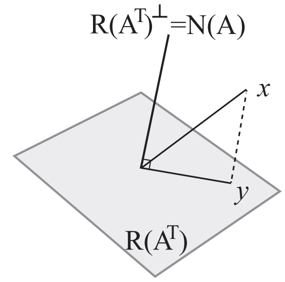
Section 5.3 Homework Solutions



§ **5.2**, (11) Given $A : \mathbb{R}^n \to \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{R}^n$, set $\mathbf{y} = \text{Proj}_{R(A^T)}(\mathbf{x})$.
Then $\mathbf{x} = \mathbf{y}+\mathbf{z}$ where $\mathbf{y} \in R(A^T)$ and $\mathbf{z} \in R(A^T)^\perp = N(A)$.
If $y = 0$ then $\mathbf{x} = \mathbf{z} \in N(A)$.
Otherwise $y \neq 0$ and

$$\mathbf{x}\cdot\mathbf{y} = (\mathbf{y}+\mathbf{z})\cdot\mathbf{x} = \mathbf{y}\cdot\mathbf{y} + \mathbf{z}\cdot\mathbf{y} = \|\mathbf{y}\|^2 + 0 \neq 0.$$

Quicker proof: Set $\mathbf{y} = A^TA\mathbf{x}$. Then $\mathbf{y} \in R(A^T)$ and
$\mathbf{x}\cdot\mathbf{y} = \mathbf{x}\cdot A^TA\mathbf{x} = A\mathbf{x}\cdot A\mathbf{x} = \|A\mathbf{x}\|^2 \neq 0$ unless $A\mathbf{x} = \mathbf{0}$.

§ **5.2**, (12) By Theorem 5.2.3, if S is a subspace of $\mathbb{R}^n$, every vector $\mathbf{x} \in \mathbb{R}^n$ can be uniquely written as $\mathbf{x} = \mathbf{y}+\mathbf{z}$ where $\mathbf{y} \in S$ and $\mathbf{z} \in S^\perp$.

(a) Take $S = N(A)$. This is a subspace of $\mathbb{R}^n$, and $S^\perp = N(A)^\perp = R(A^T)$ by Theorem 5.2.1. Therefore, every $\mathbf{x} \in \mathbb{R}^n$ can be uniquely written as $\mathbf{x} = \mathbf{y}+\mathbf{z}$ where $\mathbf{y} \in N(A)$ and $\mathbf{z} \in R(A^T)$.

§ **5.3**, (1a) Write the system as $A\mathbf{x} = \mathbf{b}$ where $A = \begin{pmatrix} 1 & 1 \\ 2 & -3 \\ 0 & 0 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$. Then $A^T\mathbf{b} = \begin{pmatrix} 1 & 2 & 0 \\ 1 & -3 & 0 \end{pmatrix}\begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$ and $A^TA = \begin{pmatrix} 1 & 2 & 0 \\ 1 & -3 & 0 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & -3 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 5 & -5 \\ -5 & 10 \end{pmatrix}$. The Least Squares solution is

$$\begin{pmatrix} x \\ y \end{pmatrix} = (A^TA)^{-1}A^T\mathbf{b} = \frac{1}{25}\begin{pmatrix} 10 & 5 \\ 5 & 5 \end{pmatrix}\begin{pmatrix} 5 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

(1b) Similarly, $A = \begin{pmatrix} -1 & 1 \\ 2 & 1 \\ 1 & -2 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} 10 \\ 5 \\ 20 \end{pmatrix}$. Then $A^T\mathbf{b} = \begin{pmatrix} -1 & 2 & 1 \\ 1 & 1 & -2 \end{pmatrix}\begin{pmatrix} 10 \\ 5 \\ 20 \end{pmatrix} = \begin{pmatrix} 20 \\ -25 \end{pmatrix}$, and $A^TA = \begin{pmatrix} 6 & -1 \\ -1 & 6 \end{pmatrix}$. The Least Squares solution is

$$\begin{pmatrix} x \\ y \end{pmatrix} = (A^TA)^{-1}A^T\mathbf{b} = \frac{1}{35}\begin{pmatrix} 6 & 1 \\ 1 & 6 \end{pmatrix}\begin{pmatrix} 20 \\ -25 \end{pmatrix} = \frac{1}{35}\begin{pmatrix} 95 \\ -130 \end{pmatrix} = \frac{1}{7}\begin{pmatrix} 19 \\ -26 \end{pmatrix}.$$

