Best Approximation

Let W be a subspace of $\mathcal{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

Best Approximation

Let W be a subspace of $\mathcal{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

- DEFINITION: For any vector $\mathbf{y} \in \mathcal{R}^n$, define **orthogonal projection** as

$$\mathbf{proj}_W \mathbf{y} = \widehat{\mathbf{y}} = \frac{\mathbf{y}^T \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}^T \mathbf{u}_p}{\mathbf{u}_p^T \mathbf{u}_p}\mathbf{u}_p$$

Best Approximation

Let W be a subspace of $\mathcal{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

- DEFINITION: For any vector $\mathbf{y} \in \mathcal{R}^n$, define **orthogonal projection** as

$$\mathbf{proj}_W\mathbf{y} = \widehat{\mathbf{y}} = \frac{\mathbf{y}^T\mathbf{u}_1}{\mathbf{u}_1^T\mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}^T\mathbf{u}_p}{\mathbf{u}_p^T\mathbf{u}_p}\mathbf{u}_p$$

- **Thm**: For all $\mathbf{v} \in W, \mathbf{v} \neq \widehat{\mathbf{y}}$,

$$\|\mathbf{y} - \widehat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|.$$

Best Approximation

Let W be a subspace of $\mathcal{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

- Definition: For any vector $\mathbf{y} \in \mathcal{R}^n$, define **orthogonal projection** as

$$\mathbf{proj}_W \mathbf{y} = \widehat{\mathbf{y}} = \frac{\mathbf{y}^T \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}^T \mathbf{u}_p}{\mathbf{u}_p^T \mathbf{u}_p}\mathbf{u}_p$$

- **Thm**: For all $\mathbf{v} \in W, \mathbf{v} \neq \widehat{\mathbf{y}}$,

$$\|\mathbf{y} - \widehat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|.$$

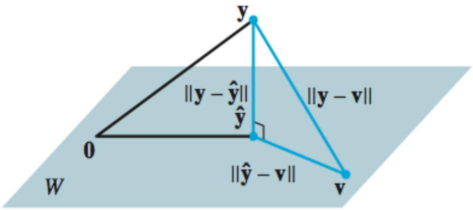


FIGURE 4 The orthogonal projection of **y** onto W is the closest point in W to **y**.

Best Approximation

Let W be a subspace of $\mathcal{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

- DEFINITION: For any vector $\mathbf{y} \in \mathcal{R}^n$, define **orthogonal projection** as

$$\mathbf{proj}_W \mathbf{y} = \widehat{\mathbf{y}} = \frac{\mathbf{y}^T \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1} \mathbf{u}_1 + \cdots + \frac{\mathbf{y}^T \mathbf{u}_p}{\mathbf{u}_p^T \mathbf{u}_p} \mathbf{u}_p$$

- **Thm**: For all $\mathbf{v} \in W, \mathbf{v} \neq \widehat{\mathbf{y}}$,

$$\|\mathbf{y} - \widehat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|.$$

Best Approximation

Let W be a subspace of $\mathcal{R}^n$ with orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

- DEFINITION: For any vector $\mathbf{y} \in \mathcal{R}^n$, define **orthogonal projection** as

$$\mathbf{proj}_W \mathbf{y} = \widehat{\mathbf{y}} = \frac{\mathbf{y}^T \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}^T \mathbf{u}_p}{\mathbf{u}_p^T \mathbf{u}_p}\mathbf{u}_p$$

- **Thm**: For all $\mathbf{v} \in W, \mathbf{v} \neq \widehat{\mathbf{y}}$,

$$\|\mathbf{y} - \widehat{\mathbf{y}}\| < \|\mathbf{y} - \mathbf{v}\|.$$

PROOF: It is clear that $\mathbf{z} \overset{def}{=} \mathbf{y} - \widehat{\mathbf{y}} \in W^{\perp}$. Then

$$\mathbf{y} - \mathbf{v} = \mathbf{z} + (\widehat{\mathbf{y}} - \mathbf{v}),$$

with $\mathbf{z} \in W^{\perp}$, $\widehat{\mathbf{y}} - \mathbf{v} \in W$. By Pythagorean Thm,

$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{z}\|^2 + \|\widehat{\mathbf{y}} - \mathbf{v}\|^2 > \|\mathbf{z}\|^2 = \|\mathbf{y} - \widehat{\mathbf{y}}\|^2. \quad \square$$

Let W be a subspace of $\mathcal{R}^3$ with orthogonal basis $\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2\}$.

$$\text{For any vector } \mathbf{y} \in \mathcal{R}^3, \ \ \mathbf{proj}_W \mathbf{y} \stackrel{def}{=} \widehat{\mathbf{y}} = \frac{\mathbf{y}^T \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1} \mathbf{u}_1 + \frac{\mathbf{y}^T \mathbf{u}_2}{\mathbf{u}_2^T \mathbf{u}_2} \mathbf{u}_2$$

$$\textbf{Thm}: \quad \|\mathbf{y} - \widehat{\mathbf{y}}\| = \textbf{minimize}_{\ \mathbf{v} \in W} \ \|\mathbf{y} - \mathbf{v}\| .$$

Let W be a subspace of $\mathcal{R}^3$ with orthogonal basis $\mathcal{S} = \{\mathbf{u}_1, \mathbf{u}_2\}$.

$$\text{For any vector } \mathbf{y} \in \mathcal{R}^3, \quad \mathbf{proj}_W \mathbf{y} \stackrel{def}{=} \widehat{\mathbf{y}} = \frac{\mathbf{y}^T \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}^T \mathbf{u}_2}{\mathbf{u}_2^T \mathbf{u}_2}\mathbf{u}_2$$

Thm: $\|\mathbf{y} - \widehat{\mathbf{y}}\| = \textbf{minimize}_{\ \mathbf{v} \in W} \ \|\mathbf{y} - \mathbf{v}\|$.

EXAMPLE: If $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Then

$$\widehat{\mathbf{y}} = \frac{\mathbf{y}^T \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1}\mathbf{u}_1 + \frac{\mathbf{y}^T \mathbf{u}_2}{\mathbf{u}_2^T \mathbf{u}_2}\mathbf{u}_2 = \frac{1}{5}\begin{bmatrix} -2 \\ 10 \\ 1 \end{bmatrix},$$

$$\textbf{minimize}_{\ \mathbf{v} \in W} \ \|\mathbf{y} - \mathbf{v}\| = \|\mathbf{y} - \widehat{\mathbf{y}}\| = \frac{7}{\sqrt{5}}.$$

REVIEW:

- Matrix $U = (\mathbf{u}_1, \cdots, \mathbf{u}_p)$ has orthonormal columns $\iff U^T U = I$.

