6.4.20 Suppose that $A = QR$ where $R$ is an invertible matrix. Show that $A$ and $Q$ have the same columns space.

Well, given a matrix $B$, $y \in \text{Col}(B)$ if and only there is an $x$ such that $Bx = y$. So suppose that $y \in \text{Col}(A)$. Then there is an $x$ such that $Ax = y$, and we can rewrite this equation as $Q(Rx) = y$. Note that $Rx$ is a vector, which we will call $z$. Then $z$ has the property that $Qz = y$, so $y \in \text{Col}(Q)$. We can conclude that $\text{Col}(A) \subseteq \text{Col}(Q)$.

Now suppose that $y \in \text{Col}(Q)$. Then there is an $x$ such that $Qx = y$. Of course $Q = AR^{-1}$, so we can rewrite this as $AR^{-1}x = y$. Here $R^{-1}x$ is simply a vector, which we will call $z$. Thus $z$ has the property that $Az = y$ and hence $y \in \text{Col}(A)$. We conclude that $\text{Col}(Q) \subseteq \text{Col}(A)$, and because we proved the opposite inclusion above, it must be that $\text{Col}(A) = \text{Col}(Q)$.

6.4.22 Let $u_1, \ldots, u_p$ be an orthogonal basis for a subspace $W \subseteq \mathbb{R}^n$, and let $T : \mathbb{R}^n \to \mathbb{R}^n$ be defined by $T(x) = \text{proj}_W(x)$. Show that $T$ is a linear transformation.

To do this, I cut and pasted quite a bit of material from the last assignment. Well, there are some things we have to check. First, for all $x, y \in \mathbb{R}^3$, note that

$$T(x + y) = \text{proj}_W(x + y) = \frac{(x + y) \cdot u_1}{u_1 \cdot u_1} u_1 + \cdots + \frac{(x + y) \cdot u_p}{u_p \cdot u_p} u_p =$$

$$\frac{x \cdot u_1 + y \cdot u_1}{u_1 \cdot u_1} u_1 + \cdots + \frac{x \cdot u_p + y \cdot u_p}{u_p \cdot u_p} u_p =$$

$$\frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \cdots + \frac{x \cdot u_p}{u_p \cdot u_p} u_p + \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \cdots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p =$$

$$\text{proj}_W(x) + \text{proj}_W(y) = T(x) + T(y).$$

For all $c \in \mathbb{R}$,

$$T(cx) = \text{proj}_W(cx) = \frac{c x \cdot u_1}{u_1 \cdot u_1} u_1 + \cdots + \frac{c x \cdot u_p}{u_p \cdot u_p} u_p =$$
\[
\frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \cdots + \frac{x \cdot u_p}{u_p \cdot u_p} u_p = c \left( \frac{x \cdot u_1}{u_1 \cdot u_1} u_1 + \cdots + \frac{x \cdot u_p}{u_p \cdot u_p} u_p \right) = c \text{proj}_W(x) = cT(x).
\]

We conclude that \( T \) is a linear transformation.

6.4.24 Use the Gram-Schmidt process to produce an orthogonal basis for the column space of
\[
A = \begin{bmatrix}
-10 & 13 & 7 & -11 \\
2 & 1 & -5 & 3 \\
-6 & 3 & 13 & -3 \\
16 & -16 & -2 & 5 \\
2 & 1 & -5 & -7
\end{bmatrix}.
\]

Label the columns \( x_1, x_2, x_3, x_4 \), and let
\[
\begin{align*}
\mathbf{u}_1 &= x_1 \\
\mathbf{u}_2 &= x_2 - \frac{\mathbf{u}_1 \cdot x_2}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 \\
\mathbf{u}_3 &= x_3 - \frac{\mathbf{u}_1 \cdot x_3}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 - \frac{\mathbf{u}_2 \cdot x_3}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 \\
\mathbf{u}_4 &= x_4 - \frac{\mathbf{u}_1 \cdot x_4}{\mathbf{u}_1 \cdot \mathbf{u}_1} \mathbf{u}_1 - \frac{\mathbf{u}_2 \cdot x_4}{\mathbf{u}_2 \cdot \mathbf{u}_2} \mathbf{u}_2 - \frac{\mathbf{u}_3 \cdot x_4}{\mathbf{u}_3 \cdot \mathbf{u}_3} \mathbf{u}_3.
\end{align*}
\]

A little hard work on the calculator (or actually I used Mathematica so that I could cut and paste), yields:
\[
\mathbf{u}_1 = \begin{bmatrix} -10 \\ 2 \\ -6 \\ 16 \\ 2 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 3 \\ 3 \\ -3 \\ 0 \\ 3 \end{bmatrix}, \quad \mathbf{u}_3 = \begin{bmatrix} 6 \\ 0 \\ 0 \\ 6 \\ 0 \end{bmatrix}, \quad \mathbf{u}_4 = \begin{bmatrix} 0 \\ 5 \\ 0 \\ 0 \\ -5 \end{bmatrix}.
\]

This is an orthogonal basis for the column space of \( A \).

