Solution for Assignment 9

1. **a)** Consider the matrix $M := AB$. We claim that it has $rank(M) = n$. To see this, observe that $\text{rank}(B) = n$ implies $\mathbf{C}(B) = \mathbb{R}^n$ because n is also the number of rows of B. Hence, we get $\mathbf{C}(M) = \mathbf{C}(A)$ (and therefore $\text{rank}(M) = \text{rank}(A) = n$). Finally, we can use Proposition 4.5.9 to get $(AB)^\dagger = M^\dagger = B^\dagger A^\dagger$.

b) Assume first that A has full column rank $n = \text{rank}(A)$. In this case, we have $A^\dagger = (A^\top A)^{-1}A^\top$ by definition of the pseudoinverse for matrices with full column rank. Moreover, notice that $A^\top$ has full row rank and hence we also get $(A^\top)^\dagger = A(A^\top A)^{-1}$ by definition of the pseudoinverse for matrices with full row rank. Hence, we get

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

We conclude that the statements holds for all matrices with full column rank.

Analogously, we can prove that the statement holds if A has full row rank $m = \text{rank}(A)$. In that case, we have $A^\dagger = A^\top(AA^\top)^{-1}$ and $(A^\top)^\dagger = (AA^\top)^{-1}A$. Hence, we indeed get

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

We conclude that the statement holds for all matrices with full row rank.

It remains to prove the general case, i.e. we do not assume anymore that A has full row rank or full column rank. Then by definition, we have $A^\dagger = R^\dagger C^\dagger$ where $A = CR$ is a CR decomposition of A. In particular, we have $C \in \mathbb{R}^{m\times r}$ and $R \in \mathbb{R}^{r\times n}$ where $r = \text{rank}(A)$. Now observe that we also have $A^\top = R^\top C^\top$ with $R^\top \in \mathbb{R}^{n\times r}$ and $C^\top \in \mathbb{R}^{r\times m}$ and of course, $r = \text{rank}(A) = \text{rank}(A^\top)$. Hence, we can use Proposition 4.5.9 to get $(A^\top)^\dagger = (C^\top)^\dagger(R^\top)^\dagger$. We conclude that

$$(A^\top)^\dagger = (C^\top)^\dagger(R^\top)^\dagger = (C^\dagger)^\top(R^\dagger)^\top = (R^\dagger C^\dagger)^\top = (A^\dagger)^\top$$

by using that C has full column rank and R has full row rank and hence $(C^\top)^\dagger = (C^\dagger)^\top$ and $(R^\top)^\dagger = (R^\dagger)^\top$.

c) Let $A = CR$ be a CR decomposition of A with $C \in \mathbb{R}^{m\times r}$ and $R \in \mathbb{R}^{r\times n}$ where $r = \text{rank}(A)$. We can rewrite

$$AA^\dagger = CR(CR)^\dagger = CRR^\dagger C^\dagger \overset{\text{Prop. 4.5.4}}{=} CIC^\dagger = C(C^\top C)^{-1}C^\top$$

and hence we conclude symmetry of $AA^\dagger$ since

$$(AA^\dagger)^\top = (C(C^\top C)^{-1}C^\top)^\top = C((C^\top C)^{-1})^\top C^\top = C((C^\top C)^\top)^{-1}C^\top = C(C^\top C)^{-1}C^\top = AA^\dagger.$$

It remains to prove that $AA^\dagger$ is the projection matrix that projects vectors onto $\mathbf{C}(A)$. For this, observe that we have $\mathbf{C}(A) = \mathbf{C}(C)$ (by the properties of CR-decomposition) and that we have already shown $AA^\dagger = C(C^\top C)^{-1}C^\top$. By Theorem 4.2.6, this is exactly the projection matrix that projects vectors onto $\mathbf{C}(C) = \mathbf{C}(A)$.

d) We proceed as in the preceding subtask. Let $A = CR$ be a CR decomposition of A with $C \in \mathbb{R}^{m\times r}$ and $R \in \mathbb{R}^{r\times n}$ where $r = \text{rank}(A)$. We can rewrite

$$A^\dagger A = (CR)^\dagger CR = R^\dagger C^\dagger CR \overset{\text{Prop. 4.5.2}}{=} R^\dagger IR = R^\top(RR^\top)^{-1}R$$

and hence we conclude symmetry of $AA^\dagger$ since

$$(A^\dagger A)^\top = (R^\top(RR^\top)^{-1}R)^\top = R^\top((RR^\top)^{-1})^\top R = R^\top((RR^\top)^\top)^{-1}R = R^\top(RR^\top)^{-1}R = A^\dagger A.$$

By Theorem 4.2.6, the matrix $R^\top(RR^\top)^{-1}R = A^\dagger A$ is exactly the projection matrix onto the subspace $\mathbf{C}(R^\top) = \mathbf{R}(R) = \mathbf{R}(A) = \mathbf{C}(A^\top)$ (the equality $\mathbf{R}(R) = \mathbf{R}(A)$ is due to the observation that R can be obtained from A through row operations and deleting 0-rows, and by recalling that row operations preserve the row space).

