Orthogonality and Least Squares

6.1 INNER PRODUCT, LENGTH, AND ORTHOGONALITY
INNER PRODUCT

- If \( \mathbf{u} \) and \( \mathbf{v} \) are vectors in \( \mathbb{R}^n \), then we regard \( \mathbf{u} \) and \( \mathbf{v} \) as \( n \times 1 \) matrices.

- The transpose \( \mathbf{u}^T \) is a \( 1 \times n \) matrix, and the matrix product \( \mathbf{u}^T \mathbf{v} \) is a \( 1 \times 1 \) matrix, which we write as a single real number (a scalar) without brackets.

- The number \( \mathbf{u}^T \mathbf{v} \) is called the **inner product** of \( \mathbf{u} \) and \( \mathbf{v} \), and it is written as \( \mathbf{u} \cdot \mathbf{v} \).

- The inner product is also referred to as a **dot product**.
If \( u \) and \( v \) are vectors,

\[
\begin{bmatrix}
  u_1 \\
  u_2 \\
  \vdots \\
  u_n \\
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
  v_1 \\
  v_2 \\
  \vdots \\
  v_n \\
\end{bmatrix},
\]

then the inner product of \( u \) and \( v \) is

\[
\begin{bmatrix}
  u_1 & u_2 & \cdots & u_n \\
\end{bmatrix}
\begin{bmatrix}
  v_1 \\
  v_2 \\
  \vdots \\
  v_n \\
\end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.
\]
Theorem 1: Let $u$, $v$, and $w$ be vectors in $\mathbb{R}^n$, and let $c$ be a scalar. Then

a. $u \cdot v = v \cdot u$

b. $(u + v) \cdot w = u \cdot w + v \cdot w$

c. $(cu) \cdot v = c(u \cdot v) = u \cdot (cv)$

d. $u \cdot u \geq 0$, and $u \cdot u = 0$ if and only if $u = 0$

Properties (b) and (c) can be combined several times to produce the following useful rule:

$$(c_1u_1 + \cdots + c_p u_p) \cdot w = c_1(u_1 \cdot w) + \cdots + c_p(u_p \cdot w)$$
THE LENGTH OF A VECTOR

- If \( \mathbf{v} \) is in \( \mathbb{R}^n \), with entries \( v_1, \ldots, v_n \), then the square root of \( \mathbf{v} \cdot \mathbf{v} \) is defined because \( \mathbf{v} \cdot \mathbf{v} \) is nonnegative.

- **Definition:** The length (or norm) of \( \mathbf{v} \) is the nonnegative scalar \( \| \mathbf{v} \| \) defined by

  \[
  \| \mathbf{v} \| = \sqrt{\mathbf{v} \cdot \mathbf{v}} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}
  \text{ and } \| \mathbf{v} \|^2 = \mathbf{v} \cdot \mathbf{v}
  \]

- Suppose \( \mathbf{v} \) is in \( \mathbb{R}^2 \), say, \( \mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix} \).
THE LENGTH OF A VECTOR

- If we identify $\mathbf{v}$ with a geometric point in the plane, as usual, then $\|\mathbf{v}\|$ coincides with the standard notion of the length of the line segment from the origin to $\mathbf{v}$.
- This follows from the Pythagorean Theorem applied to a triangle such as the one shown in the following figure.

For any scalar $c$, the length $c\mathbf{v}$ is $|c|$ times the length of $\mathbf{v}$. That is,

$$\|c\mathbf{v}\| = |c||\mathbf{v}|$$
A vector whose length is 1 is called a **unit vector**.

If we *divide* a nonzero vector \(\mathbf{v}\) by its length—that is, multiply by \(1 / \|\mathbf{v}\|\)—we obtain a unit vector \(\mathbf{u}\) because the length of \(\mathbf{u}\) is \((1 / \|\mathbf{v}\|)\|\mathbf{v}\|\).

The process of creating \(\mathbf{u}\) from \(\mathbf{v}\) is sometimes called **normalizing** \(\mathbf{v}\), and we say that \(\mathbf{u}\) is *in the same direction* as \(\mathbf{v}\).
Example 1: Let \( v = (1, -2, 2, 0) \). Find a unit vector \( u \) in the same direction as \( v \).

