Section 6.2-6.3 Orthogonal sets and orthogonal projections

Gexin Yu
gyu@wm.edu

College of William and Mary

Orthogonal sets

- A set of vectors $\{u_1, u_2, \ldots, u_p\}$ in $\mathbf{R}^n$ is said to be an orthogonal set if each pair of distinct vectors from the set is orthogonal, that is $u_i \cdot u_j = 0$ whenever $i \neq j$.

Orthogonal sets

- A set of vectors $\{u_1, u_2, \ldots, u_p\}$ in $\mathbf{R}^n$ is said to be an orthogonal set if each pair of distinct vectors from the set is orthogonal, that is $u_i \cdot u_j = 0$ whenever $i \neq j$.

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

Proof of Theorem 4

PF. Let $c_1u_1 + \ldots + c_pu_p = 0$ for some scalars $c_1, \ldots, c_p$.

Proof of Theorem 4

PF. Let $c_1 u_1 + \ldots + c_p u_p = 0$ for some scalars $c_1, \ldots, c_p$.

- Then we have

$$
\begin{aligned}
0 = 0 \cdot u_1 &= (c_1 u_1 + \ldots + c_p u_p) \cdot u_1 \\
&= (c_1 u_1) \cdot u_1 + \ldots + (c_p u_p) \cdot u_1 \\
&= c_1(u_1 \cdot u_1) + \ldots + c_p(u_p \cdot u_1) = c_1(u_1 \cdot u_1)
\end{aligned}
$$

because u_1 is orthogonal to $u_2, \ldots, u_p$.

Proof of Theorem 4

PF. Let $c_1 u_1 + \ldots + c_p u_p = 0$ for some scalars $c_1, \ldots, c_p$.

- Then we have

$$\begin{aligned} 0 = 0 \cdot u_1 &= (c_1 u_1 + \ldots + c_p u_p) \cdot u_1 \\ &= (c_1 u_1) \cdot u_1 + \ldots + (c_p u_p) \cdot u_1 \\ &= c_1(u_1 \cdot u_1) + \ldots + c_p(u_p \cdot u_1) = c_1(u_1 \cdot u_1) \end{aligned}$$

  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 $c_1 = 0$.

Proof of Theorem 4

PF. Let $c_1 u_1 + \ldots + c_p u_p = 0$ for some scalars $c_1, \ldots, c_p$.

- Then we have

$$\begin{aligned} 0 = 0 \cdot u_1 &= (c_1 u_1 + \ldots + c_p u_p) \cdot u_1 \\ &= (c_1 u_1) \cdot u_1 + \ldots + (c_p u_p) \cdot u_1 \\ &= c_1(u_1 \cdot u_1) + \ldots + c_p(u_p \cdot u_1) = c_1(u_1 \cdot u_1) \end{aligned}$$

  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.

Proof of Theorem 4

PF. Let $c_1 u_1 + \ldots + c_p u_p = 0$ for some scalars $c_1, \ldots, c_p$.

- Then we have

$$\begin{aligned} 0 = 0 \cdot u_1 &= (c_1 u_1 + \ldots + c_p u_p) \cdot u_1 \\ &= (c_1 u_1) \cdot u_1 + \ldots + (c_p u_p) \cdot u_1 \\ &= c_1(u_1 \cdot u_1) + \ldots + c_p(u_p \cdot u_1) = c_1(u_1 \cdot u_1) \end{aligned}$$

  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.

Orthogonal basis

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

$$\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix} \text{ basis for } \mathbb{R}^3$$

Orthogonal basis

- An orthogonal basis for a subspace W of $\mathbf{R}^n$ 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 $\mathbf{R}^n$. For each y in W, the weights in the linear combination $y = c_1 u_1 + \ldots + c_p u_p$ are given by $c_j = \frac{y \cdot u_j}{u_j \cdot u_j}$ for $j = 1, \ldots, p$.

Remark: easy to find $[y]_U$.

Orthogonal basis

- An orthogonal basis for a subspace W of $\mathbf{R}^n$ 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 $\mathbf{R}^n$. For each y in W, the weights in the linear combination $y = c_1 u_1 + \ldots + c_p u_p$ are given by $c_j = \frac{y \cdot u_j}{u_j \cdot u_j}$ for $j = 1, \ldots, p$.

