Solution to Homework Assignment No. 3

1. First, we apply elimination to transform $\boldsymbol{A}$ into the reduced row echelon (RRE) form:

$$\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 4 & 6 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix} \quad \Longrightarrow \quad \boldsymbol{R} = \begin{bmatrix} 0 & 1 & 2 & 0 & -2 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We can have $\boldsymbol{R} = \boldsymbol{E}\boldsymbol{A}$, where

$$\boldsymbol{E} = \begin{bmatrix} 1 & 0 & -3 \\ -1 & 1 & 0 \\ 1 & -1 & 1 \end{bmatrix}.$$

A basis for the row space is given by

$$\beta_{\text{row}} = \left\{ \begin{bmatrix} 0 \\ 1 \\ 2 \\ 0 \\ -2 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 2 \end{bmatrix} \right\}.$$

Since columns 2 and 4 are the pivot columns of $\boldsymbol{R}$, we know in class that a basis for the column space can be formed by columns 2 and 4 of $\boldsymbol{A}$, i.e.,

$$\beta_{\text{column}} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 4 \\ 1 \end{bmatrix} \right\}.$$

On the other hand, the vectors in the nullspace satisfy

$$\begin{bmatrix} 0 & 1 & 2 & 0 & -2 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Since x_2, x_4 are pivot variables and x_1, x_3, x_5 are free variables, a basis for the nullspace can be given by the three special solutions:

$$\beta_{\text{null}} = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \\ 0 \\ -2 \\ 1 \end{bmatrix} \right\}.$$

Finally, since the last row of $\boldsymbol{R}$ is a zero row, a basis for the left nullspace can be given by the last row of $\boldsymbol{E}$:

$$\beta_{\text{left}} = \left\{ \begin{bmatrix} 1 \\ -1 \\ 1 \end{bmatrix} \right\}.$$

2. (a) Let $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ be $n \times n$ matrices with $\boldsymbol{a}_k$, $\boldsymbol{b}_k$, and $\boldsymbol{c}_k$ denoting the kth row of $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$, respectively. Now since $\boldsymbol{C} = \boldsymbol{AB}$, we have

$$\boldsymbol{c}_i = \boldsymbol{a}_i \boldsymbol{B} = \sum_{j=1}^{n} a_{ij} \boldsymbol{b}_j \quad \text{for any } 1 \le i \le n$$

which shows that the rows of $\boldsymbol{C}$ are linear combinations of the rows of $\boldsymbol{B}$. Therefore, it follows that

$$\boldsymbol{c}_i^T \in \mathcal{C}(\boldsymbol{B}^T) \quad \text{for any } 1 \le i \le n. \tag{1}$$

On the other hand, the rank of $\boldsymbol{C}$ is the maximum number of linearly independent rows in $\boldsymbol{C}$, which of course cannot exceed the dimension of $\mathcal{C}(\boldsymbol{B}^T)$ because of (1). As a result, we have

$$\text{rank}(\boldsymbol{C}) \le \dim(\mathcal{C}(\boldsymbol{B}^T)) = \text{rank}(\boldsymbol{B}).$$

(b) From $\boldsymbol{C} = \boldsymbol{AB}$ we may obtain $\boldsymbol{C}^T = \boldsymbol{B}^T \boldsymbol{A}^T$ by taking transpose on both sides. It now follows from part (a) that

$$\text{rank}(\boldsymbol{C}^T) \le \text{rank}(\boldsymbol{A}^T).$$

Together with the fact that

$$\text{rank}(\boldsymbol{C}) = \text{rank}(\boldsymbol{C}^T) \quad \text{and} \quad \text{rank}(\boldsymbol{A}^T) = \text{rank}(\boldsymbol{A})$$

we finally arrive at

$$\text{rank}(\boldsymbol{C}) = \text{rank}(\boldsymbol{C}^T) \le \text{rank}(\boldsymbol{A}^T) = \text{rank}(\boldsymbol{A}).$$

3. (a) Let

$$\boldsymbol{A} \triangleq \begin{bmatrix} 1 & 2 & 2 & 3 \\ 1 & 3 & 3 & 2 \end{bmatrix}$$

so that $S = \mathcal{C}(\boldsymbol{A}^T)$. In class we know that

$$S^\perp = \mathcal{C}(\boldsymbol{A}^T)^\perp = \mathcal{N}(\boldsymbol{A}).$$

Hence we solve the system of linear equations:

$$\begin{bmatrix} 1 & 2 & 2 & 3 \\ 1 & 3 & 3 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which leads to two vectors $(0, -1, 1, 0)$ and $(-5, 1, 0, 1)$ spanning $S^\perp$.

