - \underline{a}_i\|^2}{2\sigma^2}\right\} = \sum_{i=1}^{K} w_i \underline{a}_i \tag{1}$$

as we know from the expected value of a Gaussian distributed random variable.
For the second moment, let $f_i(\underline{x}) = (2\pi\sigma^2)^{-D/2} \int\limits_{\underline{x}} \exp\left\{-\frac{\|\underline{x}-\underline{a}_i\|^2}{2\sigma^2}\right\}$. Then, we have:

$$\mathbb{E}[\underline{x}\underline{x}^T] = \sum_{i=1}^{K} w_i (2\pi\sigma^2)^{-D/2} \int\limits_{\underline{x}} \exp\left\{-\frac{\|\underline{x} - \underline{a}_i\|^2}{2\sigma^2}\right\} = \sum_{i=1}^{K} w_i (\sigma^2 I_D + \mathbb{E}_{f_i}[\underline{x}]\underline{a}_i^T + \underline{a}_i \mathbb{E}_{f_i}[\underline{x}^T] - \underline{a}_i\underline{a}_i^T) =$$

$$= \sum_{i=1}^{K} w_i (\sigma^2 I_D + \underline{a}_i \underline{a}_i^T) = \sigma^2 I_D + \sum_{i=1}^{K} w_i \underline{a}_i \underline{a}_i^T \tag{2}$$

For the third moment, we have to take each element $T_{klm} = x_k x_l x_m$ of the tensor and compute its expected value. For $k < l < m$, we have:

$$\mathbb{E}_i[T_{klm}] = (w\pi\sigma^2)^{-D/2} \int\limits_{x_1} \exp\left\{-\frac{\|x_1 - a_{i,1}\|^2}{2\sigma^2}\right\} dx_1 ... \int\limits_{x_k} x_k \exp\left\{-\frac{\|x_k - a_{i,k}\|^2}{2\sigma^2}\right\} dx_k ...$$

$$... \int\limits_{x_l} x_l \exp\left\{-\frac{\|x_l - a_{i,l}\|^2}{2\sigma^2}\right\} dx_l ... \int\limits_{x_m} x_m \exp\left\{-\frac{\|x_m - a_{i,m}\|^2}{2\sigma^2}\right\} dx_m ... \int\limits_{x_D} \exp\left\{-\frac{\|x_K - a_{i,K}\|^2}{2\sigma^2}\right\} dx_K =$$

$$= a_{i,k} a_{i,l} a_{i,m} \tag{3}$$

If two of the indices k, l, m are equal, then:

$$\mathbb{E}_i[T_{kkl}] = \mathbb{E}_i[T_{klk}] = \mathbb{E}_i[T_{lkk}] = (2\pi\sigma^2)^{-1} \int\limits_{x_k} x_k^2 \exp\left\{-\frac{\|x_k - a_{i,k}\|^2}{2\sigma^2}\right\} dx_k$$

$$\int\limits_{x_l} x_l^2 \exp\left\{-\frac{\|x_l - a_{i,l}\|^2}{2\sigma^2}\right\} dx_l = (\sigma^2 + a_{i,k}^2)a_{i,l} \quad (4)$$

If all the indices are equal, then:

$$\mathbb{E}_i[T_{kkk}] = (2\pi\sigma^2)^{-1/2} \int\limits_{-\infty}^{\infty} x_k^3 \exp\left\{-\frac{\|x_k - a_{i,k}\|^2}{2\sigma^2}\right\} dx_k = (2\pi\sigma^2) \int\limits_{-\infty}^{\infty} (t+a_{i,k})^3 \exp\left\{-\frac{t^2}{2\sigma^2}\right\} dt =$$

$$= (2\pi\sigma^2)^{-1/2} \int\limits_{-\infty}^{\infty} t^3 \exp\left\{-\frac{t^2}{2\sigma^2}\right\} dt + 3a_{i,k}(2\pi\sigma^2)^{-1/2} \int\limits_{-\infty}^{\infty} t^2 \exp\left\{-\frac{t^2}{2\sigma^2}\right\} dt +$$

$$+ 3a_{i,k}^2(2\pi\sigma^2)^{-1/2} \int\limits_{-\infty}^{\infty} t \exp\left\{-\frac{t^2}{2\sigma^2}\right\} dt + (2\pi\sigma^2)^{-1/2} \int\limits_{-\infty}^{\infty} \exp\left\{-\frac{t^2}{2\sigma^2}\right\} dt =$$

