5.3.1 Find least-squares solutions:

(a) \[
\begin{align*}
 x_1 + x_2 &= 3 \\
 2x_1 - 3x_2 &= 1 \\
 0x_1 + 0x_2 &= 2
\end{align*}
\]

**Solution:** We are trying to solve \( A\hat{x} = b \), where \( A = \begin{bmatrix} 1 & 1 \\ 2 & -3 \\ 0 & 0 \end{bmatrix} \) and \( b = (3, 1, 2)^T \). Clearly \( b \notin R(A) \). So we use the normal equation, \( A^T A \hat{x} = A^T b \), which becomes
\[
\begin{bmatrix} 5 & -5 \\ -5 & 10 \end{bmatrix} \hat{x} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}.
\]
The solution (unique, since \( A \) has rank 2) is \( \hat{x} = (2, 1)^T \).

(c) \[
\begin{align*}
 x_1 + x_2 + x_3 &= 4 \\
 -x_1 + x_2 + x_3 &= 0 \\
 -x_2 + x_3 &= 1 \\
 x_1 + x_3 &= 2
\end{align*}
\]

**Solution:** The matrix equation is
\[
\begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & -1 & 1 \\ 1 & 0 & 1 \end{bmatrix} x = \begin{bmatrix} 4 \\ 0 \\ 1 \\ 2 \end{bmatrix},
\]
which is inconsistent. The normal equations lead to the matrix equation,
\[
\begin{bmatrix} 3 & 0 & 1 \\ 0 & 3 & 1 \\ 1 & 1 & 4 \end{bmatrix} \hat{x} = \begin{bmatrix} 6 \\ 3 \\ 7 \end{bmatrix},
\]
so the solution is \( \hat{x} = (1.6, 0.6, 1.2)^T \).
5.3.2 For each solution $\hat{x}$ in exercise 5.3.1,

1. Determine $p = A\hat{x}$.
2. Calculate $r(\hat{x})$.
3. Verify that $r(\hat{x}) \in N(A^T)$.

For item (c) in 5.3.1, we have

$$p = A\hat{x} = \begin{bmatrix}
1 & 1 & 1 \\
-1 & 1 & 1 \\
0 & -1 & 1 \\
1 & 0 & 1
\end{bmatrix} \begin{bmatrix}
1.6 \\
0.6 \\
1.2
\end{bmatrix} = \begin{bmatrix}
3.4 \\
0.2 \\
0.6 \\
2.8
\end{bmatrix}, \text{ so } r(\hat{x}) = \begin{bmatrix}
0.6 \\
-0.2 \\
0.4 \\
-0.8
\end{bmatrix}.$$  

We easily verify that

$$A^T r(\hat{x}) = \begin{bmatrix}
1 & 0 & 1 \\
1 & 1 & 0 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{bmatrix} \begin{bmatrix}
0.6 \\
-0.2 \\
0.4 \\
-0.8
\end{bmatrix} = \begin{bmatrix}
0 \\
0 \\
0 \\
0
\end{bmatrix}.$$

5.3.3a Find all least squares solutions to $Ax = b$, where $A = \begin{bmatrix}
1 & 2 & 2 \\
2 & 4 & -1 \\
-1 & 4 & -2
\end{bmatrix}$ and $b = (3, 2, 1)^T$.

**Solution:** First note that the columns of $A$ are linearly dependent, so $A$ (and hence $A^T A$) has a nontrivial nullspace and we anticipate multiple solutions. Solving

$$\begin{bmatrix}
6 & 12 \\
12 & 24
\end{bmatrix} \hat{x} = \begin{bmatrix}
6 \\
12
\end{bmatrix},$$

we find that all solutions have the form $x_2 = s$ and $x_1 = 1 - 2s$. That is,

$$\hat{x} = \begin{bmatrix}
1 - 2s \\
s
\end{bmatrix} = \begin{bmatrix}
1 \\
0
\end{bmatrix} + s \begin{bmatrix}
-2 \\
1
\end{bmatrix}.$$  

Note that $s \begin{bmatrix}
-2 \\
1
\end{bmatrix} \in N(A)$ since $s \begin{bmatrix}
-2 \\
1
\end{bmatrix}$ is orthogonal to each row of $A$, i.e. orthogonal to $RS(A)$

5.3.4a For the system in Exercise 3, we want the projection $p$ of $b$ onto $R(A)$, and the verification that $b - p$ is orthogonal to each of the columns of $A$.

**Solution:** Continuing with the previous problem, the projection is

$$p = A \left( \begin{bmatrix}
1 \\
0
\end{bmatrix} + s \begin{bmatrix}
-2 \\
1
\end{bmatrix} \right) = A \begin{bmatrix}
1 \\
0
\end{bmatrix} = \begin{bmatrix}
1 \\
2 \\
-1
\end{bmatrix}.$$
The second equality above is true since \( s \left[ \begin{array}{c} -2 \\ 1 \end{array} \right] \in N(A) \).