PF. The orthogonality of $\{u_1, \ldots, u_p\}$ shows that

$$y \cdot u_1 = (c_1 u_1 + \ldots + c_p u_p) \cdot u_1 = c_1(u_1 \cdot u_1)$$

Orthogonal basis

- An orthogonal basis for a subspace W of $\mathbf{R}^n$ 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 $\mathbf{R}^n$. For each y in W, the weights in the linear combination $y = c_1 u_1 + \ldots + c_p u_p$ are given by $c_j = \frac{y \cdot u_j}{u_j \cdot u_j}$ for $j = 1, \ldots, p$.

PF. The orthogonality of $\{u_1, \ldots, u_p\}$ shows that

$$y \cdot u_1 = (c_1 u_1 + \ldots + c_p u_p) \cdot u_1 = c_1(u_1 \cdot u_1)$$

- Since $u_1 \cdot u_1$ is not zero, the equation above can be solved for c_1.

Orthogonal basis

- An orthogonal basis for a subspace W of $\mathbf{R}^n$ 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 $\mathbf{R}^n$. For each y in W, the weights in the linear combination $y = c_1 u_1 + \ldots + c_p u_p$ are given by $c_j = \frac{y \cdot u_j}{u_j \cdot u_j}$ for $j = 1, \ldots, p$.

PF. The orthogonality of $\{u_1, \ldots, u_p\}$ shows that

$$y \cdot u_1 = (c_1 u_1 + \ldots + c_p u_p) \cdot u_1 = c_1(u_1 \cdot u_1)$$

- Since $u_1 \cdot u_1$ is not zero, the equation above can be solved for c_1.

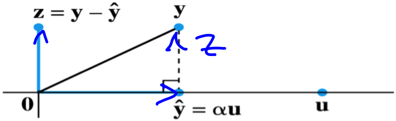
- To find c_j, we can similarly compute $y \cdot u_j$ and solve for c_j.

Orthogonal projection

- Given a nonzero vector u in $\mathbf{R}^n$, consider the problem of decomposing a vector y in $\mathbf{R}^n$ into the sum of two vectors, one a multiplier of u and the other orthogonal to u.

Orthogonal projection

- Given a nonzero vector u in $\mathbf{R}^n$, consider the problem of decomposing a vector y in $\mathbf{R}^n$ into the sum of two vectors, one a multiplier of u and the other orthogonal to u.
- We wish to write $y = \hat{y} + z$, where $\hat{y} = \alpha u$ for some scalar α and z is some vector orthogonal to u.



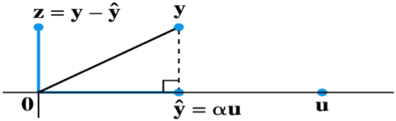
Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

$z \perp u$; $\quad (y - \hat{y}) \cdot u = 0 \quad y \cdot u - \hat{y} \cdot u = 0$

$y \cdot u - \alpha u \cdot u = 0 \quad \alpha u \cdot u = y \cdot u \Rightarrow \alpha = \frac{y \cdot u}{u \cdot u}$

Orthogonal projection

- Given a nonzero vector u in $\mathbf{R}^n$, consider the problem of decomposing a vector y in $\mathbf{R}^n$ into the sum of two vectors, one a multiplier of u and the other orthogonal to u.
- We wish to write $y = \hat{y} + z$, where $\hat{y} = \alpha u$ for some scalar α and z is some vector orthogonal to u.

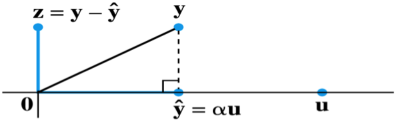


Finding α to make $\mathbf{y} - \hat{\mathbf{y}}$ orthogonal to $\mathbf{u}$.

- Given any scalar α, let $z = y - \alpha u$. Then $y - \hat{y}$ is orthogonal to u if and only if $0 = (y - \alpha u) \cdot u = y \cdot u - \alpha u \cdot u$.

Orthogonal projection

- Given a nonzero vector u in $\mathbf{R}^n$, consider the problem of decomposing a vector y in $\mathbf{R}^n$ into the sum of two vectors, one a multiplier of u and the other orthogonal to u.
- We wish to write $y = \hat{y} + z$, where $\hat{y} = \alpha u$ for some scalar α and z is some vector orthogonal to u.