(b) Let

$$\boldsymbol{B} \triangleq \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}$$

so that $P = \mathcal{N}(\boldsymbol{B})$. We also learn in class that

$$P^\perp = \mathcal{N}(\boldsymbol{B})^\perp = \mathcal{C}(\boldsymbol{B}^T).$$

Since $\boldsymbol{B}$ contains a single row, we know that $(1, 1, 1, 1)$ is a basis for $P^\perp$.

4. Applying elimination to the system of linear equations

$$\begin{bmatrix} 1 & 0 & 2 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

leads to a basis for $\mathcal{N}(\boldsymbol{A})$ given by $\boldsymbol{w} = (-2, -2, 1)^T$. The orthogonality between $\boldsymbol{w}$ and $\mathcal{C}(\boldsymbol{A}^T)$ can be verified by

$$\begin{bmatrix} 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix} = 0 \quad \text{and} \quad \begin{bmatrix} 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix} = 0.$$

To split $\boldsymbol{x} = (3, 3, 3)^T$ into $\boldsymbol{x}_r + \boldsymbol{x}_n$, we first project $\boldsymbol{x}$ onto $\mathcal{N}(\boldsymbol{A})$ so that

$$\boldsymbol{x}_n = \frac{\boldsymbol{w}^T\boldsymbol{x}}{\boldsymbol{w}^T\boldsymbol{w}}\boldsymbol{w} = -\boldsymbol{w} = \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix}.$$

Finally, it follows that

$$\boldsymbol{x}_r = \boldsymbol{x} - \boldsymbol{x}_n = \begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 4 \end{bmatrix}.$$

5. (a) In class we know the projection of $\boldsymbol{b}$ onto the column space of $\boldsymbol{A}$ is given by

$$\boldsymbol{p} = \boldsymbol{A}\hat{\boldsymbol{x}} \tag{2}$$

where $\hat{\boldsymbol{x}}$ is the solution to

$$\boldsymbol{A}^T\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^T\boldsymbol{b}. \tag{3}$$

In our case, (3) is explicitly given by

$$\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 2 \end{bmatrix}$$

and hence

$$\hat{\boldsymbol{x}} = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}.$$

It now follows from (2) that

$$\boldsymbol{p} = \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ -1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix}$$

and the error is thus

$$\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix}.$$

The orthogonality between $\boldsymbol{e}$ and the columns of $\boldsymbol{A}$ can be verified by

$$\begin{bmatrix} 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix} = 0 \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix} = 0.$$

(b) Applying (3) and (2) we obtain

$$\hat{\boldsymbol{x}} = \begin{bmatrix} -2 \\ 6 \end{bmatrix} \quad \text{and} \quad \boldsymbol{p} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -2 \\ 6 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ 4 \end{bmatrix}.$$

The error is now given by

$$\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p} = \begin{bmatrix} 4 \\ 6 \\ 4 \end{bmatrix} - \begin{bmatrix} 4 \\ 6 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

Obviously $\boldsymbol{e}$ is orthogonal to the columns of $\boldsymbol{A}$. Note that actually $\boldsymbol{b} \in \mathcal{C}(\boldsymbol{A})$ because

$$\begin{bmatrix} 4 \\ 6 \\ 4 \end{bmatrix} = -2 \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} + 6 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Hence projecting $\boldsymbol{b}$ onto $\mathcal{C}(\boldsymbol{A})$ results in $\boldsymbol{b}$ itself.

