Chapter 6 Orthogonality and Least Squares

Section 6.1: Inner Product, Length, and Orthogonality

Section 6.2: Orthogonal Sets

Section 6.3: Orthogonal Projections

Section 6.4: The Gram-Schmidt Process

Section 6.5: Least-Squares Problems

Inner Product

- 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**.

Inner Product

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

- 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.

Inner Product

Theorem 1: Let **u**, **v**, and **w** be vectors in $\mathbb{R}^n$, and let *c* be a scalar. Then

(a) $\mathrm{u} \cdot \mathrm{v} = \mathrm{v} \cdot \mathrm{u}$

(b) $(\mathrm{u} + \mathrm{v}) \cdot \mathrm{w} = \mathrm{u} \cdot \mathrm{w} + \mathrm{v} \cdot \mathrm{w}$

(c) $(c\mathrm{u}) \cdot \mathrm{v} = c(\mathrm{u} \cdot \mathrm{v}) = \mathrm{u} \cdot (c\mathrm{v})$

(b) $\mathrm{u} \cdot \mathrm{u} \geq 0, \text{and } \mathbf{u} \cdot \mathbf{u} = 0$ if and only if **u** = **0**

- Properties and can be combined several times to produce the following useful rule:

$$(c_1\mathrm{u}_1 + \cdots + c_p\mathrm{u}_p) \cdot \mathrm{w} = c_1(\mathrm{u}_1 \cdot \mathrm{w}) + \cdots + c_p(\mathrm{u}_p \cdot \mathrm{w})$$

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 $\|\mathrm{v}\|$ defined by

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

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

The Length of a Vector

- If we identify **v** with a geometric point in the plane, as usual, then $\|v\|$ 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.

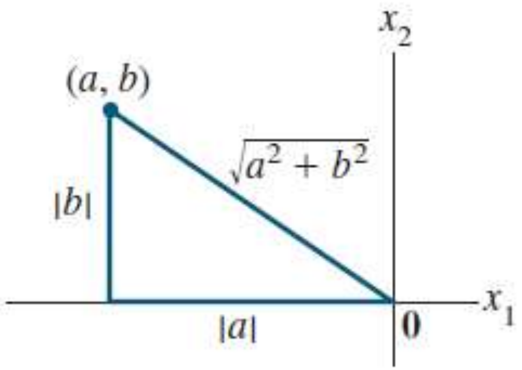


FIGURE 1

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

The Length of a Vector

- For any scalar *c*, the length *c***v** is $|c|$ times the length of **v**. That is, $\|c\mathrm{v}\| = |c|\|\mathrm{v}\|$.
- 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 $1/\|\mathrm{v}\|$ - we obtain a unit vector **u** because the length of **u** is $(1/\|\mathrm{v}\|)\|\mathrm{v}\|$.
- 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:** Let $\text{v} = (1, -2, 2, 0)$. Find a unit vector **u** in the same direction as **v**.
- **Solution:** First, compute the length of **v**:

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

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

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

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

Distance in $\mathbb{R}^n$

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

$$\|\mathrm{u}\|^2 = \mathrm{u} \cdot \mathrm{u} = \left(\frac{1}{3}\right)^2 + \left(-\frac{2}{3}\right)^2 + \left(\frac{2}{3}\right)^2 + (0)^2$$
$$= \frac{1}{9} + \frac{4}{9} + \frac{4}{9} + 0 = 1$$

- **Definition:** For **u** and **v** in $\mathbb{R}^n$, the **distance between u and v**, written as 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:** Compute the distance between the vectors $\mathbf{u} = (7,1)$ and $\mathbf{v} = (3,2)$.
- **Solution:** Calculate

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

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

When the vector $\mathbf{u} - \mathbf{v}$ is added to **v**, the result is **u**.

