Orthogonality and Least Squares Approximation (4.1-4.4)
QUESTION:

Suppose $Ax = b$ has no solution...
Then what do we do?
Can we find an approximate solution?
Say a police man is interested in clocking the speed of a vehicle by using measurements of its relative distance. At different times $t_i$ we measure distance $b_i$. 

\[ b_i = \alpha + \beta t_i \]

The error between the measured value $b_i$ and the value predicted by the function is

\[ e_i = b_i - (\alpha + \beta t_i) \]

We can write it as

\[ e = b - Ax \]

where $x = (\alpha, \beta)$. $e$ is the error vector, $b$ is the data vector, and $A$ is an $m \times 2$ matrix.

We seek the line that minimizes the total squared error or Euclidean norm

\[ \|e\| = \sqrt{\sum_{i=1}^{m} e_i^2} \]
Say a police man is interested in clocking the speed of a vehicle by using measurements of its relative distance. At different times $t_i$ we measure distance $b_i$.

Assuming the vehicle is traveling at constant speed we know a linear formula for this, but there are errors!

Suppose we expect the output $b_i$ to be a linear function of the input $t_i = \alpha + \beta t_i$, but we need to determine $\alpha, \beta$.
Say a police man is interested in clocking the speed of a vehicle by using measurements of its relative distance. At different times $t_i$ we measure distance $b_i$.

Assuming the vehicle is traveling at constant speed we know a linear formula for this, but there are errors!

Suppose we expect the output $b$ to be a linear function of the input $t$ $b = \alpha + t\beta$, but we need to determine $\alpha, \beta$. 
Say a police man is interested in clocking the speed of a vehicle by using measurements of its relative distance. At different times $t_i$ we measure distance $b_i$.

Assuming the vehicle is traveling at constant speed we know a linear formula for this, but there are errors!

Suppose we expect the output $b$ to be a linear function of the input $t$ $b = \alpha + t\beta$, but we need to determine $\alpha, \beta$.

At $t = t_i$, the error between the measured value $b_i$ and the value predicted by the function is $e_i = b_i - (\alpha + \beta t_i)$. 
Say a police man is interested in clocking the speed of a vehicle by using measurements of its relative distance. At different times $t_i$ we measure distance $b_i$.

Assuming the vehicle is traveling at constant speed we know a linear formula for this, but there are errors!

Suppose we expect the output $b$ to be a linear function of the input $t$: $b = \alpha + t\beta$, but we need to determine $\alpha, \beta$.

At $t = t_i$, the error between the measured value $b_i$ and the value predicted by the function is $e_i = b_i - (\alpha + \beta t_i)$.

We can write it as $e = b - Ax$ where $x = (\alpha, \beta)$. $e$ is the error vector, $b$ is the data vector. $A$ is an $m \times 2$ matrix.
Say a police man is interested in clocking the speed of a vehicle by using measurements of its relative distance. At different times $t_i$ we measure distance $b_i$.

Assuming the vehicle is traveling at constant speed we know a linear formula for this, but there are errors!

Suppose we expect the output $b$ to be a linear function of the input $t$, $b = \alpha + t\beta$, but we need to determine $\alpha, \beta$.

At $t = t_i$, the error between the measured value $b_i$ and the value predicted by the function is $e_i = b_i - (\alpha + \beta t_i)$.

We can write it as $e = b - Ax$ where $x = (\alpha, \beta)$. $e$ is the error vector, $b$ is the data vector. $A$ is an $m \times 2$ matrix.

We seek the line that minimizes the total squared error or Euclidean norm $\|e\| = \sqrt{\sum_{i=1}^{m} e_i^2}$. 
KEY GOAL  We are looking for $x$ that minimizes $\|b - Ax\|$. 
If \( b \in C(A) \) then we can make the value \( ||b - Ax|| = 0 \).

Note that \( ||b - Ax|| \) is the distance from \( b \) to the point \( Ax \), which is an element of the column space!

The error vector \( e = b - Ax \) is perpendicular to the column space of \( A \).

Thus for each column \( a_i \) we have:
\[
a_i^T(b - Ax) = 0.
\]

Thus in matrix notation:
\[
A^T(b - Ax) = 0.
\]

This gives:
\[
A^TAx = A^Tb.
\]

We are going to study this system ALOT!

**KEY PROPERTY**
If \( A \) has independent columns, then \( A^T A \) is square, symmetric and invertible.
If \( b \in C(A) \) then we can make the value \( \|b - Ax\| = 0 \).

Note that \( \|b - Ax\| \) is the distance from \( b \) to the point \( Ax \), which is an element of the column space!
If \( b \in C(A) \) then we can make the value \( \|b - Ax\| = 0 \).

Note that \( \|b - Ax\| \) is the distance from \( b \) to the point \( Ax \), which is an element of the column space!

The error vector \( e = b - Ax \) is **perpendicular to the column space of** \( A \).
If \( b \in C(A) \) then we can make the value \( ||b - Ax|| = 0. \)

Note that \( ||b - Ax|| \) is the distance from \( b \) to the point \( Ax \), which is an element of the column space!

The error vector \( e = b - Ax \) is perpendicular to the column space of \( A \).

Thus for each column \( a_i \) we have \( a_i^T(b - Ax) = 0. \) Thus in matrix notation: \( A^T(b - Ax) = 0. \) This gives

\[
A^T Ax = A^T b
\]
• If \( b \in C(A) \) then we can make the value \( \|b - Ax\| = 0 \).

• Note that \( \|b - Ax\| \) is the distance from \( b \) to the point \( Ax \), which is an element of the column space!

• The error vector \( e = b - Ax \) is **perpendicular to the column space of** \( A \).

• Thus for each column \( a_i \) we have \( a_i^T(b - Ax) = 0 \). Thus in matrix notation: \( A^T(b - Ax) = 0 \). This gives

\[
A^T A x = A^T b
\]

• We are going to study this system ALOT!
If $b \in C(A)$ then we can make the value $||b - Ax|| = 0$.

Note that $||b - Ax||$ is the distance from $b$ to the point $Ax$, which is an element of the column space!

The error vector $e = b - Ax$ is **perpendicular to the column space of $A$**.

Thus for each column $a_i$ we have $a_i^T (b - Ax) = 0$. Thus in matrix notation: $A^T (b - Ax) = 0$. This gives

$$A^T Ax = A^T b$$

We are going to study this system ALOT!