$$= 3a_{i,k}\sigma^2 + a_{i,k}^3 \quad (5)$$

Combining all these results, we get that:

$$\mathbb{E}[\underline{x}\otimes\underline{x}\otimes\underline{x}] = \sum_{i=1}^{K} w_i \underline{a}_i\otimes\underline{a}_i\otimes\underline{a}_i + \sigma^2 \sum_{j=1}^{D}\sum_{i=1}^{K} w_i(\underline{a}_i\otimes\underline{e}_j\otimes\underline{e}_j + \underline{e}_j\otimes\underline{a}_i\otimes\underline{e}_j + \underline{e}_j\otimes\underline{e}_j\otimes\underline{a}_i) \quad (6)$$

2) Using the formula for the second moment from the previous task, we have that:

$$\mathbb{E}'[\underline{x}\underline{x}^T] = \sigma^2 I_D + \sum_{i=1}^{K} w_i \underline{a}_i'(\underline{a}_i')^T = \sigma^2 I_D + \sum_{i=1}^{K} w_i \sum_{j=1}^{K}\sum_{k=1}^{K} \tilde{R}_{ij}\underline{a}_j \tilde{R}_{ik}(\underline{a}_k)^T =$$

$$= \sigma^2 I_D + \sum_{j=1}^{K}\sum_{k=1}^{K} \sqrt{w_j w_k}\underline{a}_j\underline{a}_k^T \sum_{i=1}^{K} R_{ij}R_{ik} \quad (7)$$

The matrix is orthogonal and as it only performs a rotation, it is orthonormal, so $\sum\limits_{i=1}^{K} R_{ij}R_{ik}$ is 1 for $j = k$ and 0 for $j \neq k$. Therefore, $\sum\limits_{j=1}^{K}\sum\limits_{k=1}^{K} \sqrt{w_j w_k}\underline{a}_j\underline{a}_k^T \sum\limits_{i=1}^{K} R_{ij}R_{ik} = \sum\limits_{j=1}^{K} w_j\underline{a}_j\underline{a}_j^T$. This leads to:

$$\mathbb{E}'[\underline{x}\underline{x}^T] = \sigma^2 I_D + \sum_{j=1}^{K} w_j\underline{a}_j\underline{a}_j^T \quad (8)$$

Therefore, we obtain the same second moment as in the previous task.

2 Exercise 2

1) The two-dimensional multiarrays are:

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, P = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, E = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \tag{9}$$

The three-dimensional multiarrays have the following frontal slices:

$$\begin{aligned} G_1 &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, G_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \\ W_1 &= \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, W_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \tag{10}$$

The matrices for the three modes of G and W are:

$$\begin{aligned} G_{(1)} &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, G_{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, G_{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ W_{(1)} &= \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, W_{(2)} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}, W_{(3)} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{aligned} \tag{11}$$

2) For the two-dimensional tensors, we can simply see that the determinants are $det(B) = 1$, $det(P) = 0$, $det(E) = 1$, so the ranks are $\mathbf{rank(B) = 2}$, $\mathbf{rank(P) = 1}$, $\mathbf{rank(E) = 2}$.
For G, we consider matrices $A_G = [e_0, e_1]$, $B_G = [e_0, e_1]$, $C_G = [e_0, e_1]$ and write $G = \sum_{r=1}^{2} a_{G,r} \otimes b_{G,r} \otimes c_{G,r}$. By Jennrich's algorithm, this is the unique factorization up to a scaling factor, so $\mathbf{rank(G) = 2}$.
For W, assume that we can write $W = a \otimes b \otimes c + d \otimes e \otimes f$. This is equivalent to:

$$\begin{cases} a_0b_0c_0 + d_0e_0f_0 = 0 \\ a_0b_0c_1 + d_0e_0f_1 = 1 \\ a_0b_1c_0 + d_0e_1f_0 = 1 \\ a_0b_1c_1 + d_0e_1f_1 = 0 \end{cases} \quad \begin{cases} a_1b_0c_0 + d_1e_0f_0 = 1 \\ a_1b_0c_1 + d_1e_0f_1 = 0 \\ a_1b_1c_0 + d_1e_1f_0 = 0 \\ a_1b_1c_1 + d_1e_1f_1 = 0 \end{cases} \tag{12}$$