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

- Given any scalar α, let $z = y - \alpha u$. Then $y - \hat{y}$ is orthogonal to u if and only if $0 = (y - \alpha u) \cdot u = y \cdot u - \alpha u \cdot u$.
- That is, $y = \hat{y} + z$ with z orthogonal to u if and only if $\alpha = \frac{y \cdot u}{u \cdot u}$ and $\hat{y} = \frac{y \cdot u}{u \cdot u} u$.

- 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.

- 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.
- 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.

- 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.

- 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).

- 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.
- 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 $proj_L y$ and is called the orthogonal projection of y onto L.

- 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.
- 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 $proj_L y$ and is called the orthogonal projection of y onto L.
- That is,

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

Example

Ex. Let $y = [7\ 6]^T$ and $u = [4\ 2]^T$. 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.

$$\text{Proj}_L y = \frac{y \cdot u}{u \cdot u} u = \frac{4\cdot 7 + 6\cdot 2}{4\cdot 4 + 2\cdot 2}\begin{bmatrix} 4 \\ 2 \end{bmatrix} = 2\begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

$$= \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

$$z = y - \text{Proj}_L y = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

Example

Ex. Let $y = [7\ 6]^T$ and $u = [4\ 2]^T$. 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.

Sol. Compute $y \cdot u = 40$ and $u \cdot u = 20$.

Example

Ex. Let $y = [7\ 6]^T$ and $u = [4\ 2]^T$. 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.

Sol. Compute $y \cdot u = 40$ and $u \cdot u = 20$.

- The orthogonal projection of y onto u is $\hat{y} = \frac{y \cdot u}{u \cdot u} u = 2u = [8\ 4]^T$.

Example

Ex. Let $y = [7\ 6]^T$ and $u = [4\ 2]^T$. 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.

Sol. Compute $y \cdot u = 40$ and $u \cdot u = 20$.

- The orthogonal projection of y onto u is $\hat{y} = \frac{y \cdot u}{u \cdot u} u = 2u = [8\ 4]^T$.
- And the component of y orthogonal to u is $y - \hat{y} = [7\ 6]^T - [8\ 4]^T = [-1\ 2]^T$.

Example

Ex. Let $y = [7\ 6]^T$ and $u = [4\ 2]^T$. 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.

Sol. Compute $y \cdot u = 40$ and $u \cdot u = 20$.

- The orthogonal projection of y onto u is $\hat{y} = \frac{y \cdot u}{u \cdot u} u = 2u = [8\ 4]^T$.
- And the component of y orthogonal to u is $y - \hat{y} = [7\ 6]^T - [8\ 4]^T = [-1\ 2]^T$.
- That is, $[7\ 6] = [8\ 4] + [-1\ 2]$.

Orthonormal sets

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

Orthonormal sets

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

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

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 $\mathbf{R}^n$.

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 $\mathbf{R}^n$.
- Any nonempty subset of $\{e_1, \cdots, e_n\}$ is orthonormal, too

Orthonormal columns of a matrix

- Theorem 6: an $n \times m$ matrix U has orthonormal columns if and only if $U^T U = I$.

To check $u_1, u_2, \ldots, u_m$ to be orthonormal,

$$U = [u_1\ u_2\ \cdots\ u_n]$$

check $U^T U = I$

Orthonormal columns of a matrix

- Theorem 6: an $n \times m$ matrix U has orthonormal columns if and only if $U^T U = I$.

PF. To simplify notation, we suppose U has only three columns, $U = [u_1 \ u_2 \ u_2]$.

Orthonormal columns of a matrix

- Theorem 6: an $n \times m$ matrix U has orthonormal columns if and only if $U^T U = I$.

PF. To simplify notation, we suppose U has only three columns, $U = [u_1 \ u_2 \ u_2]$.

- Then

$$U^T U = \begin{bmatrix} u_1^T \\ u_2^T \\ u_3^T \end{bmatrix} [u_1 \ u_2 \ u_3] = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & u_1^T u_3 \\ u_2^T u_1 & u_2^T u_2 & u_2^T u_3 \\ u_3^T u_1 & u_3^T u_2 & u_3^T u_3 \end{bmatrix}$$

Orthonormal columns of a matrix

- Theorem 6: an $n \times m$ matrix U has orthonormal columns if and only if $U^T U = I$.

PF. To simplify notation, we suppose U has only three columns, $U = [u_1 \ u_2 \ u_2]$.