Solution: First, compute the length of \( v \):

\[
\|v\|^2 = v \cdot v = (1)^2 + (-2)^2 + (2)^2 + (0)^2 = 9
\]

\[
\|v\| = \sqrt{9} = 3
\]

Then, multiply \( v \) by \( 1 / \|v\| \) to obtain

\[
u = \frac{1}{\|v\|} v = \frac{1}{3} v = \frac{1}{3} \begin{bmatrix} 1 \\ -2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \\ 0 \end{bmatrix}
\]
DISTANCE IN $\mathbb{R}^n$

- To check that $\|u\| = 1$, it suffices to show that $\|u\|^2 = 1$.

\[
\|u\|^2 = ugu = \left( \frac{1}{3} \right)^2 + \left( -\frac{2}{3} \right)^2 + \left( \frac{2}{3} \right)^2 + (0)^2
\]

\[
= \frac{1}{9} + \frac{4}{9} + \frac{4}{9} + 0 = 1
\]

- **Definition:** For $u$ and $v$ in $\mathbb{R}^n$, the distance between $u$ and $v$, written as $\text{dist} \ (u, v)$, is the length of the vector $u - v$. That is,

\[
\text{dist} \ (u,v) = \|u - v\|
\]
Example 2: Compute the distance between the vectors $u = (7,1)$ and $v = (3,2)$.

Solution: Calculate

$$u - v = \begin{bmatrix} 7 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

$$\|u - v\| = \sqrt{4^2 + (-1)^2} = \sqrt{17}$$

The vectors $u$, $v$, and $u - v$ are shown in the figure on the next slide.

When the vector $u - v$ is added to $v$, the result is $u$. 
The distance between \( \mathbf{u} \) and \( \mathbf{v} \) is the length of \( \mathbf{u} - \mathbf{v} \).

- Notice that the parallelogram in the above figure shows that the distance from \( \mathbf{u} \) to \( \mathbf{v} \) is the same as the distance from \( \mathbf{u} - \mathbf{v} \) to \( \mathbf{0} \).
ORTHOGONAL VECTORS

- Consider $\mathbb{R}^2$ or $\mathbb{R}^3$ and two lines through the origin determined by vectors $\mathbf{u}$ and $\mathbf{v}$.

- See the figure below. The two lines shown in the figure are geometrically perpendicular if and only if the distance from $\mathbf{u}$ to $\mathbf{v}$ is the same as the distance from $\mathbf{u}$ to $-\mathbf{v}$.

- This is the same as requiring the squares of the distances to be the same.
ORTHOGONAL VECTORS

- Now

\[
\left[ \text{dist}(u, -v) \right]^2 = \|u - (-v)\|^2 = \|u + v\|^2
\]

\[
= (u + v) \cdot (u + v)
\]

\[
= u \cdot (u + v) + v \cdot (u + v) \quad \text{Theorem 1(b)}
\]

\[
= u \cdot u + u \cdot v + v \cdot u + v \cdot v \quad \text{Theorem 1(a), (b)}
\]

\[
= \|u\|^2 + \|v\|^2 + 2u \cdot v \quad \text{Theorem 1(a)}
\]

- The same calculations with \(v\) and \(-v\) interchanged show that

\[
\left[ \text{dist}(u,v) \right]^2 = \|u\|^2 + \|-v\|^2 + 2u \cdot (-v)
\]

\[
= \|u\|^2 + \|v\|^2 - 2u \cdot v
\]
The two squared distances are equal if and only if $2u \cdot v = -2u \cdot v$, which happens if and only if $u \cdot v = 0$.

This calculation shows that when vectors $u$ and $v$ are identified with geometric points, the corresponding lines through the points and the origin are perpendicular if and only if $u \cdot v = 0$.

**Definition:** Two vectors $u$ and $v$ in $\mathbb{R}^n$ are orthogonal (to each other) if $u \cdot v = 0$.

The zero vector is orthogonal to every vector in $\mathbb{R}^n$ because $0^T v = 0$ for all $v$. 