(3a) Here $A^TA = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 2 & 4 \\ -1 & -2 \end{pmatrix} = \begin{pmatrix} 6 & 12 \\ 12 & 24 \end{pmatrix}$. Then $\det A^TA = 0$, so A^TA is not invertible. In this situation, the Least Squares solution is not given by the formula $(A^TA)^{-1}A^T\mathbf{b}$, and is not unique. Instead, it is the set of all solutions of the equation $(A^TA)\mathbf{x} = A^T\mathbf{b}$ with $A^T\mathbf{b} = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 12 \end{pmatrix}$. Solving by row operations:

$$\left(\begin{array}{cc|c} 6 & 12 & 6 \\ 12 & 24 & 12 \end{array}\right) \approx \left(\begin{array}{cc|c} 1 & 2 & 1 \\ 1 & 2 & 1 \end{array}\right) \approx \left(\begin{array}{cc|c} 1 & 2 & 1 \\ 0 & 0 & 0 \end{array}\right).$$

The solution set is $\begin{pmatrix} x \\ y \end{pmatrix} = \left\{ \begin{pmatrix} 1-2\alpha \\ \alpha \end{pmatrix} \,\middle|\, \alpha \in \mathbb{R} \right\}$.

(3b) Using the web to calculate, one finds that $A^T\mathbf{b} = \begin{pmatrix} 6 \\ 14 \\ 26 \end{pmatrix}$ and $A^TA = \begin{pmatrix} 3 & 0 & 6 \\ 0 & 14 & 14 \\ 6 & 14 & 26 \end{pmatrix}$, and that this matrix is also singular. Solving $(A^TA)\mathbf{x} = A^T\mathbf{b}$ by row operations,

$$\left(\begin{array}{ccc|c} 3 & 0 & 6 & 6 \\ 0 & 14 & 14 & 14 \\ 0 & 14 & 26 & 26 \end{array}\right) \approx \left(\begin{array}{ccc|c} 1 & 0 & 2 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 14 & 14 & 14 \end{array}\right) \approx \left(\begin{array}{ccc|c} 1 & 0 & 2 & 2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right) \Rightarrow \begin{array}{l} x = 2-2\alpha \\ y = 1-\alpha \\ z = \alpha \end{array} \quad \text{for all } \alpha \in \mathbb{R}$$

(4a) As in Problem 9 below, the projection of $\mathbf{b}$ onto $R(A)$ is given by $P\mathbf{b} = A(A^TA)^{-1}A^T\mathbf{b}$ and this is $A\mathbf{x}_0$ where $\mathbf{x}_0$ is the solution. Using the calculations of Problem 3a above, this gives

$$p \;=\; P\mathbf{b} \;=\; A\mathbf{x}_0 \;=\; \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & -2 \end{pmatrix} \begin{pmatrix} 1-2\alpha \\ \alpha \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}.$$

One then checks that the the vector $\mathbf{b} - p = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 2 \end{pmatrix}$ is perpendicular to the two columns $\begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ 4 \\ -2 \end{pmatrix}$ of the matrix A.

(5) Write the best-fitting line as $y = ax + b$. Each of the data points $(-1,0), (0,1), (1,3)$ and $(2,9)$ gives a linear condition on a, b, and hence a linear system $A\mathbf{x} = \mathbf{b}$:

$$\begin{array}{ll} \text{thru } (-1,0): & -a+b=0 \\ \text{thru } (0,1): & b=1 \\ \text{thru } (1,3): & a+b=3 \\ \text{thru } (2,4): & 2a+b=9 \end{array} \qquad A = \begin{pmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 9 \end{pmatrix}.$$

