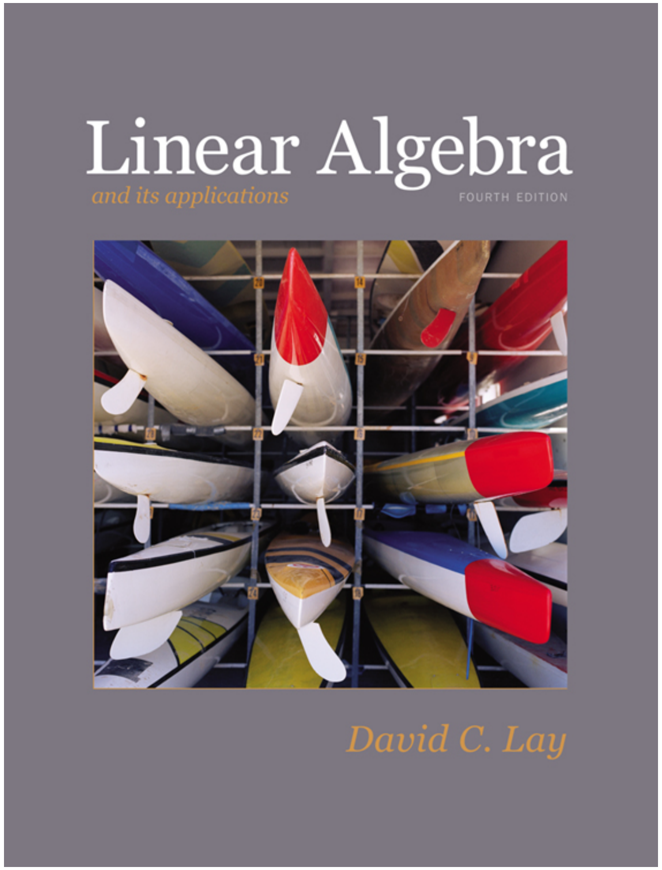
6 Orthogonality and Least Squares

6.1

INNER PRODUCT, LENGTH, AND ORTHOGONALITY



PEARSON

INNER PRODUCT

- If **u** and **v** are vectors in $\mathbb{R}^n$, then we regard **u** and **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 1×1 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 **u** and **v**, and it is written as $\mathbf{u} \cdot \mathbf{v}$.

- The inner product is also referred to as a **dot product**.

INNER PRODUCT

- If $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$,

  then the inner product of **u** and **v** is

$$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \ .$$

INNER PRODUCT

- **Theorem 1:** Let **u**, **v**, and **w** be vectors in $\mathbb{R}^n$, and let c be a scalar. Then
  a. $\mathrm{u} \bullet \mathrm{v} = \mathrm{v} \bullet \mathrm{u}$
  b. $(\mathrm{u} + \mathrm{v}) \bullet \mathrm{w} = \mathrm{u} \bullet \mathrm{w} + \mathrm{v} \bullet \mathrm{w}$
  c. $(c\mathrm{u}) \bullet \mathrm{v} = c(\mathrm{u} \bullet \mathrm{v}) = \mathrm{u} \bullet (c\mathrm{v})$
  d. $\mathrm{u} \bullet \mathrm{u} \geq 0$, and $\mathrm{u} \bullet \mathrm{u} = 0$ if and only if $\mathrm{u} = 0$

- Properties (b) and (c) can be combined several times to produce the following useful rule:
$$(c_1\mathrm{u}_1 + \cdots + c_p\mathrm{u}_p) \bullet \mathrm{w} = c_1(\mathrm{u}_1 \bullet \mathrm{w}) + \cdots + c_p(\mathrm{u}_p \bullet \mathrm{w})$$

THE LENGTH OF A VECTOR

- If **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.

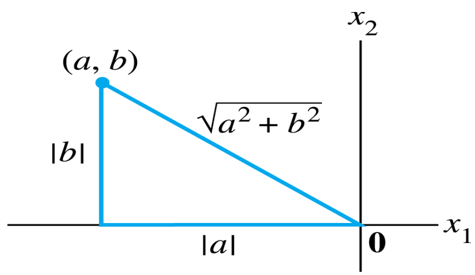
- **Definition:** The **length** (or **norm**) of **v** is the nonnegative scalar $\|\mathbf{v}\|$ defined by

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

- Suppose **v** is in $\mathbb{R}^2$, say, $\mathbf{v} = \begin{bmatrix} a \\ b \end{bmatrix}$.

THE LENGTH OF A VECTOR

- If we identify **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 **v**.
- This follows from the Pythagorean Theorem applied to a triangle such as the one shown in the following figure.



Interpretation of $\|\mathbf{v}\|$ as length.

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

$$\|c\mathbf{v}\| = |c|\|\mathbf{v}\|$$

THE LENGTH OF A VECTOR

- A vector whose length is 1 is called a **unit vector**.
- If we *divide* a nonzero vector **v** by its length—that is, multiply by $1/\|\mathbf{v}\|$—we obtain a unit vector **u** because the length of **u** is $(1/\|\mathbf{v}\|)\|\mathbf{v}\|$.
- The process of creating **u** from **v** is sometimes called **normalizing v**, and we say that **u** is *in the same direction* as **v**.