- Then

$$U^T U = \begin{bmatrix} u_1^T \\ u_2^T \\ u_3^T \end{bmatrix} [u_1 \ u_2 \ u_3] = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & u_1^T u_3 \\ u_2^T u_1 & u_2^T u_2 & u_2^T u_3 \\ u_3^T u_1 & u_3^T u_2 & u_3^T u_3 \end{bmatrix}$$

- The columns of U are orthogonal if and only if $u_i^T u_j = 0$ for $i \neq j$.

Orthonormal columns of a matrix

- Theorem 6: an $n \times m$ matrix U has orthonormal columns if and only if $U^T U = I$.

PF. To simplify notation, we suppose U has only three columns, $U = [u_1 \ u_2 \ u_2]$.

- Then

$$U^T U = \begin{bmatrix} u_1^T \\ u_2^T \\ u_3^T \end{bmatrix} [u_1 \ u_2 \ u_3] = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & u_1^T u_3 \\ u_2^T u_1 & u_2^T u_2 & u_2^T u_3 \\ u_3^T u_1 & u_3^T u_2 & u_3^T u_3 \end{bmatrix}$$

- The columns of U are orthogonal if and only if $u_i^T u_j = 0$ for $i \neq j$.
- The columns of U all have unit length if and only if $u_i^T u_i = 1$ for $i = 1, 2, 3$.

Orthonormal columns of a matrix

- Theorem 6: an $n \times m$ matrix U has orthonormal columns if and only if $U^T U = I$.

PF. To simplify notation, we suppose U has only three columns, $U = [u_1 \ u_2 \ u_2]$.

- Then

$$U^T U = \begin{bmatrix} u_1^T \\ u_2^T \\ u_3^T \end{bmatrix} [u_1 \ u_2 \ u_3] = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & u_1^T u_3 \\ u_2^T u_1 & u_2^T u_2 & u_2^T u_3 \\ u_3^T u_1 & u_3^T u_2 & u_3^T u_3 \end{bmatrix}$$

- The columns of U are orthogonal if and only if $u_i^T u_j = 0$ for $i \neq j$.
- The columns of U all have unit length if and only if $u_i^T u_i = 1$ for $i = 1, 2, 3$.
- From above conditions, the theorem follows immediately.

Orthonormal sets

- Theorem 7: Let U be an $m \times n$ matrix with orthonormal columns, and let x and y be in $\mathbf{R}^n$. Then

Orthonormal sets

- Theorem 7: Let U be an $m \times n$ matrix with orthonormal columns, and let x and y be in $\mathbf{R}^n$. Then
  - $||Ux|| = ||x||$

Orthonormal sets

- Theorem 7: Let U be an $m \times n$ matrix with orthonormal columns, and let x and y be in $\mathbf{R}^n$. Then
  - $||Ux|| = ||x||$
  - $(Ux) \cdot (Uy) = x \cdot y$

Orthonormal sets

- Theorem 7: Let U be an $m \times n$ matrix with orthonormal columns, and let x and y be in $\mathbf{R}^n$. Then
  - $||Ux|| = ||x||$
  - $(Ux) \cdot (Uy) = x \cdot y$
  - $(Ux) \cdot (Uy) = 0$ if and only if $x \cdot y = 0$.

Orthonormal sets

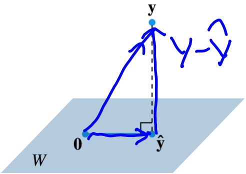
- Theorem 7: Let U be an $m \times n$ matrix with orthonormal columns, and let x and y be in $\mathbf{R}^n$. Then
  - $||Ux|| = ||x||$
  - $(Ux) \cdot (Uy) = x \cdot y$
  - $(Ux) \cdot (Uy) = 0$ if and only if $x \cdot y = 0$.

- The above properties say that the linear mapping $x \to Ux$ preserves lengths and orthogonality.

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

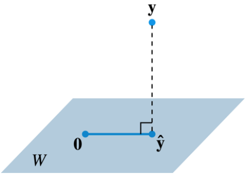
- The orthogonal projection of a point in $\mathbf{R}^2$ onto a line through the origin has an important analogue in $\mathbf{R}^n$.

- Given a vector y and a subspace W in $\mathbf{R}^n$, 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.



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

- Given a vector y and a subspace W in $\mathbf{R}^n$, 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.



