Chapter 5 Orthogonality

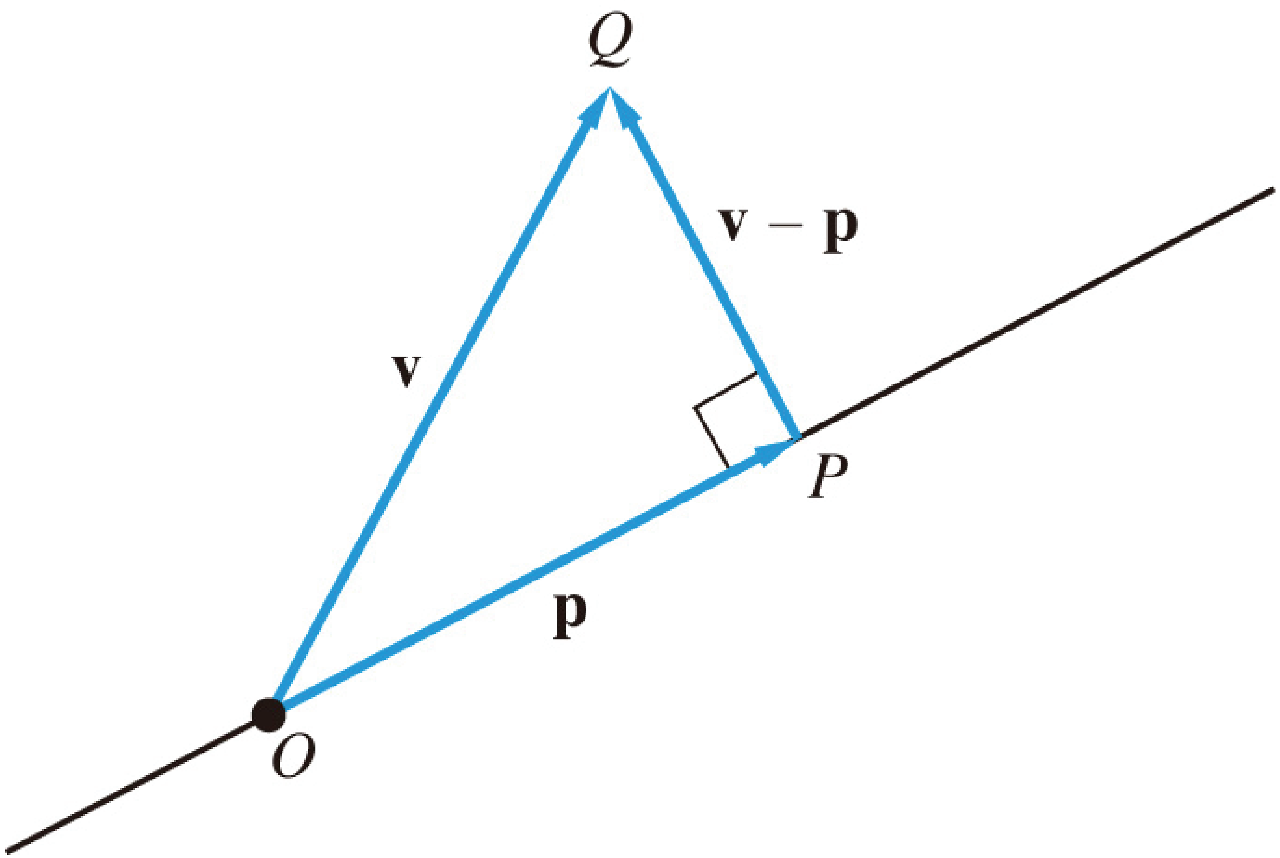
Outlines

- The Scalar Product in R^n
- Orthogonal Subspaces (skip)
- Least Squares problem (skip)
- Inner product Spaces
- Orthonormal Sets
- The Gram-Schmidt Orthogonalization Process (*)
- Orthogonal Polynomials (skip)

Introduction

- The **scalar product** between two vectors $\mathbf{x}$ and $\mathbf{y}$ can be defined as $\mathbf{x}^T\mathbf{y}$.
- A vector in can $\boldsymbol{R}^2$ be thought of as a directed line segment initiating at the origin.
- In $\boldsymbol{R}^2$, the angle between two line segments will be a right angle if and only if the scalar product of the corresponding vectors is 0.
- In general, if V is a vector space with a scalar product, then two vectors in V are said to be **orthogonal** if their scalar product is 0.