THE LENGTH OF A VECTOR

- **Example 1:** Let $\mathrm{v} = (1, -2, 2, 0)$. Find a unit vector **u** in the same direction as **v**.
- **Solution:** First, compute the length of **v**:

$$\|\mathrm{v}\|^2 = \mathrm{v} \cdot \mathrm{v} = (1)^2 + (-2)^2 + (2)^2 + (0)^2 = 9$$

$$\|\mathrm{v}\| = \sqrt{9} = 3$$

- Then, multiply **v** by $1 / \|\mathrm{v}\|$ to obtain

$$\mathrm{u} = \frac{1}{\|\mathrm{v}\|}\mathrm{v} = \frac{1}{3}\mathrm{v} = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \\ 0 \end{bmatrix}$$

DISTANCE IN $\mathbb{R}^n$

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

$$\|\mathbf{u}\|^2 = \mathbf{u}\cdot\mathbf{u} = \left(\frac{1}{3}\right)^2 + \left(-\frac{2}{3}\right) + \left(\frac{2}{3}\right)^2 + (0)^2$$

$$= \frac{1}{9} + \frac{4}{9} + \frac{4}{9} + 0 = 1$$

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

$$\text{dist}\,(\mathbf{u},\mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

DISTANCE IN $\mathbb{R}^n$

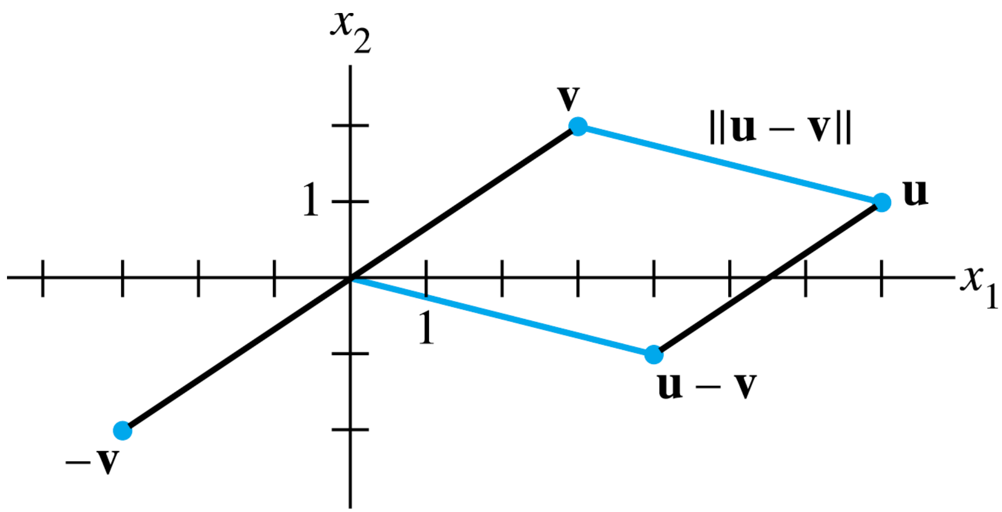
- **Example 2:** Compute the distance between the vectors $\mathbf{u} = (7,1)$ and $\mathbf{v} = (3,2)$.
- **Solution:** Calculate

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

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

- The vectors **u**, **v**, and $\mathbf{u} - \mathbf{v}$ are shown in the figure on the next slide.
- When the vector $\mathbf{u} - \mathbf{v}$ is added to **v**, the result is **u**.

DISTANCE IN $\mathbb{R}^n$

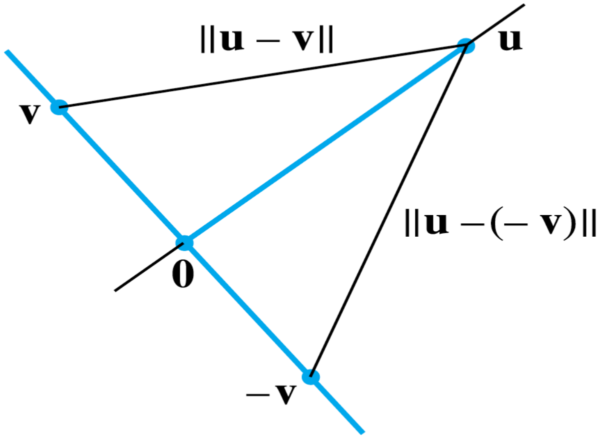


The distance between **u** and **v** is the length of **u** – **v**.

- Notice that the parallelogram in the above figure shows that the distance from **u** to **v** is the same as the distance from u – v to **0**.

ORTHOGONAL VECTORS

- Consider $\mathbb{R}^2$ or $\mathbb{R}^3$ and two lines through the origin determined by vectors **u** and **v**.
- See the figure below. The two lines shown in the figure are geometrically perpendicular if and only if the distance from **u** to **v** is the same as the distance from **u** to $-$v.