- 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 $\mathbf{R}^n$. Then each y in $\mathbf{R}^n$ can be written uniquely in the form $y = \hat{y} + z$, where $\hat{y}$ is in W and z is in $W^{\perp}$.

- Theorem 8: Let W be a subspace of $\mathbf{R}^n$. Then each y in $\mathbf{R}^n$ 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

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \ldots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p \text{ and } z = y - \hat{y}$$

- Theorem 8: Let W be a subspace of $\mathbf{R}^n$. Then each y in $\mathbf{R}^n$ 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 $\{u_1 \dots, u_p\}$ is any orthogonal basis of W, then

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \ldots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p \text{ and } z = y - \hat{y}$$

PF. Similar to the proof of Theorem 5, we get $\hat{y}$.

- Theorem 8: Let W be a subspace of $\mathbf{R}^n$. Then each y in $\mathbf{R}^n$ 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

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \ldots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p \text{ and } z = y - \hat{y}$$

PF. Similar to the proof of Theorem 5, we get $\hat{y}$.

- To see that $z = y - \hat{y}$ is in $W^{\perp}$, we observe that z is orthogonal to each u_j in the basis for W, thus to every vector in W.

$z \perp u_i$ $\quad (y - \hat{y}) \cdot u_i = 0$

$y = c_1 u_1 + c_2 u_2 + \cdots + c_p u_p$

$y \cdot u_i - \hat{y} \cdot u_i = 0$

$y \cdot u_i = c_i u_i \cdot u_i$

- Theorem 8: Let W be a subspace of $\mathbf{R}^n$. Then each y in $\mathbf{R}^n$ 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

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \ldots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p \text{ and } z = y - \hat{y}$$

PF. Similar to the proof of Theorem 5, we get $\hat{y}$.

- To see that $z = y - \hat{y}$ is in $W^{\perp}$, we observe that z is orthogonal to each u_j in the basis for W, thus to every vector in W.
- To show the the decomposition 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 $W^{\perp}$.

- Theorem 8: Let W be a subspace of $\mathbf{R}^n$. Then each y in $\mathbf{R}^n$ 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

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \ldots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p \text{ and } z = y - \hat{y}$$

PF. Similar to the proof of Theorem 5, we get $\hat{y}$.

- To see that $z = y - \hat{y}$ is in $W^\perp$, we observe that z is orthogonal to each u_j in the basis for W, thus to every vector in W.
- To show the the decomposition 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 $y = \hat{y} + z = \hat{y}_1 + z_1$, and so $\hat{y} - \hat{y}_1 = z_1 - z$.

- Theorem 8: Let W be a subspace of $\mathbf{R}^n$. Then each y in $\mathbf{R}^n$ 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

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \ldots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p \text{ and } z = y - \hat{y}$$

PF. Similar to the proof of Theorem 5, we get $\hat{y}$.

- To see that $z = y - \hat{y}$ is in $W^\perp$, we observe that z is orthogonal to each u_j in the basis for W, thus to every vector in W.
- To show the the decomposition 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 $y = \hat{y} + z = \hat{y}_1 + z_1$, and so $\hat{y} - \hat{y}_1 = z_1 - z$.
- So vector $v = \hat{y} - \hat{y}_1$ is in W, but also in $W^\perp$, as $z - z_1$ is in $W^\perp$.

- Theorem 8: Let W be a subspace of $\mathbf{R}^n$. Then each y in $\mathbf{R}^n$ 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

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \ldots + \frac{y \cdot u_p}{u_p \cdot u_p} u_p \text{ and } z = y - \hat{y}$$

PF. Similar to the proof of Theorem 5, we get $\hat{y}$.

- To see that $z = y - \hat{y}$ is in $W^\perp$, we observe that z is orthogonal to each u_j in the basis for W, thus to every vector in W.
- To show the the decomposition 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 $y = \hat{y} + z = \hat{y}_1 + z_1$, and so $\hat{y} - \hat{y}_1 = z_1 - z$.
- So vector $v = \hat{y} - \hat{y}_1$ is in W, but also in $W^\perp$, as $z - z_1$ is in $W^\perp$.
- Hence $v \cdot v = 0$, and implies that $v = 0$. So $\hat{y} = \hat{y}_1$ and $z = z_1$.

Example

Ex. 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 $\{u_1, u_2\}$ is an orthogonal basis for $W = Span\{u_1, u_2\}$. Write y as the sum of a vector in W and a vector orthogonal to W.