**KEY PROPERTY** If $A$ has independent columns, then $A^T A$ is square, symmetric and invertible.
We say vectors $x, y$ are **perpendicular** when they create a 90 degree angle. When that happens the triangle they define is a right triangle!
Recall

- We say vectors $x, y$ are **perpendicular** when they create a 90 degree angle. When that happens the triangle they define is a right triangle!

- **Lemma** Two vectors $x, y$ in $\mathbb{R}^n$ are perpendicular if and only if
  \[
  x_1y_1 + \cdots + x_ny_n = x^T y = 0
  \]
  When this last equation holds we say that $x, y$ are **orthogonal**.
Recall

- We say vectors $x, y$ are **perpendicular** when they create a 90 degree angle. When that happens the triangle they define is a right triangle!

- **Lemma** Two vectors $x, y$ in $\mathbb{R}^n$ are perpendicular if and only if

$$x_1y_1 + \cdots + x_ny_n = x^Ty = 0$$

When this last equation holds we say that $x, y$ are **orthogonal**.

- **Orthogonal Bases**: A basis $u_1, \ldots, u_n$ of $V$ is orthogonal if $\langle u_i, u_j \rangle = 0$ for all $i \neq j$. 
Recall

We say vectors \( x, y \) are **perpendicular** when they create a 90 degree angle. When that happens the triangle they define is a right triangle!

**Lemma** Two vectors \( x, y \) in \( \mathbb{R}^n \) are perpendicular if and only if

\[
x_1 y_1 + \cdots + x_n y_n = x^T y = 0
\]

When this last equation holds we say that \( x, y \) are **orthogonal**.

**Orthogonal Bases:** A basis \( u_1, \ldots, u_n \) of \( V \) is orthogonal if

\[
\langle u_i, u_j \rangle = 0 \quad \text{for all} \quad i \neq j.
\]

**Lemma** If \( v_1, v_2, \ldots, v_k \) are orthogonal then they are linearly independent.
Definition We say two subspaces $V$, $W$ of $\mathbb{R}^n$ are orthogonal if for $u \in V$ and $w \in W$ we have $u^T w = 0$. Can you see a way to detect when two subspaces are orthogonal? Through their bases!

Theorem: The row space and the nullspace are orthogonal. Similarly, the column space is orthogonal to the left nullspace.

proof: The dot product between the rows of $A^T$ and the respective entries in the vector $y$ is zero. Therefore, the rows of $A^T$ are perpendicular to any $y \in N(A^T)$.

\[
A^T y = \begin{bmatrix}
\text{Column 1 of } A \\
\cdots \\
\text{Column n of } A
\end{bmatrix}
\begin{bmatrix}
y_1 \\
\cdots \\
y_n
\end{bmatrix}
= \begin{bmatrix}
0 \\
\cdots \\
0
\end{bmatrix}
\]

where $y \in N(A^T)$. 

MATH 22A: LINEAR ALGEBRA Chapter 4
The Orthogonality of the Subspaces

- **Definition** We say two subspaces $V$, $W$ of $\mathbb{R}^n$ are **orthogonal** if for $u \in V$ and $w \in W$ we have $u^T w = 0$.

- Can you see a way to detect when two subspaces are orthogonal?? **Through their bases!**
The Orthogonality of the Subspaces

- **Definition**: We say two subspaces $V, W$ of $\mathbb{R}^n$ are **orthogonal** if for $u \in V$ and $w \in W$ we have $u^T w = 0$.

- Can you see a way to detect when two subspaces are orthogonal?? **Through their bases**!

- **Theorem**: The row space and the nullspace are orthogonal. Similarly the column space is orthogonal to the left nullspace.
The Orthogonality of the Subspaces

- **Definition** We say two subspaces $V, W$ of $\mathbb{R}^n$ are **orthogonal** if for $u \in V$ and $w \in W$ we have $u^T w = 0$.

- Can you see a way to detect when two subspaces are orthogonal?? **Through their bases!**

- **Theorem:** The row space and the nullspace are orthogonal. Similarly, the column space is orthogonal to the left nullspace.

- **proof:** The dot product between the rows of $A^T$ and the respective entries in the vector $y$ is zero.
**Definition** We say two subspaces $V, W$ of $\mathbb{R}^n$ are **orthogonal** if for $u \in V$ and $w \in W$ we have $u^T w = 0$.

Can you see a way to detect when two subspaces are orthogonal?? **Through their bases!**

**Theorem:** The row space and the nullspace are orthogonal. Similarly the column space is orthogonal to the left nullspace.

**proof:** The dot product between the rows of $A^T$ and the respective entries in the vector $y$ is zero.

Therefore the rows of $A^T$ are perpendicular to any $y \in N(A^T)$

\[
A^T y = \begin{bmatrix}
\text{Column 1 of } A \\
\vdots \\
\text{Column n of } A
\end{bmatrix}
\begin{bmatrix}
y_1 \\
\vdots \\
y_n
\end{bmatrix} = \begin{bmatrix}
0 \\
\vdots \\
0
\end{bmatrix}
\]

where $y \in N(A^T)$. 

---

MATH 22A: LINEAR ALGEBRA  Chapter 4
There is a stronger relation, for a subspace $V$ of $\mathbb{R}^n$ the set of all vectors orthogonal to $V$ is the **orthogonal complement** of $V$, denoted $V^\perp$. 

**Exercise**

Let $W$ be a subspace, its orthogonal complement is a subspace, and $W \cap W^\perp = 0$.

**Exercise**

If $V \subset W$ are subspaces, then $W^\perp \subset V^\perp$. 
There is a stronger relation, for a subspace $V$ of $\mathbb{R}^n$ the set of all vectors orthogonal to $V$ is the **orthogonal complement** of $V$, denoted $V^\perp$.

**Warning** Spaces can be orthogonal without being complements!
There is a stronger relation, for a subspace $V$ of $\mathbb{R}^n$ the set of all vectors orthogonal to $V$ is the **orthogonal complement** of $V$, denoted $V^\perp$.

**Warning** Spaces can be orthogonal without being complements!

**Exercise** Let $W$ be a subspace, its orthogonal complement is a subspace, and $W \cap W^\perp = 0$. 

There is a stronger relation, for a subspace $V$ of $\mathbb{R}^n$ the set of all vectors orthogonal to $V$ is the \textbf{orthogonal complement} of $V$, denoted $V^\perp$.

\textbf{Warning} Spaces can be orthogonal without being complements!