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

ORTHOGONAL VECTORS

- Now

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

$$= (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v})$$

$$= \mathbf{u} \cdot (\mathbf{u} + \mathbf{v}) + \mathbf{v} \cdot (\mathbf{u} + \mathbf{v}) \qquad \text{Theorem 1(b)}$$

$$= \mathbf{u} \cdot \mathbf{u} + \mathbf{u} \cdot \mathbf{v} + \mathbf{v} \cdot \mathbf{u} + \mathbf{v} \cdot \mathbf{v} \qquad \text{Theorem 1(a), (b)}$$

$$= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\mathbf{u} \cdot \mathbf{v} \qquad \text{Theorem 1(a)}$$

- The same calculations with **v** and $-\mathbf{v}$ interchanged show that $\left[\operatorname{dist}(\mathbf{u}, \mathbf{v})\right]^2 = \|\mathbf{u}\|^2 + \|-\mathbf{v}\|^2 + 2\mathbf{u} \cdot (-\mathbf{v})$

$$= \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 - 2\mathbf{u} \cdot \mathbf{v}$$

ORTHOGONAL VECTORS

- The two squared distances are equal if and only if $2\mathbf{u} \cdot \mathbf{v} = -2\mathbf{u} \cdot \mathbf{v}$, which happens if and only if $\mathbf{u} \cdot \mathbf{v} = 0$.
- 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 $\mathbf{u} \cdot \mathbf{v} = 0$.

- **Definition:** Two vectors **u** and **v** in $\mathbb{R}^n$ are **orthogonal** (to each other) if $\mathbf{u} \cdot \mathbf{v} = 0$.
- The zero vector is orthogonal to every vector in $\mathbb{R}^n$ because $\mathbf{0}^T \mathbf{v} = 0$ for all **v**.

THE PYTHOGOREAN THEOREM

- **Theorem 2:** Two vectors **u** and **v** are orthogonal if and only if $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

- **Orthogonal Complements**
- If a vector **z** is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then **z** is said to be **orthogonal to** W.
- The set of all vectors **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 **x** is in $W^{\perp}$ if and only if **x** is orthogonal to every vector in a set that spans W.
2. $W^{\perp}$ is a subspace of $\mathbb{R}^n$.

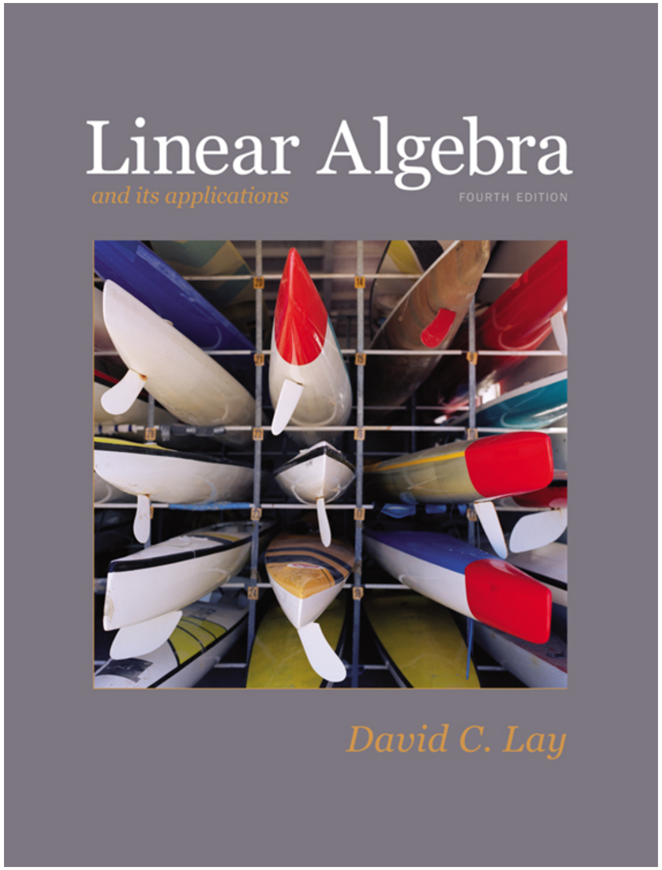
- **Theorem 3:** Let A be an $m \times n$ matrix. The orthogonal complement of the column space of A is the null space of A^T

$$(\text{Col } A)^{\perp} = \text{Nul } A^T$$

6 Orthogonality and Least Squares

6.2

ORTHOGONAL SETS



ORTHOGONAL SETS

- A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ 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 $\mathbf{u}_i \cdot \mathbf{u}_j = 0$ whenever $i \neq j$.

- **Theorem 4:** If $S = \{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ 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 $0 = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ for some scalars $c_1, \ldots, c_p$, then

$$0 = 0 \cdot \mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p) \cdot \mathbf{u}_1$$

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

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

$$= c_1(\mathbf{u}_1 \cdot \mathbf{u}_1)$$

  because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \ldots, \mathbf{u}_p$.

- Since $\mathbf{u}_1$ is nonzero, $\mathbf{u}_1 \cdot \mathbf{u}_1$ is not zero and so $c_1 = 0$.
- Similarly, $c_2, \ldots, c_p$ must be zero.

ORTHOGONAL SETS

- 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 $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination

$$\mathrm{y} = c_1\mathrm{u}_1 + \cdots + c_p\mathrm{u}_p$$

are given by

$$c_j = \frac{\mathrm{y}\bullet\mathrm{u}_j}{\mathrm{u}_j\bullet\mathrm{u}_j} \qquad (j = 1,\ldots,p)$$

ORTHOGONAL SETS

- **Proof:** The orthogonality of $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ shows that

$$\mathbf{y} \bullet \mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p) \bullet \mathbf{u}_1 = c_1(\mathbf{u}_1 \bullet \mathbf{u}_1)$$

- Since $\mathbf{u}_1 \bullet \mathbf{u}_1$ is not zero, the equation above can be solved for c_1.

