Solution for Assignment 7

1. **a)** Observe that since both V and W are subspaces, we must have $\mathbf{0} \in V \cap W$ and it remains to prove that this is the only element of $V \cap W$. Since V and W are orthogonal, we must have $\mathbf{v}^\top \mathbf{w} = 0$ for all $\mathbf{v} \in V$ and $\mathbf{w} \in W$. Now assume for a contradiction that there exists a vector $\mathbf{u} \in V \cap W$ with $\mathbf{u} \neq \mathbf{0}$. From $\mathbf{u} \in V \cap W$ we get that $\mathbf{u}$ belongs to both V and W and hence we must have $\mathbf{u}^\top \mathbf{u} = 0$. On the other hand, we get $0 < \|\mathbf{u}\|^2 = \mathbf{u}^\top \mathbf{u}$ from $\mathbf{u} \neq \mathbf{0}$. In conclusion, we have shown

$$0 < \mathbf{u}^\top \mathbf{u} = 0,$$

a contradiction.

b) First of all, we observe that $V^\perp$ is not empty since $\mathbf{0} \cdot \mathbf{v} = 0$ for all $\mathbf{v} \in V$ and hence $\mathbf{0} \in V^\perp$. Thus, it remains to show that $V^\perp$ is closed under vector addition and scalar multiplication.

Let $\mathbf{u}, \mathbf{w} \in V^\perp$ and $c \in \mathbb{R}$ be arbitrary. We want to prove $\mathbf{u} + \mathbf{w} \in V^\perp$ and $c\mathbf{w} \in V^\perp$. Let $\mathbf{v} \in V$ be arbitrary. We first calculate

$$(\mathbf{u} + \mathbf{w}) \cdot \mathbf{v} = \mathbf{u} \cdot \mathbf{v} + \mathbf{w} \cdot \mathbf{v} = 0 + 0 = 0$$

by using $\mathbf{u} \in V^\perp$ and $\mathbf{w} \in V^\perp$. Similarly, we also get

$$(c\mathbf{w}) \cdot \mathbf{v} = c(\mathbf{w} \cdot \mathbf{v}) = c0 = 0$$

by using $\mathbf{w} \in V^\perp$. Since this works for any vector $\mathbf{v} \in V$, we conclude that the vectors $\mathbf{u} + \mathbf{w}$ and $c\mathbf{w}$ are indeed orthogonal to all vectors from V and hence belong to $V^\perp$, i.e. $\mathbf{u} + \mathbf{w} \in V^\perp$ and $c\mathbf{w} \in V^\perp$.

We have shown that $V^\perp$ is non-empty and closed under vector addition and scalar multiplication. Hence, we conclude that it is a subspace of $\mathbb{R}^n$.

2. Unless stated otherwise, references to the lecture notes refer to the notes of the second part of the course.

a) In the proof of Proposition 4.2.4 it was shown that A and $A^\top A$ have the same nullspace. From this, we concluded that A has independent columns if and only if $A^\top A$ is invertible. But in fact, having $\mathbf{N}(A) = \mathbf{N}(A^\top A)$ also implies $\mathrm{rank}(A) = \mathrm{rank}(A^\top A)$.

b) We know from the lecture that the rank of a matrix does not change if we transpose it. In particular, we have $\mathrm{rank}(A^\top) = \mathrm{rank}(A)$ and hence $\mathrm{rank}(A^\top) = \mathrm{rank}(A^\top A)$ by using a).

For the second part, observe that the statement from a) holds for any matrix. In particular, it also holds for the matrix $B := A^\top$, i.e. we have $\mathrm{rank}(B) = \mathrm{rank}(B^\top B)$. Plugging in $B = A^\top$ yields $\mathrm{rank}(A^\top) = \mathrm{rank}(AA^\top)$. In conclusion, we have

$$\mathrm{rank}(A) = \mathrm{rank}(A^\top) = \mathrm{rank}(A^\top A) = \mathrm{rank}(AA^\top).$$

c) We will use the following three facts:

1. For any matrix $B \in \mathbb{R}^{k \times \ell}$, we have $\mathbf{R}(B) = \mathbf{C}(B^\top) = \mathbf{N}(B)^\perp$. This was discussed in Section 4.1 of the lecture notes (first part). In particular, we can use this with $B := A$ to get $\mathbf{C}(A^\top) = \mathbf{N}(A)^\perp$ and for $B := A^\top A$ to get $\mathbf{C}((A^\top A)^\top) = \mathbf{N}(A^\top A)^\perp$.

2. The matrix $A^\top A$ is symmetric. This is stated in Corollary 4.2.5 for the case where A has independent columns. But it holds in general since $(A^\top A)^\top = A^\top (A^\top)^\top = A^\top A$.

