4.1.9. Consider the four data points \((-1, 0), (0, 1), (1, 3), (2, 5)\).

(a) Find the “least squares horizontal line” \(y = a\) fitting the data points. Check that the sum of the errors is 0.

We are trying to solve the obviously inconsistent system

\[
\begin{align*}
    a &= 0 \\
    a &= 1 \\
    a &= 3 \\
    a &= 5,
\end{align*}
\]

i.e.

\[
\begin{bmatrix}
1 \\
1 \\
1 \\
1
\end{bmatrix} \begin{bmatrix}
a \\
b
\end{bmatrix} = \begin{bmatrix}
0 \\
1 \\
3 \\
5
\end{bmatrix}.
\]

The least squares equation is

\[
\begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1
\end{bmatrix} \begin{bmatrix}
a \\
b
\end{bmatrix} = \begin{bmatrix}
0 \\
1 & 1 & 1 & 1 \\
1 \\
3 \\
5
\end{bmatrix}
\]

\[
\begin{align*}
4a &= 9 \\
a &= 9/4.
\end{align*}
\]

This is just a fancy way of saying that the \(y\) coordinate of the least squares horizontal line is the average of the \(y\) coordinates of the data points! We can check explicitly: \((9/4 - 0) + (9/4 - 1) + (9/4 - 3) + (9/4 - 5) = 0\).

(b) Find the least squares line \(y = ax + b\) fitting the data points. Check that the sum of errors is 0. We are trying to solve the inconsistent system

\[
\begin{bmatrix}
-1 & 1 \\
0 & 1 \\
1 & 1 \\
2 & 1
\end{bmatrix} \begin{bmatrix}
a \\
b
\end{bmatrix} = \begin{bmatrix}
0 \\
1 \\
3 \\
5
\end{bmatrix}.
\]
The least squares equation is
\[
\begin{bmatrix}
-1 & 0 & 1 & 2 \\
1 & 1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
a \\
b
\end{bmatrix} = 
\begin{bmatrix}
0 \\
1
\end{bmatrix}
\begin{bmatrix}
0 \\
1
\end{bmatrix}
\begin{bmatrix}
0 \\
1
\end{bmatrix}
\begin{bmatrix}
-1 & 0 & 1 & 2 \\
1 & 1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
3 \\
5
\end{bmatrix}
\]

\[
\begin{bmatrix}
6 & 2 \\
2 & 4
\end{bmatrix}
\begin{bmatrix}
a \\
b
\end{bmatrix} = 
\begin{bmatrix}
13 \\
9
\end{bmatrix}
\]

Solving, we get \( a = 1.7 \), \( b = 1.4 \). The error vector is thus
\[
\begin{bmatrix}
0 \\
1 \\
3 \\
5
\end{bmatrix} - \begin{bmatrix}
-1 & 1 \\
0 & 1 \\
1 & 1 \\
2 & 1
\end{bmatrix}
\begin{bmatrix}
1.7 \\
1.4 \\
1.4 \\
1.4
\end{bmatrix} = 
\begin{bmatrix}
-1.3 \\
1.4 \\
3.1 \\
4.8
\end{bmatrix} - \begin{bmatrix}
0.3 \\
0.4 \\
-0.4 \\
-0.2
\end{bmatrix}
\]

and we see explicitly that the sum of the errors is 0.

**(c)** Find the “least squares parabola” \( y = ax^2 + bx + c \) fitting the data points. What is true of the sum of the errors in this case? We are trying to solve the system
\[
\begin{bmatrix}
1 & -1 & 1 \\
0 & 0 & 1 \\
1 & 1 & 1 \\
4 & 2 & 1
\end{bmatrix}
\begin{bmatrix}
a \\
b \\
c
\end{bmatrix} = 
\begin{bmatrix}
0 \\
1 \\
3 \\
5
\end{bmatrix}
\]

The least squares equation is
\[
\begin{bmatrix}
1 & 0 & 1 & 4 \\
-1 & 0 & 1 & 2 \\
1 & 1 & 1 & 1 \\
4 & 2 & 1
\end{bmatrix}
\begin{bmatrix}
a \\
b \\
c
\end{bmatrix} = 
\begin{bmatrix}
1 & 0 & 1 & 4 \\
-1 & 0 & 1 & 2 \\
1 & 1 & 1 & 1 \\
18 & 8 & 6 & 2
\end{bmatrix}
\begin{bmatrix}
0 \\
1 \\
3 \\
5
\end{bmatrix}
\]