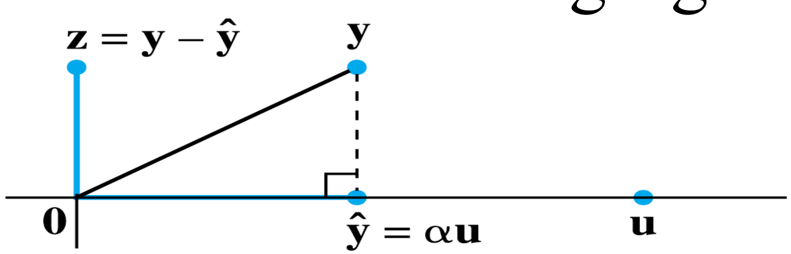
- To find c_j for $j = 2, \ldots, p$, compute $\mathbf{y} \bullet \mathbf{u}_j$ and solve for c_j.

AN ORTHOGONAL PROJECTION

- Given a nonzero vector **u** in $\mathbb{R}^n$, consider the problem of decomposing a vector **y** in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of **u** and the other orthogonal to **u**.
- We wish to write

$$\mathrm{y} = \hat{\mathrm{y}} + \mathrm{z} \qquad \text{----(1)}$$

where $\hat{\mathrm{y}} = \alpha \mathrm{u}$ for some scalar α and **z** is some vector orthogonal to **u**. See the following figure.



Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

AN ORTHOGONAL PROJECTION

- Given any scalar α, let $z = y - \alpha u$, so that (1) is satisfied.
- Then $y - \hat{y}$ is orthogonal to **u** if an only if

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

- That is, (1) is satisfied with **z** orthogonal to **u** if and only if $\alpha = \dfrac{y \bullet u}{u \bullet u}$ and $\hat{y} = \dfrac{y \bullet u}{u \bullet u} u$.
- 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 $\mathbf{u}$ is replaced by $c\mathbf{u}$ in the definition of $\hat{\mathbf{y}}$, then the orthogonal projection of $\mathbf{y}$ onto $c\mathbf{u}$ is exactly the same as the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$.
- Hence this projection is determined by the *subspace L* spanned by $\mathbf{u}$ (the line through $\mathbf{u}$ and $\mathbf{0}$).
- Sometimes $\hat{\mathbf{y}}$ is denoted by $\text{proj}_L \mathbf{y}$ and is called the **orthogonal projection of y onto** L.
- That is,

$$\hat{\mathbf{y}} = \text{proj}_L \mathbf{y} = \frac{\mathbf{y} \bullet \mathbf{u}}{\mathbf{u} \bullet \mathbf{u}} \mathbf{u} \qquad \text{----(2)}$$

AN ORTHOGONAL PROJECTION

- **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 **y** onto **u**. Then write **y** as the sum of two orthogonal vectors, one in Span $\{\mathbf{u}\}$ and one orthogonal to **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

$$\hat{y} = \frac{y \bullet u}{u \bullet u} u = \frac{40}{20} u = 2\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

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

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

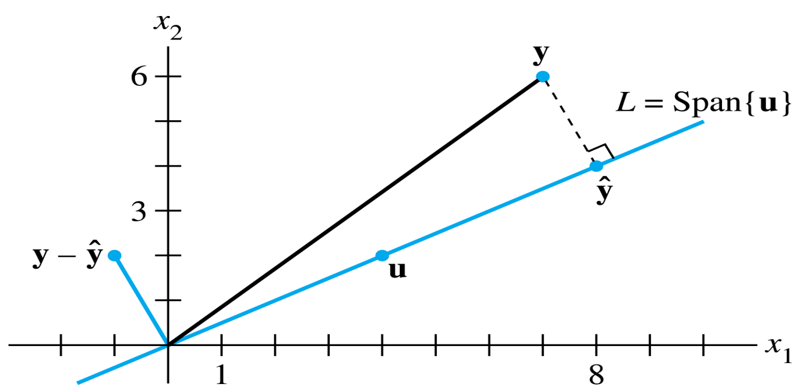
- The sum of these two vectors is **y**.

AN ORTHOGONAL PROJECTION

- That is,

$$\underset{\mathbf{y}}{\begin{bmatrix} 7 \\ 6 \end{bmatrix}} = \underset{\hat{\mathbf{y}}}{\begin{bmatrix} 8 \\ 4 \end{bmatrix}} + \underset{(\mathbf{y}-\hat{\mathbf{y}})}{\begin{bmatrix} -1 \\ 2 \end{bmatrix}}$$

- 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

$$\hat{y} \bullet (y - \hat{y}) = \begin{bmatrix} 8 \\ 4 \end{bmatrix} \bullet \begin{bmatrix} -1 \\ 2 \end{bmatrix} = -8 + 8 = 0$$

- 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 $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors.

- If W is the subspace spanned by such a set, then $\{\mathbf{u}_1, \ldots,\mathbf{u}_p\}$ 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 $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ for $\mathbb{R}^n$.

- Any nonempty subset of $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ is orthonormal, too.

ORTHONORMAL SETS

- **Example 2:** Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of $\mathbb{R}^3$, where

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

- **Solution:** Compute

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

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

ORTHONORMAL SETS

$$v_2 \cdot v_3 = 1/\sqrt{396} - 8/\sqrt{396} + 7/\sqrt{396} = 0$$

- Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set.
- Also, $\quad v_1 \bullet v_1 = 9/11 + 1/11 + 1/11 = 0$

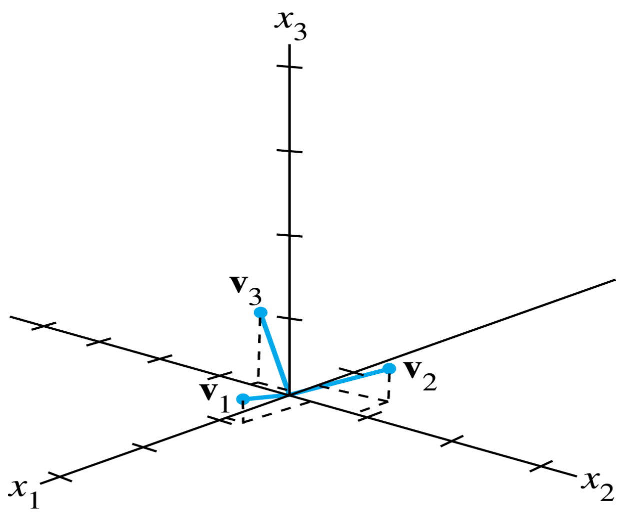
$$v_2 \bullet v_2 = 1/6 + 4/6 + 1/6 = 1$$

$$v_3 \bullet v_3 = 1/66 + 16/66 + 49/66 = 1$$

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.

ORTHONORMAL SETS

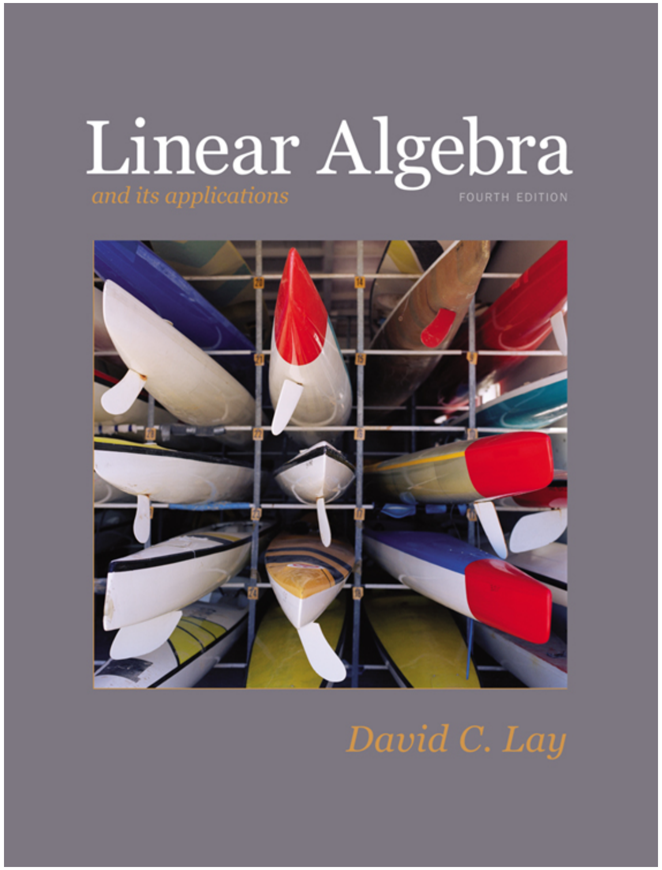


- When the vectors in an orthogonal set of nonzero vectors are *normalized* to have unit length, the new vectors will still be orthogonal, and hence the new set will be an orthonormal set.

6 Orthogonality and Least Squares

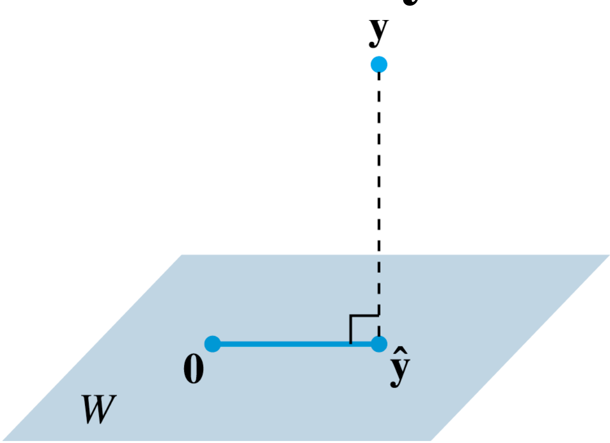
6.3

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 $\mathbf{y}$ and a subspace W in $\mathbb{R}^n$, there is a vector $\hat{\mathbf{y}}$ in W such that (1) $\hat{\mathbf{y}}$ is the unique vector in W for which $\mathbf{y}-\hat{\mathbf{y}}$ is orthogonal to W, and (2) $\hat{\mathbf{y}}$ is the unique vector in W closest to $\mathbf{y}$. See the following figure.



THE ORTHOGONAL DECOMPOSITION THEOREM

- 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 $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

$$y = \hat{y} + z \qquad \text{----(1)}$$

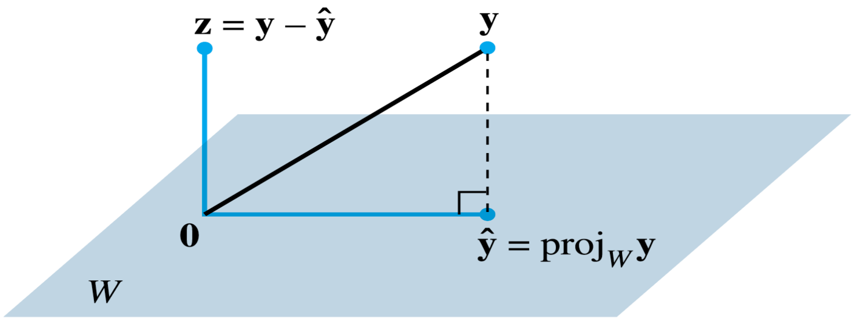
where $\hat{y}$ is in W and $\mathbf{z}$ is in $W^{\perp}$.
- In fact, if $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ is any orthogonal basis of W, then

$$\hat{y} = \frac{y \bullet u_1}{u_1 \bullet u_1} u_1 + \cdots + \frac{y \bullet u_p}{u_p \bullet u_p} u_p \qquad \text{----(2)}$$

and $z = y - \hat{y}$.

THE ORTHOGONAL DECOMPOSITION THEOREM

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



The orthogonal projection of **y** onto W.

- **Proof:** Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ 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 $\mathbf{u}_1, \ldots, \mathbf{u}_p$.

THE ORTHOGONAL DECOMPOSITION THEOREM

- Let $z = y - \hat{y}$.
- Since $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2, \ldots, \mathbf{u}_p$, it follows from (2) that

$$z \bullet u_1 = (y - \hat{y}) \bullet u_1 = y \bullet u_1 - \left(\frac{y \bullet u_1}{u_1 \bullet u_1} \right) u_1 \bullet u_1 - 0 - \cdots - 0$$

$$= y \bullet u_1 - y \bullet u_1 = 0$$

- Thus $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$.
- Similarly, $\mathbf{z}$ is orthogonal to each $\mathbf{u}_j$ in the basis for W.
- Hence $\mathbf{z}$ is orthogonal to every vector in W.
- That is, $\mathbf{z}$ 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 $\mathbf{z}_1$ 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 $\mathbf{z}_1$ and $\mathbf{z}$ are both in $W^{\perp}$, and $W^{\perp}$ is a subspace).
- Hence $v \bullet v = 0$, which shows that $v = 0$.
- This proves that $\hat{y} = \hat{y}_1$ and also $z_1 = z$.

THE ORTHOGONAL DECOMPOSITION THEOREM

- 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 $u_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, u_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

  Observe that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{u_1, u_2\}$. Write **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

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

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

- Also

$$y - \hat{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

THE ORTHOGONAL DECOMPOSITION THEOREM

- Theorem 8 ensures that $y - \hat{y}$ is in $W^{\perp}$.

- To check the calculations, verify that $y - \hat{y}$ is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$ and hence to all of W.

- The desired decomposition of $\mathbf{y}$ is

$$y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix} + \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$$

PROPERTIES OF ORTHOGONAL PROJECTIONS

- If $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ is an orthogonal basis for W and if $\mathbf{y}$ happens to be in W, then the formula for $\text{proj}_W\mathbf{y}$ is exactly the same as the representation of $\mathbf{y}$ given in Theorem 5 in Section 6.2.

- In this case, $\text{proj}_W \mathrm{y} = \mathrm{y}$.

- If $\mathbf{y}$ is in $W = \text{Span}\{\mathrm{u}_1,\ldots,\mathrm{u}_p\}$, then $\text{proj}_W \mathrm{y} = \mathrm{y}$.

THE BEST APPROXIMATION THEOREM

- **Theorem 9:** Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathrm{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathrm{y}}$ is the closest point in W to $\mathbf{y}$, in the sense that

$$\|\mathrm{y} - \hat{\mathrm{y}}\| < \|\mathrm{y} - \mathrm{v}\| \qquad \text{----(3)}$$

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

- The vector $\hat{\mathrm{y}}$ in Theorem 9 is called **the best approximation to y by elements of** W.
- The distance from $\mathbf{y}$ to $\mathbf{v}$, given by $\|\mathrm{y} - \mathrm{v}\|$, can be regarded as the "error" of using $\mathbf{v}$ in place of $\mathbf{y}$.
- Theorem 9 says that this error is minimized when $\mathrm{v} = \hat{\mathrm{y}}$.