First, assume that $a_0b_0c_0 = 0$. Then $d_0e_0f_0 = 0$. If $a_0 = 0$, then $d_0e_0f_1 = 1$ and as $d_0e_0f_0 = 0$, it follows that $f_0 = 0$. Also, we have that $d_0e_1f_0 = 1$, so $f_0 \neq 0$, which is a contradiction. If $b_0 = 0$, then $d_0e_0f_1 = 1$ and as $d_0e_0f_0 = 0$, it follows that $f_0 = 0$. Also, we have that $d_1e_0f_0 = 1$, so $f_0 \neq 0$, which gives a contradiction. If $c_0 = 0$, then $d_0e_1f_0 = 0$, and as $d_0e_0f_0 = 0$, it follows that $e_0 = 0$. Also, we have that $d_1e_0f_0 = 1$, so $e_0 \neq 0$, which gives a contradiction. Therefore, a_0, b_0, c_0, d_0, e_0, f_0 are all nonzero.
Now, assume that $a_1b_1c_1 = 0$. Then $d_1e_1f_1 = 0$. If $a_1 = 0$, then $d_1e_0f_0 = 1$, so $d_1 \neq 0$. As $d_1e_0f_1 = 0$, $d_1e_1f_0 = 0$ and $d_1 \neq 0$, it follows that $f_1 = 0$ and $e_1 = 0$. From $f_1 = 0$, we get $a_0b_1c_1 = 0$ and $a_0b_0c_1 = 1$ and from these $b_1 = 0$. As $b_1 = 0$, it follows that $d_0e_1f_0 = 1$, so $e_1 \neq 0$, which is a contradiction. If $b_1 = 0$, then $d_0e_1f_0 = 0$, so $e_1 \neq 0$. Also $d_1e_1f_0 = 0$ and as $e_1 \neq 0$, it follows

that $d_1 = 0$. From this $a_1 b_1 c_0 = 1$, which implies $a_1 \neq 0$, and $a_1 b_0 c_1 = 0$, so $c_1 = 0$. From this, $d_0 e_0 f_1 = 1$ and $d_0 e_1 f_1 = 0$, but the first one implies $f_1 \neq 0$ and the second one $f_1 = 0$, so we get a contradiction. If $c_1 = 0$, then $d_0 e_0 f_1 = 1$, so $f_1 \neq 0$. Also $d_0 e_1 f_1 = 0$ and also using $f_1 \neq 0$, it follows that $e_1 = 0$. From this, $a_0 b_1 c_0 = 1$, which implies $b_1 \neq 0$, and $a_1 b_1 c_0 = 0$, so $a_1 = 0$. From $a_1 = 0$, we get $d_1 e_0 f_0 = 1$ and $d_1 e_0 f_1 = 0$, but the first one implies that $d_1 \neq 0$ and the second one $d_1 = 0$, so we get a contradiction. Therefore, a_1, b_1, c_1, d_1, e_1, f_1 are all nonzero.
Now, in these conditions, we have $a_0 b_1 c_1 = -d_0 e_1 f_1$ and $a_1 b_1 c_1 = -d_1 e_1 f_1$, so $\frac{a_0}{a_1} = \frac{d_0}{d_1}$. Similarly, $\frac{b_0}{b_1} = \frac{e_0}{e_1}$ and $\frac{c_0}{c_1} = \frac{f_0}{f_1}$. This gives us $1 = a_0 b_0 c_0 \frac{f_1}{f_0} + d_0 e_0 f_0 \frac{c_1}{c_0} = a_0 b_0 c_0 \left(\frac{f_1}{f_0} - \frac{c_1}{c_0}\right)$, but given that $\frac{f_1}{f_0} = \frac{c_1}{c_0}$, this is impossible.
In conclusion, we can not write $W = a \otimes b \otimes c + d \otimes e \otimes f$, so $\mathbf{rank(W) = 3}$.
3) Let $O = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be the orthonormal matrix i.e. $\begin{bmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{bmatrix} = I$. Then,

$$(Oe_0) \otimes (Oe_0) + (Oe_1) \otimes (Oe_1) = \begin{bmatrix} a \\ c \end{bmatrix} \otimes \begin{bmatrix} a \\ c \end{bmatrix} + \begin{bmatrix} b \\ d \end{bmatrix} \otimes \begin{bmatrix} b \\ d \end{bmatrix} = \begin{bmatrix} a^2 + b^2 & ac + bd \\ ac + bd & c^2 + d^2 \end{bmatrix} = I \tag{13}$$