Now calculate

$$A^T\mathbf{b} = \begin{pmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 3 \\ 9 \end{pmatrix} = \begin{pmatrix} 21 \\ 13 \end{pmatrix} \qquad A^TA = \begin{pmatrix} -1 & 0 & 1 & 2 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} -1 & 1 \\ 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 2 \\ 2 & 4 \end{pmatrix}$$

so the Least Square solution is

$$\begin{pmatrix} a \\ b \end{pmatrix} \;=\; (A^TA)^{-1}A^T\mathbf{b} \;=\; \frac{1}{20} \begin{pmatrix} 4 & -2 \\ -2 & 6 \end{pmatrix} \begin{pmatrix} 21 \\ 13 \end{pmatrix} \;=\; \frac{1}{10} \begin{pmatrix} 29 \\ 18 \end{pmatrix} \;=\; \begin{pmatrix} 2.9 \\ 1.8 \end{pmatrix}$$

Thus the best-fit line is $y = 2.9x + 1.8$.

(6) Write the best-fitting parabola as $y = ax^2 + bx + c$ and follow the steps of Problem 5 to get:

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 9 \end{pmatrix} \qquad A^T\mathbf{b} = \begin{pmatrix} 39 \\ 21 \\ 13 \end{pmatrix} \qquad A^TA = \begin{pmatrix} 18 & 8 & 6 \\ 8 & 6 & 2 \\ 6 & 2 & 4 \end{pmatrix}.$$

Using a web calculator, one finds the inverse of A^TA, and hence

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \;=\; \begin{pmatrix} 0.25 & -0.25 & -0.25 \\ -.25 & .450 & .15 \\ -.25 & .15 & .55 \end{pmatrix} \;=\; \begin{pmatrix} 1.25 \\ 1.65 \\ 0.55 \end{pmatrix}.$$

Thus the best-fit parabola is $y = 1.25x^2 + 1.65x + .55$.

(9) Set $P = A(A^TA)^{-1}A^T$. This is interpreted as a projection from $\mathbb{R}^m$ to the subspace $R(A)$.

(a) If $\mathbf{b} \in R(A)$, then $\mathbf{b} = A\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$. Then $P\mathbf{b} = A(A^TA)^{-1}A^T(A\mathbf{x}) = A(A^TA)^{-1}(A^TA)\mathbf{x} = A\mathbf{x} = \mathbf{b}$. This is compatible with the projection interpretation: a projection onto a subspace does not change vectors in that subspace.

(b) If $\mathbf{b} \in R(A)^\perp$, then note that $R(A)^\perp = N(A^T)$ by Theorem 5.2.1. Thus $\mathbf{b} \in N(A^T)$, i.e. $A^T\mathbf{b} = 0$. But then

$$P\mathbf{b} = A(A^TA)^{-1}(A^T\mathbf{b}) = A(A^TA)^{-1} \cdot 0 = 0.$$

Alternative proof: Compute the dot product of $P\mathbf{b}$ with itself, and use our lemma that $\mathbf{x} \cdot Ay = A^T\mathbf{x} \cdot \mathbf{y}$ for all $\mathbf{x}, \mathbf{y}$:

$$\begin{aligned}
\|P\mathbf{b}\|^2 \;=\; P\mathbf{b} \cdot P\mathbf{b} &= A(A^TA)^{-1}A^T\mathbf{b} \cdot A(A^TA)^{-1}A^T\mathbf{b} && \text{subsitution} \\
&= A^TA(A^TA)^{-1}A^T\mathbf{b} \cdot (A^TA)^{-1}A^T\mathbf{b} && \mathbf{x} \cdot Ay = A^T\mathbf{x} \cdot \mathbf{y} \\
&= A^T\mathbf{b} \cdot \mathbf{z} && \text{cancel and set } \mathbf{z} = (A^TA)^{-1}A^T\mathbf{b} \\
&= \mathbf{b} \cdot A\mathbf{z} && A^T\mathbf{x} \cdot y = \mathbf{x} \cdot A\mathbf{y} \\
&= 0 && \text{Assumption } \mathbf{b} \in R(A)^{\perp}.
\end{aligned}$$

This shows that $\|P\mathbf{b}\| = 0$, so $P\mathbf{b} = 0$.