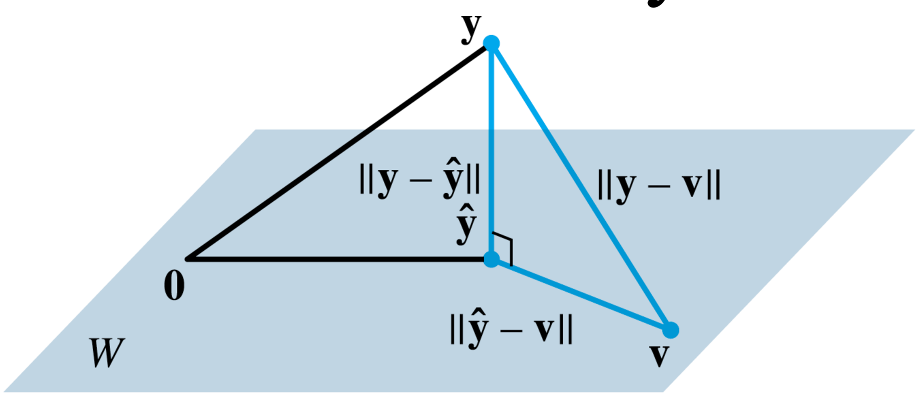
THE BEST APPROXIMATION THEOREM

- Inequality (3) leads to a new proof that $\hat{\mathrm{y}}$ 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 $\mathbf{y}$, then this projection would also be the closest point in W to $\mathbf{y}$, namely, $\hat{\mathrm{y}}$.

THE BEST APPROXIMATION THEOREM

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



The orthogonal projection of $\mathbf{y}$ onto W is the closest point in W to $\mathbf{y}$.

- 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).

THE BEST APPROXIMATION THEOREM

- 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.

PROPERTIES OF ORTHOGONAL PROJECTIONS

- **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 $W = \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, where

$$\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

- **Solution:** By the Best Approximation Theorem, the distance from **y** to W is $\|\mathbf{y} - \hat{\mathbf{y}}\|$, where $\hat{\mathbf{y}} = \mathrm{proj}_W \mathbf{y}$.

PROPERTIES OF ORTHOGONAL PROJECTIONS

- Since $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for W,

$$\hat{y} = \frac{15}{30}u_1 + \frac{-21}{6}u_2 = \frac{1}{2}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{7}{2}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

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

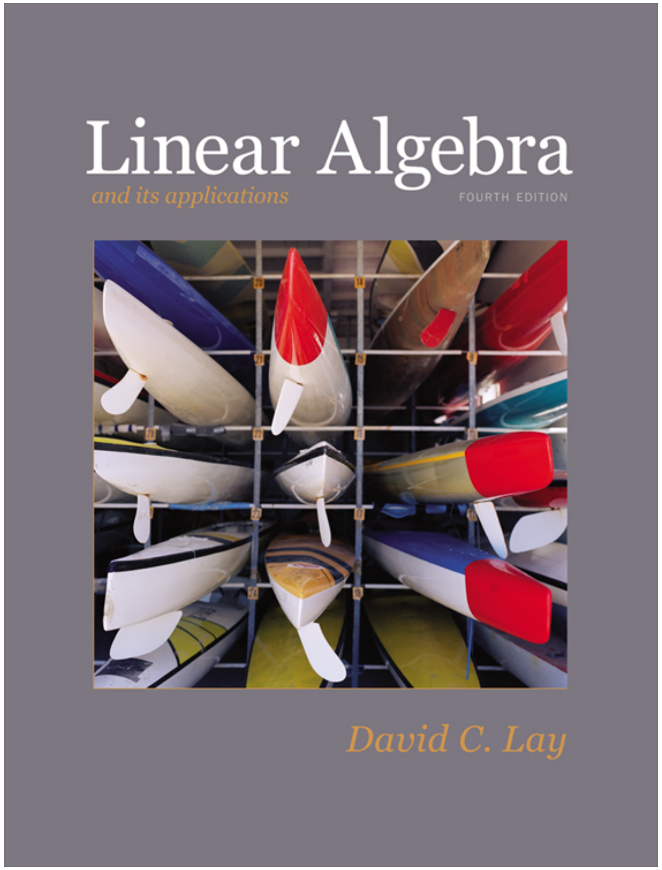
$$\|y - \hat{y}\|^2 = 3^2 + 6^2 = 45$$

- The distance from $\mathbf{y}$ to W is $\sqrt{45} = 3\sqrt{5}$.

6 Orthogonality and Least Squares

6.5

LEAST-SQUARES PROBLEMS



PEARSON

LEAST-SQUARES PROBLEMS

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

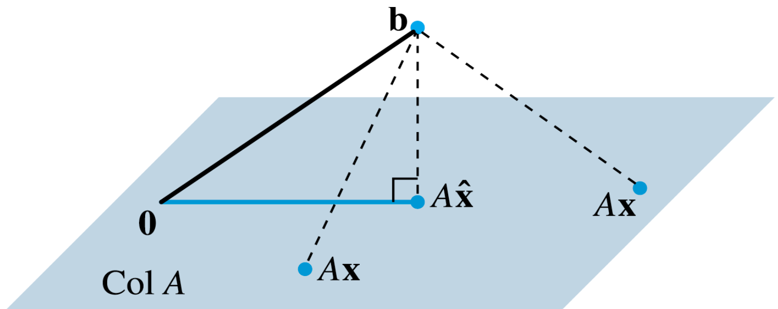
$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|$$

  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 $A\mathbf{x}$ will necessarily be in the column space, Col A.

- So we seek an **x** that makes $A\mathbf{x}$ the closest point in Col A to **b**. See the figure on the next slide.

LEAST-SQUARES PROBLEMS



The vector **b** is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other **x**.

