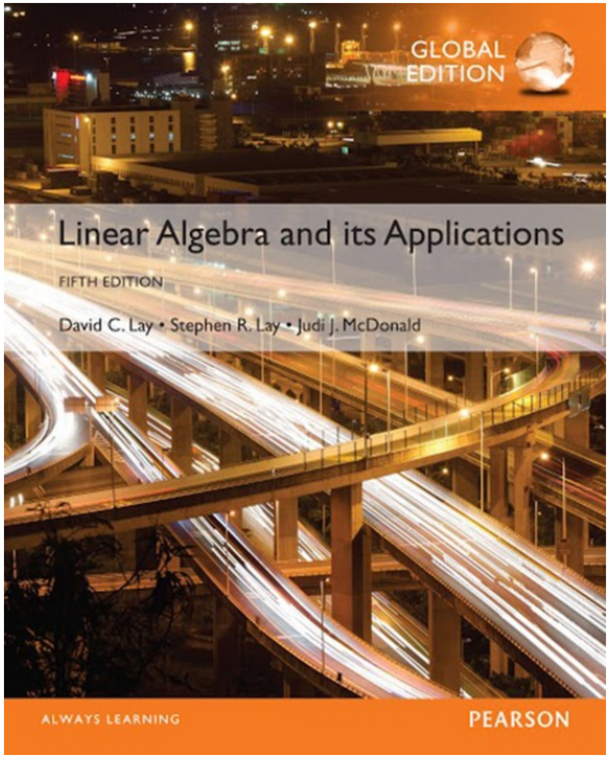
6 Orthogonality and Least Squares

6.3

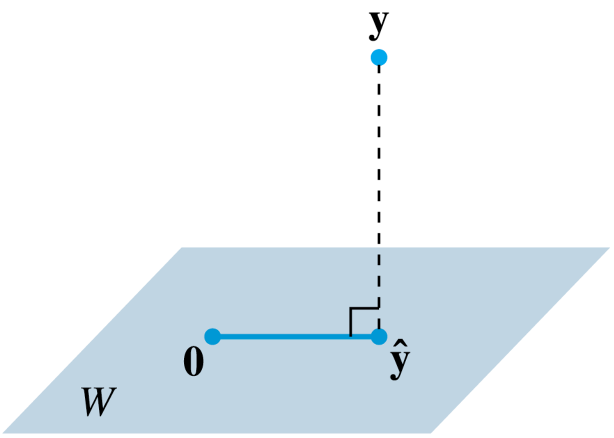
ORTHOGONAL PROJECTIONS



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 $\hat{y}$ 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 $\mathbf{y}$. 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

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

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

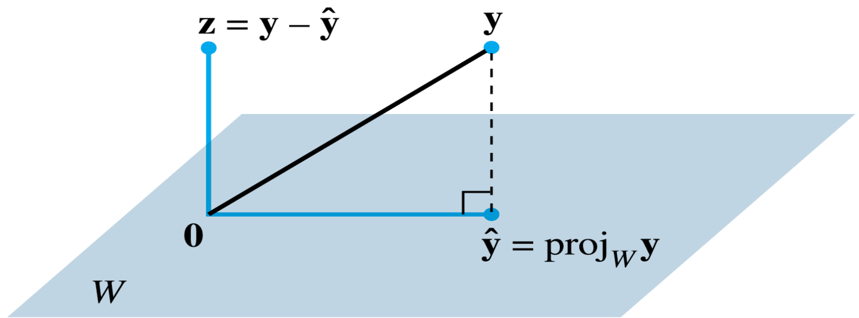
- In fact, if $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is any orthogonal basis of W, then

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

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

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

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

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

- 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 $v \cdot v = 0$, 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 $\mathrm{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \mathrm{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathrm{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

  Observe that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \mathrm{Span}\{\mathrm{u}_1, \mathrm{u}_2\}$. 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