\textbf{Exercise} Let $W$ be a subspace, its orthogonal complement is a subspace, and $W \cap W^\perp = 0$.

\textbf{Exercise} If $V \subset W$ are subspaces, then $W^\perp \subset V^\perp$. 
**Theorem** (Fundamental theorem part II)

\[ C(A^T)^\perp = N(A) \text{ and } N(A)^\perp = C(A^T). \]

**proof:** The first equation holds because \( x \) is orthogonal to all vectors of the row space \( \iff x \) is orthogonal to each of the rows \( \iff x \in N(A). \) The other equality follows from the above exercises.

**Corollary**

Given an \( m \times n \) matrix \( A, \) the nullspace is the orthogonal complement of the row space in \( \mathbb{R}^n. \) Similarly, the left nullspace is the orthogonal complement of the column space inside \( \mathbb{R}^m. \)
Theorem (Fundamental theorem part II)
\[ C(A^T) \perp = N(A) \text{ and } N(A) \perp = C(A^T). \]

proof: The first equation holds because \( x \) is orthogonal to all vectors of the row space \( \iff \) \( x \) is orthogonal to each of the rows \( \iff \) \( x \in N(A) \). The other equality follows from the above exercises.

Corollary
Given an \( m \times n \) matrix \( A \), the nullspace is the orthogonal complement of the row space in \( \mathbb{R}^n \). Similarly, the left nullspace is the orthogonal complement of the column space inside \( \mathbb{R}^m \).

Why is this such a big deal?

MATH 22A: LINEAR ALGEBRA Chapter 4
Theorem (Fundamental theorem part II)

$C(A^T) \perp = N(A)$ and $N(A) \perp = C(A^T)$.

proof: The first equation holds because $x$ is orthogonal to all vectors of the row space $\iff x$ is orthogonal to each of the rows $\iff x \in N(A)$. The other equality follows from the above exercises.

Corollary Given an $m \times n$ matrix $A$, the nullspace is the orthogonal complement of the row space in $\mathbb{R}^n$. Similarly, the left nullspace is the orthogonal complement of the column space inside $\mathbb{R}^m$. 
Theorem (Fundamental theorem part II)
\[ C(A^T) \perp = N(A) \text{ and } N(A) \perp = C(A^T). \]

proof: The first equation holds because \( x \) is orthogonal to all vectors of the row space \( \Leftrightarrow x \) is orthogonal to each of the rows \( \Leftrightarrow x \in N(A) \). The other equality follows from the above exercises.

Corollary Given an \( m \times n \) matrix \( A \), the nullspace is the orthogonal complement of the row space in \( \mathbb{R}^n \). Similarly, the left nullspace is the orthogonal complement of the column space inside \( \mathbb{R}^m \).

WHY is this such a big deal?
Theorem  Given an $m \times n$ matrix $A$, every vector $x$ in $\mathbb{R}^n$ can be written in a unique way as $x = x_n + x_r$ where $x_n$ is in the nullspace and $x_r$ is in the row space of $A$. 

Proof Pick $x_n$ to be the orthogonal projection of $x$ into $N(A)$ and $x_r$ to be the orthogonal projection into $C(A^T)$. Clearly $x$ is a sum of both, but why are they unique? If $x_n + x_r = x_n' + x_r'$, then $x_n - x_n' = x_r - x_r'$. Thus this must be the zero vector because $N(A)$ is orthogonal to $C(A^T)$.

This has a beautiful consequence: Every matrix $A$, when we think of it as a linear map, transforms the row space into its column space!
Theorem Given an $m \times n$ matrix $A$, every vector $x$ in $\mathbb{R}^n$ can be written in a unique way as $x = x_n + x_r$ where $x_n$ is in the nullspace and $x_r$ is in the row space of $A$.

proof Pick $x_n$ to be the orthogonal projection of $x$ into $N(A)$ and $x_r$ to be the orthogonal projection into $C(A^T)$. Clearly $x$ is a sum of both, but why are they unique?
Theorem  Given an $m \times n$ matrix $A$, every vector $x$ in $\mathbb{R}^n$ can be written in a unique way as $x = x_n + x_r$ where $x_n$ is in the nullspace and $x_r$ is in the row space of $A$.

proof  Pick $x_n$ to be the orthogonal projection of $x$ into $N(A)$ and $x_r$ to be the orthogonal projection into $C(A^T)$. Clearly $x$ is a sum of both, but why are they unique?

If $x_n + x_r = x_n' + x_r'$, then $x_n - x_n' = x_r - x_r'$. Thus this must be the zero vector because $N(A)$ is orthogonal to $C(A^T)$. 

This has a beautiful consequence: Every matrix $A$, when we think of it as a linear map, transforms the row space into its column space!
Theorem  Given an $m \times n$ matrix $A$, every vector $x$ in $\mathbb{R}^n$ can be written in a unique way as $x = x_n + x_r$ where $x_n$ is in the nullspace and $x_r$ is in the row space of $A$.

proof  Pick $x_n$ to be the orthogonal projection of $x$ into $N(A)$ and $x_r$ to be the orthogonal projection into $C(A^T)$. Clearly $x$ is a sum of both, but why are they unique?

If $x_n + x_r = x_n' + x_r'$, then $x_n - x_n' = x_r - x_r'$. Thus this must be the zero vector because $N(A)$ is orthogonal to to $C(A^T)$.

This has a beautiful consequence: Every matrix $A$, when we think of it as a linear map, transforms the row space into its column space!
Figure 1. The action of $A$: Row space to column space, nullspace to zero.
Projections onto Subspaces
QUESTION: Given a subspace $S$, what is the formula for the projection $p$ of a vector $b$ into $S$?
QUESTION: Given a subspace $S$, what is the formula for the projection $p$ of a vector $b$ into $S$?

Key idea of **LEAST SQUARES** for regression analysis
QUESTION: Given a subspace $S$, what is the formula for the projection $p$ of a vector $b$ into $S$?

Key idea of LEAST SQUARES for regression analysis

Think of $b$ as data from experiments, $b$ is not in $S$, due to error of measurement.
QUESTION: Given a subspace $S$, what is the formula for the projection $p$ of a vector $b$ into $S$?

Key idea of **LEAST SQUARES** for regression analysis

Think of $b$ as data from experiments, $b$ is not in $S$, due to error of measurement.

Projection $p$ is the best choice to replace $b$. 
QUESTION: Given a subspace $S$, what is the formula for the projection $p$ of a vector $b$ into $S$?