Therefore, $B = (Oe_0) \otimes (Oe_0) + (Oe_1) \otimes (Oe_1)$.
For G, we have the orthogonal matrices A_G, B_G and C_G, as above, used to form it by tensorial product. Multiplying these by an orthonormal matrix preserves their orthogonality, so by Jennrich's theorem, we have again get G by using the vectors multiplied by an orthonormal matrix in the tensorial products. For W, the matrices used are not orthogonal to begin with, so we can not use Jennrich's theorem in this case.
4) We have $e_0 + \epsilon e_1 = \begin{bmatrix} 1 \\ \epsilon \end{bmatrix}$. By simple computation, we get:

$$\begin{cases} \lim\limits_{\epsilon \to 0} (D_\epsilon)_{000} = \lim\limits_{\epsilon \to 0} \left(\frac{1}{\epsilon} - \frac{1}{\epsilon}\right) = 0 \\ \lim\limits_{\epsilon \to 0} (D_\epsilon)_{001} = 1 \\ \lim\limits_{\epsilon \to 0} (D_\epsilon)_{010} = 1 \\ \lim\limits_{\epsilon \to 0} (D_\epsilon)_{011} = \lim\limits_{\epsilon \to 0} \epsilon = 0 \end{cases} \quad \begin{cases} \lim\limits_{\epsilon \to 0} (D_\epsilon)_{100} = 1 \\ \lim\limits_{\epsilon \to 0} (D_\epsilon)_{101} = \lim\limits_{\epsilon \to 0} \epsilon = 0 \\ \lim\limits_{\epsilon \to 0} (D_\epsilon)_{110} = \lim\limits_{\epsilon \to 0} \epsilon = 0 \\ \lim\limits_{\epsilon \to 0} (D_\epsilon)_{111} = \lim\limits_{\epsilon \to 0} \epsilon^2 = 0 \end{cases} \tag{14}$$

Therefore, $\lim\limits_{\epsilon \to 0} D_\epsilon = W$.

3 Exercise 3

1) Considering the phenomenon in exercise 2, question 4, it could be that given some order-p tensor T of rank R, the minimum of $\|T - S\|$, for S an order-p tensor of rank $k < R$, would be achieved for a tensor $\hat{T}$ that would actually have rank $> k$. Therefore, the analogous problem of the Eckart-Young theorem for order-p, $p \geq 3$, tensors is not well-posed.
2) Using Eckart-Young in our case, we have that for a matrix A of rank $(R+1)$

that has the SVD $U\Sigma V^*$, with $\Sigma_{ii} = \sigma_i$ and $\sigma_i \geq \sigma_j, \forall i < j$, there exists the rank R matrix $\hat{A} = U\tilde{\Sigma}V^*$, with $\tilde{\Sigma}_{ii} = \sigma_i, \forall i \leq R$ and $\tilde{\Sigma}_{(R+1)(R+1)} = 0$, such that $\|A-\hat{A}\|_F = \min\limits_{S:rank(S)\leq k} \|A-S\|_F$. Consider now a sequence D_n of sums of R rank-one matrices. The sum of R rank-one matrices has rank at most R. As $\|A-\hat{A}\|_F = \min\limits_{S:rank(S)\leq k} \|A-S\|_F$, it means that $\|A-\hat{A}\|_F \leq \|A-D_n\|_F, \forall n \geq 0$. Therefore, $\|A-\hat{A}\|_F \leq \lim\limits_{n\to\infty} \|A-D_n\|_F$ and as $\|A-\hat{A}\|_F > 0$, it means that $\lim\limits_{n\to\infty} D_n \neq A$, so we can not write a rank $R+1$ matrix as the limit of the sum of R rank-one matrices.

To obtain a rank $R-1$ matrix from a sequence of rank R matrices, we can simply set the smallest eigenvalue of a matrix D_n in the sequence to $\frac{1}{n}$, the other eigenvalues being kept constant. Then, when $n \to \infty$, that eigenvalue becomes 0, so we get a rank $R-1$ matrix.