We obtain \( a = 1/4 \), \( b = 29/20 \), \( c = 23/20 \). The error vector is
\[
\begin{bmatrix}
0 \\
1 \\
3 \\
5
\end{bmatrix} - \begin{bmatrix}
1 & -1 & 1 \\
0 & 0 & 1 \\
1 & 1 & 1 \\
4 & 2 & 1
\end{bmatrix}
\begin{bmatrix}
1/4 \\
29/20 \\
23/20 \\
23/20
\end{bmatrix} = 
\begin{bmatrix}
0 \\
1 \\
3 \\
5
\end{bmatrix} - \begin{bmatrix}
-1/20 \\
23/20 \\
57/20 \\
101/20
\end{bmatrix} = \begin{bmatrix}
1/20 \\
-3/20 \\
-1/20 \\
-1/20
\end{bmatrix}.
\]

The sum of the errors is again zero. Note that the error is much smaller here than in (a) or (b).
4.1.11. Find the least squares fit of the form \( y = ax^k \) to the data points \((1, 2), (2, 3), (3, 5), \) and \((5, 8)\).

The trick here is to take logs to get \( \ln y = \ln(ax^k) = k \ln x + \ln a \). This is a linear system in the variables \( k \) and \( \ln a \), namely:

\[
\begin{align*}
    k(0) + \ln a &= \ln 2 \\
    k \ln 2 + \ln a &= \ln 3 \\
    k \ln 3 + \ln a &= \ln 5 \\
    k \ln 5 + \ln a &= \ln 8
\end{align*}
\]

i.e.

\[
\begin{bmatrix}
    0 & 1 \\
    \ln 2 & 1 \\
    \ln 3 & 1 \\
    \ln 5 & 1
\end{bmatrix}
\begin{bmatrix}
    k \\
    \ln a
\end{bmatrix}
= \begin{bmatrix}
    \ln 2 \\
    \ln 3 \\
    \ln 5 \\
    \ln 8
\end{bmatrix}.
\]

The least squares equation is

\[
\begin{bmatrix}
    0 & \ln 2 & \ln 3 & \ln 5 \\
    1 & 1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
    0 & 1 \\
    \ln 2 & 1 \\
    \ln 3 & 1 \\
    \ln 5 & 1
\end{bmatrix}
\begin{bmatrix}
    k \\
    \ln a
\end{bmatrix}
= \begin{bmatrix}
    \ln 2 \\
    \ln 3 \\
    \ln 5 \\
    \ln 8
\end{bmatrix}.
\]

Solving this (using a calculator), we get \( k \approx 0.878 \), \( \ln a \approx 0.624 \), and hence \( a \approx 1.866 \).

4.1.13. Use the definition of projection on p. 192 to show that for any subspace \( V \subset \mathbb{R}^m \), \( \text{proj}_V : \mathbb{R}^m \to \mathbb{R}^m \) is a linear transformation.

First, recall the definition: Given any vector \( x \in \mathbb{R}^m \), we can uniquely write \( x = x^\parallel + x^\perp \), where \( x^\parallel \in V \) and \( x^\perp \in V^\perp \). The projection \( \text{proj}_V(x) \) is then defined to be \( x^\parallel \).

Given any vectors \( x, y \), write \( x = x^\parallel + x^\perp \) and \( y = y^\parallel + y^\perp \) as above. Then

\[
x + y = x^\parallel + x^\perp + y^\parallel + y^\perp \\
= (x^\parallel + y^\parallel) + (x^\perp + y^\perp).
\]

The second line gives a decomposition of \( x + y \) as an element of \( V \) (namely \( x^\parallel + y^\parallel \)) plus an element of \( V^\perp \) (namely \( x^\perp + y^\perp \)). Since such a decomposition is unique, by definition we have \( \text{proj}_V(x + y) = x^\parallel + y^\parallel = \text{proj}_V(x) + \text{proj}_V(y) \), as required. Similarly, for any scalar \( c \), we have

\[
cx = c(x^\parallel + x^\perp) = cx^\parallel + cx^\perp,
\]

which the sum of an element of in \( V \) and an element in \( V^\perp \), so \( \text{proj}_V(cx) = cx^\parallel = c \text{proj}_V(x) \). Thus, \( \text{proj}_V \) is a linear transformation. \( \square \)