- Orthogonality can be thought of as a generalization of **perpendicularity** to any vector space with an inner product.
- Consider the following problem: find the closest point P on l that is closest to Q which is not a point on l
- The solution is to find $\mathbf{p}$ that is orthogonal to $\mathbf{v} - \mathbf{p}$



5.1 The Scalar Product in R^n

- The **scalar product** of two $n\times 1$ matrices $\mathbf{x}$ and $\mathbf{y}$ is the 1×1 matrix $\mathbf{x}^T\mathbf{y}$, or simply regarded as a real number. That is, if $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T$, then

$$\mathbf{x}^T\mathbf{y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = x_1 y_1 + x_2 y_2 + \ldots + x_n y_n$$

Example 1

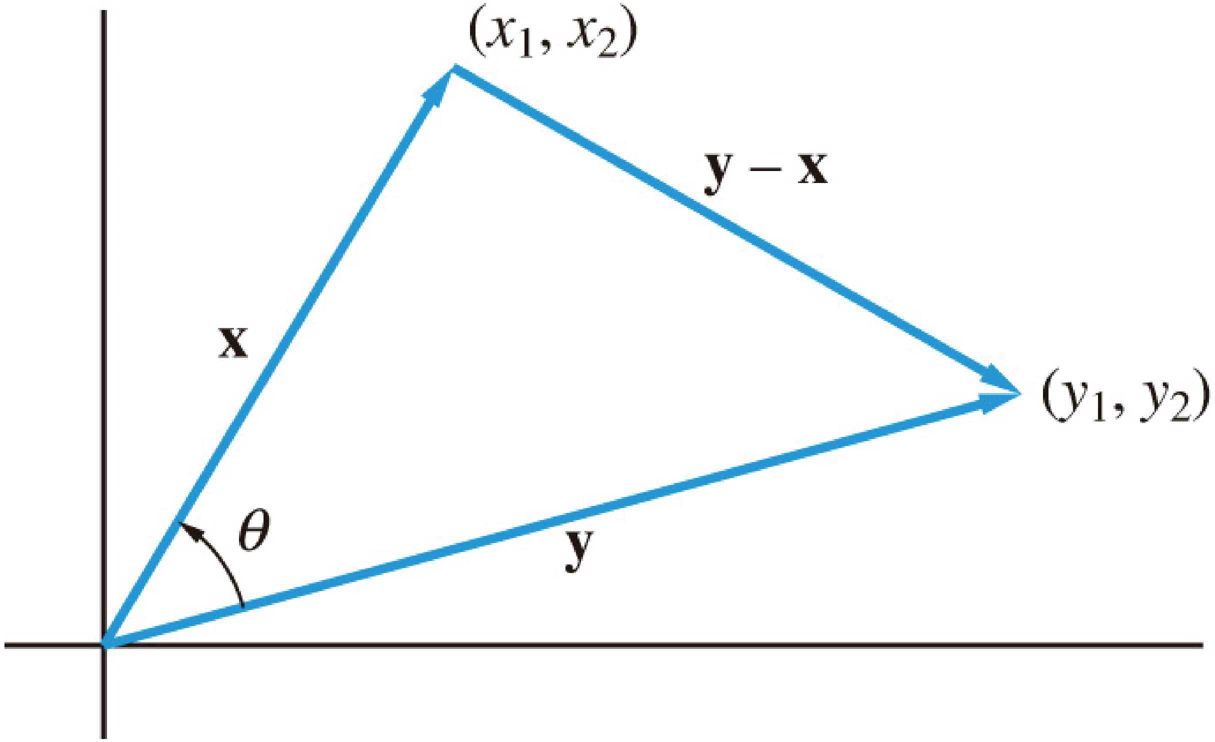
If $\mathbf{x} = \begin{bmatrix} 3 \\ -2 \\ 1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 4 \\ 3 \\ 2 \end{bmatrix}$, then $\mathbf{x}^T\mathbf{y} = 3\times 4 + (\text{-}2)\times 3 + 1\times 2 = 8$

The Scalar Product in R^2 and R^3

- A vector in $\boldsymbol{R}^2$ and $\boldsymbol{R}^3$ can be represented by directed line segment
- The **Euclidean length** of a vector **x** in either $\boldsymbol{R}^2$ or $\boldsymbol{R}^3$ can be defined in terms of the scalar product:

$$\|\mathbf{x}\| = (\mathbf{x}^T\mathbf{x})^{1/2} = \begin{cases} \sqrt{x_1^2 + x_2^2} & \text{if } \mathbf{x} \in R^2 \\ \sqrt{x_1^2 + x_2^2 + x_3^2} & \text{if } \mathbf{x} \in R^3 \end{cases}$$

- The angle between two vectors is defined as the angle θ between the line segments.
- The distance between the vectors is measured by the length of the vector joining the terminal point of **x** and the terminal point of **y**



Definition

Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in either $\boldsymbol{R}^2$ or $\boldsymbol{R}^3$. The **distance** between $\mathbf{x}$ and $\mathbf{y}$ is defined to be the number $\|\mathbf{x} - \mathbf{y}\|$.

- If $\mathbf{x} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} -1 \\ 7 \end{bmatrix}$, then the distance between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$\mathbf{y} - \mathbf{x} = \begin{bmatrix} -1 \\ 7 \end{bmatrix} - \begin{bmatrix} 3 \\ 4 \end{bmatrix} = \begin{bmatrix} -4 \\ 3 \end{bmatrix}$$

$$\|\mathbf{y} - \mathbf{x}\| = \sqrt{(-1-3)^2 + (7-4)^2} = 5$$

Theorem 5.1.1

If **x** and **y** are two nonzero vectors in either $\boldsymbol{R}^2$ or $\boldsymbol{R}^3$ and θ is the angle between them, then

$$\mathbf{x}^T\mathbf{y} = \|\mathbf{x}\|\,\|\mathbf{y}\| \cos\theta$$

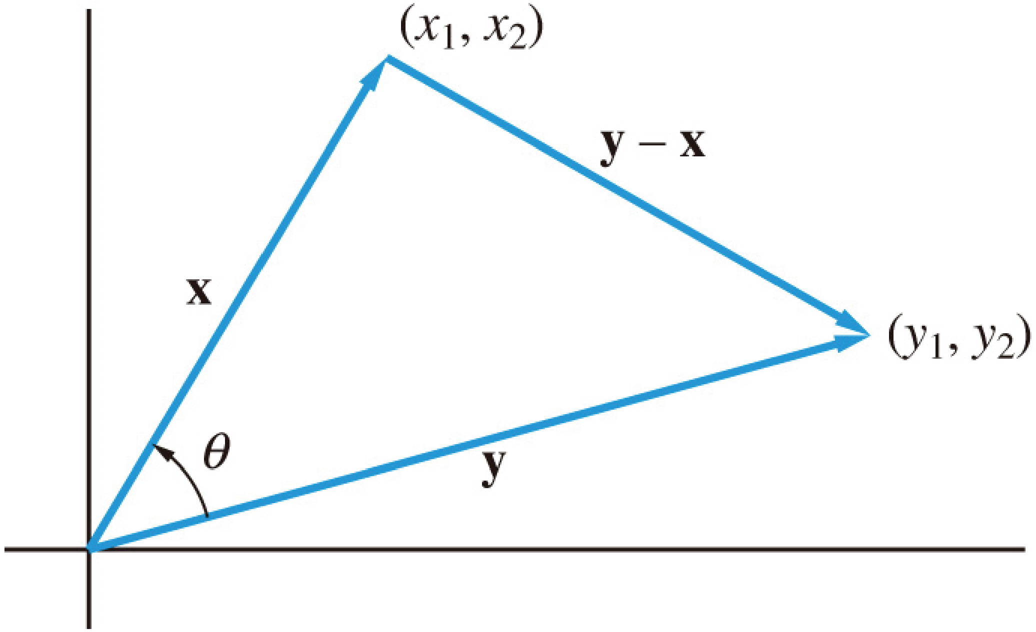


Figure 5.1.1

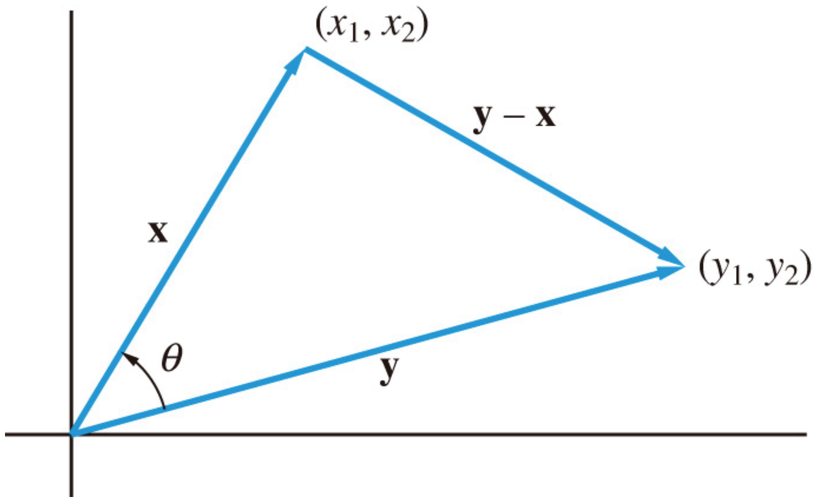
Theorem 5.1.1 proof

- By the law of cosines (餘弦定理),

$$\|\mathbf{y} - \mathbf{x}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\,\|\mathbf{x}\|\|\mathbf{y}\| \cos\theta$$