Review:

- Matrix $U = (\mathbf{u}_1, \cdots, \mathbf{u}_p)$ has orthonormal columns $\iff U^T U = I$.
- U has orthonormal columns $\implies \mathcal{S} = \{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ is an orthonormal basis for $\mathbf{Span}\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

REVIEW:

- Matrix $U = (\mathbf{u}_1, \cdots, \mathbf{u}_p)$ has orthonormal columns $\iff U^T U = I$.
- U has orthonormal columns $\implies \mathcal{S} = \{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$ is an orthonormal basis for **Span** $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

Thm 10: Let W be a subspace of $\mathcal{R}^n$ with orthonormal basis $\mathcal{S} = \{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$. Then for any vector $\mathbf{y} \in \mathcal{R}^n$,

$$\begin{aligned}
\mathbf{proj}_W \mathbf{y} &= \widehat{\mathbf{y}} = \frac{\mathbf{y}^T \mathbf{u}_1}{\mathbf{u}_1^T \mathbf{u}_1}\mathbf{u}_1 + \cdots + \frac{\mathbf{y}^T \mathbf{u}_p}{\mathbf{u}_p^T \mathbf{u}_p}\mathbf{u}_p \\
&= \mathbf{u}_1 \left(\mathbf{u}_1^T \mathbf{y}\right) + \cdots + \mathbf{u}_p \left(\mathbf{u}_p^T \mathbf{y}\right) \\
&= \left(\mathbf{u}_1 \mathbf{u}_1^T + \cdots + \mathbf{u}_p \mathbf{u}_p^T\right) \mathbf{y} \\
&= \left(U\, U^T\right) \mathbf{y}
\end{aligned}$$

$\boxed{\text{How to construct an orthonormal basis?}}$

§6.4 Gram-Schmidt Process

EXAMPLE 1: Let $W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$ with

$$\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Construct an orthogonal basis for W.

SOLUTION: Choose $\mathbf{v}_1 = \mathbf{x}_1$, and

$$\mathbf{v}_2 = \mathbf{x}_2 - \textbf{proj}_{\textbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2^T\,\mathbf{v}_1}{\mathbf{v}_1^T\,\mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} - \frac{15}{45}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

$\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal, with $\textbf{Span}\{\mathbf{v}_1, \mathbf{v}_2\} = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$.

EXAMPLE 1: Let $W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$ with

$$\mathbf{x}_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

Construct an <u>orthonormal basis</u> for W.

SOLUTION:

$$\mathbf{q}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{q}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.$$

EXAMPLE 2: Construct an orthogonal basis for $W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ with

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

EXAMPLE 2: Construct an orthogonal basis for $W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ with

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

SOLUTION: Choose $\mathbf{v}_1 = \mathbf{x}_1$, and

$$\mathbf{v}_2 = \mathbf{x}_2 - \textbf{proj}_{\textbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

$\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal, with $\textbf{Span}\{\mathbf{v}_1, \mathbf{v}_2\} = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$.

EXAMPLE 2: Construct an orthogonal basis for $W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ with

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

SOLUTION: Choose $\mathbf{v}_1 = \mathbf{x}_1$, and

$$\mathbf{v}_2 = \mathbf{x}_2 - \textbf{proj}_{\textbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

$\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal, with $\textbf{Span}\{\mathbf{v}_1, \mathbf{v}_2\} = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2\}$.

$$\mathbf{v}_3 = \mathbf{x}_3 - \textbf{proj}_{\textbf{Span}\{\mathbf{v}_1, \mathbf{v}_2\}}\mathbf{x}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3^T \mathbf{v}_2}{\mathbf{v}_2^T \mathbf{v}_2}\mathbf{v}_2 = \frac{1}{3}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}.$$

$\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ are orthogonal, with $\textbf{Span}\{\mathbf{v}_1, \mathbf{v}_2, , \mathbf{v}_3\} = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$.

EXAMPLE 2: Construct an orthonormal basis for $W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$

$$\text{with} \quad \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

EXAMPLE 2: Construct an orthonormal basis for $W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$

$$\text{with} \quad \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

SOLUTION:

$$\mathbf{q}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{q}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \frac{1}{\sqrt{12}}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\text{and} \quad \mathbf{q}_3 = \frac{1}{\|\mathbf{v}_3\|}\mathbf{v}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}.$$

Gram-Schmidt Process

Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p\}$ for subspace W of $\mathcal{R}^n$, define

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

Gram-Schmidt Process

Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p\}$ for subspace W of $\mathcal{R}^n$, define

- $\mathbf{v}_1 = \mathbf{x}_1$
- $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2$

Gram-Schmidt Process

Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p\}$ for subspace W of $\mathcal{R}^n$, define

- $\mathbf{v}_1 = \mathbf{x}_1$
- $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2$
- $\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1,\mathbf{v}_2\}}\mathbf{x}_3$

Gram-Schmidt Process

Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p\}$ for subspace W of $\mathcal{R}^n$, define

- $\mathbf{v}_1 = \mathbf{x}_1$
- $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2$
- $\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1, \mathbf{v}_2\}}\mathbf{x}_3$

$$\vdots$$

- $\mathbf{v}_p = \mathbf{x}_p - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_{p-1}\}}\mathbf{x}_p$

Gram-Schmidt Process

Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p\}$ for subspace W of $\mathcal{R}^n$, define

- $\mathbf{v}_1 = \mathbf{x}_1$
- $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2$
- $\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1,\mathbf{v}_2\}}\mathbf{x}_3$

$$\vdots$$

- $\mathbf{v}_p = \mathbf{x}_p - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1,\cdots,\mathbf{v}_{p-1}\}}\mathbf{x}_p$

Then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is an orthogonal basis for W, and

$$\mathbf{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\} = \mathbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k\}, \quad k = 1, \cdots, p.$$

Gram-Schmidt Process

Given a basis $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_p\}$ for subspace W of $\mathcal{R}^n$, define

- $\mathbf{v}_1 = \mathbf{x}_1$
- $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2$
- $\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1,\mathbf{v}_2\}}\mathbf{x}_3$

$$\vdots$$

- $\mathbf{v}_p = \mathbf{x}_p - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1,\cdots,\mathbf{v}_{p-1}\}}\mathbf{x}_p$

Then $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_p\}$ is an orthogonal basis for W, and

$$\mathbf{Span}\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\} = \mathbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_k\}, \quad k = 1, \cdots, p.$$

$$\text{Orthonormal basis} \quad \mathbf{q}_k = \frac{1}{\|\mathbf{v}_k\|}\mathbf{v}_k, \quad k = 1, \cdots, p.$$

Gram-Schmidt Process = QR Factorization (I)

- $\mathbf{v}_1 = \mathbf{x}_1 \quad \Longrightarrow \quad \mathbf{x}_1 = \mathbf{v}_1 \stackrel{def}{=} \widehat{r}_{1,1}\,\mathbf{v}_1 = [\mathbf{v}_1,\, \mathbf{v}_2,\, \cdots, \mathbf{v}_p] \begin{bmatrix} \widehat{r}_{1,1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

Gram-Schmidt Process = QR Factorization (I)

- $\mathbf{v}_1 = \mathbf{x}_1 \quad \Longrightarrow \quad \mathbf{x}_1 = \mathbf{v}_1 \stackrel{def}{=} \widehat{r}_{1,1}\,\mathbf{v}_1 = [\mathbf{v}_1,\, \mathbf{v}_2,\, \cdots,\, \mathbf{v}_p] \begin{bmatrix} \widehat{r}_{1,1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$
- $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2 \quad \Longrightarrow \quad \mathbf{x}_2 = \mathbf{v}_2 + \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2$