Distance in $\mathbb{R}^n$

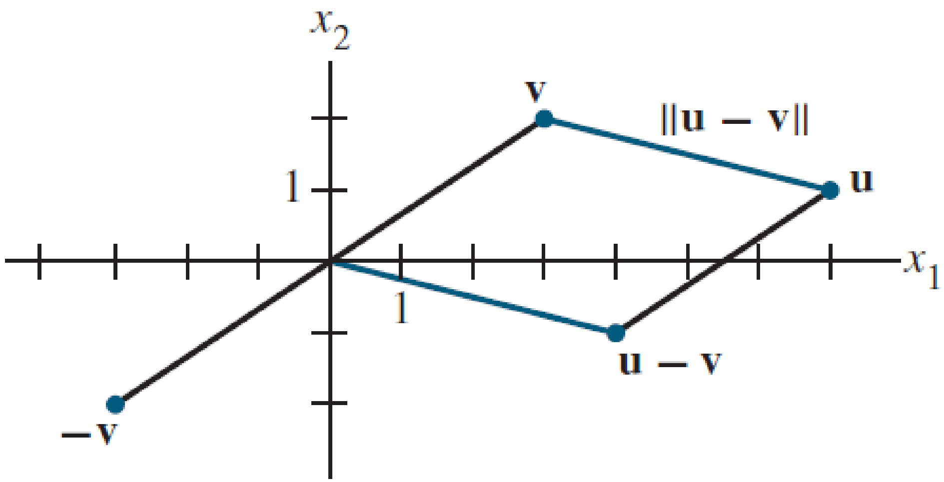
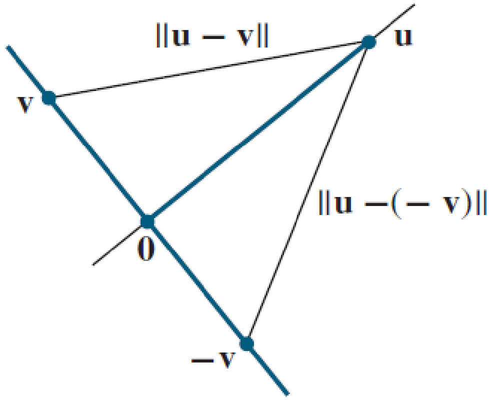


FIGURE 4 The distance between **u** and **v** is the length of $\mathbf{u} - \mathbf{v}$.

- Notice that the parallelogram in the above figure shows that the distance from **u** to **v** is the same as the distance from $\mathbf{u} - \mathbf{v}$ 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 $-\mathbf{v}$.



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

Orthogonal Vectors

$$
\begin{aligned}
[\mathrm{dist}(\mathrm{u},-\mathrm{v})]^2 &= \|\mathrm{u}-(-\mathrm{v})\|^2 = \|\mathrm{u}+\mathrm{v}\|^2 \\
&= (\mathrm{u}+\mathrm{v})\cdot(\mathrm{u}+\mathrm{v}) \\
&= \mathrm{u}\cdot(\mathrm{u}+\mathrm{v}) + \mathrm{v}\cdot(\mathrm{u}+\mathrm{v}) \\
&= \mathrm{u}\cdot\mathrm{u} + \mathrm{u}\cdot\mathrm{v} + \mathrm{v}\cdot\mathrm{u} + \mathrm{v}\cdot\mathrm{v} \\
&= \|\mathrm{u}\|^2 + \|\mathrm{v}\|^2 + 2\mathrm{u}\cdot\mathrm{v}
\end{aligned}
$$

$$
\begin{aligned}
[\mathrm{dist}(\mathrm{u},\mathrm{v})]^2 &= \|\mathrm{u}\|^2 + \|-\mathrm{v}\|^2 + 2\mathrm{u}\cdot(-\mathrm{v}) \\
&= \|\mathrm{u}\|^2 + \|\mathrm{v}\|^2 - 2\mathrm{u}\cdot\mathrm{v}
\end{aligned}
$$

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

Orthogonal Vectors

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

The Pythagorean Theorem

- **Theorem:** Two vectors **u** and **v** are orthogonal if and only if

$$\|\mathrm{u}+\mathrm{v}\|^2 = \|\mathrm{u}\|^2 + \|\mathrm{v}\|^2$$

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

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

Orthogonal Complements

- **Proof:** The row-column rule for computing $A\mathbf{x}$ shows that if $\mathbf{x}$ is in Nul A, then $\mathbf{x}$ is orthogonal to each row of A (with the rows treated as vectors in $\mathbb{R}^n$).

- Since the rows of A span the row space, $\mathbf{x}$ is orthogonal to Row A.

- Conversely, if $\mathbf{x}$ is orthogonal to Row A, then $\mathbf{x}$ is certainly orthogonal to each row of A, and hence $A\mathbf{x} = \mathbf{0}$.

- This proves the first statement of the theorem.