Key idea of **LEAST SQUARES** for regression analysis

Think of $b$ as **data from experiments**, $b$ is not in $S$, due to error of measurement.

Projection $p$ is the best choice to replace $b$.

**How to do this projection for a line?**
• \( b \) is projected onto the line \( L \) given by the vector \( a \) in this picture:

\[
p = a^T b a^T a
\]

Exercise

We wish to project the vector \( b = (2, 3, 4) \) onto the \( z \)-axis...Find it!
- $b$ is projected onto the line $L$ given by the vector $a$ in this picture:

- The projection of vector $b$ onto the line in the direction $a$ is

$$p = \frac{a^T b}{a^T a}.$$
- $b$ is projected onto the line $L$ given by the vector $a$ in this picture:

$$p = \frac{a^T b}{a^T a}.$$

- The projection of vector $b$ onto the line in the direction $a$ is

- **Exercise** We wish to project the vector $b = (2, 3, 4)$ onto the $z$-axis...Find it!
Suppose the subspace $S = C(A)$ is the column space of $A$. Now $b$ is a vector that is outside the column space!
Suppose the subspace $S = C(A)$ is the column space of $A$. Now $b$ is a vector that is outside the column space!

We want $x$ that minimizes $\|b - Ax\|$. Then $p = Ax$ is the projection of $b$ onto $C(A)$.
Suppose the subspace $S = C(A)$ is the column space of $A$. Now $b$ is a vector that is outside the column space!

We want $x$ that minimizes $\|b - Ax\|$. Then $p = Ax$ is the projection of $b$ onto $C(A)$

Note that $\|b - Ax\|$ is the distance from $b$ to the point $Ax$ which is element of the column space!
Suppose the subspace \( S = C(A) \) is the column space of \( A \). Now \( b \) is a vector that is outside the column space!

We want \( x \) that minimizes \( \|b - Ax\| \). Then \( p = Ax \) is the projection of \( b \) onto \( C(A) \).

Note that \( \|b - Ax\| \) is the distance from \( b \) to the point \( Ax \) which is element of the column space!

The vector \( w = b - Ax \) must be perpendicular to the column space (like before).
Suppose the subspace $S = C(A)$ is the column space of $A$. Now $b$ is a vector that is outside the column space!

We want $x$ that minimizes $\|b - Ax\|$. Then $p = Ax$ is the projection of $b$ onto $C(A)$.

Note that $\|b - Ax\|$ is the distance from $b$ to the point $Ax$ which is element of the column space!

The vector $w = b - Ax$ must be perpendicular to the column space (like before).

For each column $a_i$ we have $a_i^T(b - Ax) = 0$. 

Thus in matrix notation:

$$A^T(b - Ax) = 0.$$ 

This gives the normal equation or least-squares equation:

$$A^TAx = A^Tb.$$
Suppose the subspace $S = C(A)$ is the column space of $A$. Now $b$ is a vector that is outside the column space!

We want $x$ that minimizes $\|b - Ax\|$. Then $p = Ax$ is the projection of $b$ onto $C(A)$.

Note that $\|b - Ax\|$ is the distance from $b$ to the point $Ax$ which is element of the column space!

The vector $w = b - Ax$ must be perpendicular to the column space (like before).

For each column $a_i$ we have $a_i^T(b - Ax) = 0$.

Thus in matrix notation: $A^T(b - Ax) = 0$. This gives the normal equation or least-squares equation:

$$A^T A x = A^T b$$
*Theorem* The solution \( x = (A^T A)^{-1} A^T b \) gives the coordinates of the projection \( p \) in terms of the columns of \( A \). The projection of \( b \) into \( C(A) \) is

\[
p = A((A^T A)^{-1} A^T) b
\]
Theorem The solution $x = (A^T A)^{-1} A^T b$ gives the coordinates of the projection $p$ in terms of the columns of $A$. The projection of $b$ into $C(A)$ is

$$p = A((A^T A)^{-1} A^T )b$$

Theorem The matrix $P = A((A^T A)^{-1} A^T )$ is a projection matrix. It has the properties $P^T = P$, and $P^2 = P$. 

**Lemma** $A^T A$ is a symmetric matrix. $A^T A$ has the same Nullspace as $A$. 

**Proof** if $x \in N(A)$, then clearly $A^T Ax = 0$. Conversely, if $A^T Ax = 0$ then $x^T A^T Ax = \|Ax\| = 0$, thus $Ax = 0$. 

**Corollary** If $A$ has independent columns, then $A^T A$ is square, symmetric and invertible. 

**Example 1** We wish to project the vector $b = (2, 3, 4, 1)$ into the subspace $x + y = 0$. What is the distance?
**Theorem** The solution \( x = (A^T A)^{-1} A^T b \) gives the coordinates of the projection \( p \) in terms of the columns of \( A \). The projection of \( b \) into \( C(A) \) is

\[
p = A((A^T A)^{-1} A^T) b
\]

**Theorem** The matrix \( P = A((A^T A)^{-1} A^T) \) is a projection matrix. It has the properties \( P^T = P \), and \( P^2 = P \).

**Lemma** \( A^T A \) is a symmetric matrix. \( A^T A \) has the same Nullspace as \( A \).
\textbf{Theorem} The solution \( x = (A^T A)^{-1} A^T b \) gives the coordinates of the projection \( p \) in terms of the columns of \( A \). The projection of \( b \) into \( C(A) \) is

\[ p = A((A^T A)^{-1} A^T) b \]

\textbf{Theorem} The matrix \( P = A((A^T A)^{-1} A^T) \) is a \textit{projection matrix}. It has the properties \( P^T = P \), and \( P^2 = P \).

\textbf{Lemma} \( A^T A \) is a symmetric matrix. \( A^T A \) has the same Nullspace as \( A \). \textbf{proof} if \( x \in N(A) \), then clearly \( A^T Ax = 0 \). Conversely, if \( A^T Ax = 0 \) then \( x^T A^T Ax = \|Ax\| = 0 \), thus \( Ax = 0 \).
Theorem The solution \( x = (A^T A)^{-1} A^T b \) gives the coordinates of the projection \( p \) in terms of the columns of \( A \). The projection of \( b \) into \( C(A) \) is

\[
p = A((A^T A)^{-1} A^T) b
\]

Theorem The matrix \( P = A((A^T A)^{-1} A^T) \) is a projection matrix. It has the properties \( P^T = P \), and \( P^2 = P \).