6.5.16 Use the factorization \( A = QR \), where
\[
A = \begin{bmatrix}
1 & -1 \\
1 & 4 \\
1 & -1 \\
1 & 4
\end{bmatrix}, \quad Q = \begin{bmatrix}
1/2 & -1/2 \\
1/2 & 1/2 \\
1/2 & -1/2 \\
1/2 & 1/2
\end{bmatrix}, \quad \text{and} \quad R = \begin{bmatrix}
2 & 3 \\
0 & 5
\end{bmatrix}.
\]

to find the least-squares solutions of \( Ax = b \) for \( b = \begin{bmatrix} -1 \\ 6 \\ 5 \\ 7 \end{bmatrix} \).
According to theorem 15, if $A$ has linearly independent columns and $A = QR$ is a $QR$-factorization, then the solution is $\hat{x} = R^{-1}Q^TB$. It is clear that the columns of $A$ are linearly independent (there are only two, so it is enough to check that one is not a multiple of the other, which is true). In order for $QR$ to be a $QR$-factorization we need to know that $R$ is an invertible upper triangular matrix with positive diagonal entries (which is true by inspection), and that the columns of $Q$ give an orthonormal basis of $\text{Col}(A)$. Let $u_1, u_2$ be the columns of $Q$. Then $u_1$ and $u_2$ are obviously linearly independent, and $u_1 \cdot u_1 = 1 = u_2 \cdot u_2$, so they give an orthonormal basis. To show that they give an orthonormal basis of $\text{Col}(A)$ it is enough to demonstrate that $\text{Col}(A) \subseteq \text{Col}(Q)$ because they both have dimension 2. Writing $R = [r_1 \ r_2]$, then $QR = [Qr_1 \ Qr_2] = [2u_1, 3u_1 + 5u_2]$, that is, each of the columns of $A$ is a linear combination of the columns of $Q$. So $Q$ has the desired properties and $A = QR$ is a $QR$-factorization.

These things being checked, we turn the crank and get

$$\hat{x} = \begin{bmatrix} 5 & -3 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -1/2 & 1/2 & -1/2 & 1/2 \end{bmatrix} \begin{bmatrix} -1 \\ 6 \\ 5 \\ 7 \end{bmatrix} = \begin{bmatrix} 29/10 \\ 9/10 \end{bmatrix}.$$ 

### 6.5.20
Let $A$ be an $m \times n$ matrix such that $A^TA$ is invertible. Show that the columns of $A$ are linearly independent.

We know that the columns of $A$ are linearly independent if and only if $Ax = 0$ has only the trivial solution. So suppose that $Ax = 0$. Then we need to prove that $x = 0$. Of course, $A^TAx = A^T0 = 0$, so $x \in \text{Nul}(A^TA)$. But $A^TA$ is invertible, so by the IMT, the only vector in its null space is $0$. This implies that $x = 0$ as required.

### 6.5.22
Show that $\text{rank}(A^TA) = \text{rank}(A)$.

Denote the number of columns of $A$ by $n$. Then $A^TA$ also has $n$ columns, so by the Rank Theorem, $\text{rank}(A) + \text{dim}(\text{Nul}(A)) = n = \text{rank}(A^TA) + \text{dim}(\text{Nul}(A^TA))$. Thus to prove that $\text{rank}(A^TA) = \text{rank}(A)$ it is enough to show that $\text{Nul}(A) = \text{Nul}(A^TA)$.