Orthogonal Complements

- Since this statement is true for any matrix, it is true for A^T.
- That is, the orthogonal complement of the row space of A^T is the null space of A^T.
- This proves the second statement, because

$$\mathrm{Row}\,A^T = \mathrm{Col}\,A$$

Section 6.1: Inner Product, Length, and Orthogonality

Section 6.2: Orthogonal Sets

Section 6.3: Orthogonal Projections

Section 6.4: The Gram-Schmidt Process

Section 6.5: Least-Squares Problems

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:** 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

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

- **Theorem:** 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 are given by

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

where

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

An Orthogonal Projection

- Write $\mathrm{y} = \hat{\mathrm{y}} + \mathrm{z}$

  where $\hat{\mathrm{y}} = \alpha\mathrm{u}$ for some scalar α and **z** is some vector orthogonal to **u** (a nonzero vector in $\mathbb{R}^n$).

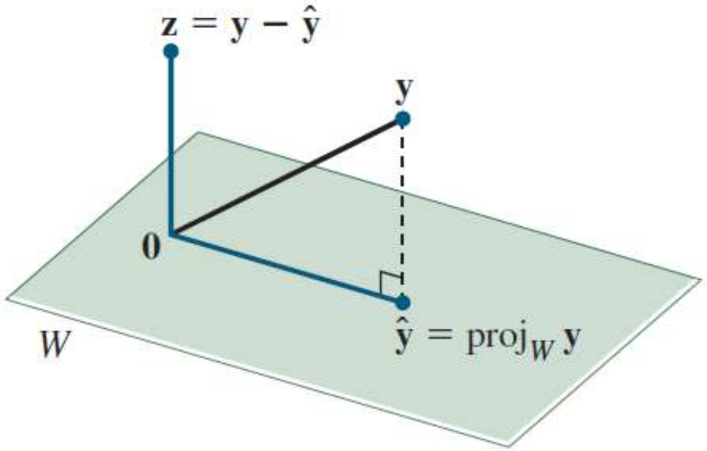


FIGURE 2
Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

$$\alpha = \frac{\mathrm{y} \cdot \mathrm{u}}{\mathrm{u} \cdot \mathrm{u}}$$

$$\hat{\mathrm{y}} = \frac{\mathrm{y} \cdot \mathrm{u}}{\mathrm{u} \cdot \mathrm{u}}\mathrm{u}$$

The vector $\hat{\mathbf{y}}$ 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

- **Example:** 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 {**u**} and one orthogonal to **u**.
- **Solution:** Compute

$$\mathrm{y} \cdot \mathrm{u} = 40$$
$$\mathrm{u} \cdot \mathrm{u} = 20$$

An Orthogonal Projection

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

$$\hat{y} = \frac{y \cdot u}{u \cdot u} u = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

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,

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

- The decomposition of **y**:

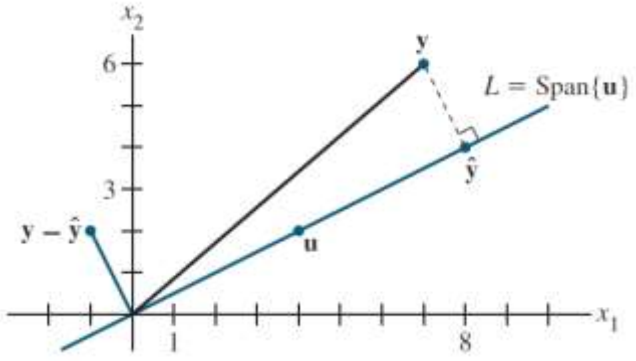


FIGURE 3 The orthogonal projection of **y** onto a line L through the origin.

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

Orthogonal 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.

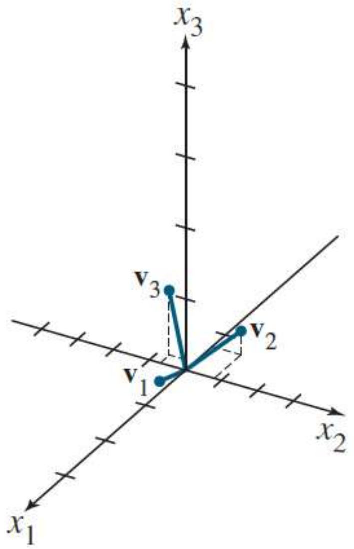
- The simplest example of an orthonormal set is the standard basis $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ for $\mathbb{R}^n$. Any nonempty subset of $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ is orthonormal.