Lemma \( A^T A \) is a symmetric matrix. \( A^T A \) has the same Nullspace as \( A \). \textbf{proof} if \( x \in N(A) \), then clearly \( A^T Ax = 0 \). Conversely, if \( A^T Ax = 0 \) then \( x^T A^T Ax = \|Ax\| = 0 \), thus \( Ax = 0 \).

Corollary If \( A \) has independent columns, then \( A^T A \) is square, symmetric and invertible.
Theorem The solution \( x = (A^T A)^{-1} A^T b \) gives the coordinates of the projection \( p \) in terms of the columns of \( A \). The projection of \( b \) into \( C(A) \) is

\[
p = A((A^T A)^{-1} A^T )b
\]

.

Theorem The matrix \( P = A((A^T A)^{-1} A^T ) \) is a projection matrix. It has the properties \( P^T = P \), and \( P^2 = P \).

Lemma \( A^T A \) is a symmetric matrix. \( A^T A \) has the same Nullspace as \( A \). proof if \( x \in N(A) \), then clearly \( A^T Ax = 0 \). Conversely, if \( A^T Ax = 0 \) then \( x^T A^T Ax = \|Ax\| = 0 \), thus \( Ax = 0 \).

Corollary If \( A \) has independent columns, then \( A^T A \) is square, symmetric and invertible.

Example 1 We wish to project the vector \( b = (2, 3, 4, 1) \) into the subspace \( x + y = 0 \). What is the distance?
Example 2 Consider the problem $Ax = b$ with

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 3 & -1 & 1 \\ -1 & 2 & 1 \\ 1 & -1 & -2 \\ 2 & 1 & -1 \end{bmatrix} \quad b^T = (1, 0, -1, 2, 2).$$

There is no exact solution to $Ax = b$, so we use the "normal equation" with

$$A^T A = \begin{bmatrix} 16 & -2 & -2 \\ -2 & 11 & 2 \\ -2 & 2 & 7 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 8 \\ 0 \\ -7 \end{bmatrix}$$

Solving $A^T A x = A^T b$ we get the least squares solution $x^* \approx (0.4119, 0.2482, -0.9532)^T$ with error $\|b - Ax^*\| \approx 0.1799$. 
Example 2 Consider the problem $Ax = b$ with

$$A = \begin{bmatrix}
1 & 2 & 0 \\
3 & -1 & 1 \\
-1 & 2 & 1 \\
1 & -1 & -2 \\
2 & 1 & -1
\end{bmatrix} \quad \quad b^T = (1, 0, -1, 2, 2).$$

There is no exact solution to $Ax = b$, so we use the "NORMAL EQUATION" with

$$A^T A = \begin{bmatrix}
16 & -2 & -2 \\
-2 & 11 & 2 \\
-2 & 2 & 7
\end{bmatrix} \quad \quad A^T b = \begin{bmatrix}
8 \\
0 \\
-7
\end{bmatrix}$$

Solving $A^T A x = A^T b$ we get the least squares solution $x^* \approx (0.4119, 0.2482, -0.9532)^T$ with error $\|b - Ax^*\| \approx 0.1799$. 

MATH 22A: LINEAR ALGEBRA Chapter 4
Example 2 Consider the problem $Ax = b$ with

$$A = \begin{bmatrix}
1 & 2 & 0 \\
3 & -1 & 1 \\
-1 & 2 & 1 \\
1 & -1 & -2 \\
2 & 1 & -1 \\
\end{bmatrix} \quad b^T = (1, 0, -1, 2, 2).$$

There is no EXACT solution to $Ax = b$, so we use the "NORMAL EQUATION" with

$$A^T A = \begin{bmatrix}
16 & -2 & -2 \\
-2 & 11 & 2 \\
-2 & 2 & 7 \\
\end{bmatrix} \quad A^T b = \begin{bmatrix}
8 \\
0 \\
-7 \\
\end{bmatrix}$$

Solving $A^T Ax = A^T b$ we get the least squares solution $x^* \approx (0.4119, 0.2482, -0.9532)^T$ with error $\|b - Ax^*\| \approx 0.1799.$
Example 3  A sample of lead-210 measured the following radioactivity data at the given times (time in days). Can YOU predict how long will it take until one percent of the original amount remains?

<table>
<thead>
<tr>
<th>time in days</th>
<th>0</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>14</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>mg</td>
<td>10</td>
<td>8.8</td>
<td>7.8</td>
<td>7.3</td>
<td>6.4</td>
<td>6.4</td>
</tr>
</tbody>
</table>

A linear model does not work here!!

There is an exponential decay on the material $m(t) = m_0 e^{\beta t}$, where $m_0$ is the initial radioactive material and $\beta$ the decay rate.

Taking logarithms, we get

$$y(t) = \log(m(t)) = \log(m_0) + \beta t$$

We can use the usual least squares to fit to the logarithms $y_i = \log(m_i)$ of the radioactive mass data.
Example 3 A sample of lead-210 measured the following radioactivity data at the given times (time in days). Can YOU predict how long will it take until one percent of the original amount remains?

<table>
<thead>
<tr>
<th>time in days</th>
<th>0</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>14</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>mg</td>
<td>10</td>
<td>8.8</td>
<td>7.8</td>
<td>7.3</td>
<td>6.4</td>
<td>6.4</td>
</tr>
</tbody>
</table>

A linear model does not work here!!
Example 3 A sample of lead-210 measured the following radioactivity data at the given times (time in days). Can YOU predict how long will it take until one percent of the original amount remains?

<table>
<thead>
<tr>
<th>time in days</th>
<th>0</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>14</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>mg</td>
<td>10</td>
<td>8.8</td>
<td>7.8</td>
<td>7.3</td>
<td>6.4</td>
<td>6.4</td>
</tr>
</tbody>
</table>

A linear model does not work here!!

There is an exponential decay on the material $m(t) = m_0 e^{\beta t}$, where $m_0$ is the initial radioactive material and $\beta$ the decay rate.
Example 3 A sample of lead-210 measured the following radioactivity data at the given times (time in days). Can YOU predict how long will it take until one percent of the original amount remains?

<table>
<thead>
<tr>
<th>time in days</th>
<th>0</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>14</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>mg</td>
<td>10</td>
<td>8.8</td>
<td>7.8</td>
<td>7.3</td>
<td>6.4</td>
<td>6.4</td>
</tr>
</tbody>
</table>

A linear model does not work here!!