So suppose that $x \in \text{Nul}(A)$. Then $Ax = 0$ and thus $A^TAx = A^T0 = 0$, so $x \in \text{Nul}(A^TA)$. We conclude that $\text{Nul}(A) \subseteq \text{Nul}(A^TA)$. If, on the other hand, $x \in \text{Nul}(A^TA)$, then $A^TAx = 0$, so $x^TA^TAx = x^T0 = 0$, and finally
\[(Ax)^T Ax = 0.\] But the left side of this equation is equal to the dot product of \(Ax\) with itself, that is \((Ax)^T Ax = (Ax) \cdot (Ax) = 0\), and we know that this implies \(Ax = 0\). So \(\text{Nul}(A^T A) \subseteq \text{Nul}(A)\), and given the inclusion above, this implies that \(\text{Nul}(A) = \text{Nul}(A^T A)\) as required.

6.6.2 Find the equation \(y = B_0 + B_1 x\) of the least-squares line that best fits the given data points: \((1, 0), (2, 1), (4, 2), (5, 3)\).

Finding the least-squares line that best fits the given data points is equivalent to solving the least squares problem for the system:

\[
\begin{align*}
B_0 + 1B_1 &= 0 \\
B_0 + 2B_1 &= 1 \\
B_0 + 4B_1 &= 2 \\
B_0 + 5B_1 &= 3
\end{align*}
\]

or

\[
\begin{bmatrix}
1 & 1 \\
1 & 2 \\
1 & 4 \\
1 & 5
\end{bmatrix}
\begin{bmatrix}
B_0 \\
B_1
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
1 \\
2 \\
3
\end{bmatrix}.
\]

It is easy to see that the columns of

\[
A = 
\begin{bmatrix}
1 & 1 \\
1 & 2 \\
1 & 4 \\
1 & 5
\end{bmatrix}
\]

are linearly independent, so according to theorem 14, page 413, the answer is

\[
\hat{x} = (A^T A)^{-1} A^T b = 
\begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & 2 & 4 & 5
\end{bmatrix}
\begin{bmatrix}
1 & 1 \\
1 & 2 \\
1 & 4 \\
1 & 5
\end{bmatrix}^{-1}
\begin{bmatrix}
0 \\
1 \\
2 \\
3
\end{bmatrix} = 
\begin{bmatrix}
-6 \\
.7
\end{bmatrix}.
\]

I used the calculator for the last step.

6.6.6 Let \(X\) be the design matrix \(X = \begin{bmatrix}
1 & x_1 & x_1^2 \\
1 & x_2 & x_2^2 \\
\vdots & \vdots & \vdots \\
1 & x_n & x_n^2
\end{bmatrix}\) corresponding to a least-squares fit of a parabola to data \((x_1, y_1), \ldots, (x_n, y_n)\). Suppose \(x_1, x_2, x_3\) are distinct.
Explain why there is only one parabola that fits the data best, in a least-squares sense.

The least-squares fit to the given data is \( y = B_0 + B_1 x + B_2 x^2 \) where the \( B_i \) are obtained by finding the least-squares solution to the equation

\[
\begin{bmatrix}
1 & x_1 & x_1^2 \\
1 & x_2 & x_2^2 \\
\vdots & \vdots & \vdots \\
1 & x_n & x_n^2
\end{bmatrix}
\begin{bmatrix}
B_0 \\
B_1 \\
B_2
\end{bmatrix} =
\begin{bmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{bmatrix}.
\]

We know that this has a least squares solution if \( A^T A \) is invertible, and that this will happen if and only if the columns of \( A \) are linearly independent (theorem 14, page 413). So there will be a unique solution \( X \) has a pivot in every column (i.e., if the columns are linearly independent). To check if this is the case, it is enough to find a pivot in each of the first three rows, that is, it is enough to check that the matrix \( X' = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ 1 & x_3 & x_3^2 \end{bmatrix} \) has three pivot rows.

First suppose that \( x_1 \neq 0 \). Then there are constants \( a = x_2/x_1 \) and \( b = x_3/x_1 \) such that \( ax_1 = x_2 \) and \( bx_1 = x_3 \). Our matrix now becomes

\[
\begin{bmatrix}
1 & x_1 & x_1^2 \\
1 & ax_1 & a^2 x_1^2 \\
1 & bx_1 & b^2 x_1^2
\end{bmatrix}.
\]

We begin row reducing:

\[
\begin{bmatrix}
1 & x_1 & x_1^2 \\
1 & ax_1 & a^2 x_1^2 \\
1 & bx_1 & b^2 x_1^2
\end{bmatrix} \leftrightarrow
\begin{bmatrix}
1 & x_1 & x_1^2 \\
0 & ax_1 - x_1 & a^2 x_1^2 - x_1^2 \\
0 & bx_1 - x_1 & b^2 x_1^2 - x_1^2
\end{bmatrix}.
\]