Orthogonal Sets

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

$$\mathrm{v}_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \quad \mathrm{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \quad \mathrm{v}_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

Orthogonal 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.

Orthogonal Sets

- **Theorem:** An $m \times n$ matrix *U* has orthonormal columns if and only if $U^T U = I$.

- **Proof:** To simplify notation, we suppose that *U* has only three columns, each a vector in $\mathbb{R}^m$.
- Let $U = [\mathrm{u}_1 \quad \mathrm{u}_2 \quad \mathrm{u}_3]$ and compute

$$U^T U = \begin{bmatrix} \mathrm{u}_1^T \\ \mathrm{u}_2^T \\ \mathrm{u}_3^T \end{bmatrix} [\mathrm{u}_1 \quad \mathrm{u}_2 \quad \mathrm{u}_3] = \begin{bmatrix} \mathrm{u}_1^T \mathrm{u}_1 & \mathrm{u}_1^T \mathrm{u}_2 & \mathrm{u}_1^T \mathrm{u}_3 \\ \mathrm{u}_2^T \mathrm{u}_1 & \mathrm{u}_2^T \mathrm{u}_2 & \mathrm{u}_2^T \mathrm{u}_3 \\ \mathrm{u}_3^T \mathrm{u}_1 & \mathrm{u}_3^T \mathrm{u}_2 & \mathrm{u}_3^T \mathrm{u}_3 \end{bmatrix} \quad (3)$$

Orthogonal Sets

- The entries in the matrix at the right are inner products, using transpose notation.
- The columns of *U* are orthogonal if and only if

$$\mathbf{u}_1^T\mathbf{u}_2 = \mathbf{u}_2^T\mathbf{u}_1 = 0, \quad \mathbf{u}_1^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_1 = 0, \quad \mathbf{u}_2^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_2 = 0 \qquad (4)$$

- The columns of *U* all have unit length if and only if

$$\mathbf{u}_1^T\mathbf{u}_1 = 1, \quad \mathbf{u}_2^T\mathbf{u}_2 = 1, \quad \mathbf{u}_3^T\mathbf{u}_3 = 1 \qquad (5)$$

- The theorem follows immediately from (3)–(5).

Orthogonal Sets

- **Theorem:** Let U be an $m \times n$ matrix with orthonormal columns, and let **x** and **y** be in $\mathbb{R}^n$.

  Then

  a. $\|U\mathrm{x}\| = \|\mathrm{x}\|$

  b. $\mathrm{Ux} \cdot \mathrm{Uy} = \mathrm{x} \cdot \mathrm{y}$

  c. $\mathrm{Ux} \cdot \mathrm{Uy} = 0$ if and only if $\mathrm{x} \cdot \mathrm{y} = 0$

- Properties (a) and (c) say that the linear mapping preserves lengths and orthogonality.

$$\mathbf{x} \mapsto U\mathbf{x}$$

Section 6.1: Inner Product, Length, and Orthogonality

Section 6.2: Orthogonal Sets

Section 6.3: Orthogonal Projections

Section 6.4: The Gram-Schmidt Process

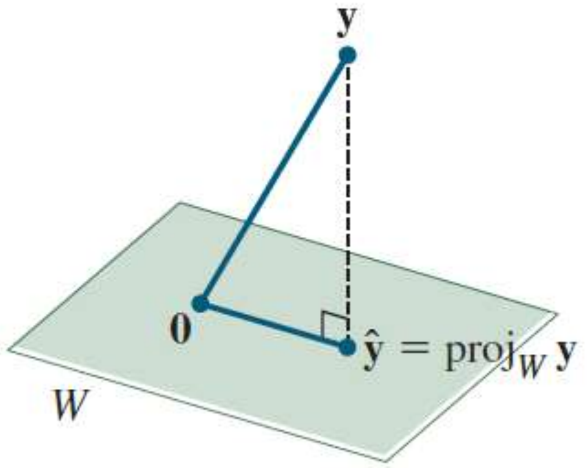
Section 6.5: Least-Squares Problems

Orthogonal Projections

- Given a vector **y** and a subspace *W* in $\mathbb{R}^2$, there is a vector $\hat{\mathbf{y}}$ in *W* such that

a. $\hat{\mathbf{y}}$ is the unique vector in *W* for which $\mathbf{y}-\hat{\mathbf{y}}$ is orthogonal to *W*, and

b. $\hat{\mathbf{y}}$ is the unique vector in *W* closest to **y**.