6. (a) In class we know the projection matrix projecting a vector onto the column space of $\boldsymbol{A}$ is given by

$$\boldsymbol{P} = \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T \tag{4}$$

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

$$\boldsymbol{A} = \begin{bmatrix} 2 & 4 & 4 \\ 5 & 10 & 10 \end{bmatrix}$$

does not have full column rank because its columns are linearly dependent. As a result we cannot apply (4) directly. However, a closer look at $\boldsymbol{A}$ reveals that its column space is actually spanned by a single vector, say

$$\boldsymbol{v}_C = \begin{bmatrix} 2 \\ 5 \end{bmatrix}.$$

Since $\boldsymbol{v}_C$ has full column rank, we may apply (4) and obtain

$$\boldsymbol{P}_C = \begin{bmatrix} 2 \\ 5 \end{bmatrix} \cdot \frac{1}{29} \cdot \begin{bmatrix} 2 & 5 \end{bmatrix} = \begin{bmatrix} 4/29 & 10/29 \\ 10/29 & 25/29 \end{bmatrix}.$$

(b) Let $\boldsymbol{B} \triangleq \boldsymbol{A}^T = \begin{bmatrix} 2 & 5 \\ 4 & 10 \\ 4 & 10 \end{bmatrix}$ so that $\mathcal{C}(\boldsymbol{A}^T) = \mathcal{C}(\boldsymbol{B})$. Since the column space of $\boldsymbol{B}$ is spanned by a single vector $\boldsymbol{v}_R = \begin{bmatrix} 2 \\ 4 \\ 4 \end{bmatrix}$, we may apply (4) to obtain

$$\boldsymbol{P}_R = \begin{bmatrix} 2 \\ 4 \\ 4 \end{bmatrix} \cdot \frac{1}{36} \cdot \begin{bmatrix} 2 & 4 & 4 \end{bmatrix} = \begin{bmatrix} 1/9 & 2/9 & 2/9 \\ 2/9 & 4/9 & 4/9 \\ 2/9 & 4/9 & 4/9 \end{bmatrix}.$$

After some calculations we discover that

$$\boldsymbol{P}_C\boldsymbol{A}\boldsymbol{P}_R = \begin{bmatrix} 2 & 4 & 4 \\ 5 & 10 & 10 \end{bmatrix} = \boldsymbol{A}.$$

This result follows from the facts that

$$\boldsymbol{P}_C\boldsymbol{A} = \boldsymbol{A} \quad \text{and} \quad \boldsymbol{A}\boldsymbol{P}_R = \boldsymbol{A}.$$

To explain why, we let $\boldsymbol{x}$ be a vector. Since $\boldsymbol{A}\boldsymbol{x} \in \mathcal{C}(\boldsymbol{A})$ and $\boldsymbol{A}^T\boldsymbol{x} \in \mathcal{C}(\boldsymbol{A}^T)$, it follows that

$$(\boldsymbol{P}_C\boldsymbol{A})\boldsymbol{x} = \boldsymbol{P}_C(\boldsymbol{A}\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}$$

and

$$(\boldsymbol{P}_R^T\boldsymbol{A}^T)\boldsymbol{x} = \boldsymbol{P}_R^T(\boldsymbol{A}^T\boldsymbol{x}) = \boldsymbol{P}_R(\boldsymbol{A}^T\boldsymbol{x}) = \boldsymbol{A}^T\boldsymbol{x}.$$

Since $\boldsymbol{x}$ is arbitrary, we must have

$$\boldsymbol{P}_C\boldsymbol{A} = \boldsymbol{A} \quad \text{and} \quad \boldsymbol{P}_R^T\boldsymbol{A}^T = \boldsymbol{A}^T \text{ (or } \boldsymbol{A}\boldsymbol{P}_R = \boldsymbol{A}).$$

Using these two facts, we may obtain

$$\boldsymbol{P}_C\boldsymbol{A}\boldsymbol{P}_R = (\boldsymbol{P}_C\boldsymbol{A})\boldsymbol{P}_R = \boldsymbol{A}\boldsymbol{P}_R = \boldsymbol{A}.$$