We can make a similar construct for order-p tensors. Say we have a sequence of rank R tensors with elements decomposed as $D_n = \sum\limits_{i=1}^{R-1} \underline{a}_{i,1} \otimes \cdots \otimes \underline{a}_{i,p} + \frac{1}{n}\underline{a}_{R,1} \otimes \cdots \otimes \underline{a}_{R,p}$. Then, as $n \to \infty$, the last term becomes 0, so $\lim\limits_{n\to\infty} D_n$ is rank $R-1$.

3) In what follows, we limit ourselves to the case of real values. Using a Tucker decomposition, we have that $T = \sum\limits_{p=1}^{r_1}\sum\limits_{q=1}^{r_2}\sum\limits_{r=1}^{r_3} G_{pqr}\underline{u}_p \otimes \underline{v}_q \otimes \underline{w}_r$. Therefore, $T^{\alpha\beta\gamma} = \sum\limits_{p=1}^{r_1}\sum\limits_{q=1}^{r_2}\sum\limits_{r=1}^{r_3} G_{pqr}u_{p,\alpha}v_{q,\beta}w_{r,\gamma}$. The Frobenius norm of the T is:

$$\begin{gathered}\|T\|_F^2 = \sum_{\alpha,\beta,\gamma} |\sum_{p=1}^{r_1}\sum_{q=1}^{r_2}\sum_{r=1}^{r_3} G_{pqr}u_{p,\alpha}v_{q,\beta}w_{r,\gamma}|^2 = \\ = \sum_{p=1}^{r_1}\sum_{q=1}^{r_2}\sum_{r=1}^{r_3}\sum_{k=1}^{r_1}\sum_{l=1}^{r_2}\sum_{m=1}^{r_3} G_{pqr}G_{klm}\sum_{\alpha} u_{p,\alpha}u_{k,\alpha}\sum_{\beta} v_{q,\beta}v_{l,\beta}\sum_{\gamma} w_{r,\gamma}w_{m,\gamma}\end{gathered} \tag{15}$$

Due to the matrices U, V, W being orthogonal, we have $\sum\limits_{\alpha} u_{p,\alpha}u_{k,\alpha} = \delta_{pk}\sum\limits_{\alpha} u_{p,\alpha}^2$ and the same for V and W. Therefore

$$\|T\|_F^2 = \sum_{p=1}^{r_1}\sum_{q=1}^{r_2}\sum_{r=1}^{r_3} G_{pqr}^2\sum_{\alpha} u_{p,\alpha}^2\sum_{\beta} v_{q,\beta}^2\sum_{\gamma} u_{r,\gamma}^2 = \sum_{p=1}^{r_1}\sum_{q=1}^{r_2}\sum_{r=1}^{r_3} G_{pqr}^2\|u_p\|_F^2\|v_q\|_F^2\|w_r\|_F^2 \tag{16}$$

For $T(R_1, R_2, R_3)$, we have that:

$$\begin{gathered}T(R_1, R_2, R_3)^{\alpha\beta\gamma} = \sum_{\eta,\theta,\iota} R_1^{\alpha\eta}R_2^{\beta\theta}R_3^{\gamma\iota}\sum_{p=1}^{r_1}\sum_{q=1}^{r_2}\sum_{r=1}^{r_3} G_{pqr}u_{p,\eta}v_{q,\theta}w_{r,\iota} = \\ = \sum_{p=1}^{r_1}\sum_{q=1}^{r_2}\sum_{r=1}^{r_3}(\sum_{\eta} R_1^{\alpha\eta}u_{p,\eta})(\sum_{\theta} R_2^{\gamma\theta}v_{q,\theta})(\sum_{\iota} R_3^{\gamma\iota}w_{r,\iota})G_{pqr}\end{gathered} \tag{17}$$