There is an exponential decay on the material $m(t) = m_0 e^{\beta t}$, where $m_0$ is the initial radioactive material and $\beta$ the decay rate.

Taking logarithms, we get

$$y(t) = \log(m(t)) = \log(m_0) + \beta t$$
Example 3 A sample of lead-210 measured the following radioactivity data at the given times (time in days). Can YOU predict how long will it take until one percent of the original amount remains?

<table>
<thead>
<tr>
<th>time in days</th>
<th>0</th>
<th>4</th>
<th>8</th>
<th>10</th>
<th>14</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>mg</td>
<td>10</td>
<td>8.8</td>
<td>7.8</td>
<td>7.3</td>
<td>6.4</td>
<td>6.4</td>
</tr>
</tbody>
</table>

A linear model does not work here!!

There is an exponential decay on the material \( m(t) = m_0 e^{\beta t} \), where \( m_0 \) is the initial radioactive material and \( \beta \) the decay rate.

Taking logarithms, we get

\[
y(t) = \log(m(t)) = \log(m_0) + \beta t
\]

We can use the usual least squares to fit to the logarithms \( y_i = \log(m_i) \) of the radioactive mass data.
In this case we have

\[ A^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 4 & 8 & 10 & 14 & 18 \end{bmatrix} \]

\[ y(t) = b^T = [2.30258, 2.17475, 2.05412, 1.98787, 1.856297, 1.85629] \]
In this case we have

\[
A^T = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 \\
0 & 4 & 8 & 10 & 14 & 18
\end{bmatrix}
\]

\[y(t) = b^T = [2.30258, 2.17475, 2.05412, 1.98787, 1.856297, 1.85629]\]

Thus \(A^T A = \begin{bmatrix}
6 & 54 \\
54 & 700
\end{bmatrix} \).
In this case we have

\[
A^T = \begin{bmatrix}
1 & 1 & 1 & 1 & 1 & 1 \\
0 & 4 & 8 & 10 & 14 & 18
\end{bmatrix}
\]

\[
y(t) = b^T = [2.30258, 2.17475, 2.05412, 1.98787, 1.856297, 1.85629]
\]

Thus \( A^T A = \begin{bmatrix}
6 & 54 \\
54 & 700
\end{bmatrix}. \)

Solving the NORMAL system we get \( \log(m_0) = 2.277327661 \)
and \( \beta = -0.0265191683 \)
In this case we have

$$A^T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 4 & 8 & 10 & 14 & 18 \end{bmatrix}$$

$$y(t) = b^T = [2.30258, 2.17475, 2.05412, 1.98787, 1.85629, 1.85629]$$

Thus $$A^T A = \begin{bmatrix} 6 & 54 \\ 54 & 700 \end{bmatrix}$$.

Solving the NORMAL system we get $$\log(m_0) = 2.277327661$$ and $$\beta = -0.0265191683$$

The original amount was 10 mg. After 173 days it will be below one percent of the radioactive material.
Orthogonal Bases and Gram-Schmidt
Not all bases of a vector space are created equal! Some are better than others!
Not all bases of a vector space are created equal! Some are better than others!

A basis $u_1, \ldots, u_n$ of a vector space $V$ is orthonormal if it is orthogonal and each vector has unit length.
Not all bases of a vector space are created equal! Some are better than others!

A basis \( u_1, \ldots, u_n \) of a vector space \( V \) is **orthonormal** if it is orthogonal and each vector has unit length.

**Observation** If the vectors \( u_1, \ldots, u_n \) are an orthogonal basis, their normalizations \( \frac{u_i}{\|u_i\|} \) form an orthonormal basis.
Not all bases of a vector space are created equal! Some are better than others!

A basis $u_1, \ldots, u_n$ of a vector space $V$ is orthonormal if it is orthogonal and each vector has unit length.

**Observation** If the vectors $u_1, \ldots, u_n$ are an orthogonal basis, their normalizations $\frac{u_i}{\|u_i\|}$ form an orthonormal basis.

**Example** The standard unit vectors are orthonormal.
Not all bases of a vector space are created equal! Some are better than others!

A basis $u_1, \ldots, u_n$ of a vector space $V$ is **orthonormal** if it is orthogonal and each vector has unit length.

**Observation** If the vectors $u_1, \ldots, u_n$ are an orthogonal basis, their normalizations $\frac{u_i}{\|u_i\|}$ form an orthonormal basis.

**Example** The standard unit vectors are orthonormal.

**Example** The vectors

$$
\begin{pmatrix}
1 \\
2 \\
-1
\end{pmatrix},
\begin{pmatrix}
0 \\
1 \\
2
\end{pmatrix},
\begin{pmatrix}
5 \\
-2 \\
1
\end{pmatrix}
$$

are an orthogonal basis of $\mathbb{R}^3$. 

Why do we care about orthonormal bases?

- **Theorem** Let $u_1, \ldots, u_n$ be an orthonormal bases for a vector space with inner product $V$. The one can write any element $v \in V$ as a linear combination $v = c_1 u_1 + \cdots + c_n u_n$ where $c_i = \langle v, u_i \rangle$, for $i = 1, \ldots, n$. 

Example Let us rewrite the vector $v = (1, 1, 1)^T$ in terms of the orthonormal basis $u_1 = (\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}})^T$, $u_2 = (0, \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}})^T$, $u_3 = (\frac{5}{\sqrt{30}}, -\frac{2}{\sqrt{30}}, \frac{1}{\sqrt{30}})^T$.

Computing the dot products $v^T u_1 = 2\sqrt{6}$, $v^T u_2 = 3\sqrt{5}$, and $v^T u_3 = 4\sqrt{30}$, we get $v = 2\sqrt{6} u_1 + 3\sqrt{5} u_2 + 4\sqrt{30} u_3$.

The best reason to like matrices that have orthonormal vectors: The least-squares equations are even nicer!
Why do we care about orthonormal bases?

- **Theorem** Let $u_1, \ldots, u_n$ be an orthonormal bases for a vector space with inner product $V$. The one can write any element $v \in V$ as a linear combination $v = c_1 u_1 + \cdots + c_n u_n$ where $c_i = \langle v, u_i \rangle$, for $i = 1, \ldots, n$.