2. We solve this by using our knowledge on pseudoinverses. Consider the function $f^{-1} : \mathbf{C}(A) \to \mathbf{C}(A^\top)$ given by $f^{-1}(\mathbf{x}) = A^\dagger\mathbf{x}$ for all $\mathbf{x} \in \mathbf{C}(A)$. Observe that the composition $f^{-1} \circ f$ is the identity: we know from Exercise 1 that $A^\dagger A$ is the projection matrix that projects vectors onto the subspace $\mathbf{C}(A^\top)$, and hence we have

$$f^{-1}(f(\mathbf{x})) = A^\dagger A\mathbf{x} = \mathbf{x}$$

for all $\mathbf{x} \in \mathbf{C}(A^\top)$. This already implies that f is injective. Observe that with an analogous argument we get

$$f(f^{-1}(\mathbf{x})) = AA^\dagger\mathbf{x} = \mathbf{x}$$

for all $\mathbf{x} \in \mathbf{C}(A)$. Hence, f^{-1} is injective as well which implies that both f and f^{-1} are bijective.

Note that the matrix $A^\dagger A$ is in general not the identity matrix. It is crucial that the function f is only defined on $\mathbf{C}(A^\top)$ and not on all of $\mathbb{R}^n$.

3. **a)** The linear transformation given by $A = I - 2\mathbf{v}\mathbf{v}^\top$ corresponds to reflection along the hyperplane $H = \{\mathbf{x} \in \mathbb{R}^2 : \mathbf{x} \cdot \mathbf{v} = 0\}$. To see this, recall that the projection matrix for projection onto $\mathsf{Span}(\mathbf{v})$ is given by $B = \frac{\mathbf{v}\mathbf{v}^\top}{||\mathbf{v}||^2}$. Since $\mathbf{v}$ is a unit vector, this simplifies to $B = \mathbf{v}\mathbf{v}^\top$. In particular, this means that $A = I - 2B$ applied to some vector $\mathbf{x} \in \mathbb{R}^2$ will subtract the projection of $\mathbf{x}$ onto $\mathsf{Span}(\mathbf{v})$ twice from $\mathbf{x}$ itself. In other words, assume we split $\mathbf{x}$ into two parts $\mathbf{x} = \mathbf{x}_| + \mathbf{x}_\perp$ with $\mathbf{x}_| = B\mathbf{x}$ and $\mathbf{x}_\perp \cdot \mathbf{v} = 0$. Then the transformation given by matrix A maps $\mathbf{x}$ to the vector

$$A\mathbf{x} = (I-2B)(\mathbf{x}_|+\mathbf{x}_\perp) = \mathbf{x}_|+\mathbf{x}_\perp-2B\mathbf{x}_|-2B\mathbf{x}_\perp = \mathbf{x}_|+\mathbf{x}_\perp-2\mathbf{x}_|-2\mathbf{v}\mathbf{v}^\top\mathbf{x}_\perp = \mathbf{x}_|-\mathbf{x}_\perp-\mathbf{0} = \mathbf{x}_|-\mathbf{x}_\perp.$$

A picture of this is provided in Figure 1.

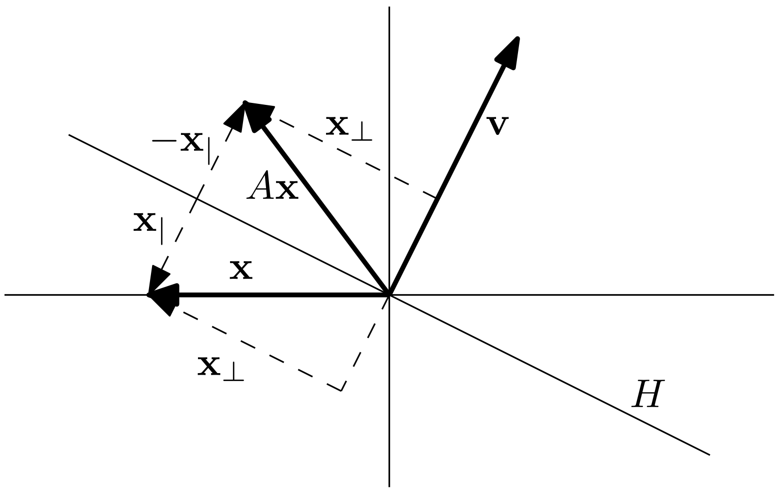

Figure 1: A sketch of the situation in subtask a).

b) Recall from the lecture (notes) that the 2×2 matrix

$$A' = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

is a rotation matrix. In particular, it rotates vectors in $\mathbb{R}^2$ by $\frac{\pi}{4}$ in counter-clockwise direction. Observe that the matrix A contains A' as a submatrix, i.e. we can obtain A' by removing the second column and second row from A.