We now consider the Frobenius norm of $T(R_1, R_2, R_3)$:

$$\|T(R_1,R_2,R_3)\|_F^2 = \sum_{\alpha,\beta,\gamma} |T(R_1,R_2,R_3)^{\alpha\beta\gamma}|^2 = \sum_{\alpha,\beta,\gamma} |\sum_{\delta,\epsilon,\zeta} R_1^{\alpha\delta} R_2^{\beta\epsilon} R_3^{\gamma\zeta} T^{\delta\epsilon\zeta}|^2 =$$
$$= \sum_{\alpha,\beta,\gamma}\sum_{\delta,\epsilon,\zeta}\sum_{\eta,\theta,\iota} R_1^{\alpha\delta} R_2^{\beta\epsilon} R_3^{\gamma\zeta} (\sum_{p=1}^{r_1}\sum_{q=1}^{r_2}\sum_{r=1}^{r_3} G_{pqr} u_{p,\delta} v_{q,\epsilon} w_{r,\zeta}) R_1^{\alpha\eta} R_2^{\beta\theta} R_3^{\gamma\iota} (\sum_{k=1}^{r_1}\sum_{l=1}^{r_2}\sum_{m=1}^{r_3} G_{klm} u_{k,\iota} v_{l,\theta} w_{m,\iota}) =$$
$$= \sum_{p=1}^{r_1}\sum_{q=1}^{r_2}\sum_{r=1}^{r_3}\sum_{k=1}^{r_1}\sum_{l=1}^{r_2}\sum_{m=1}^{r_3} G_{pqr} G_{klm} (\sum_{\alpha,\delta,\eta} R_1^{\alpha,\delta} R_1^{\alpha,\eta} u_{p,\delta} u_{k,\eta})(\sum_{\beta,\epsilon,\theta} R_2^{\beta,\epsilon} R_2^{\beta,\theta} v_{q,\epsilon} v_{l,\theta})(\sum_{\gamma,\zeta,\iota} R_3^{\gamma,\zeta} R_3^{\gamma,\iota} w_{r,\zeta} w_{m,\iota}) \tag{18}$$

Rotating two orthogonal vectors with the same rotation matrix keeps them orthogonal. Using this, we have

$$\|T(R_1,R_2,R_3)\|_F^2 = \sum_{p=1}^{r_1}\sum_{q=1}^{r_2}\sum_{r=1}^{r_3} G_{pqr}^2 (\sum_{\alpha,\delta,\eta} R_1^{\alpha,\delta} R_1^{\alpha,\eta} u_{p,\delta} u_{p,\eta})(\sum_{\beta,\epsilon,\theta} R_2^{\beta,\epsilon} R_2^{\beta,\theta} v_{q,\epsilon} v_{q,\theta})(\sum_{\gamma,\zeta,\iota} R_3^{\gamma,\zeta} R_3^{\gamma,\iota} w_{r,\zeta} w_{r,\iota}) =$$
$$= \sum_{p=1}^{r_1}\sum_{q=1}^{r_2}\sum_{r=1}^{r_3} G_{pqr}^2 \|R_1 u_p\|_F^2 \|R_2 v_q\|_F^2 \|R_3 w_r\|_F^2 \tag{19}$$

Vectors maintain the same Frobenius norm under rotation, therefore we finally get that $\|T\|_F^2 = \|T(R_1,R_2,R_3)\|_F^2$.

4 Exercise 4

1) Let's find the solution of the system $(A \odot_{KhR} B)\underline{\gamma} = 0$. Expanding the equation, we get:

$$\begin{gathered}
\gamma_1 a_{11} b_{11} + \cdots + \gamma_R a_{1R} b_{1R} = 0 \\
\gamma_1 a_{11} b_{21} + \cdots + \gamma_R a_{1R} b_{2R} = 0 \\
\vdots \\
\gamma_1 a_{11} b_{I_2 1} + \cdots + \gamma_R a_{1R} b_{I_2 R} = 0 \\
\vdots \\
\gamma_1 a_{I_1 1} b_{I_2 1} + \cdots + \gamma_R a_{I_1 R} b_{I_2 R} = 0
\end{gathered} \tag{20}$$

Grouping the equations above by the rows of B, we get:

$$\begin{gathered}
\gamma_1 \sum_{i=1}^{I_1} a_{i1} b_{11} + \cdots + \gamma_R \sum_{i=1}^{I_1} a_{iR} b_{1R} = 0 \\
\vdots \\
\gamma_1 \sum_{i=1}^{I_1} a_{i1} b_{I_2 1} + \cdots + \gamma_R \sum_{i=1}^{I_1} a_{iR} b_{I_2 R} = 0
\end{gathered} \tag{21}$$