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2 = \frac{9}{30}\begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$$

Example

Ex. 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 $\{u_1, u_2\}$ is an orthogonal basis for $W = Span\{u_1, u_2\}$. Write y as the sum of a vector in W and a vector orthogonal to W.

Sol. Let the vector in W be $\hat{y}$ and the vector orthogonal to W be $z = y - \hat{y}$.

Example

Ex. 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 $\{u_1, u_2\}$ is an orthogonal basis for $W = Span\{u_1, u_2\}$. Write y as the sum of a vector in W and a vector orthogonal to W.

Sol. Let the vector in W be $\hat{y}$ and the vector orthogonal to W be $z = y - \hat{y}$.

- Then

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2 = \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

Example

Ex. 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 $\{u_1, u_2\}$ is an orthogonal basis for $W = Span\{u_1, u_2\}$. Write y as the sum of a vector in W and a vector orthogonal to W.

Sol. Let the vector in W be $\hat{y}$ and the vector orthogonal to W be $z = y - \hat{y}$.

- Then

$$\hat{y} = \frac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{y \cdot u_2}{u_2 \cdot u_2} u_2 = \frac{9}{30} \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix} + \frac{3}{6} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 2 \\ 1/5 \end{bmatrix}$$

- And $z = y - \hat{y} = \begin{bmatrix} 7/5 \\ 0 \\ 14/5 \end{bmatrix}$

Properties of orthogonal projections

- If $\{u_1, \ldots, u_p\}$ is an orthogonal basis for W and if y happens to be in W, then the formula for $proj_W y$ is exactly the same as the representation of y given in Theorem 5.

Properties of orthogonal projections

- If $\{u_1, \ldots, u_p\}$ is an orthogonal basis for W and if y happens to be in W, then the formula for $proj_W y$ is exactly the same as the representation of y given in Theorem 5.

- In this case, $proj_W y = y$. In particular, if y is in $W = Span\{u_1, \ldots, u_p\}$, then $proj_W y = y$.

The best approximation theorem

- Theorem 9. Let W be a subspace of $\mathbf{R}^n$, let y be any vector in $\mathbf{R}^n$, 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 best approximation theorem

- Theorem 9. Let W be a subspace of $\mathbf{R}^n$, let y be any vector in $\mathbf{R}^n$, 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 best approximation theorem

- Theorem 9. Let W be a subspace of $\mathbf{R}^n$, let y be any vector in $\mathbf{R}^n$, 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 $v = \hat{y}$.

The best approximation theorem

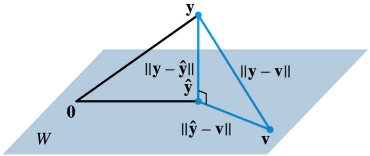
- Theorem 9. Let W be a subspace of $\mathbf{R}^n$, let y be any vector in $\mathbf{R}^n$, and let $\hat{y}$ be the orthogonal projection of y onto W. Then $\hat{y}$ is the closest point in W to y, in the sense that $||y - \hat{y}|| < ||y - v||$ for all v in W distinct from $\hat{y}$.
- The vector $\hat{y}$ 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 $||y - \hat{y}|| < ||y - v||$ leads to a new proof that $\hat{y}$ does not depend on the particular orthogonal basis used to compute it.

The best approximation theorem

- Theorem 9. Let W be a subspace of $\mathbf{R}^n$, let y be any vector in $\mathbf{R}^n$, 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 $||y - \hat{y}|| < ||y - v||$ 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, $\hat{y}$.

Proof of Theorem 9

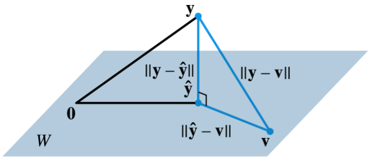
PF. Take v in W distinct from $\hat{y}$. Then $\hat{y} - v$ is in W.



The orthogonal projection of **y** onto W is the closest point in W to **y**.

Proof of Theorem 9

PF. Take v in W distinct from $\hat{y}$. Then $\hat{y} - v$ is in W.

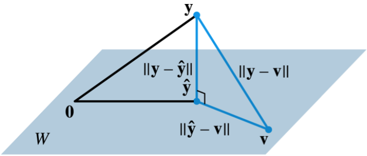


The orthogonal projection of **y** onto W is the closest point in W to **y**.