Now consider an arbitrary vector $\mathbf{v} \in \mathbb{R}^3$ that is spanned by $\mathbf{e}_1$ and $\mathbf{e}_3$, i.e. $\mathbf{v} = v_1\mathbf{e}_1 + v_3\mathbf{e}_3$. Then applying A to $\mathbf{v}$ has no effect on the second coordinate, but it will rotate that $\mathbf{v}$ in the plane spanned by $\mathbf{e}_1$ and $\mathbf{e}_3$. A better way to say this, is that applying A to $\mathbf{v}$ corresponds to a rotation around the axis $\mathbf{e}_2$ (the vecor orthogonal to the plane spanned by $\mathbf{e}_1$ and $\mathbf{e}_3$).

Finally, observe that this is still true even if we don't have $v_2 = 0$. Since the second column of A is simply $\mathbf{e}_2$, the second coordinate of $\mathbf{v}$ remains unchanged when applying A to it.

In conclusion, the linear transformation given by matrix A is a rotation by $\frac{\pi}{4}$ around the axis spanned by $\mathbf{e}_2$.

c) Note that we did not specify a direction for the rotation in the exercise. The exercise is still well defined because we want to rotate by π and hence the direction does not matter.

We start by thinking about what such a transformation would do to the standard unit vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$. The vector $\mathbf{e}_1$ should be mapped to $\mathbf{e}_2$, the vector $\mathbf{e}_2$ should be mapped to $\mathbf{e}_1$, and the vector $\mathbf{e}_3$ should be mapped to $-\mathbf{e}_3$. In particular, we want

$$A \begin{bmatrix} | & | & | \\ \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ | & | & | \end{bmatrix} \overset{!}{=} \begin{bmatrix} | & | & | \\ \mathbf{e}_2 & \mathbf{e}_1 & -\mathbf{e}_3 \\ | & | & | \end{bmatrix}.$$

From this, we conclude that A has to be the matrix

$$A = \begin{bmatrix} | & | & | \\ \mathbf{e}_2 & \mathbf{e}_1 & -\mathbf{e}_3 \\ | & | & | \end{bmatrix}.$$

4. **a)** Let $\mathbf{x} \in T$ be arbitrary. By definition, there exist $c_1, c_2, c_3 \in \mathbb{R}_0^+$ with $c_1 + c_2 + c_3 = 1$ such that $\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3$. We compute

$$A\mathbf{x} = A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3) = c_1A\mathbf{v}_1 + c_2A\mathbf{v}_2 + c_3A\mathbf{v}_3 = c_1\mathbf{v}_1' + c_2\mathbf{v}_2' + c_3\mathbf{v}_3'$$

where $\mathbf{v}_1', \mathbf{v}_2', \mathbf{v}_3'$ are the vertices of T' given in the exercise. We conclude that indeed we have $A\mathbf{x} \in T'$.

b) Since T' is not a single point, two of its vertices must be distinct. Without loss of generality, assume it is $\mathbf{v}_1'$ and $\mathbf{v}_2'$, i.e. $\mathbf{v}_1' \neq \mathbf{v}_2'$. We also know that T' is not a triangle. Hence, by definition of a triangle, T is either a line segment or the vertices $\mathbf{v}_1', \mathbf{v}_2', \mathbf{v}_3'$ are not distinct. If the former is the case, we are done. Thus, assume now the latter and without loss of generality assume $\mathbf{v}_3' = \mathbf{v}_2'$. But then T' can be described by just using $\mathbf{v}_1'$ and $\mathbf{v}_2'$ as

$$\begin{aligned} T' &= \{c_1\mathbf{v}_1' + c_2\mathbf{v}_2' + c_3\mathbf{v}_3' : c_1, c_2, c_3 \in \mathbb{R}_0^+, c_1 + c_2 + c_3 = 1\} \\ &= \{c_1\mathbf{v}_1' + (c_2 + c_3)\mathbf{v}_2' : c_1, c_2, c_3 \in \mathbb{R}_0^+, c_1 + (c_2 + c_3) = 1\} \\ &= \{c_1\mathbf{v}_1' + c_{23}\mathbf{v}_2' : c_1, c_{23} \in \mathbb{R}_0^+, c_1 + c_{23} = 1\}. \end{aligned}$$

Notice that we still have $\mathbf{v}_1' \neq \mathbf{v}_2'$ and hence T' is a line segment.

c) We prove both directions individually.