As B is full-column rank, it follows that the only solution to the system above is that all the coefficients of the elements in B are zero. Given that this must hold for any full-column rank matrix A, we can use such a matrix that does not have a zero sum on any of the columns. Therefore, the only solution is that $\gamma_1 = \cdots = \gamma_R = 0$. Therefore, $A \odot_{KhR} B$ is also a full-column rank matrix.
2) After computing matrices A and B from the eigendecompositions, we need to compute C. We make use of the fact that A and B are full-column rank in this computation as follows: We compute each row of the matrix C by slicing through the tensor with canonical basis vectors e_1, e_2, ..., e_{I_3}. For slice i, we have the value $ADiag(\langle c_r, e_i \rangle)B^T$, so we can multiply it by the Moore-Penrose pseudo-inverses of A and B^T to get the elements in the first row of matrix C. This last step is possible due to the full-column rank property of matrices A and B.

5 Exercise 5

1) Using the property from the previous exercise that the Khatri-Rao product of two full-column rank matrices is also full-column rank, we have that $C \odot_{KhR} D$ is full-column rank. As A, B and $C \odot_{KhR} D$ are all column rank, we can apply Jennrich's algorithm to find the factorization of $\tilde{T}$.
2) Using Jennrich's algorithm, we can uniquely determine R, A, B and $C \odot_{KhR} D$ from the flattened version of T, $\tilde{T}$. To find C and D, we can consider $c_{r,1} = 1, \forall 1 \leq r \leq R$ and then compute the other values in c_r and d_r for each value $1 \leq r \leq R$. Namely, take a column from $C \odot_{KhR} D$, $\left[v_1, v_2, \ldots, v_{I_3 I_4}\right] = \left[c_1 d_1, c_1 d_2, \ldots, c_1 d_{I_4}, \ldots, c_{I_3} d_{I_4}\right]$. We set $c_1 = 1$ and compute $d_i = v_i$ then $c_i = \frac{v_{i I_3 + 1}}{d_1}, 2 \leq i \leq R$.

6 Exercise 6

1) Due to the uniqueness of the solution, we can simply verify that the matrix $\Sigma^\dagger$ proposed in the statement is indeed the pseudoinverse. We assume without loss of generality that $M \leq N$.

$$\Sigma\Sigma^\dagger = \begin{bmatrix} \Sigma_{11} & \dots & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \dots & \Sigma_{MM} & \dots & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\Sigma_{11}} & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & \frac{1}{\Sigma_{MM}} \\ \vdots & \dots & \vdots \\ 0 & \dots & 0 \end{bmatrix} = I_{M\times M}$$

$$\Sigma^\dagger\Sigma = \begin{bmatrix} \frac{1}{\Sigma_{11}} & \dots & 0 \\ \vdots & \dots & \vdots \\ 0 & \dots & \frac{1}{\Sigma_{MM}} \\ \vdots & \dots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \dots & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \dots & \vdots \\ 0 & \dots & \Sigma_{MM} & \dots & 0 \end{bmatrix} = \begin{bmatrix} I_{M\times M} & O_{M\times(N-M)} \\ O_{(N-M)\times M} & O_{(N-M)\times(N-M)} \end{bmatrix} \tag{22}$$

Therefore,

$$\begin{gathered} \Sigma\Sigma^\dagger\Sigma = I_{M\times M}\Sigma = \Sigma \\ \Sigma^\dagger\Sigma\Sigma^\dagger = \Sigma^\dagger I_{M\times M} = \Sigma^\dagger \\ (\Sigma\Sigma^\dagger)^* = I^*_{M\times M} = I_{M\times M} = \Sigma\Sigma^\dagger \\ (\Sigma^\dagger\Sigma)^* = \begin{bmatrix} I_{M\times M} & O_{M\times(N-M)} \\ O_{(N-M)\times M} & O_{(N-M)\times(N-M)} \end{bmatrix}^* = \begin{bmatrix} I_{M\times M} & O_{M\times(N-M)} \\ O_{(N-M)\times M} & O_{(N-M)\times(N-M)} \end{bmatrix} = \Sigma^\dagger\Sigma \end{gathered} \tag{23}$$

so $\Sigma^\dagger$ is the pseudoinverse of Σ.
2) From the previous subtask, we also see that $(\Sigma\Sigma^\dagger)^T = \Sigma^\dagger\Sigma$ and $(\Sigma^\dagger\Sigma)^T = \Sigma\Sigma^\dagger$. We consider $A^\dagger = V\Sigma^\dagger U^*$ to be the pseudoinverse and prove that this holds.

