Solution to Homework Assignment No. 3

1. (a) Since $\boldsymbol{v}^T\boldsymbol{0} = 0$, $\forall \boldsymbol{v} \in \mathcal{R}^3$, we have $S^\perp = \mathcal{R}^3$.

(b) Let $\boldsymbol{A}_1 = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$. We can have $S = \mathcal{C}(\boldsymbol{A}_1^T)$ and

$$S^\perp = \mathcal{C}(\boldsymbol{A}_1^T)^\perp = \mathcal{N}(\boldsymbol{A}_1) = \left\{ \boldsymbol{x} : \boldsymbol{x} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, x_2, x_3 \in \mathcal{R} \right\}.$$

(c) Let $\boldsymbol{A}_2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \end{bmatrix}$. We can have $S = \mathcal{C}(\boldsymbol{A}_2^T)$ and

$$S^\perp = \mathcal{C}(\boldsymbol{A}_2^T)^\perp = \mathcal{N}(\boldsymbol{A}_2) = \left\{ \boldsymbol{x} : \boldsymbol{x} = x_2 \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, x_2 \in \mathcal{R} \right\}.$$

Hence, $\left\{ \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} \right\}$ is a basis for $S^\perp$.

2. (a) We can have $\mathcal{C}(\boldsymbol{A}^T)^\perp = \mathcal{N}(\boldsymbol{A})$. Since the RRE form of $\boldsymbol{A}$ is

$$\boldsymbol{R}_A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

we can obtain that $\left\{(-1, 0, 1)^T\right\}$ is a basis for the orthogonal complement of the row space of $\boldsymbol{A}$.

(b) In class we knew that the projection matrix onto the column space of $\boldsymbol{A}$ is given by

$$\boldsymbol{P} = \boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T \tag{1}$$

where $\boldsymbol{A}$ is assumed to have full column rank so that $\left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}$ exists. Unfortunately, the matrix

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

does not have full column rank, and hence we cannot apply the formula (1) directly. Yet from $\boldsymbol{R}_A$, we can find that a basis for $\mathcal{C}(\boldsymbol{A})$ can be given by $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$. Let $\hat{\boldsymbol{A}} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$. Then we can have

$$\boldsymbol{P}_C = \hat{\boldsymbol{A}}\left(\hat{\boldsymbol{A}}^T\hat{\boldsymbol{A}}\right)^{-1}\hat{\boldsymbol{A}}^T = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{I}.$$

(c) The projection matrix $\boldsymbol{P}_R$ onto the row space of $\boldsymbol{A}$ can be obtained by replacing $\boldsymbol{A}$ in (1) with $\boldsymbol{A}^T$. Hence we can have

$$\boldsymbol{P}_R = \boldsymbol{A}^T \left(\boldsymbol{A}\boldsymbol{A}^T\right)^{-1} \boldsymbol{A} = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}.$$

(d) From (c), we can have

$$\boldsymbol{x}_r = \boldsymbol{P}_R\, \boldsymbol{x}^T = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$$

and

$$\boldsymbol{x}_n = \boldsymbol{x} - \boldsymbol{x}_r = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}.$$

(e) We can have

$$\left[\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 1 & 0 & 1 & 3 \end{array}\right] \quad \Longrightarrow \quad \left[\begin{array}{ccc|c} 1 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 \end{array}\right].$$

A particular solution $\boldsymbol{x}_p$ to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ can be given by

$$\boldsymbol{x}_p = \begin{bmatrix} 3 \\ -1 \\ 0 \end{bmatrix}.$$

Hence

$$\boldsymbol{x}_r = \boldsymbol{P}_R\, \boldsymbol{x}_p = \begin{bmatrix} 3/2 \\ -1 \\ 3/2 \end{bmatrix}.$$

3. (a) We can have

$$\begin{cases} C + D + E = 3 \\ C + 3E = 6 \\ C + 2D + E = 5 \\ C = 0 \end{cases} \quad \Longrightarrow \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

where

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 3 \\ 1 & 2 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{x} = \begin{bmatrix} C \\ D \\ E \end{bmatrix}, \quad \boldsymbol{b} = \begin{bmatrix} 3 \\ 6 \\ 5 \\ 0 \end{bmatrix}.$$