(c) If $R(A)$ is a plane in $\mathbb{R}^3$, then P is the perpendicular projection onto $R(A)$, which takes $R(A)^{\perp}$ to the origin.

(11) For $P = A(A^TA)^{-1}A^T$, we have

(a) $P^2 = A(A^TA)^{-1}(A^TA)(A^TA)^{-1}A^T = A(A^TA)^{-1}(Id.)A^T = P$.

(b) Clearly $P^1 = P$. Proceed by induction: suppose that $P^k = P$ for some $k \geq 1$. Then

$$\begin{aligned}
P^{k+1} \;=\; P^k \cdot P &= P \cdot P && \text{by the induction hypothesis} \\
&= P && \text{by Part (a).}
\end{aligned}$$

By induction, we conclude that $P^k = P$ for all $k \geq 1$.

(c) The matrix P is symmetric if $P^T = P$. And in fact, using the hint and the fact that $(CD)^T = D^TC^T$, we have $[(A^TA)^{-1}]^T = [(A^TA)^T]^{-1} = [A^TA]^{-1}$, and therefore

$$P^T \;=\; [A \cdot (A^TA)^{-1} \cdot A^T]^T \;=\; (A^T)^T \cdot [(A^TA)^{-1}]^T \cdot A^T \;=\; A \cdot [A^TA]^{-1} \cdot A^T \;=\; P.$$

Supplemental Problem 12: Given n points $(x_1, y_1), \ldots, (x_n, y_n)$ in $\mathbb{R}^2$, find the line $y = ax + b$ of best fit.

Solution. Set $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$ and $\overline{x} = \frac{1}{n}\sum_i x_i$ and $\overline{y} = \frac{1}{n}\sum_i y_i$. Then $y = ax + b$ passes through (x_i, y_i) if $x_i a + b = y_i$. Thus, as in Problem 5 above, we are seeking the Least Squares solution to the over-determined linear system $A\mathbf{x}_0 = \mathbf{b}$ where

$$A \;=\; \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \qquad \mathbf{b} \;=\; \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Then

$$A^T\mathbf{b} \;=\; \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ 1 & 1 & \ldots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \mathbf{x} \cdot \mathbf{y} \\ n\overline{y} \end{pmatrix} \qquad A^TA \;=\; \begin{pmatrix} x_1 & x_2 & \ldots & x_n \\ 1 & 1 & \ldots & 1 \end{pmatrix} \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} = \begin{pmatrix} \mathbf{x} \cdot \mathbf{x} & n\overline{x} \\ n\overline{x} & n \end{pmatrix}$$

Hence the line of best fit is $y = ax + b$ where

$$\begin{pmatrix} a \\ b \end{pmatrix} \;=\; (A^TA)^{-1}A^T\mathbf{b} \;=\; \frac{1}{n\big(\mathbf{x} \cdot \mathbf{x} - n\overline{x}^2\big)} \begin{pmatrix} n & -n\overline{x} \\ -n\overline{x} & \mathbf{x} \cdot \mathbf{x} \end{pmatrix} \begin{pmatrix} \mathbf{x} \cdot \mathbf{y} \\ n\overline{y} \end{pmatrix} \;=\; \frac{1}{\big(\mathbf{x} \cdot \mathbf{x} - n\overline{x}^2\big)} \begin{pmatrix} \mathbf{x} \cdot \mathbf{y} - n\overline{x}\,\overline{y} \\ \overline{y}\mathbf{x} \cdot \mathbf{x} - \overline{x}(\mathbf{x} \cdot \mathbf{y}) \end{pmatrix}$$

This can also be written as

$$y = ax + b \qquad \text{where} \qquad a = \frac{\mathbf{x} \cdot \mathbf{y} - n\overline{x}\,\overline{y}}{\mathbf{x} \cdot \mathbf{x} - n\overline{x}^2} \qquad \text{and} \qquad b = \overline{y} - a\overline{x}.$$