Gram-Schmidt Process = QR Factorization (I)

- $\mathbf{v}_1 = \mathbf{x}_1 \quad \Longrightarrow \quad \mathbf{x}_1 = \mathbf{v}_1 \stackrel{def}{=} \widehat{r}_{1,1}\,\mathbf{v}_1 = [\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_p] \begin{bmatrix} \widehat{r}_{1,1} \\ 0 \\ \vdots \\ 0 \end{bmatrix}$
- $\mathbf{v}_2 = \mathbf{x}_2 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2 \quad \Longrightarrow \quad \mathbf{x}_2 = \mathbf{v}_2 + \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1\}}\mathbf{x}_2$

$$\begin{aligned} \mathbf{x}_2 &= \mathbf{v}_2 + \frac{\mathbf{x}_2^T\,\mathbf{v}_1}{\mathbf{v}_1^T\,\mathbf{v}_1}\mathbf{v}_1 \stackrel{def}{=} \widehat{r}_{1,2}\,\mathbf{v}_1 + \widehat{r}_{2,2}\,\mathbf{v}_2 \\ &= [\mathbf{v}_1,\ \mathbf{v}_2,\ \cdots,\mathbf{v}_p] \begin{bmatrix} \widehat{r}_{1,2} \\ \widehat{r}_{2,2} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \end{aligned}$$

Gram-Schmidt Process = QR Factorization (II)

- $\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1, \mathbf{v}_2\}} \mathbf{x}_3 \implies \mathbf{x}_3 = \mathbf{v}_3 + \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1, \mathbf{v}_2\}} \mathbf{x}_3$

Gram-Schmidt Process = QR Factorization (II)

- $\mathbf{v}_3 = \mathbf{x}_3 - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1,\mathbf{v}_2\}}\mathbf{x}_3 \implies \mathbf{x}_3 = \mathbf{v}_3 + \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1,\mathbf{v}_2\}}\mathbf{x}_3$

$$\begin{aligned}
\mathbf{x}_3 &= \mathbf{v}_3 + \frac{\mathbf{x}_3^T\,\mathbf{v}_1}{\mathbf{v}_1^T\,\mathbf{v}_1}\,\mathbf{v}_1 + \frac{\mathbf{x}_3^T\,\mathbf{v}_2}{\mathbf{v}_2^T\,\mathbf{v}_2}\,\mathbf{v}_2 \stackrel{def}{=} \widehat{r}_{1,3}\,\mathbf{v}_1 + \widehat{r}_{2,3}\,\mathbf{v}_2 + \widehat{r}_{3,3}\,\mathbf{v}_3 \\
&= [\mathbf{v}_1,\,\mathbf{v}_2,\,\mathbf{v}_3,\,\cdots,\mathbf{v}_p] \begin{bmatrix} \widehat{r}_{1,3} \\ \widehat{r}_{2,3} \\ \widehat{r}_{3,3} \\ 0 \\ \vdots \\ 0 \end{bmatrix}
\end{aligned}$$

-

$$\vdots$$

Gram-Schmidt Process = QR Factorization (III)

$$\mathbf{v}_p = \mathbf{x}_p - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_{p-1}\}} \mathbf{x}_p \implies \mathbf{v}_p = \mathbf{x}_p + \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1, , \cdots, \mathbf{v}_{p-1}\}} \mathbf{x}_p$$

Gram-Schmidt Process = QR Factorization (III)

$$\mathbf{v}_p = \mathbf{x}_p - \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1,\cdots,\mathbf{v}_{p-1}\}}\mathbf{x}_p \implies \mathbf{v}_p = \mathbf{x}_p + \mathbf{proj}_{\mathbf{Span}\{\mathbf{v}_1,,\cdots,\mathbf{v}_{p-1}\}}\mathbf{x}_p$$

$$\begin{aligned}
\mathbf{x}_p &= \mathbf{v}_p + \frac{\mathbf{x}_p^T\,\mathbf{v}_1}{\mathbf{v}_1^T\,\mathbf{v}_1}\,\mathbf{v}_1 + \cdots + \frac{\mathbf{x}_p^T\,\mathbf{v}_{p-1}}{\mathbf{v}_{p-1}^T\,\mathbf{v}_{p-1}}\,\mathbf{v}_{p-1} \\
&\stackrel{def}{=} \widehat{r}_{1,p}\,\mathbf{v}_1 + \cdots + \widehat{r}_{p-1,p}\,\mathbf{v}_{p-1} + \widehat{r}_{p,p}\,\mathbf{v}_p \\
&= [\mathbf{v}_1,\,\mathbf{v}_2,\,\cdots,\mathbf{v}_p]\begin{bmatrix}\widehat{r}_{1,p}\\ \widehat{r}_{2,p}\\ \vdots\\ \widehat{r}_{p,p}\end{bmatrix} \\
[\mathbf{x}_1,\,\mathbf{x}_2,\,\cdots,\mathbf{x}_p] &= [\mathbf{v}_1,\,\mathbf{v}_2,\,\cdots,\mathbf{v}_p]\begin{bmatrix}\widehat{r}_{1,1} & \widehat{r}_{1,2} & \cdots & \widehat{r}_{1,p}\\ & \widehat{r}_{2,2} & \cdots & \widehat{r}_{2,p}\\ & & \ddots & \vdots\\ & & & \widehat{r}_{p,p}\end{bmatrix}
\end{aligned}$$

Gram-Schmidt Process = QR Factorization (IV)

Normalizing $\mathbf{v}_1, \cdots, \mathbf{v}_p$: $\mathbf{q}_j = \frac{\mathbf{v}_j}{\|\mathbf{v}_j\|}$ for $j = 1, \cdots, p$:

$$[\mathbf{x}_1,\, \mathbf{x}_2,\, \cdots, \mathbf{x}_p] = [\mathbf{v}_1,\, \mathbf{v}_2,\, \cdots, \mathbf{v}_p] \begin{bmatrix} \widehat{r}_{1,1} & \widehat{r}_{1,2} & \cdots & \widehat{r}_{1,p} \\ & \widehat{r}_{2,2} & \cdots & \widehat{r}_{2,p} \\ & & \ddots & \vdots \\ & & & \widehat{r}_{p,p} \end{bmatrix}$$

$$= [\mathbf{q}_1,\, \mathbf{q}_2,\, \cdots, \mathbf{q}_p] \left(\begin{bmatrix} \|\mathbf{v}_1\| & & & \\ & \|\mathbf{v}_2\| & & \\ & & \ddots & \\ & & & \|\mathbf{v}_p\| \end{bmatrix} \begin{bmatrix} \widehat{r}_{1,1} & \widehat{r}_{1,2} & \cdots & \widehat{r}_{1,p} \\ & \widehat{r}_{2,2} & \cdots & \widehat{r}_{2,p} \\ & & \ddots & \vdots \\ & & & \widehat{r}_{p,p} \end{bmatrix} \right)$$

$$\overset{def}{=} \quad Q\,R = Q \left(\quad \right), \qquad Q^T\,Q = I.$$

$\boxed{\text{Revisit}}$ EXAMPLE 2: Construct an orthogonal basis for

$$W = \mathbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} \text{ with } \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

$\boxed{\text{Revisit}}$ EXAMPLE 2: Construct an orthogonal basis for

$$W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} \text{ with } \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

SOLUTION: $\mathbf{v}_1 = \mathbf{x}_1$, and

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1}\mathbf{v}_1 = \frac{1}{4}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{with} \quad \widehat{r}_{1,2} = \frac{3}{4}.$$

$\boxed{\text{Revisit}}$ EXAMPLE 2: Construct an orthogonal basis for

$$W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} \text{ with } \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

SOLUTION: $\mathbf{v}_1 = \mathbf{x}_1$, and

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1}\mathbf{v}_1 = \frac{1}{4}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{with} \quad \widehat{r}_{1,2} = \frac{3}{4}.$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3^T \mathbf{v}_2}{\mathbf{v}_2^T \mathbf{v}_2}\mathbf{v}_2 = \frac{1}{3}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}, \quad \text{with} \quad \widehat{r}_{1,3} = \frac{1}{2}, \ \widehat{r}_{2,3} = \frac{2}{3}.$$

$$[\mathbf{x}_1, \ \mathbf{x}_2, \ \mathbf{x}_3] = [\mathbf{v}_1, \ \mathbf{v}_2, \ \mathbf{v}_3] \begin{bmatrix} 1 & \frac{3}{4} & \frac{1}{2} \\ & 1 & \frac{2}{3} \\ & & 1 \end{bmatrix}$$

$\boxed{\text{Revisit}}$ EXAMPLE 2: Construct orthonormal basis for

$$W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} \text{ with } \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

$\boxed{\text{Revisit}}$ EXAMPLE 2: Construct orthonormal basis for

$$W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} \text{ with } \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

SOLUTION: $\mathbf{v}_1 = \mathbf{x}_1$, and

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1}\mathbf{v}_1 = \frac{1}{4}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{with} \quad \widehat{r}_{1,2} = \frac{3}{4}.$$

$\boxed{\text{Revisit}}$ EXAMPLE 2: Construct orthonormal basis for

$$W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} \text{ with } \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

SOLUTION: $\mathbf{v}_1 = \mathbf{x}_1$, and

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1}\mathbf{v}_1 = \frac{1}{4}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \text{with} \quad \widehat{r}_{1,2} = \frac{3}{4}.$$

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3^T \mathbf{v}_1}{\mathbf{v}_1^T \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3^T \mathbf{v}_2}{\mathbf{v}_2^T \mathbf{v}_2}\mathbf{v}_2 = \frac{1}{3}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}, \quad \text{with} \quad \widehat{r}_{1,3} = \frac{1}{2}, \ \widehat{r}_{2,3} = \frac{2}{3}.$$

$$[\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] = [\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3]\begin{bmatrix} 1 & \frac{3}{4} & \frac{1}{2} \\ & 1 & \frac{2}{3} \\ & & 1 \end{bmatrix}$$

EXAMPLE 2: Construct an orthonormal basis for $W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$

$$\text{with}\quad \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix},\ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix},\ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

EXAMPLE 2: Construct an orthonormal basis for $W = \textbf{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$

$$\text{with} \quad \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

SOLUTION: $\|\mathbf{v}_1\| = 2, \ \|\mathbf{v}_2\| = \frac{\sqrt{3}}{2}, \ \|\mathbf{v}_3\| = \frac{\sqrt{6}}{3}$

$$\mathbf{q}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \ \mathbf{q}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \frac{1}{\sqrt{12}}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\text{and} \quad \mathbf{q}_3 = \frac{1}{\|\mathbf{v}_3\|}\mathbf{v}_3 = \frac{1}{\sqrt{6}}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}.$$

QR Factorization

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3] = Q\,R, \quad \text{where}$$

$$Q = [\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3] = \begin{bmatrix} \frac{1}{2} & -\frac{3}{\sqrt{12}} & 0 \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & -\frac{2}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \\ \frac{1}{2} & \frac{1}{\sqrt{12}} & \frac{1}{\sqrt{6}} \end{bmatrix}$$

$$R = \begin{bmatrix} 2 & & \\ & \frac{\sqrt{3}}{2} & \\ & & \frac{\sqrt{6}}{3} \end{bmatrix} \begin{bmatrix} 1 & \frac{3}{4} & \frac{1}{2} \\ & 1 & \frac{2}{3} \\ & & 1 \end{bmatrix} = \begin{bmatrix} 2 & \frac{3}{2} & 1 \\ & \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{3} \\ & & \frac{\sqrt{6}}{3} \end{bmatrix}$$

$$\boxed{\text{Gram-Schmidt} \implies \text{QR}, \quad \underline{\text{but}}\ \text{QR} \;\not\Longrightarrow\; \text{Gram-Schmidt}}$$

§6.5 Least Squares Approximation

Example Problem:

- **Given:** Decennial census data since 1610 on US population.
- **Predict:** US population in next twenty years.

Real focus is on prediction

§6.5 Least Squares Approximation

Example Problem:

- **Given:** Decennial census data since 1610 on US population.
- **Predict:** US population in next twenty years.

$\boxed{\text{Real focus is on prediction}}$

Example Models:

- LINEAR MODEL:

$$\mathcal{P}\left(\mathit{Year}\right) \approx \alpha + \beta \times \mathit{Year}.$$

- LOG-LINEAR MODEL:

$$\mathcal{P}\left(\mathit{Year}\right) \approx \mathbf{exp}\left(\alpha + \beta \times \mathit{Year}\right).$$

§6.5 Least Squares Approximation

Example Problem:

- **Given:** Decennial census data since 1610 on US population.
- **Predict:** US population in next twenty years.