The best least squares fit can be derived by solving $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$. Hence one can obtain

$$\hat{\boldsymbol{x}} = \begin{bmatrix} C \\ D \\ E \end{bmatrix} = \begin{bmatrix} -3/25 \\ 73/50 \\ 101/50 \end{bmatrix}.$$

(b) We can have

$$\begin{cases} C &= y_1 \\ C &= y_2 \\ &\vdots \\ C &= y_m \end{cases} \quad \Longrightarrow \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

where

$$\boldsymbol{A} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \boldsymbol{x} = C, \quad \boldsymbol{b} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}.$$

The best least squares fit can be found by solving $\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}$. Hence, we can obtain

$$mC = y_1 + y_2 + \cdots + y_m \quad \Longrightarrow \quad C = \frac{y_1 + y_2 + \cdots + y_m}{m}.$$

4. (a) Let $\boldsymbol{y} \triangleq \boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{z} \triangleq \boldsymbol{A}^T\boldsymbol{y}$. Since

$$\frac{\partial}{\partial x_k}\|\boldsymbol{A}\boldsymbol{x}\|^2 = \frac{\partial}{\partial x_k}\|\boldsymbol{y}\|^2 = \frac{\partial}{\partial x_k}\sum_{i=1}^{m} y_i^2 = \sum_{i=1}^{m} 2y_i\frac{\partial y_i}{\partial x_k}$$

and

$$\frac{\partial y_i}{\partial x_k} = \frac{\partial}{\partial x_k}\sum_{j=1}^{n} A_{ij}x_j = A_{ik} = A_{ki}^T$$

we have

$$\frac{\partial}{\partial x_k}\|\boldsymbol{A}\boldsymbol{x}\|^2 = 2\sum_{i=1}^{m} A_{ki}^T y_i = 2z_k.$$

Collecting the partial derivatives yields

$$\begin{bmatrix} \frac{\partial}{\partial x_1}\|\boldsymbol{A}\boldsymbol{x}\|^2 \\ \vdots \\ \frac{\partial}{\partial x_n}\|\boldsymbol{A}\boldsymbol{x}\|^2 \end{bmatrix} = \begin{bmatrix} 2z_1 \\ \vdots \\ 2z_n \end{bmatrix} = 2\boldsymbol{z} = 2\boldsymbol{A}^T\boldsymbol{y} = 2\boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x}.$$

(b) Let $\boldsymbol{w} \triangleq \boldsymbol{A}^T\boldsymbol{b}$. Then we have

$$\frac{\partial}{\partial x_k}\left(2\boldsymbol{b}^T\boldsymbol{A}\boldsymbol{x}\right) = \frac{\partial}{\partial x_k}\left(2\sum_{i=1}^{m} b_i y_i\right) = 2\sum_{i=1}^{m} A_{ki}^T b_i = 2w_k.$$

Collecting the partial derivatives yields

$$\begin{bmatrix} \frac{\partial}{\partial x_1}\left(2\boldsymbol{b}^T\boldsymbol{A}\boldsymbol{x}\right) \\ \vdots \\ \frac{\partial}{\partial x_n}\left(2\boldsymbol{b}^T\boldsymbol{A}\boldsymbol{x}\right) \end{bmatrix} = \begin{bmatrix} 2w_1 \\ \vdots \\ 2w_n \end{bmatrix} = 2\boldsymbol{w} = 2\boldsymbol{A}^T\boldsymbol{b}.$$

(c) Finally, we can have

$$\frac{\partial}{\partial x_k}\|\boldsymbol{Ax}-\boldsymbol{b}\|^2=\frac{\partial}{\partial x_k}\|\boldsymbol{Ax}\|^2-\frac{\partial}{\partial x_k}\left(2\boldsymbol{b}^T\boldsymbol{Ax}\right)=2z_k-2w_k.$$