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

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

- Also

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

THE ORTHOGONAL DECOMPOSITION THEOREM

- Theorem 8 ensures that $\mathbf{y} - \hat{\mathbf{y}}$ is in $W^{\perp}$.
- To check the calculations, verify that $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to both $\mathbf{u}_1$ and $\mathbf{u}_2$ and hence to all of W.
- The desired decomposition of $\mathbf{y}$ is

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

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 $\hat{\mathbf{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathbf{y}}$ is the closest point in W to $\mathbf{y}$, in the sense that

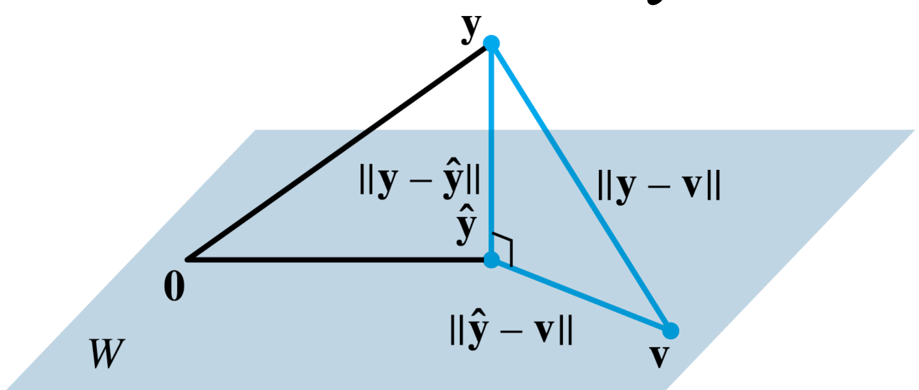
(3) $$\|\mathbf{y}-\hat{\mathbf{y}}\| < \|\mathbf{y}-\mathbf{v}\|$$

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

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

THE BEST APPROXIMATION THEOREM

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



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

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

THE BEST APPROXIMATION THEOREM

- Since
$$\mathbf{y}-\mathbf{v}=(\mathbf{y}-\hat{\mathbf{y}})+(\hat{\mathbf{y}}-\mathbf{v})$$
the Pythagorean Theorem gives
$$\|\mathbf{y}-\mathbf{v}\|^2=\|\mathbf{y}-\hat{\mathbf{y}}\|^2+\|\hat{\mathbf{y}}-\mathbf{v}\|^2$$
- (See the colored right triangle in the figure on the previous slide. The length of each side is labeled.)
- Now $\|\hat{\mathbf{y}}-\mathbf{v}\|^2>0$ because $\hat{\mathbf{y}}-\mathbf{v}\neq 0$, and so inequality (3) follows immediately.

PROPERTIES OF ORTHOGONAL PROJECTIONS

- **Example 4:** 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

$W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, 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 $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$.

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 $\sqrt{45} = 3\sqrt{5}$.

PROPERTIES OF ORTHOGONAL PROJECTIONS

- **Theorem 10:**

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ is an orthonormal basis for a subspace W of $\mathbb{R}^n$, then

$$\text{proj}_W \mathbf{y} = (\mathbf{y} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{y} \cdot \mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{y} \cdot \mathbf{u}_p)\mathbf{u}_p \tag{4}$$

If $U = [\,\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_p\,]$, then

$$\text{proj}_W \mathbf{y} = UU^T\mathbf{y} \quad \text{for all } \mathbf{y} \text{ in } \mathbb{R}^n \tag{5}$$

Proof:

- Formula (4) follows immediately from (2) in Theorem 8.

PROPERTIES OF ORTHOGONAL PROJECTIONS

- Also, (4) shows that $\text{proj}_W \mathbf{y}$ is a linear combination of the columns of U using the weights.

- The weights can be written as $\mathbf{u}_1^T\mathbf{y}, \mathbf{u}_2^T\mathbf{y}, \ldots, \mathbf{u}_p^T\mathbf{y}$
  - showing that they are the entries in $U^T\mathbf{y}$ and justifying (5).