- By the orthogonal decomposition theorem, $y - \hat{y}$ is orthogonal to W.

Proof of Theorem 9

PF. Take v in W distinct from $\hat{y}$. Then $\hat{y} - v$ is in W.

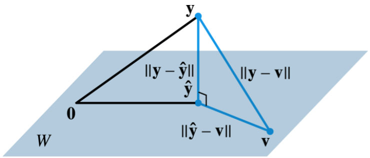


The orthogonal projection of **y** onto W is the closest point in W to **y**.

- 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).

Proof of Theorem 9

PF. Take v in W distinct from $\hat{y}$. Then $\hat{y} - v$ is in W.

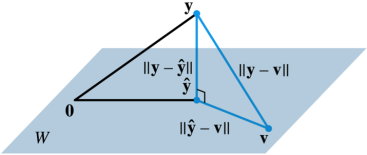


The orthogonal projection of **y** onto W is the closest point in W to **y**.

- 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 $||y - v||^2 = ||y - \hat{y}||^2 + ||\hat{y} - v||^2$.

Proof of Theorem 9

PF. Take v in W distinct from $\hat{y}$. Then $\hat{y} - v$ is in W.



The orthogonal projection of **y** onto W is the closest point in W to **y**.

- 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 $||y - v||^2 = ||y - \hat{y}||^2 + ||\hat{y} - v||^2$.
- Now $||\hat{y} - v||^2 > 0$ because $\hat{y} - v \neq 0$, and so inequality follows immediately.

Example

- The distance from a point y in $\mathbf{R}^n$ to a subspace W is defined as the distance from y to the nearest point in W.

Example

- The distance from a point y in $\mathbf{R}^n$ to a subspace W is defined as the distance from y to the nearest point in W.

Ex. Find the distance from y to $W = Span\{u_1, u_2\}$, where

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

$u_1 \cdot u_2 = 0 \Rightarrow u_1 \perp u_2$

Example

- The distance from a point y in $\mathbf{R}^n$ to a subspace W is defined as the distance from y to the nearest point in W.

Ex. Find the distance from y to $W = Span\{u_1, u_2\}$, where

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

Sol. By the Best Approximation Theorem, the distance from y to W is $||y - \hat{y}||$, where $\hat{y} = proj_W y$.

Example

- The distance from a point y in $\mathbf{R}^n$ to a subspace W is defined as the distance from y to the nearest point in W.

Ex. Find the distance from y to $W = Span\{u_1, u_2\}$, where

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

Sol. By the Best Approximation Theorem, the distance from y to W is $||y - \hat{y}||$, where $\hat{y} = proj_W y$.

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

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

Example

- The distance from a point y in $\mathbf{R}^n$ to a subspace W is defined as the distance from y to the nearest point in W.

Ex. Find the distance from y to $W = Span\{u_1, u_2\}$, where

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

Sol. By the Best Approximation Theorem, the distance from y to W is $||y - \hat{y}||$, where $\hat{y} = proj_W y$.

- So $\hat{y} = \frac{15}{30}u_1 + \frac{-21}{6}u_2 = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$
- It follows that $y - \hat{y} = [0\ 3\ 6]^T$

Example

- The distance from a point y in $\mathbf{R}^n$ to a subspace W is defined as the distance from y to the nearest point in W.

Ex. Find the distance from y to $W = Span\{u_1, u_2\}$, where

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

Sol. By the Best Approximation Theorem, the distance from y to W is $||y - \hat{y}||$, where $\hat{y} = proj_W y$.

- So $\hat{y} = \frac{15}{30} u_1 + \frac{-21}{6} u_2 = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$
- It follows that $y - \hat{y} = [0\ 3\ 6]^T$
- So the distance is $||y - \hat{y}|| = \sqrt{(y - \hat{y}) \cdot (y - \hat{y})} = \sqrt{45}$.

Properties of Orthogonormal projections

- Theorem 10. If $\{u_1, \ldots, u_p\}$ is an orthogonormal basis for a subspace W of $\mathbf{R}^n$, then

$$proj_W y = (y \cdot u_1)u_1 + (y \cdot u_2)u_2 + \ldots + (y \cdot u_p)u_p$$

If $U = [u_1 \; u_2 \; \ldots \; u_p]$, then $proj_W y = UU^T y$ for all $y \in \mathbf{R}^n$.