Because \( x_1 \neq x_2 \), we know that \( a \neq 1 \), so we can divide the second row by \((1 - a)x_1 \) yielding:

\[
\begin{bmatrix}
1 & x_1 & x_1^2 \\
0 & 1 & (a + 1)x_1 \\
0 & (b - 1)x_1 & (b^2 - 1)x_1^2
\end{bmatrix} \leftrightarrow
\begin{bmatrix}
1 & x_1 & x_1^2 \\
0 & 1 & (a + 1)x_1 \\
0 & 0 & (b^2 - 1)x_1^2 - (a + 1)(b - 1)x_1^2
\end{bmatrix}.
\]

Simplifying the entry in the lower right hand corner, we obtain

\((b^2 - 1)x_1^2 - (a + 1)(b - 1)x_1^2 = (b - 1)x_1^2((b + 1 - a - 1) = (b - 1)(b - a)x_1^2.\)

We know that \( b \neq 1 \) (else \( x_1 = x_3 \)), \( a \neq b \) (else \( x_2 = x_3 \)), and \( x_1 \neq 0 \). So the entry in the lower right hand corner is nonzero and \( X' \) has three pivot rows.

If, on the other hand, \( x_1 = 0 \), then \( x_2 \neq x_3 \) and

\[
X' =
\begin{bmatrix}
1 & 0 & 0 \\
1 & x_2 & x_2^2 \\
1 & x_3 & x_3^2
\end{bmatrix} \leftrightarrow
\begin{bmatrix}
1 & 0 & 0 \\
0 & x_2 & x_2^2 \\
0 & x_3 & x_3^2
\end{bmatrix} \leftrightarrow
\begin{bmatrix}
1 & 0 & 0 \\
0 & 1 & x_2 \\
0 & 0 & x_3 - x_2
\end{bmatrix}.
\]

So \( X' \) has three pivot rows because \( x_2 \neq x_3 \).
A simple curve that often makes a good model for the variable costs of a company, as a function of the sales level \( x \), has the form
\[
y = B_1 x + B_2 x^2 + B_3 x^3.
\]
There is no constant term because the fixed costs are not included.

(a) Give the design matrix and the parameter vector for the linear model that leads to a least-squares fit of the equation above, with data \((x_1, y_1), \ldots, (x_n, y_n)\).

Let
\[
A = \begin{bmatrix}
x_1 & x_1^2 & x_1^3 \\
x_2 & x_2^2 & x_2^3 \\
\vdots & \vdots & \vdots \\
x_n & x_n^2 & x_n^3
\end{bmatrix},
\]
and let \( x = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} \). Then we are trying to find the least squares solution to the system
\[
\begin{bmatrix}
x_1 & x_1^2 & x_1^3 \\
x_2 & x_2^2 & x_2^3 \\
\vdots & \vdots & \vdots \\
x_n & x_n^2 & x_n^3
\end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.
\]
Here \( A \) is the design matrix and \( x \) is the parameter vector.

(b) Find the least-squares curve of the form above to fit the data \((4, 1.58), (6, 2.08), (8, 2.5), (10, 2.8), (12, 3.1), (14, 3.4), (16, 3.8), (18, 4.32)\), with values in thousands. If possible, produce a graph that shows the data points and the graph of the cubic approximation.

Use the calculator to solve the system (from part a)
\[
A^T Ax = A^T y = \begin{bmatrix} 4 & 6 & 8 & 10 & 12 & 14 & 16 & 18 \\ 16 & 36 & 64 & 100 & 144 & 196 & 256 & 324 \\ 64 & 216 & 512 & 1000 & 1728 & 2744 & 4096 & 5832 \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} 1.58 \\ 2.08 \\ 2.5 \\ 2.8 \\ 3.1 \\ 3.4 \\ 3.8 \\ 4.32 \end{bmatrix},
\]
The row reduced echelon form of the augmented matrix for this monstrous equation is

\[
\begin{bmatrix}
1 & 0 & 0 & 0.513216 \\
0 & 1 & 0 & -0.0334782 \\
0 & 0 & 1 & 0.00101595
\end{bmatrix}
\]

We conclude that \( x = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} = \begin{bmatrix} 0.513216 \\ -0.0334782 \\ 0.00101595 \end{bmatrix} \), and thus the curve in question is \( y = 0.513216x - 0.0334782x^2 + 0.0169341x^3 \). The picture (generated on Mathematica) is: