Solution 8 (4^{th} graded homework): 21 Mai 2019
CS-526 Learning Theory

Problem 1

1) For every $i \in [K]$, $\underline{d}_i$ is the i^{th} canonical basis vector of $\mathbb{R}^K$ and we define the latent random vector $\underline{h} \in \{\underline{d}_i : i \in [K]\}$ whose distribution is $\forall i \in [K] : \mathbb{P}(\underline{h} = \underline{d}_i) = w_i$. Finally, let $\underline{x} = \sum_{i=1}^K h_i \underline{a}_i + \underline{z}$ where $\underline{z} \sim \mathcal{N}(0, \sigma^2 I_{D\times D})$ is independent of $\underline{h}$. The random vector $\underline{x}$ has a probability density function $p(\cdot)$. We have:

$$
\begin{aligned}
\mathbb{E}[\underline{x}] &= \sum_{i=1}^K \mathbb{E}[h_i]\underline{a}_i + \mathbb{E}[\underline{z}] = \sum_{i=1}^K w_i\, \underline{a}_i \quad ; \\
\mathbb{E}[\underline{x}\underline{x}^T] &= \mathbb{E}[\underline{z}\underline{z}^T] + \sum_{i=1}^K \mathbb{E}[h_i] \underbrace{\mathbb{E}[\underline{z}]}_{=0} \underline{a}_i^T + \mathbb{E}[h_i]\underline{a}_i \mathbb{E}[\underline{z}]^T + \sum_{i,j=1}^K \underbrace{\mathbb{E}[h_i h_j]}_{=w_i\delta_{ij}} \underline{a}_i \underline{a}_j^T \\
&= \sigma^2 I_{D\times D} + \sum_{i=1}^K w_i\, \underline{a}_i \underline{a}_i^T \; .
\end{aligned}
$$

Finally, to compute the third moment tensor, note that $\mathbb{E}[\underline{z} \otimes \underline{z} \otimes \underline{z}] = 0$ and that for every $(i,j) \in [K]^2$: $\mathbb{E}[\underline{a}_i \otimes \underline{a}_j \otimes \underline{z}] = \mathbb{E}[\underline{a}_i \otimes \underline{z} \otimes \underline{a}_j] = \mathbb{E}[\underline{z} \otimes \underline{a}_i \otimes \underline{a}_j] = 0$. Hence:

$$
\begin{aligned}
\mathbb{E}[\underline{x} \otimes \underline{x} \otimes \underline{x}] = &\sum_{i,j,k=1}^K \underbrace{\mathbb{E}[h_i h_j h_k]}_{=w_i\delta_{ij}\delta_{ik}} \underline{a}_i \otimes \underline{a}_j \otimes \underline{a}_k \\
&+ \sum_{i=1}^K \mathbb{E}[h_i]\mathbb{E}[\underline{a}_i \otimes \underline{z} \otimes \underline{z}] + \mathbb{E}[h_i]\mathbb{E}[\underline{z} \otimes \underline{a}_i \otimes \underline{z}] + \mathbb{E}[h_i]\mathbb{E}[\underline{z} \otimes \underline{z} \otimes \underline{a}_i] \\
= &\sum_{i=1}^K w_i\, \underline{a}_i \otimes \underline{a}_i \otimes \underline{a}_i + \sigma^2 \sum_{j=1}^D \sum_{i=1}^K w_i (\underline{a}_i \otimes \underline{e}_j \otimes \underline{e}_j + \underline{e}_j \otimes \underline{e}_j \otimes \underline{a}_i + \underline{e}_j \otimes \underline{a}_i \otimes \underline{e}_j) \, .
\end{aligned}
$$

2) Let $A = [\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_K] \in \mathbb{R}^{D\times K}$ and $A' = [\underline{a}'_1, \underline{a}'_2, \ldots, \underline{a}'_K] \in \mathbb{R}^{D\times K}$. By definition, $\widetilde{R} = \Sigma^{-1} R \Sigma$ where Σ is the diagonal matrix such that $\Sigma_{ii} = \sqrt{w_i}$ and $A' = A\widetilde{R}^T$. We can directly apply the formula of question 1) to compute the second moment matrix of the new mixture of Gaussians:

$$
\begin{aligned}
\mathbb{E}[\underline{x}\underline{x}^T] &= \sigma^2 I_{D\times D} + A'\Sigma^2 A'^T = \sigma^2 I_{D\times D} + A\widetilde{R}^T \Sigma^2 \widetilde{R} A^T \\
&= \sigma^2 I_{D\times D} + A\Sigma R^T R \Sigma A^T = \sigma^2 I_{D\times D} + A\Sigma^2 A^T \; .
\end{aligned}
$$

Problem 2: Examples of tensors and their rank

1) The matrices corresponding to B, P, E are:

$$
B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \; ; \; P = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \; ; \; E = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}
$$

The frontal slices of G and W are:

$$G_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} , G_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \ ; \ W_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} , W_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} .$$

The matricizations of G and W are:

$$G_{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \ ; \ G_{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \ ; \ G_{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \ ;$$
$$W_{(1)} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \ ; \ W_{(2)} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \ ; \ W_{(3)} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} .$$

2) B and E are clearly rank-2 matrices, while $P = (e_0 + e_1) \otimes (e_0 + e_1)$ is a rank-1 matrix.
By its definition, G is at most rank 2. Assume it is rank 1: $G = a \otimes b \otimes c$ with $a, b, c \in \mathbb{R}^2$. We have $a_1 b_1 c_1 = G_{111} = 1$ and $a_2 b_1 c_1 = G_{211} = 0$ so we must have $a_2 = 0$. Besides, $a_2 b_2 c_2 = G_{222} = 1$ and $a_1 b_2 c_2 = G_{122} = 0$ so $a_1 = 0$. Hence $a^T = (0, 0)$ and G is the all-zero tensor. This is a contradiction and we conclude that G is rank 2.
By its definition, W is at most rank 3. To prove the rank cannot be smaller than 3, we will proceed by contradiction:

- Assume W is rank 1: $W = a \otimes b \otimes c$ with $a, b, c \in \mathbb{R}^2$. We have $a_1 b_1 c_1 = W_{111} = 0$ and $a_2 b_1 c_1 = W_{211} = 1$ so $a_1 = 0$. Besides, $a_1 b_1 c_2 = W_{112} = 1$ and $a_2 b_1 c_2 = W_{212} = 0$ so $a_2 = 0$. Then $a = (0, 0)^T$ and W is the all-zero tensor, which is a contradiction.

- Assume W is rank 2: $W = a \otimes b \otimes c + d \otimes e \otimes f$. We claim that *$a$ and d must be linearly independent*. Indeed, suppose they are parallel and take a vector x perpendicular to both a and d. Then
$$W(x, I, I) = (x^T a) b \otimes c + (x^T d) e \otimes f = 0$$
but also
$$W(x, I, I) = (x^T e_0) e_0 \otimes e_1 + (x^T e_0) e_1 \otimes e_0 + (x^T e_1) e_0 \otimes e_0 = \begin{bmatrix} x^T e_1 & x^T e_0 \\ x^T e_0 & 0 \end{bmatrix}$$
which cannot be zero since x cannot be perpendicular to both e_0 and e_1. Now, we take x perpendicular to d. We have
$$W(x, I, I) = (x^T a) b \otimes c$$
which is rank one. Therefore, we must have $x^T e_0 = 0$ which implies that x is parallel to e_1 and thus d parallel to e_0. Now, if we take x perpendicular to a, the matrix
$$W(x, I, I) = (x^T d) e \otimes f$$
is rank one and, once again, we must have $x^T e_0 = 0$, which implies x parallel to e_1 and thus a parallel to e_0. Hence, we have shown that a and d are linearly independent but also that both are parallel to e_0. This is a contradiction.

3) Writing everything in terms of matrix product, it comes:

$$(Oe_0) \otimes (Oe_0) + (Oe_1) \otimes (Oe_1) = Oe_0 e_0^T O^T + Oe_1 e_1^T O^T = OO^T = B .$$

so B does not have a unique decomposition.

For G we have $G = \underline{a}_1 \otimes \underline{b}_1 \otimes \underline{c}_1 + \underline{a}_2 \otimes \underline{b}_2 \otimes \underline{c}_2$ with

$$A = [\underline{a}_1, \underline{a}_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \ ; \ B = [\underline{b}_1, \underline{b}_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \ ; \ C = [\underline{c}_1, \underline{c}_2] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} .$$

A, B, C are full column rank and G has rank 2: by Jennrich's algorithm, the decomposition is unique (up to trivial rank permutation and feature scaling).

For W we have $W = \underline{a}_1 \otimes \underline{b}_1 \otimes \underline{c}_1 + \underline{a}_2 \otimes \underline{b}_2 \otimes \underline{c}_2 + \underline{a}_3 \otimes \underline{b}_3 \otimes \underline{c}_3$ with

$$A = [\underline{a}_1, \underline{a}_2, \underline{a}_3] = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \; ; \; B = [\underline{b}_1, \underline{b}_2, \underline{b}_3] = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \; ; \; C = [\underline{c}_1, \underline{c}_2, \underline{c}_3] = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} .$$

A, B, C are not full column rank: Jennrich's theorem does not allow to conclude that the decomposition of W is unique.

4) We expand the tensor products in the definition of D_ϵ:

$$\begin{aligned} D_\epsilon &= \frac{1}{\epsilon}\Big[(e_0 + \epsilon e_1) \otimes (e_0 + \epsilon e_1) \otimes (e_0 + \epsilon e_1) - e_0 \otimes e_0 \otimes e_0\Big] \\ &= \frac{1}{\epsilon}\Big[e_0 \otimes e_0 \otimes e_0 + \epsilon\, e_0 \otimes e_0 \otimes e_1 + \epsilon\, e_0 \otimes e_1 \otimes e_0 + \epsilon\, e_1 \otimes e_0 \otimes e_0 \\ &\qquad + \epsilon^2\, e_1 \otimes e_1 \otimes e_0 + \epsilon^2\, e_1 \otimes e_0 \otimes e_1 + \epsilon^2\, e_0 \otimes e_1 \otimes e_1 + \epsilon^3\, e_1 \otimes e_1 \otimes e_1 - e_0 \otimes e_0 \otimes e_0\Big] \\ &= e_0 \otimes e_0 \otimes e_1 + e_0 \otimes e_1 \otimes e_0 + e_1 \otimes e_0 \otimes e_0 \\ &\qquad + \epsilon(e_1 \otimes e_1 \otimes e_0 + e_1 \otimes e_0 \otimes e_1 + e_0 \otimes e_1 \otimes e_1) + \epsilon^2\, e_1 \otimes e_1 \otimes e_1 \\ &= W + \epsilon(e_1 \otimes e_1 \otimes e_0 + e_1 \otimes e_0 \otimes e_1 + e_0 \otimes e_1 \otimes e_1) + \epsilon^2\, e_1 \otimes e_1 \otimes e_1 . \end{aligned}$$

Hence $\lim_{\epsilon \to 0} D_\epsilon = 0$.

Problem 3

1) There cannot be an analogous general result for tensors. Indeed, the order-3 tensor W of Problem 2 is rank 3 and we show in 4) that $\lim_{\epsilon \to 0} \|W - D_\epsilon\|_F = 0$. So there is no minimum attained in the space of rank 2 tensors. In this sense, there is simply no *best* rank-two approximation of W.

2) Let M a matrix of rank $R+1$ with singular values $\sigma_1 \geq \sigma_2 \cdots \geq \sigma_R \geq \sigma_{R+1} > 0$. By the Eckart-Young-Mirsky theorem, the minimum of $\|M - \widehat{M}\|_F$ over rank R matrices $\widehat{M}$ is equal to $\sigma_{R+1} > 0$. Therefore, there cannot be a sequence of matrices M_n given by a sum of R rank-one matrices such that $\lim_{n \to +\infty} \|M - M_n\|_F = 0$.

4) In the real-valued case, we have:

$$|T(R_1, R_2, R_3)^{\alpha\beta\gamma}|^2 = \sum_{\delta,\epsilon,\zeta,\delta',\epsilon',\delta'} R_1^{\delta\alpha} R_1^{\delta'\alpha} R_2^{\epsilon\beta} R_2^{\epsilon'\beta} R_3^{\zeta\gamma} R_3^{\zeta'\gamma} T^{\delta\epsilon\zeta} T^{\delta'\epsilon'\zeta'} .$$

Summing over α, β, γ and using the orthogonality of rotation matrices, we find:

$$\sum_\alpha R_1^{\delta\alpha} R_1^{\delta'\alpha} = \delta_{\delta\delta'}, \quad \sum_\beta R_2^{\epsilon\beta} R_2^{\epsilon'\beta} = \delta_{\beta\beta'}, \quad \sum_\gamma R_3^{\zeta\gamma} R_3^{\zeta'\gamma} = \delta_{\zeta\zeta'} .$$

The result directly follows:

$$\|T(R_1, R_2, R_3)\|_F^2 = \sum_{\delta\epsilon\zeta} |T(R_1, R_2, R_3)^{\alpha\beta\gamma}|^2 = \sum_{\delta\epsilon\zeta} |T^{\delta\epsilon\zeta}|^2 = \|T\|_F^2 .$$

Problem 4

1) To show that $A \odot_{KhR} B$ is full column rank, we have to prove that the kernel of the linear application $\underline{x} \mapsto (A \odot_{KhR} B)\underline{x}$ 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, $\sum_{r=1}^{R} a_r^\alpha x^r \underline{b}_r = 0$ implies that $\forall r \in [R] : a_r^\alpha x^r = 0$. Note that:

$$\forall \alpha \in [I_1], \forall r \in [R] : a_k^\alpha x^r = 0 \Leftrightarrow A\underline{x} = 0 \,.$$

A is full column rank and $A\underline{x} = 0$, hence $\underline{x} = 0$. $A \odot_{KhR} B$ 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 (1) 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:

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

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:

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

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

Problem 5

1) To apply Jennrich's algorithm we need to prove that the matrix $E = [\underline{c}_1 \otimes_{Kro} \underline{d}_1, \ldots, \underline{c}_R \otimes_{Kro} \underline{d}_R]$ is full column rank (A, B are full column rank by assumption). Note that teh 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 Jennsen'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 $\alpha \in [I_3]$ such that $c_r^\alpha \neq 0$. As $c_r^\alpha \neq 0$, we can use the I_4-dimensional column vector $c_r^\alpha \underline{d}_r$ contained in the r^{th} column of E to recover $\underline{d}_r$. Doing this for every $r \in [R]$ we recover the matrix D. Finally, for every $r \in R$, pick $\beta \in I_4$ such that $d_r^\beta \neq 0$ (such β exists because D is full column rank) and use the entries $c_r^\alpha d_r^\beta$, $\alpha \in [I_3]$, to recover $\underline{c}_r$. C has then been recovered.

Problem 6

1) Define $\Sigma^\dagger$ as the $N \times M$ diagonal matrix with diagonal entries:

$$\forall i \in \{1, 2, \ldots, \min\{M, N\}\} : (\Sigma^\dagger)_{ii} \begin{cases} 1/\Sigma_{ii} & \text{if} \quad \Sigma_{ii} \neq 0 \; ; \\ 0 & \text{otherwise.} \end{cases}$$

Then both $\Sigma^\dagger\Sigma \in \mathbb{C}^{N\times N}$ and $\Sigma\Sigma^\dagger \in \mathbb{C}^{M\times M}$ are diagonal square matrices with diagonal entries:

$$\forall i \in [N] : (\Sigma^\dagger\Sigma)_{ii} = \begin{cases} 1 & \text{if} \quad i \leq \min\{M, N\} \text{ and } \Sigma_{ii} \neq 0 \; ; \\ 0 & \text{otherwise.} \end{cases}$$

$$\forall i \in [M] : (\Sigma\Sigma^\dagger)_{ii} = \begin{cases} 1 & \text{if} \quad i \leq \min\{M, N\} \text{ and } \Sigma_{ii} \neq 0 \; ; \\ 0 & \text{otherwise.} \end{cases}$$

It is then easy to check that $\Sigma^\dagger$ satisfies the first two conditions of the Moore-Penrose pseudoinverse: $\Sigma\Sigma^\dagger\Sigma = \Sigma$ and $\Sigma^\dagger\Sigma\Sigma^\dagger = \Sigma^\dagger$. Besides, $\Sigma^\dagger\Sigma$ and $\Sigma\Sigma^\dagger$ being real diagonal matrices, the last two conditions are clearly satisfied too.

2) We can check that the matrix $V\Sigma^\dagger U^*$ satisfies the four conditions of the Moore-Penrose pseudoinverse, i.e., $A^\dagger = V\Sigma^\dagger U^*$:

$$\begin{aligned} A[V\Sigma^\dagger U^*]A &= U\Sigma(V^*V)\Sigma^\dagger(U^*U)\Sigma V^* = U\Sigma\Sigma^\dagger\Sigma V^* = U\Sigma V^* = A \; ; \\ [V\Sigma^\dagger U^*]A[V\Sigma^\dagger U^*] &= V\Sigma^\dagger(U^*U)\Sigma(V^*V)\Sigma^\dagger U^* = V\Sigma^\dagger\Sigma\Sigma^\dagger U^* = V\Sigma^\dagger U^* \; ; \\ (AV\Sigma^\dagger U^*)^* &= (U\Sigma\Sigma^\dagger U^*)* = U(\Sigma\Sigma^\dagger)^*U^* = U\Sigma\Sigma^\dagger U^* = AV\Sigma^\dagger U^* \; ; \\ (V\Sigma^\dagger U^*A)^* &= (V\Sigma^\dagger\Sigma V^*)* = V(\Sigma^\dagger\Sigma)^*V^* = V\Sigma^\dagger\Sigma V^* = V\Sigma^\dagger U^*A \, . \end{aligned}$$

3) A is full column rank, therefore A^*A is a full rank $N \times N$ matrix and has a unique inverse $(A^*A)^{-1}$. The matrix $(A^*A)^{-1}A^*$ satisfies the four conditions:

$$A[(A^*A)^{-1}A^*]A = A \; ; \; [(A^*A)^{-1}A^*]A[(A^*A)^{-1}A^*] = (A^*A)^{-1}A^* \; ;$$
$$(A[(A^*A)^{-1}A^*])^* = A[(A^*A)^{-1}A^*] \; ; \; ([(A^*A)^{-1}A^*]A)^* = A^*A(A^*A)^{-1} = I_{N\times N} = ([(A^*A)^{-1}A^*]A \, .$$

Hence $A^\dagger = (A^*A)^{-1}A^*$.

4) A is full row rank, therefore AA^* is a full rank $M \times M$ matrix and has a unique inverse $(AA^*)^{-1}$. The matrix $A^*(AA^*)^{-1}$ satisfies the four conditions:

$$A[A^*(AA^*)^{-1}]A = A \; ; \; [A^*(AA^*)^{-1}]A[A^*(AA^*)^{-1}] = A^*(AA^*)^{-1} \; ;$$
$$(A[A^*(AA^*)^{-1}])^* = (AA^*)^{-1}AA^* = I_{M\times M} = AA^\dagger \; ; \; ([A^*(AA^*)^{-1}]A)^* = A^*(AA^*)^{-1}A \, .$$

Hence $A^\dagger = A^*(AA^*)^{-1}$.

5) We have $AA^{-1}A = A$, $A^{-1}AA^{-1} = A^{-1}$, $(AA^{-1})^* = I_{M\times M} = AA^{-1}$, $(A^{-1}A)^* = I_{N\times N} = A^{-1}A$. Hence $A^\dagger = A^{-1}$.

6) A is full column rank so $A^\dagger A = I_{M\times M}$ and B is full column rank so $BB^\dagger = I_{N\times N}$. Therefore:

$$\begin{aligned}
(AB)(B^\dagger A^\dagger)(AB) &= A(BB^\dagger)(A^\dagger A)B = AI_{M\times M}I_{N\times N}B = AB \ ; \\
(B^\dagger A^\dagger)(AB)(B^\dagger A^\dagger) &= B^\dagger(A^\dagger A)(BB^\dagger)A^\dagger = B^\dagger I_{N\times N}I_{M\times M}A^\dagger = B^\dagger A^\dagger \ ; \\
(ABB^\dagger A^\dagger)^* &= (AI_{N\times N}A^\dagger)^* = (AA^\dagger)^* = AA^\dagger = (AB)(B^\dagger A^\dagger) \ ; \\
(B^\dagger A^\dagger AB)^* &= (B^\dagger I_{M\times M}B)^* = (B^\dagger B)^* = B^\dagger B = (B^\dagger A^\dagger)(AB) \ .
\end{aligned}$$

Hence $(AB)^\dagger = B^\dagger A^\dagger$.