© 2012 Pearson Education, Inc.
The Pythagorean Theorem

- Theorem 2: Two vectors \( \mathbf{u} \) and \( \mathbf{v} \) are orthogonal if and only if
  \[
  \| \mathbf{u} + \mathbf{v} \|^2 = \| \mathbf{u} \|^2 + \| \mathbf{v} \|^2.
  \]

- Orthogonal Complements
  - If a vector \( \mathbf{z} \) is orthogonal to every vector in a subspace \( W \) of \( \mathbb{R}^n \), then \( \mathbf{z} \) is said to be orthogonal to \( W \).
  - The set of all vectors \( \mathbf{z} \) that are orthogonal to \( W \) is called the orthogonal complement of \( W \) and is denoted by \( W^\perp \) (and read as “\( W \) perpendicular” or simply “\( W \) perp”).
ORTHOGONAL COMPLEMENTS

1. A vector $\mathbf{x}$ is in $W^\perp$ if and only if $\mathbf{x}$ is orthogonal to every vector in a set that spans $W$.

2. $W^\perp$ is a subspace of $\mathbb{R}^n$.

Theorem 3: Let $A$ be an $m \times n$ matrix. The orthogonal complement of the column space of $A$ is the null space of $A^T$

$$(\text{Col } A)^\perp = \text{Nul } A^T$$
Orthogonality and Least Squares

6.2 ORTHOGONAL SETS
ORTHOGONAL SETS

- A set of vectors \( \{u_1, \ldots, u_p\} \) in \( \mathbb{R}^n \) is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is, if \( u_i \cdot u_j = 0 \) whenever \( i \neq j \).

- **Theorem 4:** If \( S = \{u_1, \ldots, u_p\} \) is an orthogonal set of nonzero vectors in \( \mathbb{R}^n \), then \( S \) is linearly independent and hence is a basis for the subspace spanned by \( S \).
ORTHOGONAL SETS

- **Proof:** If \( 0 = c_1 u_1 + \cdots + c_p u_p \) for some scalars \( c_1, \ldots, c_p \), then

\[
0 = 0 \cdot u_1 = (c_1 u_1 + c_2 u_2 + \cdots + c_p u_p) \cdot u_1
\]

\[
= (c_1 u_1) \cdot u_1 + (c_2 u_2) \cdot u_1 + \cdots + (c_p u_p) \cdot u_1
\]

\[
= c_1 (u_1 \cdot u_1) + c_2 (u_2 \cdot u_1) + \cdots + c_p (u_p \cdot u_1)
\]

\[
= c_1 (u_1 \cdot u_1)
\]

because \( u_1 \) is orthogonal to \( u_2, \ldots, u_p \).

- Since \( u_1 \) is nonzero, \( u_1 \cdot u_1 \) is not zero and so \( c_1 = 0 \).

- Similarly, \( c_2, \ldots, c_p \) must be zero.
Thus $S$ is linearly independent.

**Definition:** An **orthogonal basis** for a subspace $W$ of $\mathbb{R}^n$ is a basis for $W$ that is also an orthogonal set.

**Theorem 5:** Let $\{u_1, \ldots, u_p\}$ be an orthogonal basis for a subspace $W$ of $\mathbb{R}^n$. For each $y$ in $W$, the weights in the linear combination

$$y = c_1 u_1 + \cdots + c_p u_p$$

are given by

$$c_j = \frac{y \cdot u_j}{u_j \cdot u_j} \quad (j = 1, K, p)$$
Proof: The orthogonality of \( \{\mathbf{u}_1, \ldots, \mathbf{u}_p\} \) shows that

\[
y \cdot \mathbf{u}_1 = (c_1 \mathbf{u}_1 + c_2 \mathbf{u}_2 + \cdots + c_p \mathbf{u}_p) \cdot \mathbf{u}_1 = c_1 (\mathbf{u}_1 \cdot \mathbf{u}_1)
\]

Since \( \mathbf{u}_1 \cdot \mathbf{u}_1 \) is not zero, the equation above can be solved for \( c_1 \).