“ $\Longrightarrow$ ” Assume that A has rank 0 and hence $A = 0$. Then $\mathbf{v}_1' = \mathbf{v}_2' = \mathbf{v}_3' = \mathbf{0}$ and hence $T' = \{\mathbf{0}\}$ is a single point.

“ $\Longleftarrow$ ” Assume now T' is a single point. Then we must have $\mathbf{v}_1' = \mathbf{v}_2' = \mathbf{v}_3'$. Consider the vectors $\mathbf{w}_1 = \mathbf{v}_1 - \mathbf{v}_3$ and $\mathbf{w}_2 = \mathbf{v}_2 - \mathbf{v}_3$. Since T is a triangle, these two vectors must be linearly independent (otherwise they would be collinear and we get that T is actually a line segment, and not a triangle). We have $A\mathbf{w}_1 = A\mathbf{v}_1 - A\mathbf{v}_3 = \mathbf{0}$ and also $A\mathbf{w}_2 = A\mathbf{v}_2 - A\mathbf{v}_3 = \mathbf{0}$. Hence, the nullspace of A has dimension two and consequently the rank of A (dimension of its column space) must be zero.

d) We prove both directions individually.

“ $\Longrightarrow$ ” Assume that A has rank 2. Consider again the two linearly independent vectors $\mathbf{w}_1 = \mathbf{v}_1 - \mathbf{v}_3$ and $\mathbf{w}_2 = \mathbf{v}_2 - \mathbf{v}_3$. Since A has full rank, the vectors $A\mathbf{w}_1 = \mathbf{v}_1' - \mathbf{v}_3'$ and $A\mathbf{w}_2 = \mathbf{v}_2' - \mathbf{v}_3'$ must be linearly independent as well. This directly implies $\mathbf{v}_1' \neq \mathbf{v}_2' \neq \mathbf{v}_3' \neq \mathbf{v}_1'$ and it remains to argue that T' is not a line segment.

“ $\Longleftarrow$ ” Assume now T' is a triangle. Consider the vectors $\mathbf{w}_1' = \mathbf{v}_1' - \mathbf{v}_3'$ and $\mathbf{w}_2' = \mathbf{v}_2' - \mathbf{v}_3'$. Since T' is a triangle, these two vectors must be linearly independent (otherwise they would be collinear and we would get that T' is actually a line segment, and not a triangle). We have $A(\mathbf{v}_1 - \mathbf{v}_3) = A\mathbf{v}_1 - A\mathbf{v}_3 = \mathbf{w}_1'$ and also $A(\mathbf{v}_2 - \mathbf{v}_3) = A\mathbf{v}_2 - A\mathbf{v}_3 = \mathbf{w}_2'$. Hence, the column space of A has dimension two and consequently the rank of A (dimension of its column space) is two.

e) We prove both directions individually.

“ $\Longrightarrow$ ” Assume that A has rank 1. By subtasks c) and d) we get that T' cannot be a single point and it also cannot be a triangle. Hence, T' must be a line segment by subtask b).

“ $\Longleftarrow$ ” Assume that T' is a line segment. By subtasks c) and d) we know that A cannot have rank 0 and it also cannot have rank 2. We conclude that is must have rank 1.

5. Let us denote the four given points by $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4$, respectively. We want to find $r \in \mathbb{R}^+$ such that the sum

$$\sum_{i=1}^{4} (r - ||\mathbf{p}_i||)^2$$

is minimized. The key observation of this exercise is that this is the least squares objective of the linear system

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} r \end{bmatrix} = \begin{bmatrix} ||\mathbf{p}_1|| \\ ||\mathbf{p}_2|| \\ ||\mathbf{p}_3|| \\ ||\mathbf{p}_4|| \end{bmatrix} = \begin{bmatrix} 2 \\ \sqrt{2} \\ \sqrt{\frac{20}{9}} \\ \sqrt{\frac{10}{4}}. \end{bmatrix}$$

Using the normal equations to solve this we get

$$4r = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} r \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} ||\mathbf{p}_1|| \\ ||\mathbf{p}_2|| \\ ||\mathbf{p}_3|| \\ ||\mathbf{p}_4|| \end{bmatrix} = \sum_{i=1}^{4} ||\mathbf{p}_i||$$

and hence

$$r = \frac{1}{4} \sum_{i=1}^{4} ||\mathbf{p}_i|| = \frac{1}{4}(2 + \sqrt{2} + \sqrt{\frac{20}{9}} + \sqrt{\frac{10}{4}}).$$