7. (a) Let

$$\boldsymbol{A}_1 \triangleq \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix}, \quad \hat{\boldsymbol{x}}_1 \triangleq \begin{bmatrix} C_1 \\ D_1 \\ E_1 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{b}_1 \triangleq \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}.$$

In class we learn the choice of $\hat{\boldsymbol{x}}_1$ which minimizes $\|\boldsymbol{e}\|^2$ is given by solving

$$\boldsymbol{A}_1^T\boldsymbol{A}_1\hat{\boldsymbol{x}}_1 = \boldsymbol{A}_1^T\boldsymbol{b}_1$$

which yields

$$\begin{bmatrix} 4 & 8 & 26 \\ 8 & 26 & 92 \\ 26 & 92 & 338 \end{bmatrix} \begin{bmatrix} C_1 \\ D_1 \\ E_1 \end{bmatrix} = \begin{bmatrix} 36 \\ 112 \\ 400 \end{bmatrix}.$$

Therefore,

$$\hat{\boldsymbol{x}}_1 = \begin{bmatrix} C_1 \\ D_1 \\ E_1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4/3 \\ 2/3 \end{bmatrix}.$$

The closest parabola is hence $b = 2 + (4/3)t + (2/3)t^2$. The projection of $\boldsymbol{b}_1$ onto $\mathcal{C}(\boldsymbol{A}_1)$ is

$$\boldsymbol{p}_1 = \boldsymbol{A}_1\hat{\boldsymbol{x}}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \end{bmatrix} \begin{bmatrix} 2 \\ 4/3 \\ 2/3 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 12 \\ 18 \end{bmatrix}.$$

Finally, the error is given by

$$\boldsymbol{e}_1 = \boldsymbol{b}_1 - \boldsymbol{p}_1 = \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix} - \begin{bmatrix} 2 \\ 4 \\ 12 \\ 18 \end{bmatrix} = \begin{bmatrix} -2 \\ 4 \\ -4 \\ 2 \end{bmatrix}$$

with $\|\boldsymbol{e}_1\|^2 = (-2)^2 + 4^2 + (-4)^2 + 2^2 = 40$.

(b) Computations similar to part (a) can be carried out by letting

$$\boldsymbol{A}_2 \triangleq \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{bmatrix}, \quad \hat{\boldsymbol{x}}_2 \triangleq \begin{bmatrix} C_2 \\ D_2 \\ E_2 \\ F_2 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{b}_2 \triangleq \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}.$$

However, there is no need for such amount of computation in this problem! A simple check will confirm that the columns of $\boldsymbol{A}_2$ are linearly independent. Hence $\mathcal{C}(\boldsymbol{A}_2) = \mathcal{R}^4$. Now since $\boldsymbol{b}_2 \in \mathcal{R}^4 = \mathcal{C}(\boldsymbol{A}_2)$, we should be able to fit a curve without any error! The exact coefficients of the curve can be determined by solving

$$\boldsymbol{A}_2 \hat{\boldsymbol{x}}_2 = \boldsymbol{b}_2$$

or equivalently,

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 3 & 9 & 27 \\ 1 & 4 & 16 & 64 \end{bmatrix} \begin{bmatrix} C_2 \\ D_2 \\ E_2 \\ F_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 8 \\ 8 \\ 20 \end{bmatrix}.$$

Applying elimination yields

$$\hat{\boldsymbol{x}}_2 = \begin{bmatrix} C_2 \\ D_2 \\ E_2 \\ F_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 47/3 \\ -28/3 \\ 5/3 \end{bmatrix}.$$

Therefore, the closest cubic is given by $b = (47/3)t - (28/3)t^2 + (5/3)t^3$. The error in this case is of course

$$\boldsymbol{e}_2 = \boldsymbol{b}_2 - \boldsymbol{p}_2 = \boldsymbol{b}_2 - \boldsymbol{b}_2 = \boldsymbol{0} \quad \text{with} \quad \|\boldsymbol{e}_2\|^2 = 0.$$

8. (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

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

Collecting the partial derivatives yields

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

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

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

Collecting the partial derivatives yields

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

(c) Finally, we obtain

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

Collecting the partial derivatives yields

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

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