- **Example** Let us rewrite the vector $v = (1, 1, 1)^T$ in terms of the orthonormal basis

  $$
  u_1 = \left( \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}} \right)^T, u_2 = (0, \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}), u_3 = \left( \frac{5}{\sqrt{30}}, \frac{-2}{\sqrt{30}}, \frac{1}{\sqrt{30}} \right)
  $$

The best reason to like matrices that have orthonormal vectors: The least-squares equations are even nicer!
Why do we care about orthonormal bases?

- **Theorem** Let $u_1, \ldots, u_n$ be an orthonormal bases for a vector space with inner product $V$. The one can write any element $v \in V$ as a linear combination $v = c_1 u_1 + \cdots + c_n u_n$ where $c_i = \langle v, u_i \rangle$, for $i = 1, \ldots, n$.

- **Example** Let us rewrite the vector $v = (1, 1, 1)^T$ in terms of the orthonormal basis

$$u_1 = \left( \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}} \right)^T, u_2 = \left( 0, \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right), u_3 = \left( \frac{5}{\sqrt{30}}, \frac{-2}{\sqrt{30}}, \frac{1}{\sqrt{30}} \right)$$

Computing the dot products $v^T u_1 = \frac{2}{\sqrt{6}}$, $v^T u_2 = \frac{3}{\sqrt{5}}$, and $v^T u_3 = \frac{4}{\sqrt{30}}$, we get

$$v = \frac{2}{\sqrt{6}} u_1 + \frac{3}{\sqrt{5}} u_2 + \frac{4}{\sqrt{30}} u_3$$
Why do we care about orthonormal bases?

- **Theorem** Let \( u_1, \ldots, u_n \) be an orthonormal bases for a vector space with inner product \( V \). The one can write any element \( v \in V \) as a linear combination \( v = c_1 u_1 + \cdots + c_n u_n \) where \( c_i = \langle v, u_i \rangle \), for \( i = 1, \ldots, n \).

- **Example** Let us rewrite the vector \( v = (1, 1, 1)^T \) in terms of the orthonormal basis

  \[
  u_1 = \left( \frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}, -\frac{1}{\sqrt{6}} \right)^T, \quad u_2 = \left( 0, \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right), \quad u_3 = \left( \frac{5}{\sqrt{30}}, \frac{-2}{\sqrt{30}}, \frac{1}{\sqrt{30}} \right)
  \]

  Computing the dot products \( v^T u_1 = \frac{2}{\sqrt{6}} \), \( v^T u_2 = \frac{3}{\sqrt{5}} \), and \( v^T u_3 = \frac{4}{\sqrt{30}} \), we get

  \[
  v = \frac{2}{\sqrt{6}} u_1 + \frac{3}{\sqrt{5}} u_2 + \frac{4}{\sqrt{30}} u_3
  \]

- The best reason to like matrices that have orthonormal vectors: The least-squares equations are even nicer!
Lemma If $Q$ is a rectangular matrix with orthonormal columns, then the normal equations simplify because $Q^T Q = I$:

- $Q^T Q x = Q^T b$ simplifies to $x = Q^T b$
- Projection matrix simplifies to $Q(Q^T Q)^{-1} Q^T = Q I Q^T = QQ^T$.
- Thus the projection point is $p = QQ^T b$, thus
  
  \[ p = (q_1^T b)q_1 + (q_2^T b)q_2 + \cdots + (q_n^T b)q_n \]
Lemma If $Q$ is a rectangular matrix with orthonormal columns, then the normal equations simplify because $Q^T Q = I$:

- $Q^T Q x = Q^T b$ simplifies to $x = Q^T b$
- Projection matrix simplifies to $Q(Q^T Q)^{-1} Q^T = Q I Q^T = Q Q^T$.
- Thus the projection point is $p = QQ^T b$, thus

$$p = (q_1^T b)q_1 + (q_2^T b)q_2 + \cdots + (q_n^T b)q_n$$

To compute orthogonal/orthonormal bases for a space, we use the GRAM-SCHMIDT ALGORITHM.
To compute orthogonal/orthonormal bases for a space, we use the **GRAM-SCHMIDT ALGORITHM**.

- **Step 1:** $q_1 = a_1$
- **Step 2:** $q_2 = a_2 - (a_2^T q_1) q_1$
- **Step 3:** $q_3 = a_3 - (a_3^T q_1) q_1 - (a_3^T q_2) q_2$
- **Step 4:** $q_4 = a_4 - (a_4^T q_1) q_1 - (a_4^T q_2) q_2 - (a_4^T q_3) q_3$
- **Step j:** $q_j = a_j - \left( a_j^T q_1 q_1 + \cdots + a_j^T q_{j-1} q_{j-1} \right) q_{j-1}$

At the end, normalize all vectors if you wish to have unit vectors! (Divide them by their length!)
To compute orthogonal/orthonormal bases for a space, we use the GRAM-SCHMIDT ALGORITHM.

**input:** linearly independent vectors $a_1, \ldots, a_n$

**output:** orthogonal vectors $q_1, \ldots, q_n$.

- **Step 1:** $q_1 = a_1$
- **Step 2:** $q_2 = a_2 - \left( \frac{a_2^T q_1}{q_1^T q_1} \right) q_1$
- **Step 3:** $q_3 = a_3 - \left( \frac{a_3^T q_1}{q_1^T q_1} \right) q_1 - \left( \frac{a_3^T q_2}{q_2^T q_2} \right) q_2$
- **Step 4:** $q_4 = a_4 - \left( \frac{a_4^T q_1}{q_1^T q_1} \right) q_1 - \left( \frac{a_4^T q_2}{q_2^T q_2} \right) q_2 - \left( \frac{a_4^T q_3}{q_3^T q_3} \right) q_3$
  \[ \vdots \]

- **Step $j$:** $q_j = a_j - \left( \frac{a_j^T q_1}{q_1^T q_1} \right) q_1 - \left( \frac{a_j^T q_2}{q_2^T q_2} \right) q_2 - \cdots - \left( \frac{a_j^T q_{j-1}}{q_{j-1}^T q_{j-1}} \right) q_{j-1}$

At the end, normalize all vectors if you wish to have unit vectors! (Divide them by their length!)
To compute orthogonal/orthonormal bases for a space, we use the **GRAM-SCHMIDT ALGORITHM**.