These two properties of $\hat{\mathbf{y}}$ provide the key to finding the least-squares solutions of linear systems.

The Orthogonal Decomposition Theorem

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

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

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

- In fact, if $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ is any orthogonal basis of *W*, then

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u_1}}{\mathbf{u_1}\cdot\mathbf{u_1}}\mathbf{u_1} + \cdots + \frac{\mathbf{y}\cdot\mathbf{u_p}}{\mathbf{u_p}\cdot\mathbf{u_p}}\mathbf{u_p} \qquad (2)$$

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

The Orthogonal Decomposition Theorem

- **Example:** 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 $W = \mathrm{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. 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

$$\hat{\mathbf{y}} = \frac{\mathbf{y}\cdot\mathbf{u_1}}{\mathbf{u_1}\cdot\mathbf{u_1}}\mathbf{u_1} + \frac{\mathbf{y}\cdot\mathbf{u_2}}{\mathbf{u_2}\cdot\mathbf{u_2}}\mathbf{u_2} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

Also

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

The Orthogonal Decomposition Theorem

- The desired decomposition of **y** is

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

The Gram-Schmidt Process

- **Theorem: The Gram-Schmidt Process**
- Given a basis $\{\mathbf{x}_1, \ldots, \mathbf{x}_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

$$\mathbf{v}_1 = \mathbf{x}_1$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2$$

$$\vdots$$

$$\mathbf{v}_p = \mathbf{x}_p - \frac{\mathbf{x}_p \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \ldots - \frac{\mathbf{x}_p \cdot \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \cdot \mathbf{v}_{p-1}}\mathbf{v}_{p-1}$$

- Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition

$$\text{Span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \text{Span}\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \quad \text{for } 1 \leq k \leq p. \tag{1}$$

Orthonormal Bases

- **Example** Example 1 constructed the orthogonal basis

$$\mathbf{v}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

- An orthonormal basis is

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{45}}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix}$$

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Section 6.1: Inner Product, Length, and Orthogonality

Section 6.2: Orthogonal Sets

Section 6.3: Orthogonal Projections

Section 6.4: The Gram-Schmidt Process

Section 6.5: Least-Squares Problems

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

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

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

- So we seek an **x** that makes A**x** the closest point in Col A to **b**.

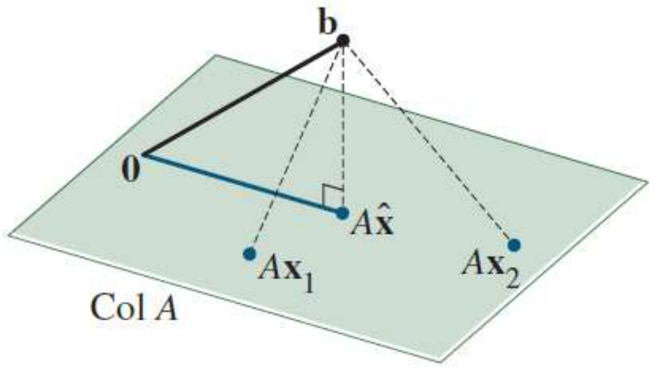


FIGURE 1 The vector **b** is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other **x**.

Solution of the General Least-Squares Problems

- **Solution of the General Least-Squares Problem**

Let $\hat{\mathbf{b}} = \text{proj}_{\text{Col } A}\mathbf{b}$, there is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that $A\hat{\mathrm{x}} = \hat{\mathrm{b}}$

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

- Such an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ is a list of weights that will build $\hat{\mathbf{b}}$ out of the columns of *A*