To find \( c_j \) for \( j = 2, \ldots, p \), compute \( y \cdot \mathbf{u}_j \) and solve for \( c_j \).
Given a nonzero vector $\mathbf{u}$ in $\mathbb{R}^n$, consider the problem of decomposing a vector $\mathbf{y}$ in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of $\mathbf{u}$ and the other orthogonal to $\mathbf{u}$.

We wish to write

$$\mathbf{y} = \hat{\mathbf{y}} + \mathbf{z} \quad ----(1)$$

where $\hat{\mathbf{y}} = \alpha \mathbf{u}$ for some scalar $\alpha$ and $\mathbf{z}$ is some vector orthogonal to $\mathbf{u}$. See the following figure.
AN ORTHOGONAL PROJECTION

- Given any scalar $\alpha$, let $z = y - \alpha u$, so that (1) is satisfied.

- Then $y - \hat{y}$ is orthogonal to $u$ if and only if

$$0 = (y - \alpha u) \cdot u = y \cdot u - (\alpha u) \cdot u = y \cdot u - \alpha (u \cdot u)$$

- That is, (1) is satisfied with $z$ orthogonal to $u$ if and only if

$$\alpha = \frac{y \cdot u}{u \cdot u} \quad \text{and} \quad \hat{y} = \frac{y \cdot u}{u \cdot u} u.$$

- The vector $\hat{y}$ is called the orthogonal projection of $y$ onto $u$, and the vector $z$ is called the component of $y$ orthogonal to $u$. 
AN ORTHOGONAL PROJECTION

- If $c$ is any nonzero scalar and if $u$ is replaced by $cu$ in the definition of $\hat{y}$, then the orthogonal projection of $y$ onto $cu$ is exactly the same as the orthogonal projection of $y$ onto $u$.

- Hence this projection is determined by the subspace $L$ spanned by $u$ (the line through $u$ and 0).

- Sometimes $\hat{y}$ is denoted by $\text{proj}_L y$ and is called the orthogonal projection of $y$ onto $L$.

- That is,

$$\hat{y} = \text{proj}_L y = \frac{y \cdot u}{u \cdot u} \quad ----(2)$$
**Example 1:** Let \( \mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} \) and \( \mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \). Find the orthogonal projection of \( \mathbf{y} \) onto \( \mathbf{u} \). Then write \( \mathbf{y} \) as the sum of two orthogonal vectors, one in \( \text{Span} \{ \mathbf{u} \} \) and one orthogonal to \( \mathbf{u} \).

**Solution:** Compute

\[
\mathbf{y} \cdot \mathbf{u} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 40
\]

\[
\mathbf{u} \cdot \mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20
\]
AN ORTHOGONAL PROJECTION

- The orthogonal projection of $y$ onto $u$ is

\[
\hat{y} = \frac{y \cdot u}{u \cdot u} u = \frac{40}{20} u = 2 \begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}
\]

and the component of $y$ orthogonal to $u$ is

\[
y - \hat{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}
\]

- The sum of these two vectors is $y$. 
AN ORTHOGONAL PROJECTION

That is,

\[
\begin{bmatrix}
7 \\
6
\end{bmatrix}
= \begin{bmatrix}
8 \\
4
\end{bmatrix} + \begin{bmatrix}
-1 \\
2
\end{bmatrix}
\]

The decomposition of \( y \) is illustrated in the following figure.

The orthogonal projection of \( y \) onto a line \( L \) through the origin.
AN ORTHOGONAL PROJECTION

- **Note:** If the calculations above are correct, then \( \{\hat{y}, y - \hat{y}\} \) will be an orthogonal set.

- As a check, compute
  \[
  \hat{y} \cdot (y - \hat{y}) = \begin{bmatrix} 8 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} = -8 + 8 = 0
  \]

- Since the line segment in the figure on the previous slide between \( y \) and \( \hat{y} \) is perpendicular to \( L \), by construction of \( \hat{y} \), the point identified with \( \hat{y} \) is the closest point of \( L \) to \( y \).
ORTHONORMAL SETS

- A set \{u_1, \ldots, u_p\} is an **orthonormal set** if it is an orthogonal set of unit vectors.

- If \(W\) is the subspace spanned by such a set, then \{u_1, \ldots, u_p\} is an **orthonormal basis** for \(W\), since the set is automatically linearly independent, by Theorem 4.

- The simplest example of an orthonormal set is the standard basis \{e_1, \ldots, e_n\} for \(\mathbb{R}^n\).

- Any nonempty subset of \{e_1, \ldots, e_n\} is orthonormal, too.
ORTHONORMAL SETS

- **Example 2:** Show that $\{v_1, v_2, v_3\}$ is an orthonormal basis of $\mathbb{R}^3$, where

$$v_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \quad v_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \quad v_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

- **Solution:** Compute

$$v_1 \cdot v_2 = -3/\sqrt{66} + 2/\sqrt{66} + 1/\sqrt{66} = 0$$

$$v_1 \cdot v_3 = -3/\sqrt{726} - 4/\sqrt{726} + 7/\sqrt{726} = 0$$
ORTHONORMAL SETS

\[ \mathbf{v}_2 \cdot \mathbf{v}_3 = \frac{1}{\sqrt{396}} - \frac{8}{\sqrt{396}} + \frac{7}{\sqrt{396}} = 0 \]

- Thus \( \{ \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \} \) is an orthogonal set.
- Also, \( \mathbf{v}_1 \cdot \mathbf{v}_1 = \frac{9}{11} + \frac{1}{11} + \frac{1}{11} = 0 \)
  \[ \mathbf{v}_2 \cdot \mathbf{v}_2 = \frac{1}{6} + \frac{4}{6} + \frac{1}{6} = 1 \]
  \[ \mathbf{v}_3 \cdot \mathbf{v}_3 = \frac{1}{66} + \frac{16}{66} + \frac{49}{66} = 1 \]
  which shows that \( \mathbf{v}_1, \mathbf{v}_2, \) and \( \mathbf{v}_3 \) are unit vectors.
- Thus \( \{ \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \} \) is an orthonormal set.
- Since the set is linearly independent, its three vectors form a basis for \( \mathbb{R}^3 \). See the figure on the next slide.
When the vectors in an orthogonal set of nonzero vectors are *normalized* to have unit length, the new vectors will still be orthogonal, and hence the new set will be an orthonormal set.
Orthogonality and Least Squares

6.3 ORTHOGONAL PROJECTIONS
ORTHOGONAL PROJECTIONS

- The orthogonal projection of a point in $\mathbb{R}^2$ onto a line through the origin has an important analogue in $\mathbb{R}^n$.

- Given a vector $y$ and a subspace $W$ in $\mathbb{R}^n$, there is a vector $\hat{y}$ in $W$ such that (1) $\hat{y}$ is the unique vector in $W$ for which $y - \hat{y}$ is orthogonal to $W$, and (2) $\hat{y}$ is the unique vector in $W$ closest to $y$. See the following figure.
These two properties of $\hat{y}$ provide the key to finding the least-squares solutions of linear systems.

**Theorem 8:** Let $W$ be a subspace of $\mathbb{R}^n$. Then each $y$ in $\mathbb{R}^n$ can be written uniquely in the form

$$ y = \hat{y} + z $$

where $\hat{y}$ is in $W$ and $z$ is in $W^\perp$.

In fact, if $\{u_1, \ldots, u_p\}$ is any orthogonal basis of $W$, then

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

and $z = y - \hat{y}$. 
The vector \( \hat{y} \) in (1) is called the orthogonal projection of \( y \) onto \( W \) and often is written as \( \text{proj}_W y \). See the following figure.

**Proof:** Let \( \{u_1, \ldots, u_p\} \) be any orthogonal basis for \( W \), and define \( \hat{y} \) by (2).

Then \( \hat{y} \) is in \( W \) because \( \hat{y} \) is a linear combination of the basis \( u_1, \ldots, u_p \).
THE ORTHOGONAL DECOMPOSITION THEOREM

- Let $z = y - \hat{y}$.
- Since $u_1$ is orthogonal to $u_2, \ldots, u_p$, it follows from (2) that
  
  $z \cdot u_1 = (y - \hat{y}) \cdot u_1 = y \cdot u_1 - \left( \frac{y \cdot u_1}{u_1 \cdot u_1} \right) u_1 \cdot u_1 = 0 - \cdots - 0$

  
  $= y \cdot u_1 - y \cdot u_1 = 0$

- Thus $z$ is orthogonal to $u_1$.
- Similarly, $z$ is orthogonal to each $u_j$ in the basis for $W$.
- Hence $z$ is orthogonal to every vector in $W$.
- That is, $z$ is in $W^\perp$. 
THE ORTHOGONAL DECOMPOSITION THEOREM

- To show that the decomposition in (1) is unique, suppose \( y \) can also be written as \( y = \hat{y}_1 + z_1 \), with \( \hat{y}_1 \) in \( W \) and \( z_1 \) in \( W^\perp \).
- Then \( \hat{y} + z = \hat{y}_1 + z_1 \) (since both sides equal \( y \)), and so
  \[
  \hat{y} - \hat{y}_1 = z_1 - z
  \]
- This equality shows that the vector \( v = \hat{y} - \hat{y}_1 \) is in \( W \) and in \( W^\perp \) (because \( z_1 \) and \( z \) are both in \( W^\perp \), and \( W^\perp \) is a subspace).
- Hence \( v \cdot v = 0 \), which shows that \( v = 0 \).
- This proves that \( \hat{y} = \hat{y}_1 \) and also \( z_1 = z \).
The uniqueness of the decomposition (1) shows that the orthogonal projection $\hat{y}$ depends only on $W$ and not on the particular basis used in (2).

Example 1: Let $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix},$ and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$

Observe that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}.$ Write $\mathbf{y}$ as the sum of a vector in $W$ and a vector orthogonal to $W.$
THE ORTHOGONAL DECOMPOSITION THEOREM

- **Solution:** The orthogonal projection of $y$ onto $W$ is

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2$$

$$= \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{15}{30} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

- Also

$$y - \hat{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$
Theorem 8 ensures that \( y - \hat{y} \) is in \( W^\perp \).

To check the calculations, verify that \( y - \hat{y} \) is orthogonal to both \( u_1 \) and \( u_2 \) and hence to all of \( W \).

The desired decomposition of \( y \) is

\[
y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}
\]
If \( \{u_1, \ldots, u_p\} \) is an orthogonal basis for \( W \) and if \( y \) happens to be in \( W \), then the formula for \( \text{proj}_W y \) is exactly the same as the representation of \( y \) given in Theorem 5 in Section 6.2.

In this case, \( \text{proj}_W y = y \).

If \( y \) is in \( W = \text{Span}\{u_1, \ldots, u_p\} \), then \( \text{proj}_W y = y \).
Theorem 9: Let $W$ be a subspace of $\mathbb{R}^n$, let $y$ be any vector in $\mathbb{R}^n$, and let $\hat{y}$ be the orthogonal projection of $y$ onto $W$. Then $\hat{y}$ is the closest point in $W$ to $y$, in the sense that

$$\|y - \hat{y}\| < \|y - v\|$$

for all $v$ in $W$ distinct from $\hat{y}$.

The vector $\hat{y}$ in Theorem 9 is called the best approximation to $y$ by elements of $W$.

The distance from $y$ to $v$, given by $\|y - v\|$, can be regarded as the “error” of using $v$ in place of $y$.

Theorem 9 says that this error is minimized when $v = \hat{y}$. 
Inequality (3) leads to a new proof that $\hat{y}$ does not depend on the particular orthogonal basis used to compute it.

If a different orthogonal basis for $W$ were used to construct an orthogonal projection of $y$, then this projection would also be the closest point in $W$ to $y$, namely, $\hat{y}$. 
THE BEST APPROXIMATION THEOREM

- **Proof:** Take \( v \) in \( W \) distinct from \( \hat{y} \). See the following figure.

![Diagram](image)

The orthogonal projection of \( y \) onto \( W \) is the closest point in \( W \) to \( y \).

- Then \( \hat{y} - v \) is in \( W \).
- By the Orthogonal Decomposition Theorem, \( y - \hat{y} \) is orthogonal to \( W \).
- In particular, \( y - \hat{y} \) is orthogonal to \( \hat{y} - v \) (which is in \( W \)).
Since

\[ y - v = (y - \hat{y}) + (\hat{y} - v) \]

the Pythagorean Theorem gives

\[ \|y - v\|^2 = \|y - \hat{y}\|^2 + \|\hat{y} - v\|^2 \]

(See the colored right triangle in the figure on the previous slide. The length of each side is labeled.)

Now \(\|\hat{y} - v\|^2 > 0\) because \(\hat{y} - v \neq 0\), and so inequality (3) follows immediately.
Example 2: The distance from a point $y$ in $\mathbb{R}^n$ to a subspace $W$ is defined as the distance from $y$ to the nearest point in $W$. Find the distance from $y$ to $W = \text{Span}\{u_1, u_2\}$, where

$$y = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \quad u_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \quad u_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

Solution: By the Best Approximation Theorem, the distance from $y$ to $W$ is $\|y - \hat{y}\|$, where $\hat{y} = \text{proj}_W y$. 
PROPERTIES OF ORTHOGONAL PROJECTIONS

- Since \( \{\mathbf{u}_1, \mathbf{u}_2\} \) is an orthogonal basis for \( W \),
  \[
  \hat{y} = \frac{15}{30} \mathbf{u}_1 + \frac{-21}{6} \mathbf{u}_2 = \frac{1}{2} \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{7}{2} \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \end{bmatrix}
  \]

- \( y - \hat{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix} \)

- \( \|y - \hat{y}\|^2 = 3^2 + 6^2 = 45 \)

- The distance from \( y \) to \( W \) is \( \sqrt{45} = 3\sqrt{5} \).
Orthogonality and Least Squares

6.5

LEAST-SQUARES PROBLEMS
Definition: If $A$ is $m \times n$ and $b$ is in $\mathbb{R}^m$, a least-squares solution of $Ax = b$ is an $\hat{x}$ in $\mathbb{R}^n$ such that

$$
\|b - A\hat{x}\| \leq \|b - Ax\|
$$

for all $x$ in $\mathbb{R}^n$.

The most important aspect of the least-squares problem is that no matter what $x$ we select, the vector $Ax$ will necessarily be in the column space, $\text{Col } A$.

So we seek an $x$ that makes $Ax$ the closest point in $\text{Col } A$ to $b$. See the figure on the next slide.
Solution of the General Least-Squares Problem

Given $A$ and $b$, apply the Best Approximation Theorem to the subspace $\text{Col } A$.

Let

$$\hat{b} = \text{proj}_{\text{Col } A} b$$
Because $\hat{b}$ is in the column space $A$, the equation $Ax = \hat{b}$ is consistent, and there is an $\hat{x}$ in $\mathbb{R}^n$ such that

$$A\hat{x} = \hat{b} \quad \text{----(1)}$$

Since $\hat{b}$ is the closest point in $\text{Col } A$ to $b$, a vector $\hat{x}$ is a least-squares solution of $Ax = b$ if and only if $\hat{x}$ satisfies (1).

Such an $\hat{x}$ in $\mathbb{R}^n$ is a list of weights that will build $\hat{b}$ out of the columns of $A$. See the figure on the next slide.
SOLUTION OF THE GENERAL LEAST-SQUARES PROBLEM

- Suppose \( \hat{x} \) satisfies \( A \hat{x} = \hat{b} \).
- By the Orthogonal Decomposition Theorem, the projection \( \hat{b} \) has the property that \( b - \hat{b} \) is orthogonal to \( \text{Col} \ A \), so \( b - A \hat{x} \) is orthogonal to each column of \( A \).
- If \( a_j \) is any column of \( A \), then \( a_j \cdot (b - A \hat{x}) = 0 \), and \( a_j^T (b - A \hat{x}) \).
Since each $a_j^T$ is a row of $A^T$,
\[ A^T (b - A\hat{x}) = 0 \] ------(2)

Thus
\[ A^T b - A^T A\hat{x} = 0 \]
\[ A^T A\hat{x} = A^T b \]

These calculations show that each least-squares solution of $Ax = b$ satisfies the equation
\[ A^T A x = A^T b \] ------(3)

The matrix equation (3) represents a system of equations called the **normal equations** for $Ax = b$.
A solution of (3) is often denoted by $\hat{x}$. 
Theorem 13: The set of least-squares solutions of $Ax = b$ coincides with the nonempty set of solutions of the normal equation $A^T Ax = A^T b$.