**input:** linearly independent vectors $a_1, \ldots, a_n$

**output:** orthogonal vectors $q_1, \ldots, q_n$.

- Step 1: $q_1 = a_1$
- Step 2: $q_2 = a_2 - (\frac{a_2^T q_1}{q_1^T q_1})q_1$
- Step 3: $q_3 = a_3 - (\frac{a_3^T q_1}{q_1^T q_1})q_1 - (\frac{a_3^T q_2}{q_2^T q_2})q_2$
- Step 4: $q_4 = a_4 - (\frac{a_4^T q_1}{q_1^T q_1})q_1 - (\frac{a_4^T q_2}{q_2^T q_2})q_2 - (\frac{a_4^T q_3}{q_3^T q_3})q_3$
- $\vdots$
- $\vdots$
- $\vdots$
- $\vdots$

- Step $j$: $q_j = a_j - (\frac{a_j^T q_1}{q_1^T q_1})q_1 - (\frac{a_j^T q_2}{q_2^T q_2})q_2 - \ldots - (\frac{a_j^T q_{j-1}}{q_{j-1}^T q_{j-1}})q_{j-1}$

At the end **NORMALIZE** all vectors if you wish to have unit vectors! (Divide them by their length!).
Consider the subspace $W$ spanned by $(1, -2, 0, 1)$, $(-1, 0, 0, -1)$ and $(1, 1, 0, 0)$. Find an orthonormal basis for the space $W$. 

\[
\begin{align*}
(1/\sqrt{6}, -2/\sqrt{6}, 0, 1/\sqrt{6}), \\
(-1/\sqrt{3}, -1/\sqrt{3}, 0, -1/\sqrt{3}), \\
(1/\sqrt{2}, 0, 0, -1/\sqrt{2})
\end{align*}
\]
Consider the subspace $W$ spanned by $(1, -2, 0, 1)$, $(-1, 0, 0, -1)$ and $(1, 1, 0, 0)$. Find an orthonormal basis for the space $W$.

**ANSWER:**

\[
\left( \frac{1}{\sqrt{6}}, \frac{-2}{\sqrt{6}}, 0, \frac{1}{\sqrt{6}} \right), \left( \frac{-1}{\sqrt{3}}, \frac{-1}{\sqrt{3}}, 0, \frac{-1}{\sqrt{3}} \right), \left( \frac{1}{\sqrt{2}}, 0, 0, \frac{-1}{\sqrt{2}} \right)
\]
In this way, the original basis vectors $a_1, \ldots, a_n$ can be written in a “triangular” way!

If $q_1, q_2, \ldots, q_n$ are orthogonal Just think of $r_{ij} = a_j^T q_i$

\[
a_1 = r_{11}(q_1/q_1^T q_1),
\]
\[
a_2 = r_{12}(q_1/q_1^T q_1) + r_{22}(q_2/q_2^T q_2)
\]
\[
a_3 = r_{13}(q_1/q_1^T q_1) + r_{23}(q_2/q_2^T q_2) + r_{33}(q_3/q_3^T q_3)
\]
\[
\vdots
\]
\[
a_n = r_{1n}(q_1/q_1^T q_1) + r_{2n}(q_2/q_2^T q_2) + \cdots + r_{nn}(q_n/q_n^T q_n).
\]
In this way, the original basis vectors \( a_1, \ldots, a_n \) can be written in a “triangular” way!

If \( q_1, q_2, \ldots, q_n \) are orthogonal Just think of \( r_{ij} = a_j^T q_i \)

\[
\begin{align*}
a_1 &= r_{11}(q_1 / q_1^T q_1), \\
a_2 &= r_{12}(q_1 / q_1^T q_1) + r_{22}(q_2 / q_2^T q_2) \\
a_3 &= r_{13}(q_1 / q_1^T q_1) + r_{23}(q_2 / q_2^T q_2) + r_{33}(q_3 / q_3^T q_3) \\
&\vdots \\
a_n &= r_{1n}(q_1 / q_1^T q_1) + r_{2n}(q_2 / q_2^T q_2) + \cdots + r_{nn}(q_n / q_n^T q_n).
\end{align*}
\]

Write these equations in matrix form! We obtain \( A = QR \)

where \( A = (a_1 \ \ldots \ a_n) \) and \( Q = (q_1 \ \ldots \ q_n) \) and \( R = (r_{ij}). \)
Theorem (QR decomposition) Every $m \times n$ matrix $A$ with independent columns can be factorized as $A = QR$ where the columns of $Q$ are orthonormal and $R$ is upper triangular and invertible.

NOTE: $A$ and $Q$ have the same column space. $R$ is an invertible and upper triangular matrix.

The simplest way to compute this decomposition:

1. Use Gram-Schmidt to get orthonormal vectors $q_i$.
2. Matrix $Q$ has columns $q_1, \ldots, q_n$.
3. The matrix $R$ is filled with the dot products $r_{ij} = a_j^T q_i$.

NOTE: Every matrix has two decompositions LU and QR. They are both useful for different reasons! One is for solving equations, the other good for least-squares.
**Theorem (QR decomposition)** Every $m \times n$ matrix $A$ with independent columns can be factorized as $A = QR$ where the columns of $Q$ are orthonormal and $R$ is upper triangular and invertible.

**NOTE:** $A$ and $Q$ have the same column space. $R$ is an invertible and upper triangular matrix.
Theorem (QR decomposition) Every $m \times n$ matrix $A$ with independent columns can be factorized as $A = QR$ where the columns of $Q$ are orthonormal and $R$ is upper triangular and invertible.

NOTE: $A$ and $Q$ have the same column space. $R$ is an invertible and upper triangular matrix.

The simplest way to compute this decomposition:
1. Use Gram-Schmidt to get orthonormal vectors $q_i$.
2. Matrix $Q$ has columns $q_1, \ldots, q_n$
3. The matrix $R$ is filled with the dot products $r_{ij} = a_j^T q_i$. 
Theorem (QR decomposition) Every $m \times n$ matrix $A$ with independent columns can be factorized as $A = QR$ where the columns of $Q$ are orthonormal and $R$ is upper triangular and invertible.

NOTE: $A$ and $Q$ have the same column space. $R$ is an invertible and upper triangular matrix.

The simplest way to compute this decomposition:
1. Use Gram-Schmidt to get orthonormal vectors $q_i$.
2. Matrix $Q$ has columns $q_1, \ldots, q_n$
3. The matrix $R$ is filled with the dot products $r_{ij} = a_j^T q_i$.

NOTE: Every matrix has two decompositions LU and QR.

They are both useful for different reasons! One is for solving equations, the other good for least-squares.