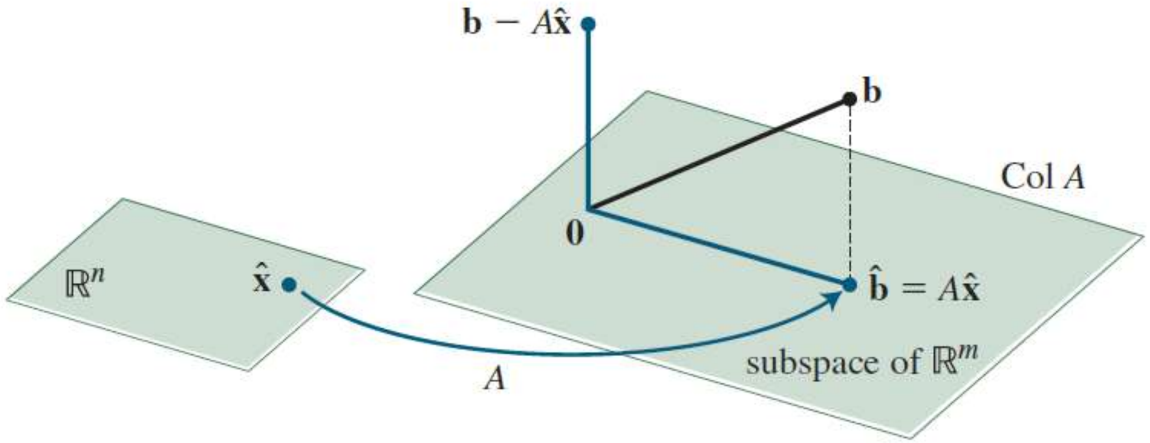


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

Solution of the General Least-Squares Problems

- Suppose $\hat{\mathbf{x}}$ satisfies $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.
- By the Orthogonal Decomposition Theorem, the projection $\hat{\mathbf{b}}$ has the property that $\mathbf{b} - \hat{\mathbf{b}}$ is orthogonal to Col *A*, so $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to each column of *A*.
- If $\mathbf{a}_j$ is any column of *A*, then $\mathrm{a}_j \cdot (\mathrm{b} - A\hat{\mathrm{x}}) = 0$, so $\mathrm{a}_j^T(\mathrm{b} - A\hat{\mathrm{x}}) = 0$.

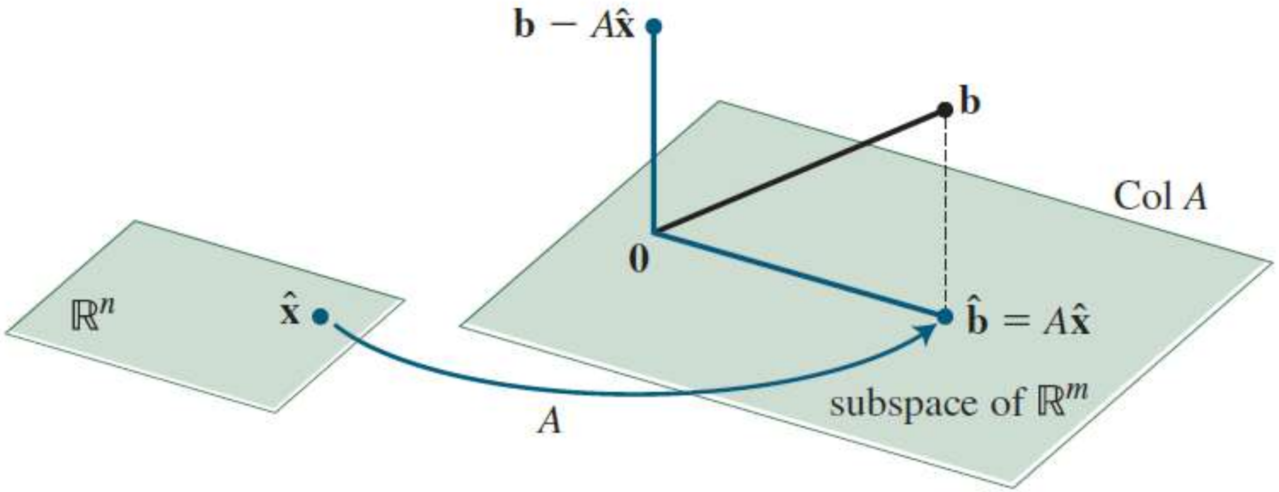


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

Solution of the General Least-Squares Problems

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

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

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 $A^T A\mathbf{x} = A^T\mathbf{b}$ (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 General Least-Squares Problems

- **Theorem:** The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equation $A^T A\mathbf{x} = A^T\mathbf{b}$.

Solution of the General Least-Squares Problems

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

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

- **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 General Least-Squares Problems

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

- 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 General Least-Squares Problems

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

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

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

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

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