3. The matrices A and $A^\top A$ have the same nullspace, i.e. $\mathbf{N}(A) = \mathbf{N}(A^\top A)$. This was shown in the proof of Proposition 4.2.4 and we also used it in subtask a).

We now prove $\mathbf{C}(A^\top) = \mathbf{C}(A^\top A)$ by combining these three facts in

$$\mathbf{C}(A^\top) \stackrel{1.}{=} \mathbf{N}(A)^\perp \stackrel{3.}{=} \mathbf{N}(A^\top A)^\perp \stackrel{1.}{=} \mathbf{C}((A^\top A)^\top) \stackrel{2.}{=} \mathbf{C}(A^\top A).$$

3. **a)** Observe (or recall from the lecture) that

$$P^2 = A(A^\top A)^{-1}A^\top A(A^\top A)^{-1}A^\top = A(A^\top A)^{-1}IA^\top = A(A^\top A)^{-1}A^\top = P.$$

Using this, we similarly get

$$(I - P)^2 = I^2 - 2IP + P^2 = I - 2P + P = I - P.$$

As for the projection matrix P, the interpretation here is that applying $(I - P)$ twice does not do more than applying $(I - P)$ once.

b) Since P projects vectors onto $\mathbf{C}(A)$ and $\mathbf{w}$ is already in $\mathbf{C}(A)$, we get that $P\mathbf{w} = \mathbf{w}$ (i.e. nothing happens to $\mathbf{w}$ if we apply P). We can derive this more formally from Theorem 4.2.6, which tells us that

$$P\mathbf{w} = \arg\min_{\mathbf{p}\in\mathbf{C}(A)} \| \mathbf{w} - \mathbf{p} \|.$$

We also have

$$\arg\min_{\mathbf{p}\in\mathbf{C}(A)} \| \mathbf{w} - \mathbf{p} \| = \mathbf{w}$$

by $\mathbf{w} \in \mathbf{C}(A)$, and putting those two things together we get $P\mathbf{w} = \mathbf{w}$, as desired. From this we also get $(I - P)\mathbf{w} = \mathbf{w} - P\mathbf{w} = \mathbf{w} - \mathbf{w} = \mathbf{0}$. In other words, we have $\mathbf{w} \in \mathbf{N}(I - P)$.

c) Since $\mathbf{v}$ is orthogonal to all columns of A, we get $A^\top \mathbf{v} = \mathbf{0}$. Hence, we also get $P\mathbf{v} = A(A^\top A)^{-1}A^\top \mathbf{v} = \mathbf{0}$ and therefore $(I - P)\mathbf{v} = \mathbf{v}$. In words, applying $(I - P)$ to $\mathbf{v}$ does nothing.

d) As is usual, there are many different ways to solve this. We present one way that heavily relies on Section 4.1 from the first part of the course.

We need the following four observations:

1. In subtask b), we essentially proved $\mathbf{C}(A) \subseteq \mathbf{C}(P)$ and hence $\dim(\mathbf{C}(A)) \leq \dim(\mathbf{C}(P))$.
2. Similarly, our arguments from c) imply $\mathbf{C}(A)^\perp \subseteq \mathbf{N}(P)$ and therefore $\dim(\mathbf{C}(A)^\perp) \leq \dim(\mathbf{N}(P))$.
3. From Section 4.1 (first part) we know that $\dim(\mathbf{C}(A)) + \dim(\mathbf{C}(A)^\perp) = m$.
4. We also have $\dim(\mathbf{C}(P)) + \dim(\mathbf{N}(P)) \leq m$. This follows from $\mathbf{C}(P) \cap \mathbf{N}(P) = \{\mathbf{0}\}$ by Section 4.1 (first part). To see why $\mathbf{C}(P) \cap \mathbf{N}(P) = \{\mathbf{0}\}$ holds, assume for a contradiction that there exists $\mathbf{v} \in \mathbf{C}(P) \cap \mathbf{N}(P)$ with $\mathbf{v} \neq \mathbf{0}$. By $\mathbf{v} \in \mathbf{C}(P)$, there exists $\mathbf{w} \in \mathbb{R}^m$ with $P\mathbf{w} = \mathbf{v}$. By $\mathbf{v} \in \mathbf{N}(P)$ we also get $P(P\mathbf{w}) = P\mathbf{v} = \mathbf{0}$. But since P is a projection matrix, we must have $P^2\mathbf{w} = P\mathbf{w}$ which is in contradiction with $P\mathbf{w} = \mathbf{v} \neq \mathbf{0} = P^2\mathbf{w}$.