$$\begin{gathered} AA^\dagger A = U\Sigma V^* V\Sigma^\dagger U^* U\Sigma V^* = U\Sigma\Sigma^\dagger\Sigma V^* = U\Sigma V^* = A \\ A^\dagger AA^\dagger = V\Sigma^\dagger U^* U\Sigma V^* V\Sigma^\dagger U^* = V\Sigma^\dagger\Sigma\Sigma^\dagger U^* = V\Sigma^\dagger U^* = A^\dagger \\ (AA^\dagger)^* = (U\Sigma V^* V\Sigma^\dagger U^*)^* = (U\Sigma\Sigma^\dagger U^*)^* = U(\Sigma^\dagger\Sigma)^T U^* = U\Sigma\Sigma^\dagger U^* = U\Sigma V^* V\Sigma^\dagger U^* = AA^\dagger \\ (A^\dagger A)^* = (V\Sigma^\dagger U^* U\Sigma V^*)^* = (V\Sigma^\dagger\Sigma V^*)^* = V(\Sigma\Sigma^\dagger)^T V^* = V\Sigma^\dagger\Sigma V^* = V\Sigma^\dagger U^* U\Sigma V^* = A^\dagger A \end{gathered} \tag{24}$$

Therefore $A^\dagger$ is indeed the pseudoinverse of A.
3) If A has full-column rank, we first note that $N \leq M$. Let us test that $A^\dagger = (A^*A)^{-1}A^*$ respects the properties of a pseudoinverse.

$$\begin{gathered} AA^\dagger A = A(A^*A)^{-1}A^*A = A \\ A^\dagger AA^\dagger = (A^*A)^{-1}A^*A(A^*A)^{-1}A^* = (A^*A)^{-1}A^* = A^\dagger \\ (AA^\dagger)^* = (A(A^*A)^{-1}A^*)* = A(A^*A)^{-1}A^* = AA^\dagger \\ (A^\dagger A)^* = ((A^*A)^{-1}A^*A)* = I_N = (A^*A)^{-1}A^*A = A^\dagger A \end{gathered} \tag{25}$$

From this, we also see that $A^\dagger A = I_N$.
4) If A has full-row rank, we first note that $M \leq N$. Let us test that $A^\dagger =$

$A^*(AA^*)^{-1}$ respects the properties of a pseudoinverse.

$$\begin{aligned}
AA^\dagger A = AA^*(AA^*)^{-1}A = A \\
A^\dagger AA^\dagger = A^*(AA^*)^{-1}AA^*(AA^*)^{-1} = A^*(AA^*)^{-1} = A^\dagger \\
(AA^\dagger)^* = (AA^*(AA^*)^{-1})* = I_M = AA^*(AA^*)^{-1} = AA^\dagger \\
(A^\dagger A)^* = (A^*(AA^*)^{-1}A)* = A^*(AA^*)^{-1}A = A^\dagger A
\end{aligned} \tag{26}$$

From this, we also see that $AA^\dagger = I_M$.
5) If A is a square matrix with full rank, then $AA^\dagger = A^\dagger A = I_M$, so $A^\dagger$ is the usual inverse of A i.e. $A^\dagger = A^{-1}$.
6) We again assume that $(AB)^\dagger = B^\dagger A^\dagger$ and check that it satisfies the pseudoinverse properties. We will make use of the properties already deduced in the previous subtasks.

$$\begin{aligned}
(AB)(AB)^\dagger(AB) = ABB^\dagger A^\dagger AB = AB \\
(AB)^\dagger(AB)(AB)^\dagger = B^\dagger A^\dagger ABB^\dagger A^\dagger = B^\dagger A^\dagger = (AB)^\dagger \\
((AB)(AB)^\dagger)^* = (ABB^\dagger A^\dagger)^* = (AA^\dagger)^* = AA^\dagger = ABB^\dagger A^\dagger = (AB)(AB)^\dagger \\
((AB)^\dagger(AB))^* = (B^\dagger A^\dagger AB)^* = (B^\dagger B)^* = B^\dagger B = B^\dagger A^\dagger AB = (AB)^\dagger(AB)
\end{aligned} \tag{27}$$

Therefore, $B^\dagger A^\dagger$ is indeed the pseudoinverse of AB.