Collecting the partial derivatives yields

$$\begin{bmatrix}\frac{\partial}{\partial x_1}\|\boldsymbol{Ax}-\boldsymbol{b}\|^2\\ \vdots\\ \frac{\partial}{\partial x_n}\|\boldsymbol{Ax}-\boldsymbol{b}\|^2\end{bmatrix}=\begin{bmatrix}2z_1-2w_1\\ \vdots\\ 2z_n-2w_n\end{bmatrix}=2(\boldsymbol{z}-\boldsymbol{w})=2(\boldsymbol{A}^T\boldsymbol{Ax}-\boldsymbol{A}^T\boldsymbol{b}).$$

Hence, the partial derivatives of $\|\boldsymbol{Ax}-\boldsymbol{b}\|^2$ are zero when $\boldsymbol{A}^T\boldsymbol{Ax}=\boldsymbol{A}^T\boldsymbol{b}$.

5. (a) We have

$$\begin{aligned}\boldsymbol{Q}^T\boldsymbol{Q}&=\left(\boldsymbol{I}-2\boldsymbol{uu}^T\right)^T\left(\boldsymbol{I}-2\boldsymbol{uu}^T\right)=\left(\boldsymbol{I}-2\boldsymbol{uu}^T\right)\left(\boldsymbol{I}-2\boldsymbol{uu}^T\right)\\&=\boldsymbol{I}-4\boldsymbol{uu}^T+4\boldsymbol{uu}^T\boldsymbol{uu}^T.\end{aligned}$$

Since $\boldsymbol{u}$ is a unit vector, we have $\boldsymbol{u}^T\boldsymbol{u}=1$. And hence

$$\boldsymbol{Q}^T\boldsymbol{Q}=\boldsymbol{I}-4\boldsymbol{uu}^T+4\boldsymbol{uu}^T\boldsymbol{uu}^T=\boldsymbol{I}-4\boldsymbol{uu}^T+4\boldsymbol{uu}^T=\boldsymbol{I}.$$

As a result, $\boldsymbol{Q}$ is an orthogonal matrix.

(b) We can have

$$\boldsymbol{Qu}=\left(\boldsymbol{I}-2\boldsymbol{uu}^T\right)\boldsymbol{u}=\boldsymbol{u}-2\boldsymbol{uu}^T\boldsymbol{u}=\boldsymbol{u}-2\boldsymbol{u}=-\boldsymbol{u}.$$

(c) We have

$$\boldsymbol{Qv}=\left(\boldsymbol{I}-2\boldsymbol{uu}^T\right)\boldsymbol{v}=\boldsymbol{v}-2\boldsymbol{uu}^T\boldsymbol{v}.$$

Since $\boldsymbol{v}$ and $\boldsymbol{u}$ are orthogonal, $\boldsymbol{u}^T\boldsymbol{v}=0$. Hence

$$\boldsymbol{Qv}=\boldsymbol{v}-2\boldsymbol{uu}^T\boldsymbol{v}=\boldsymbol{v}.$$

6. (a) Let $\boldsymbol{A}=[\,\boldsymbol{a}_1\ \ \boldsymbol{a}_2\ \ \boldsymbol{a}_3\,]$, where

$$\boldsymbol{a}_1=\begin{bmatrix}0\\0\\1\end{bmatrix},\quad \boldsymbol{a}_2=\begin{bmatrix}0\\1\\1\end{bmatrix},\quad \boldsymbol{a}_3=\begin{bmatrix}1\\1\\1\end{bmatrix}.$$

Applying the Gram-Schmidt process, we can have

$$\boldsymbol{A}_1=\boldsymbol{a}_1=\begin{bmatrix}0\\0\\1\end{bmatrix}\quad\Longrightarrow\quad \boldsymbol{q}_1=\frac{\boldsymbol{A}_1}{\|\boldsymbol{A}_1\|}=\begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$\boldsymbol{A}_2=\boldsymbol{a}_2-\left(\boldsymbol{q}_1^T\boldsymbol{a}_2\right)\boldsymbol{q}_1=\begin{bmatrix}0\\1\\0\end{bmatrix}\quad\Longrightarrow\quad \boldsymbol{q}_2=\frac{\boldsymbol{A}_2}{\|\boldsymbol{A}_2\|}=\begin{bmatrix}0\\1\\0\end{bmatrix}$$

$$\boldsymbol{A}_3=\boldsymbol{a}_3-\left(\boldsymbol{q}_1^T\boldsymbol{a}_3\right)\boldsymbol{q}_1-\left(\boldsymbol{q}_2^T\boldsymbol{a}_3\right)\boldsymbol{q}_2=\begin{bmatrix}1\\0\\0\end{bmatrix}\quad\Longrightarrow\quad \boldsymbol{q}_3=\frac{\boldsymbol{A}_3}{\|\boldsymbol{A}_3\|}=\begin{bmatrix}1\\0\\0\end{bmatrix}.$$

Hence, $\{\boldsymbol{q}_1,\boldsymbol{q}_2,\boldsymbol{q}_3\}$ forms an orthonormal basis for the column space of $\boldsymbol{A}$.

(b) From (a), we can have

$$
\begin{aligned}
\boldsymbol{A} &= \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \boldsymbol{a}_3 \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{q}_1 & \boldsymbol{q}_2 & \boldsymbol{q}_3 \end{bmatrix} \begin{bmatrix} \boldsymbol{q}_1^T\boldsymbol{a}_1 & \boldsymbol{q}_1^T\boldsymbol{a}_2 & \boldsymbol{q}_1^T\boldsymbol{a}_3 \\ 0 & \boldsymbol{q}_2^T\boldsymbol{a}_2 & \boldsymbol{q}_2^T\boldsymbol{a}_3 \\ 0 & 0 & \boldsymbol{q}_3^T\boldsymbol{a}_3 \end{bmatrix} \\
&= \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \boldsymbol{QR}
\end{aligned}
$$

where

$$
\boldsymbol{Q} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad \boldsymbol{R} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}.
$$

7. (a) Let $f_1(x) = 1$, $f_2(x) = x$, $f_3(x) = x^2$. Applying the Gram-Schmidt process, we can have

$$
\begin{aligned}
&F_1(x) = f_1(x) = 1 \implies q_1(x) = \frac{F_1(x)}{\|F_1(x)\|} = \frac{\sqrt{2}}{2} \\
&F_2(x) = f_2(x) - \langle q_1(x), f_2(x)\rangle\, q_1(x) = x \implies q_2(x) = \frac{F_2(x)}{\|F_2(x)\|} = \frac{\sqrt{6}}{2}x \\
&F_3(x) = f_3(x) - \langle q_1(x), f_3(x)\rangle\, q_1(x) - \langle q_2(x), f_3(x)\rangle\, q_2(x) = x^2 - \frac{1}{3} \\
&\qquad \implies q_3(x) = \frac{F_3(x)}{\|F_3(x)\|} = \frac{3\sqrt{10}}{4}\left(x^2 - \frac{1}{3}\right).
\end{aligned}
$$

Hence, $\{q_1(x), q_2(x), q_3(x)\}$ forms an orthonormal basis for the subspace spanned by 1, x, and x^2.

(b) The best least squares approximation to x^3 by $C+Dx+Ex^2$ is the projection of x^3 onto the subspace spanned by 1, x, and x^2. In (a), we have already derived an orthonormal basis for this subspace. Since

$$
\begin{aligned}
\langle q_1, x^3\rangle &= 0 \\
\langle q_2, x^3\rangle &= \frac{\sqrt{6}}{5} \\
\langle q_3, x^3\rangle &= 0
\end{aligned}
$$

the best least squares approximation to x^3 by $C + Dx + Ex^2$ is

$$
\langle q_1, x^3\rangle q_1 + \langle q_2, x^3\rangle q_2 + \langle q_3, x^3\rangle q_3 = \frac{3}{5}x.
$$