$\boxed{\text{Real focus is on prediction}}$

Example Models:

- LINEAR MODEL:

$$\mathcal{P}\left(\mathit{Year}\right) \approx \alpha + \beta \times \mathit{Year}.$$

- LOG-LINEAR MODEL:

$$\mathcal{P}\left(\mathit{Year}\right) \approx \mathbf{exp}\left(\alpha + \beta \times \mathit{Year}\right).$$

$\boxed{\text{Model may } \underline{\text{not}} \text{ be exact, but could have predictive value.}}$

US Population

Census year	Population
1610	350
1620	2,302
1630	4,646
1640	26,634
1650	50,368
1660	75,058
1670	111,935
1680	151,507
1690	210,372
1700	250,888
1710	331,711
1720	466,185
1730	629,445
1740	905,563
1750	1,170,760
1760	1,593,625
1770	2,148,076
1780	2,780,369
1790	3,929,214
1800	5,308,483
1810	7,239,881
1820	9,638,453
1830	12,866,020
1840	17,069,453
1850	23,191,876
1860	31,443,321
1870	38,558,371
1880	50,189,209
1890	62,979,766
1900	76,212,168
1910	92,228,496
1920	106,021,537
1930	123,202,624
1940	132,164,569
1950	151,325,798
1960	179,323,175
1970	203,211,926
1980	226,545,805
1990	248,709,873
2000	281,421,906
2010	308,745,538

Least Squares Model Solution

Given population data for $\text{year}_1, \cdots, \text{year}_n$. Define for $1 \le i \le n$, $x_i = \text{year}_i$,

$$y_i = \begin{cases} \mathcal{P}(\text{year}_i), & \textsc{Linear Model}, \\ \mathbf{log}\left(\mathcal{P}(\text{year}_i)\right), & \textsc{Log-linear Model}. \end{cases}$$

- **Least Squares Fit:**

$$\mathbf{min}_{\alpha,\beta} \sum_{i=1}^{n} \left(y_i - (\alpha + x_i\beta)\right)^2 .$$

- **Prediction:** For future year x, US population will be

$$\mathcal{P}(x) = \begin{cases} \alpha + x\beta, & \textsc{Linear Model}, \\ \mathbf{exp}(\alpha + x\beta), & \textsc{Log-linear Model}. \end{cases}$$

Least Squares Fit: $\min_{\alpha,\beta} \sum_{i=1}^{n} (y_i - (\alpha + x_i\beta))^2$

$$
\begin{aligned}
\sum_{i=1}^{n} (y_i - (\alpha + x_i\beta))^2 &= \left\| \begin{pmatrix} y_1 - (\alpha + x_1\beta) \\ \vdots \\ y_n - (\alpha + x_n\beta) \end{pmatrix} \right\|^2 \\
&= \left\| \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \right\|^2 \\
&= \|\mathbf{b} - A\,\mathbf{x}\|^2, \quad \text{where}
\end{aligned}
$$

$$
\mathbf{b} \stackrel{def}{=} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad A \stackrel{def}{=} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \mathbf{x} \stackrel{def}{=} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.
$$

Least Squares Fit becomes: FINDING $\widehat{\mathbf{x}} \in \mathcal{R}^2$ so that for all $\mathbf{x} \in \mathcal{R}^2$

$\|\mathbf{b} - A\,\widehat{\mathbf{x}}\| \leq \|\mathbf{b} - A\,\mathbf{x}\|$.

Least Squares Fit: $\mathbf{min}_{\alpha,\beta} \sum_{i=1}^{n} (y_i - (\alpha + x_i\beta))^2$

$$\sum_{i=1}^{n} (y_i - (\alpha + x_i\beta))^2 = \left\| \begin{pmatrix} y_1 - (\alpha + x_1\beta) \\ \vdots \\ y_n - (\alpha + x_n\beta) \end{pmatrix} \right\|^2$$

$$= \left\| \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} - \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \right\|^2$$

$$= \|\mathbf{b} - A\,\mathbf{x}\|^2, \quad \text{where}$$

$$\mathbf{b} \stackrel{def}{=} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad A \stackrel{def}{=} \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \mathbf{x} \stackrel{def}{=} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

Least Squares Fit becomes: FINDING $\widehat{\mathbf{x}} \in \mathcal{R}^2$ so that for all $\mathbf{x} \in \mathcal{R}^2$

$\|\mathbf{b} - A\,\widehat{\mathbf{x}}\| \leq \|\mathbf{b} - A\,\mathbf{x}\|$. In general, $A \in \mathbb{R}^{m \times n}, \mathbf{x} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$; $m \geq n$.

Least Squares Solution: $\|\mathbf{b} - A\widehat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$

Since $A\mathbf{x} \in \text{Col } A$ for any $\mathbf{x} \in \mathbb{R}^n$, we must have

- $\left\|\mathbf{b} - \mathbf{Proj}_{\text{Col } A}\mathbf{b}\right\| = \mathbf{minimize}_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\|$.

Least Squares Solution: $\|\mathbf{b} - A\widehat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$

Since $A\mathbf{x} \in \text{Col } A$ for any $\mathbf{x} \in \mathbb{R}^n$, we must have

- $\left\|\mathbf{b} - \mathbf{Proj}_{\text{Col } A}\mathbf{b}\right\| = \mathbf{minimize}_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\|$.
- $\mathbf{b} - \mathbf{Proj}_{\text{Col } A}\mathbf{b} \in (\text{Col } A)^{\perp}$.

Least Squares Solution: $\|\mathbf{b} - A\widehat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|$

Since $A\mathbf{x} \in \text{Col } A$ for any $\mathbf{x} \in \mathbb{R}^n$, we must have

- $\left\|\mathbf{b} - \mathbf{Proj}_{\text{Col } A}\mathbf{b}\right\| = \mathbf{minimize}_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\|$.
- $\mathbf{b} - \mathbf{Proj}_{\text{Col } A}\mathbf{b} \in (\text{Col } A)^{\perp}$.
- $\mathbf{Proj}_{\text{Col } A}\mathbf{b} \in \text{Col } A$, there must be an $\widehat{\mathbf{x}}$ so that $\mathbf{Proj}_{\text{Col } A}\mathbf{b} = A\widehat{\mathbf{x}}$.

$$\Longrightarrow A^T\left(\mathbf{b} - A\widehat{\mathbf{x}}\right) = \mathbf{0}.$$

Least Squares Solution: $\|\mathbf{b} - A\widehat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$

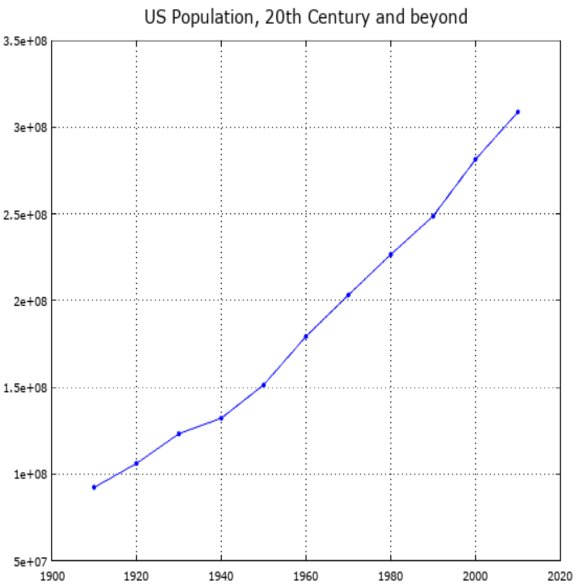
Since $A\mathbf{x} \in \text{Col } A$ for any $\mathbf{x} \in \mathbb{R}^n$, we must have

- $\left\|\mathbf{b} - \mathbf{Proj}_{\text{Col } A}\,\mathbf{b}\right\| = \mathbf{minimize}_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{b} - A\mathbf{x}\|$.
- $\mathbf{b} - \mathbf{Proj}_{\text{Col } A}\,\mathbf{b} \in (\text{Col } A)^{\perp}$.
- $\mathbf{Proj}_{\text{Col } A}\,\mathbf{b} \in \text{Col } A$, there must be an $\widehat{\mathbf{x}}$ so that $\mathbf{Proj}_{\text{Col } A}\,\mathbf{b} = A\widehat{\mathbf{x}}$.

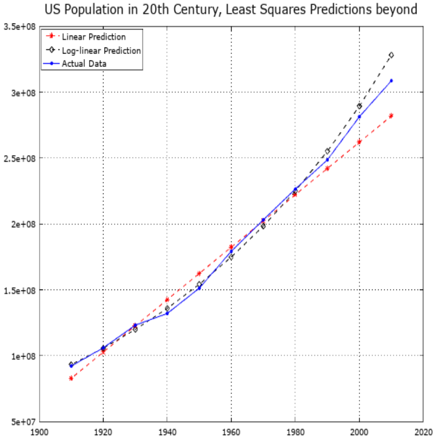
$$\Longrightarrow A^T\left(\mathbf{b} - A\widehat{\mathbf{x}}\right) = \mathbf{0}.$$

$$\boxed{\text{Least Squares solution satisfies } A^T A\mathbf{x} = A^T\mathbf{b}}$$

For any $\mathbf{x} \in \mathbb{R}^n$, $\quad \|\mathbf{b} - A\mathbf{x}\|^2 = \|\mathbf{b} - A\widehat{\mathbf{x}}\|^2 + \|A(\mathbf{x} - \widehat{\mathbf{x}})\|^2$.


US Population, 20th Century and beyond
3.5e+08
3e+08
2.5e+08
2e+08
1.5e+08
1e+08
5e+07
1900
1920
1940
1960
1980
2000
2020

Predicting years 2000, 2010 with data through 1990



- Log-linear Model predicts year 2000 better than Linear Model.
- Predictions for year 2010 are worse.

Thm: Let $A \in \mathbb{R}^{m \times n}$, and $\mathbf{b} \in \mathbb{R}^m$ with $m \geq n$.

$$\begin{array}{c} \text{Columns of } A \text{ are L.I.D.} \\ \Updownarrow \\ \text{The matrix } A^T A \text{ is invertible} \end{array} \qquad (\ell)$$

PROOF of (ℓ): Let any $\mathbf{x} \in \mathbb{R}^n$, then

$$\begin{aligned} A\mathbf{x} = \mathbf{0} \quad &\Longrightarrow \quad A^T A\mathbf{x} = \mathbf{0} \\ &\Longrightarrow \quad \mathbf{x}^T A^T A\mathbf{x} = 0 \quad \Longrightarrow \quad \|A\mathbf{x}\|^2 = 0 \\ &\Longrightarrow \quad A\mathbf{x} = \mathbf{0}. \end{aligned}$$

Thus $A^T\mathbf{x} = \mathbf{0} \iff A^T A\mathbf{x} = \mathbf{0}$, which implies (ℓ).

Thm: Let $A \in \mathbb{R}^{m \times n}$, and $\mathbf{b} \in \mathbb{R}^m$ with $m \geq n$.

$$\text{Columns of } A \text{ are L.I.D.}$$
$$\Updownarrow$$
$$\text{The matrix } A^T A \text{ is invertible}$$

If $A^T A$ is invertible, then the least squares solution is
$\widehat{\mathbf{x}} = \left(A^T A\right)^{-1} A^T \mathbf{b}$

EXAMPLE: Find least squares solution for

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

EXAMPLE: Find least squares solution for

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

SOLUTION: We first compute

$$\begin{aligned} A^T A &= \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}, \\ A^T \mathbf{b} &= \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 0 \\ 11 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix} \end{aligned}$$

so

$$\widehat{\mathbf{x}} = \left(A^T A\right)^{-1} A^T \mathbf{b} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

The QR solution to least squares

- Let $A \in \mathbb{R}^{m \times n}$ with L.I.D. columns, and $\mathbf{b} \in \mathbb{R}^m$ with $m \geq n$.
- Let $A = Q\,R$ be the QR factorization of A.

Thm: The QR solution to the least squares is $\widehat{\mathbf{x}} = R^{-1}\left(Q^T\,\widehat{\mathbf{b}}\right)$.

The QR solution to least squares

- Let $A \in \mathbb{R}^{m \times n}$ with L.I.D. columns, and $\mathbf{b} \in \mathbb{R}^m$ with $m \geq n$.
- Let $A = Q\,R$ be the QR factorization of A.

Thm: The QR solution to the least squares is $\widehat{\mathbf{x}} = R^{-1}\left(Q^T\,\widehat{\mathbf{b}}\right)$.

PROOF: Since A has L.I.D. columns, $A^T\,A$ must be invertible, so

$$\begin{aligned}
\widehat{\mathbf{x}} &= \left(A^T\,A\right)^{-1} A^T\,\mathbf{b} = \left((Q\,R)^T\,(Q\,R)\right)^{-1} (Q\,R)^T\,\mathbf{b} \\
&= \left(R^T\,R\right)^{-1} R^T\,\left(Q^T\,\mathbf{b}\right) = R^{-1}\left(Q^T\,\widehat{\mathbf{b}}\right).
\end{aligned}$$

EXAMPLE: Find least squares solution for $A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$.

EXAMPLE: Find least squares solution for $A = \begin{bmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix}$.

SOLUTION: We first compute QR factorization:

$$A = Q\,R = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}.$$

$$\text{So} \quad Q^T\,\mathbf{b} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 3 \\ 5 \\ 7 \\ -3 \end{bmatrix} = \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix}$$

$$\text{and} \quad \widehat{\mathbf{x}} = R^{-1}\left(Q^T\,\widehat{\mathbf{b}}\right) = \begin{bmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 6 \\ -6 \\ 4 \end{bmatrix} = \begin{bmatrix} 10 \\ -6 \\ 2 \end{bmatrix}.$$

What is your Inner Product?

§6.7 Inner Product Space

Let V be a vector space. **inner product** is a function

$$V \times V \longmapsto \mathcal{R} : \; < \mathbf{u}, \mathbf{v} > \; \in \mathcal{R} \quad \text{for any} \quad \mathbf{u}, \mathbf{v} \in V$$

that satisfies axioms below for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c \in \mathcal{R}$:

1. $< \mathbf{u}, \mathbf{v} > = < \mathbf{v}, \mathbf{u} >$. (<u>Symmetry with respect to **u** and **v**</u>)

§6.7 Inner Product Space

Let V be a vector space. **inner product** is a function

$$V \times V \longmapsto \mathcal{R} : \; < \mathbf{u}, \mathbf{v} > \; \in \mathcal{R} \quad \text{for any} \quad \mathbf{u}, \mathbf{v} \in V$$

that satisfies axioms below for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c \in \mathcal{R}$:

1. $< \mathbf{u}, \mathbf{v} > = < \mathbf{v}, \mathbf{u} >$. (Symmetry with respect to **u** and **v**)
2. $< \mathbf{u} + \mathbf{w}, \mathbf{v} > = < \mathbf{u}, \mathbf{v} > + < \mathbf{w}, \mathbf{v} >$.
3. $< c\,\mathbf{u}, \mathbf{v} > = c \; < \mathbf{u}, \mathbf{v} >$. (Linear transformation in **u**)