- **Solution of the General Least-Squares Problem**
- Given A and **b**, apply the Best Approximation Theorem to the subspace Col A.
- Let $$\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\,\mathbf{b}$$

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

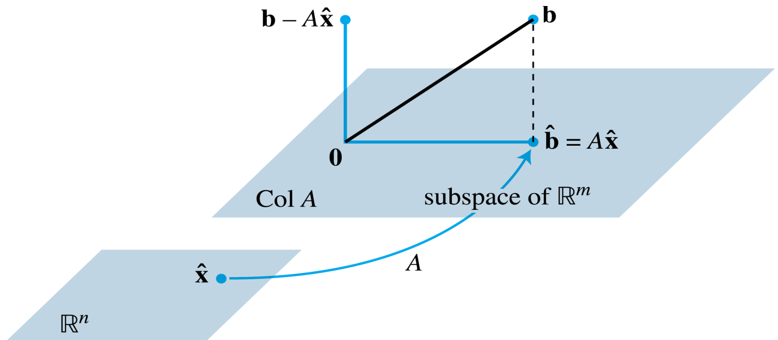
- Because $\hat{\mathrm{b}}$ is in the column space A, the equation $A\mathrm{x} = \hat{\mathrm{b}}$ *is* consistent, and there is an $\hat{\mathrm{x}}$ in $\mathbb{R}^n$ such that

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

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

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

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM



The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

- 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 Col A, so $b - A\hat{x}$ is orthogonal to each column of A.
- If $\mathbf{a}_j$ is any column of A, then $a_j \bullet (b - A\hat{x}) = 0$, and $a_j^T(b - A\hat{x})$.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- Since each $\mathbf{a}_j^T$ is a row of A^T,

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \qquad \text{----(2)}$$

- Thus

$$A^T\mathbf{b} - A^T A\hat{\mathbf{x}} = 0$$

$$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$$

- These calculations show that each least-squares solution of $A\mathbf{x} = \mathbf{b}$ satisfies the equation

$$A^T A\mathbf{x} = A^T\mathbf{b} \qquad \text{----(3)}$$

- The matrix equation (3) represents a system of equations called the **normal equations** for $A\mathbf{x} = \mathbf{b}$.
- A solution of (3) is often denoted by $\hat{\mathbf{x}}$.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- **Theorem 13:** The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equation $A^T A\mathbf{x} = A^T\mathbf{b}$.
- **Proof:** The set of least-squares solutions is nonempty and each least-squares solution $\hat{\mathbf{x}}$ satisfies the normal equations.
- Conversely, suppose $\hat{\mathbf{x}}$ satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.
- Then $\hat{\mathbf{x}}$ satisfies (2), which shows that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T and hence is orthogonal to the columns of A.
- Since the columns of A span Col A, the vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to all of Col A.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- Hence the equation

$$\mathbf{b} = A\hat{\mathbf{x}} + (\mathbf{b} - A\hat{\mathbf{x}})$$

  is a decomposition of **b** into the sum of a vector in Col A and a vector orthogonal to Col A.

- By the uniqueness of the orthogonal decomposition, $A\hat{\mathbf{x}}$ must be the orthogonal projection of **b** onto Col A.

- That is, $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$ and $\hat{\mathbf{x}}$ is a least-squares solution.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- **Example 1:** Find a least-squares solution of the inconsistent system $A\mathrm{x} = \mathrm{b}$ for

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

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

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

$$A^T\mathbf{b} = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

- Then the equation $A^T A\mathbf{x} = A^T\mathbf{b}$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- Row operations can be used to solve the system on the previous slide, but since A^TA is invertible and 2×2, it is probably faster to compute

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

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

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

$$= \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}\begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84}\begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- **Theorem 14:** Let A be an $m \times n$ matrix. The following statements are logically equivalent:
  a. The equation $A\mathrm{x} = \mathrm{b}$ 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 $\hat{\mathrm{x}}$ is given by

  $$\hat{\mathrm{x}} = (A^TA)^{-1}A^T\mathrm{b} \qquad \text{----(4)}$$

- When a least-squares solution $\hat{\mathrm{x}}$ is used to produce $A\hat{\mathrm{x}}$ as an approximation to **b**, the distance from **b** to $A\hat{\mathrm{x}}$ is called the **least-squares error** of this approximation.

ALTERNATIVE CALCULATIONS OF LEAST-SQUARES SOLUTIONS

- **Example 2:** Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

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

- **Solution:** Because the columns $\mathbf{a}_1$ and $\mathbf{a}_2$ of A are orthogonal, the orthogonal projection of $\mathbf{b}$ onto Col A is given by

$$\hat{\mathbf{b}} = \frac{\mathbf{b} \bullet \mathbf{a}_1}{\mathbf{a}_1 \bullet \mathbf{a}_1}\mathbf{a}_1 + \frac{\mathbf{b} \bullet \mathbf{a}_2}{\mathbf{a}_2 \bullet \mathbf{a}_2}\mathbf{a}_2 = \frac{8}{4}\mathbf{a}_1 + \frac{45}{90}\mathbf{a}_2 \quad \text{----(5)}$$

ALTERNATIVE CALCULATIONS OF LEAST-SQUARES SOLUTIONS

$$= \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}$$

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

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