It follows that \( b - p = (2, 0, 2)^T \), clearly orthogonal to the columns of \( A \) (which you can verify by taking the dot product).

**5.3.5a** Find the best least-squares fit by a linear function to the given data:

\[
\begin{array}{c|ccc|c}
\hline
x & -1 & 0 & 1 & 2 \\
y & 0 & 1 & 3 & 9 \\
\hline
\end{array}
\]

**Solution:** We are assuming that \( y = mx + b \), where \( m \) and \( b \) are the unknowns. Under this assumption, we have a system of equations,

\[
\begin{align*}
0 &= -m + b \\
1 &= b \\
3 &= m + b \\
9 &= 2m + b,
\end{align*}
\]

and the corresponding matrix equation is

\[
\begin{bmatrix}
-1 & 1 \\
0 & 1 \\
1 & 1 \\
2 & 1
\end{bmatrix}
\begin{bmatrix}
m \\
b
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
1 \\
3 \\
9
\end{bmatrix}.
\]

The normal equations become

\[
\begin{bmatrix}
6 & 2 \\
2 & 4
\end{bmatrix}
\begin{bmatrix}
m \\
b
\end{bmatrix}
= 
\begin{bmatrix}
21 \\
13
\end{bmatrix}.
\]

The solution: \( m = 2.9, b = 1.8 \), and so the function is \( y = 2.9x + 1.8 \).

**5.3.6** Repeat problem 5.4.5a, but this time fit a quadratic polynomial to the data.

**Solution:** We now assume that \( p(x) = ax^2 + bx + c \), where the coefficients are unknown. This leads to the equations,

\[
\begin{align*}
p(-1) &= a - b + c = 0 \\
p(0) &= c = 1 \\
p(1) &= a + b + c = 3 \\
p(2) &= 4a + 2b + c = 9.
\end{align*}
\]

The corresponding (and inconsistent) matrix equation is

\[
\begin{bmatrix}
1 & -1 & 1 \\
1 & 0 & 0 \\
1 & 1 & 1 \\
1 & 2 & 4
\end{bmatrix}
\begin{bmatrix}
c \\
b \\
a
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
1 \\
3 \\
9
\end{bmatrix}.
\]
The normal equations become

\[
\begin{bmatrix}
4 & 2 & 6 \\
2 & 6 & 8 \\
6 & 8 & 18
\end{bmatrix}
\begin{bmatrix}
c \\
b \\
a
\end{bmatrix}
=
\begin{bmatrix}
13 \\
21 \\
39
\end{bmatrix}.
\]

After elimination, we use back-substitution to find \( a = 1.25 \), \( b = 1.65 \), and \( c = 0.55 \), so \( p(x) = 1.25x^2 + 1.65x + 0.55 \).

5.3.9 Let \( A \) be an \( m \times n \) matrix of rank \( n \), and let \( P = A(A^TA)^{-1}A^T \).

(a) Show that \( Pb = b \) for every \( b \in R(A) \).

**Proof:** Let \( b \in R(A) \). Then \( b = Ax \) for some \( x \in \mathbb{R}^n \). It follows that

\[
Pb = A(A^TA)^{-1}A^Tb = A(A^TA)^{-1}A^T(Ax) = A(A^TA)^{-1}(A^TA)x = Ax = b,
\]

which is what we needed to show.

(b) If \( b \in R(A)^\perp \), show that \( Pb = 0 \).

**Proof:** Let \( b \in R(A)^\perp \), then \( b \in N(A^T) \) and so \( A^Tb = 0 \). Now, when we multiply \( Pb = (A(A^TA)^{-1}A^T)b = A(A^TA)^{-1}(A^Tb) = A(A^TA)^{-1}0 = 0 \), as desired.

(c) Give a geometric illustration of parts (a) and (b) if \( R(A) \) is a plane through the origin in \( \mathbb{R}^3 \).

**Solution:** Intuitively, think of \( P \) as the projection matrix onto \( R(A) \). So given any vector \( v \), then \( Pv \) gives the projection of \( v \) onto \( R(A) \). In particular, part (a) says that the projection of \( b \in R(A) \) onto \( R(A) \) must be \( b \) itself. Part (b) says that if \( b \in R(A)^\perp \), then the projection of \( b \) on \( R(A) \) is the zero vector.

\[
\begin{align*}
(\text{a}) & \quad \overrightarrow{b} \in R(A) \\
(\text{b}) & \quad \overrightarrow{b} \perp R(A)
\end{align*}
\]