§6.7 Inner Product Space

Let V be a vector space. **inner product** is a function

$$V \times V \longmapsto \mathcal{R}: \ <\mathbf{u}, \mathbf{v}> \ \in \mathcal{R} \quad \text{for any} \quad \mathbf{u}, \mathbf{v} \in V$$

that satisfies axioms below for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c \in \mathcal{R}$:

1. $<\mathbf{u}, \mathbf{v}> = <\mathbf{v}, \mathbf{u}>$. (Symmetry with respect to $\mathbf{u}$ and $\mathbf{v}$)
2. $<\mathbf{u} + \mathbf{w}, \mathbf{v}> = <\mathbf{u}, \mathbf{v}> + <\mathbf{w}, \mathbf{v}>$.
3. $<c\,\mathbf{u}, \mathbf{v}> = c\ <\mathbf{u}, \mathbf{v}>$. (Linear transformation in $\mathbf{u}$)
4. $<\mathbf{u}, \mathbf{u}> \ \geq 0$, and $<\mathbf{u}, \mathbf{u}> \ = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

§6.7 Inner Product Space

Let V be a vector space. **inner product** is a function

$$V \times V \longmapsto \mathcal{R} : \ < \mathbf{u}, \mathbf{v} > \ \in \mathcal{R} \quad \text{for any} \quad \mathbf{u}, \mathbf{v} \in V$$

that satisfies axioms below for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c \in \mathcal{R}$:

1. $< \mathbf{u}, \mathbf{v} > = < \mathbf{v}, \mathbf{u} >$. (Symmetry with respect to $\mathbf{u}$ and $\mathbf{v}$)
2. $< \mathbf{u} + \mathbf{w}, \mathbf{v} > = < \mathbf{u}, \mathbf{v} > + < \mathbf{w}, \mathbf{v} >$.
3. $< c\,\mathbf{u}, \mathbf{v} > = c \ < \mathbf{u}, \mathbf{v} >$. (Linear transformation in $\mathbf{u}$)
4. $< \mathbf{u}, \mathbf{u} > \ \geq 0$, and $< \mathbf{u}, \mathbf{u} > \ = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

EXAMPLE: For any $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathcal{R}^3$, function

$< \mathbf{u}, \mathbf{v} > \stackrel{def}{=} 5\, u_1\, v_1 + 3\, u_2\, v_2 + u_3\, v_3$ is an inner product on $\mathcal{R}^3$.

§6.7 Inner Product Space

Let V be a vector space. **inner product** is a function

$$V \times V \longmapsto \mathcal{R}: \ < \mathbf{u}, \mathbf{v} > \ \in \mathcal{R} \quad \text{for any} \quad \mathbf{u}, \mathbf{v} \in V$$

that satisfies axioms below for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c \in \mathcal{R}$:

1. $< \mathbf{u}, \mathbf{v} > = < \mathbf{v}, \mathbf{u} >$. (Symmetry with respect to $\mathbf{u}$ and $\mathbf{v}$)
2. $< \mathbf{u} + \mathbf{w}, \mathbf{v} > = < \mathbf{u}, \mathbf{v} > + < \mathbf{w}, \mathbf{v} >$.
3. $< c\,\mathbf{u}, \mathbf{v} > = c \ < \mathbf{u}, \mathbf{v} >$. (Linear transformation in $\mathbf{u}$)
4. $< \mathbf{u}, \mathbf{u} > \ \geq 0$, and $< \mathbf{u}, \mathbf{u} > \ = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

EXAMPLE: For any $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}, \ \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathcal{R}^3$, function

$< \mathbf{u}, \mathbf{v} > \stackrel{def}{=} 5\,u_1\,v_1 - 3\,u_2\,v_2 + u_3\,v_3$ is <u>not</u> an inner product on $\mathcal{R}^3$.

§6.7 Inner Product Space

Let V be a vector space. **inner product** is a function

$$V \times V \longmapsto \mathcal{R} : \; < \mathbf{u}, \mathbf{v} > \; \in \mathcal{R} \quad \text{for any} \quad \mathbf{u}, \mathbf{v} \in V$$

that satisfies axioms below for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $c \in \mathcal{R}$:

1. $< \mathbf{u}, \mathbf{v} > = < \mathbf{v}, \mathbf{u} >$. (<u>Symmetry with respect to **u** and **v**</u>)
2. $< \mathbf{u} + \mathbf{w}, \mathbf{v} > = < \mathbf{u}, \mathbf{v} > + < \mathbf{w}, \mathbf{v} >$.
3. $< c\,\mathbf{u}, \mathbf{v} > = c \; < \mathbf{u}, \mathbf{v} >$. (<u>Linear transformation in **u**</u>)
4. $< \mathbf{u}, \mathbf{u} > \; \geq 0$, and $< \mathbf{u}, \mathbf{u} > \; = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

Example: For any $f(x), g(x) \in C[-1, 1]$, then

$$< f, g > \stackrel{def}{=} \int_{-1}^{1} \left(1 + x^2\right) \; f(x)\, g(x)\, dx \quad \text{is an inner product on } C[-1, 1].$$

- **length** of $\mathbf{u}$ (denoted $\|\mathbf{u}\|$) $\overset{def}{=} \sqrt{< \mathbf{u}, \mathbf{u} >}$

- **length** of $\mathbf{u}$ (denoted $\|\mathbf{u}\|$) $\stackrel{def}{=} \sqrt{<\mathbf{u}, \mathbf{u}>}$
- **distance** between $\mathbf{u}$ and $\mathbf{v}$ (denoted $\mathbf{dist}\,(\mathbf{u}, \mathbf{v})$) $\stackrel{def}{=} \|\mathbf{u} - \mathbf{v}\|$

- **length** of $\mathbf{u}$ (denoted $\|\mathbf{u}\|$) $\stackrel{def}{=} \sqrt{< \mathbf{u}, \mathbf{u} >}$
- **distance** between $\mathbf{u}$ and $\mathbf{v}$ (denoted $\mathbf{dist}\,(\mathbf{u}, \mathbf{v})$) $\stackrel{def}{=} \|\mathbf{u} - \mathbf{v}\|$
- $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $< \mathbf{u}, \mathbf{v} > = 0$

- **length** of $\mathbf{u}$ (denoted $\|\mathbf{u}\|$) $\stackrel{def}{=} \sqrt{<\mathbf{u},\,\mathbf{u}>}$
- **distance** between $\mathbf{u}$ and $\mathbf{v}$ (denoted $\mathbf{dist}\,(\mathbf{u},\,\mathbf{v})$) $\stackrel{def}{=} \|\mathbf{u}-\mathbf{v}\|$
- $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $<\mathbf{u},\,\mathbf{v}> \;= 0$