Proof: The set of least-squares solutions is nonempty and each least-squares solution $\hat{x}$ satisfies the normal equations.

Conversely, suppose $\hat{x}$ satisfies $A^T A\hat{x} = A^T b$.

Then $\hat{x}$ satisfies (2), which shows that $b - A\hat{x}$ is orthogonal to the rows of $A^T$ and hence is orthogonal to the columns of $A$.

Since the columns of $A$ span $\text{Col } A$, the vector $b - A\hat{x}$ is orthogonal to all of $\text{Col } A$. 
Hence the equation
\[ b = A\hat{x} + (b - A\hat{x}) \]
is a decomposition of \( b \) into the sum of a vector in \( \text{Col} \ A \) and a vector orthogonal to \( \text{Col} \ A \).

By the uniqueness of the orthogonal decomposition, \( A\hat{x} \) must be the orthogonal projection of \( b \) onto \( \text{Col} \ A \).

That is, \( A\hat{x} = \hat{b} \) and \( \hat{x} \) is a least-squares solution.
Example 1: Find a least-squares solution of the inconsistent system $Ax = b$ for

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix}$$

Solution: To use normal equations (3), compute:

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$
Then the equation $A^T A x = A^T b$ becomes

$$
\begin{bmatrix}
17 & 1 \\
1 & 5
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2
\end{bmatrix}
= 
\begin{bmatrix}
19 \\
11
\end{bmatrix}
$$
Row operations can be used to solve the system on the previous slide, but since $A^T A$ is invertible and $2 \times 2$, it is probably faster to compute

$$(A^T A)^{-1} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

and then solve $A^T A \hat{x} = A^T b$ as

$$\hat{x} = (A^T A)^{-1} A^T b$$

$$= \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$
Theorem 14: Let $A$ be an $m \times n$ matrix. The following statements are logically equivalent:

a. The equation $Ax = b$ has a unique least-squares solution for each $b$ in $\mathbb{R}^m$.

b. The columns of $A$ are linearly independent.

c. The matrix $A^TA$ is invertible.

When these statements are true, the least-squares solution $\hat{x}$ is given by

$$\hat{x} = (A^T A)^{-1} A^T b$$

When a least-squares solution $\hat{x}$ is used to produce $A\hat{x}$ as an approximation to $b$, the distance from $b$ to $A\hat{x}$ is called the least-squares error of this approximation.
**Example 2:** Find a least-squares solution of $Ax = b$ for

$$
A = \begin{bmatrix}
1 & -6 \\
1 & -2 \\
1 & 1 \\
1 & 7
\end{bmatrix},
\quad
b = \begin{bmatrix}
-1 \\
2 \\
1 \\
6
\end{bmatrix}
$$

**Solution:** Because the columns $a_1$ and $a_2$ of $A$ are orthogonal, the orthogonal projection of $b$ onto $\text{Col } A$ is given by

$$
\hat{b} = \frac{b \cdot a_1}{a_1 \cdot a_1} a_1 + \frac{b \cdot a_2}{a_2 \cdot a_2} a_2
= \frac{8}{4} a_1 + \frac{45}{90} a_2
= \frac{8}{4} a_1 + \frac{1}{2} a_2
$$

----(5)
ALTERNATIVE CALCULATIONS OF LEAST-SQUARES SOLUTIONS

\[ \hat{\mathbf{b}} = \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix} \]

- Now that \( \hat{\mathbf{b}} \) is known, we can solve \( A\hat{\mathbf{x}} = \hat{\mathbf{b}} \).
- But this is trivial, since we already know weights to place on the columns of \( A \) to produce \( \hat{\mathbf{b}} \).
- It is clear from (5) that

\[
\hat{\mathbf{x}} = \begin{bmatrix} 8/4 \\ 45/90 \end{bmatrix} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}
\]