4.1.15. Prove that if \( A^2 = A \) and \( A = AT \), then \( A \) is a projection matrix.

Let \( V = \text{C}(A) \); we claim that for any \( x \in \mathbb{R}^n \), \( Ax = \text{proj}_V(x) \). Clearly, \( Ax \in V \), so we just have to show that \( x - Ax \in V^\perp \). And \( V^\perp = \text{C}(A)^\perp = N(AT) \), so this is equivalent to showing that \( x - Ax \in N(AT) \). We compute:

\[
AT(x - Ax) = Ax - A^T Ax = Ax - A^2 x = Ax - Ax = 0,
\]

as required.
Thus, any vector $x$ can be written as $x = Ax + (x - Ax)$, which is an element of $V$ plus an element of $V^\perp$. This shows that $\text{proj}_V(x) = Ax$, as required.

4.2.7. According to Proposition 4.10 of Chapter 3, if $A$ is an $m \times n$ matrix, then for each $b \in C(A)$, there is a unique $x \in R(A)$ with $Ax = b$. In each case, give a formula for that $x$.

(a) $A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{bmatrix}$.

It is easy to see that a basis for $C(A)$ consists of the vector $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$; that is, any element of $C(A)$ is of the form $\begin{bmatrix} b \\ b \end{bmatrix}$ for $b \in \mathbb{R}$. Also, a basis for $R(A)$ consists of the vector $\begin{bmatrix} 1/4 \\ 0 \\ 1/4 \end{bmatrix}$, and we have $A \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 14 \\ 14 \\ 14 \end{bmatrix}$. Thus, for any $b = \begin{bmatrix} b \\ b \end{bmatrix} \in C(A)$, if we set $x = \begin{bmatrix} b/14 \\ 2b/14 \\ 3b/14 \end{bmatrix}$, then $x$ is the unique element of $R(A)$ with $Ax = b$.

(b) $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \end{bmatrix}$

This matrix has rank equal to the number of rows, so every vector $b \in \mathbb{R}^2$ is in $C(A)$. For any $b$, let’s first find an arbitrary vector $x \in \mathbb{R}^3$ with $Ax = b$:

$$
\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} b_1 - b_2 \\ b_2 \end{bmatrix}
$$

We obtain one possible solution by taking $x = \begin{bmatrix} b_1 - b_2 \\ b_2 \\ 0 \end{bmatrix}$.

Now, to find an element $y \in R(A)$ with $Ay = Ax = b$, it suffices to take $y$ to be the projection of $x$ to $R(A)$. (The reason is that $x - y \in R(A)\perp = N(A)$, so $Ax = Ay$.) This projection is easy to compute because the rows of $A$ (call them $v_1, v_2$) are orthogonal. Thus, we have:

$$
y = \text{proj}_{R(A)}(x) = \frac{x \cdot v_1}{v_1 \cdot v_1} v_1 + \frac{x \cdot v_2}{v_1 \cdot v_2} v_2 = \frac{b_1}{3} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + \frac{b_2}{2} \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}
$$

It is easy to check explicitly that $Ay = b$ as required.

4.2.11. Let $A$ be an $n \times n$ matrix, and let $a_1, \ldots, a_n$ denote its column vectors.

(a) Suppose $a_1, \ldots, a_n$ form an orthonormal set. Show that $A^{-1} = A^T$.

In general, the $(i, j)$ entry in $A^T A$ is the dot product $a_i \cdot a_j$. In this case, this dot product equals 0 when $i \neq j$ and 1 when $i = j$, so $A^T A = I_n$. Since $A$ is a square matrix, this implies that $A$ is invertible. Multiplying through on the right by $A^{-1}$, we find $A^T = A^{-1}$. 


(b) Suppose \( \mathbf{a}_1, \ldots, \mathbf{a}_n \) form an orthogonal set and each is nonzero. Find the appropriate formula for \( A^{-1} \).

Let \( \mathbf{B} \) be the matrix whose \( i \)th row is \( \frac{1}{\mathbf{a}_i} \mathbf{a}_i^T \). Then the \((i, j)\) entry of \( \mathbf{B} \) is equal to \( \frac{\mathbf{a}_i \cdot \mathbf{a}_j}{\mathbf{a}_i \cdot \mathbf{a}_i} \), which is 0 when \( i \neq j \) and 1 when \( i = j \). Hence \( \mathbf{B} \mathbf{A} = \mathbf{I}_n \), so \( A^{-1} = \mathbf{B} \).