EX: Let $\mathbf{u} = \begin{bmatrix} 3 \\ -6 \\ 3 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \in \mathcal{R}^3$, with

$$\text{inner product} \quad <\mathbf{u},\mathbf{v}> \stackrel{def}{=} 5\,u_1\,v_1 + 3\,u_2\,v_2 + u_3\,v_3.$$

- **length** $\|\mathbf{u}\| = \sqrt{5\cdot 3^2 + 3\cdot(-6)^2 + 3^2} = \sqrt{162}$
- **distance** between $\mathbf{u}$ and $\mathbf{v}$:

$$\|\mathbf{u}-\mathbf{v}\| = \sqrt{5\cdot(3-1)^2 + 3\cdot(-6-1)^2 + (3-1)^2} = \sqrt{171}.$$

- $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal**:

$$<\mathbf{u},\mathbf{v}> \;= 5\cdot 3\cdot 1 + 3\cdot(-6)\cdot 1 + 3\cdot 1 = 0.$$

Best Approximation

Let W be a subspace of inner product space V, with orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

Best Approximation

Let W be a subspace of inner product space V, with orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

- DEFINITION: For any vector $\mathbf{y} \in V$, define **orthogonal projection** as

$$\mathbf{proj}_W \mathbf{y} = \widehat{\mathbf{y}} = \frac{< \mathbf{y}, \mathbf{u}_1 >}{< \mathbf{u}_1, \mathbf{u}_1 >}\mathbf{u}_1 + \cdots + \frac{< \mathbf{y}, \mathbf{u}_p >}{< \mathbf{u}_p, \mathbf{u}_p >}\mathbf{u}_p$$

Best Approximation

Let W be a subspace of inner product space V, with orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

- DEFINITION: For any vector $\mathbf{y} \in V$, define **orthogonal projection** as
$$\mathbf{proj}_W \mathbf{y} = \widehat{\mathbf{y}} = \frac{< \mathbf{y}, \mathbf{u}_1 >}{< \mathbf{u}_1, \mathbf{u}_1 >}\mathbf{u}_1 + \cdots + \frac{< \mathbf{y}, \mathbf{u}_p >}{< \mathbf{u}_p, \mathbf{u}_p >}\mathbf{u}_p$$

- **Pythagorean Thm**: For all $\mathbf{v} \in W$,
$$\|\mathbf{y} - \mathbf{v}\|^2 = \|\mathbf{y} - \widehat{\mathbf{y}}\|^2 + \|\widehat{\mathbf{y}} - \mathbf{v}\|^2 .$$

Best Approximation

Let W be a subspace of inner product space V, with orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_p\}$.

- DEFINITION: For any vector $\mathbf{y} \in V$, define **orthogonal projection** as

$$\mathbf{proj}_W \mathbf{y} = \widehat{\mathbf{y}} = \frac{< \mathbf{y}, \mathbf{u}_1 >}{< \mathbf{u}_1, \mathbf{u}_1 >} \mathbf{u}_1 + \cdots + \frac{< \mathbf{y}, \mathbf{u}_p >}{< \mathbf{u}_p, \mathbf{u}_p >} \mathbf{u}_p$$

- **Pythagorean Thm**: For all $\mathbf{v} \in W$,

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

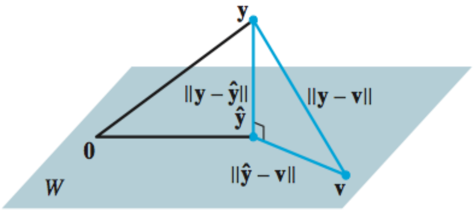


FIGURE 4 The orthogonal projection of y

Cauchy-Schwarz Inequality

Thm: Let $\mathbf{u}, \mathbf{v}$ be non-zero vectors from inner product space V,

$$\text{then,} \quad |<\mathbf{v}, \mathbf{u}>| \leq \|\mathbf{u}\| \, \|\mathbf{v}\| \, .$$

Proof: By Pythagorean Thm:

$$\begin{aligned} \|\mathbf{v}\|^2 &= \left\|\mathbf{v} - \mathbf{proj}_{\mathbf{Span}\{\mathbf{u}\}}\mathbf{v}\right\|^2 + \left\|\mathbf{proj}_{\mathbf{Span}\{\mathbf{u}\}}\mathbf{v}\right\|^2 \\ &\geq \left\|\mathbf{proj}_{\mathbf{Span}\{\mathbf{u}\}}\mathbf{v}\right\|^2 \\ &= \left\|\frac{<\mathbf{v}, \mathbf{u}>}{<\mathbf{u}, \mathbf{u}>}\mathbf{u}\right\|^2 = \frac{|<\mathbf{v}, \mathbf{u}>|^2}{\|\mathbf{u}\|^2} \quad \square \end{aligned}$$

The Triangle Inequality

Thm: Let $\mathbf{u}, \mathbf{v}$ be vectors from inner product space V,

$$\text{then,} \quad \|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|\,.$$

PROOF: By Cauchy-Schwarz Inequality:

$$\begin{aligned}
\|\mathbf{u}+\mathbf{v}\|^2 &= \ < \mathbf{u}+\mathbf{v},\ \mathbf{u}+\mathbf{v} > \\
&= \ < \mathbf{u},\ \mathbf{u} > + < \mathbf{v},\ \mathbf{v} > +2 < \mathbf{u},\ \mathbf{v} > \\
&\leq \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 + 2\ \|\mathbf{u}\|\ \|\mathbf{v}\| \\
&= (\|\ \|\mathbf{u}\| + \|\mathbf{v}\|)^2\,. \quad \square
\end{aligned}$$

Gram-Schmidt Process: **Example**

EXAMPLE: Let $W = \textbf{Span}\left\{1, x, x^2\right\}$ be a subspace of $C[-1, 1]$. Find an orthogonal basis for W using the inner product

$$< f, g > \overset{def}{=} \int_{-1}^{1} f(x)\, g(x)\, dx.$$

SOLUTION: Let $\mathbf{p}_1(x) = 1, \mathbf{p}_2(x) = x^2, \mathbf{p}_3(x) = x^2$.
Choose $\mathbf{q}_1(x) = 1$,

$$\begin{aligned}
\mathbf{q}_2(x) &= \mathbf{p}_2(x) - \textbf{proj}_{\textbf{Span}\{\mathbf{q}_1\}}\mathbf{p}_2 = x - \frac{\int_{-1}^{1} 1 \cdot x\, dx}{\int_{-1}^{1} 1 \cdot 1\, dx} 1 = x, \\
\mathbf{q}_3(x) &= \mathbf{p}_3(x) - \textbf{proj}_{\textbf{Span}\{\mathbf{q}_1, \mathbf{q}_2\}}\mathbf{p}_3 \\
&= x^2 - \frac{\int_{-1}^{1} 1 \cdot x^2\, dx}{\int_{-1}^{1} 1 \cdot 1\, dx} 1 - \frac{\int_{-1}^{1} x \cdot x^2\, dx}{\int_{-1}^{1} x \cdot x\, dx} x = x^2 - \frac{1}{3}. \quad \square
\end{aligned}$$