Finally, we plug together all the pieces to obtain

$$\dim(\mathbf{C}(A)) \stackrel{1.}{\leq} \dim(\mathbf{C}(P)) \stackrel{4.}{\leq} m - \dim(\mathbf{N}(P)) \stackrel{2.}{\leq} m - \dim(\mathbf{C}(A)^\perp) \stackrel{3.}{=} \dim(\mathbf{C}(A))$$

which implies $\mathrm{rank}(P) = \dim(\mathbf{C}(P)) = \dim(\mathbf{C}(A))$.

e) Before we turn our attention to $I - P$, observe that from subtask d), we can also deduce $\mathbf{C}(A) = \mathbf{C}(P)$ since we have $\mathbf{C}(A) \subseteq \mathbf{C}(P)$ and $\dim(\mathbf{C}(A)) = \dim(\mathbf{C}(P))$. Analogously, we get $\mathbf{C}(A)^\perp = \mathbf{N}(P)$.

We claim that $\mathbf{C}(I - P) = \mathbf{C}(A)^\perp$ and hence $\operatorname{rank}(I - P) = \dim(\mathbf{C}(I - P)) = \dim(\mathbf{C}(A)^\perp)$. From subtask c) we already know that $\mathbf{C}(A)^\perp \subseteq \mathbf{C}(I - P)$. Hence, it remains to prove $\mathbf{C}(I - P) \subseteq \mathbf{C}(A)^\perp$. To this end, let $\mathbf{v} \in \mathbf{C}(I - P)$ be arbitrary. Then there exists $\mathbf{w} \in \mathbb{R}^m$ with $(I - P)\mathbf{w} = \mathbf{v}$. Hence, we get $(I - P)\mathbf{v} = (I - P)^2 \mathbf{w} = (I - P)\mathbf{w} = \mathbf{v}$ and therefore $P\mathbf{v} = \mathbf{0}$. This again implies $\mathbf{v} \in \mathbf{N}(P) = \mathbf{C}(A)^\perp$, as desired.

We conclude $\mathbf{C}(I - P) = \mathbf{C}(A)^\perp$ and hence $\operatorname{rank}(I - P) = \dim(\mathbf{C}(I - P)) = \dim(\mathbf{C}(A)^\perp)$.

f) Given a point $\mathbf{p} \in \mathbb{R}^m$, we know from the lecture that the projection $\operatorname{proj}_{\mathbf{C}(A)^\perp}(\mathbf{p})$ of $\mathbf{p}$ onto $\mathbf{C}(A)^\perp$ is given by the unique point $\mathbf{q} \in \mathbf{C}(A)^\perp$ satisfying $\mathbf{v}^\top(\mathbf{p} - \mathbf{q}) = 0$ for all $\mathbf{v} \in \mathbf{C}(A)^\perp$. Now let $\mathbf{v} \in \mathbf{C}(A)^\perp$ be arbitrary. Choosing $\mathbf{q} = (I - P)\mathbf{p}$ indeed yields

$$\mathbf{v}^\top(\mathbf{p} - \mathbf{q}) = \mathbf{v}^\top(P\mathbf{p}) = 0$$

since $(P\mathbf{p}) \in \mathbf{C}(P) = \mathbf{C}(A)$ and $\mathbf{v} \in \mathbf{C}(A)^\perp$.

4. **a)** For the first datapoint we get $x_1 = 1$ and $y_1 = 2$. Hence, we get the equation $a + b = 2$ from $f(x_1) = y_1$. We proceed analogously for all $i \in [5]$ and obtain the system of linear equations

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 5 \\ 6 \\ 8 \end{bmatrix}.$$

b) Since there are 5 equations and only 2 unknowns, we generally don't expect this system to have a solution. In particular, if our data comes from a real world process we cannot expect the data to exactly lie on a line (even if this is a good model for the real world process).

c) Let A be the system matrix from the linear system above and let $\mathbf{b}$ be the right-hand side. We then get

$$A^\top A = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 4 & 1 \\ 5 & 1 \end{bmatrix} = \begin{bmatrix} 55 & 15 \\ 15 & 5 \end{bmatrix}$$

and

$$A^\top \mathbf{b} = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 5 \\ 6 \\ 8 \end{bmatrix} = \begin{bmatrix} 87 \\ 24 \end{bmatrix}.$$

Hence, we get the normal equations

$$\begin{bmatrix} 55 & 15 \\ 15 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 87 \\ 24 \end{bmatrix}.$$

We can now solve this for a and b using Gauss-elimination or other techniques. One way to do it in this particular case, is to subtract the second row 3 times from the first row to get

$$\begin{bmatrix} 10 & 0 \\ 15 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 15 \\ 24 \end{bmatrix}.$$

We then conclude $a = \frac{3}{2}$ and $b = \frac{3}{10}$. In particular, we conclude that the line $f(x) = \frac{3}{2}x + \frac{3}{10}$ explains our data best.