or $\|\mathbf{x}\|\|\mathbf{y}\| \cos\theta$

$$= \tfrac{1}{2}(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - \|\mathbf{y} - \mathbf{x}\|^2)$$
$$= \tfrac{1}{2}(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - (\mathbf{y} - \mathbf{x})^T(\mathbf{y} - \mathbf{x}))$$
$$= \tfrac{1}{2}(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - (\mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{x} - \mathbf{x}^T\mathbf{y} + \mathbf{x}^T\mathbf{x}))$$
$$= \tfrac{1}{2}(2\,\mathbf{x}^T\mathbf{y})$$
$$= \mathbf{x}^T\mathbf{y}$$



Unit Vector

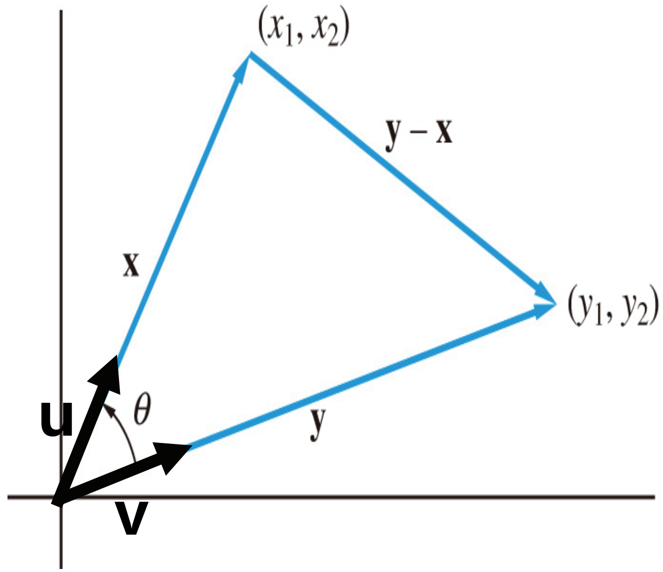
- If $\mathbf{x}$ is a nonzero vector, then we can form the **unit vector** $\mathbf{u}$ of $\mathbf{x}$ as $\mathbf{u} = \frac{1}{\|\mathbf{x}\|}\mathbf{x}$

- If $\mathbf{x}$ and $\mathbf{y}$ are two nonzero vectors, $\mathbf{u}$ and $\mathbf{v}$ are the unit vectors of $\mathbf{x}$ and $\mathbf{y}$:

$$\mathbf{u} = \frac{1}{\|\mathbf{x}\|}\mathbf{x} \quad \text{and } \mathbf{v} = \frac{1}{\|\mathbf{y}\|}\mathbf{y}$$

then the angle θ between $\mathbf{x}$ and $\mathbf{y}$ is

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \mathbf{u}^T\mathbf{v}$$



Example 3

- If $\mathbf{x} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} -1 \\ 7 \end{bmatrix}$, then the directions of these two vectors are given by the unit vectors:

$$\mathbf{u} = \frac{1}{\|\mathbf{x}\|}\mathbf{x} = \begin{bmatrix} \frac{3}{5} \\ \frac{4}{5} \end{bmatrix} \text{ and } \mathbf{v} = \frac{1}{\|\mathbf{y}\|}\mathbf{y} = \begin{bmatrix} \frac{-1}{5\sqrt{2}} \\ \frac{7}{5\sqrt{2}} \end{bmatrix}$$

$$\cos\theta = \mathbf{u}^T\mathbf{v} = \frac{3}{5}\times\frac{-1}{5\sqrt{2}} + \frac{4}{5}\times\frac{7}{5\sqrt{2}} = \frac{25}{25\sqrt{2}} = \frac{1}{\sqrt{2}} \Rightarrow \theta = \pi/4$$

Corollary 5.1.2

(Cauchy-Schwarz Inequality)

If $\mathbf{x}$ and $\mathbf{y}$ are vectors in either $\boldsymbol{R}^2$ or $\boldsymbol{R}^3$, then

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|\ \|\mathbf{y}\|$$

with equality holding if and only if one of the vectors is 0 or one vector is a multiple of the other.

- Proof: hint: $\mathbf{x}^T\mathbf{y} = \|\mathbf{x}\|\ \|\mathbf{y}\| \cos\theta$

$$-1 \leq \cos\theta \leq 1$$

Definition

The vectors $\mathbf{x}$ and $\mathbf{y}$ are in $\boldsymbol{R}^2$ (or $\boldsymbol{R}^3$) are said to be **orthogonal** if $\mathbf{x}^T\mathbf{y} = 0$.

Example 4

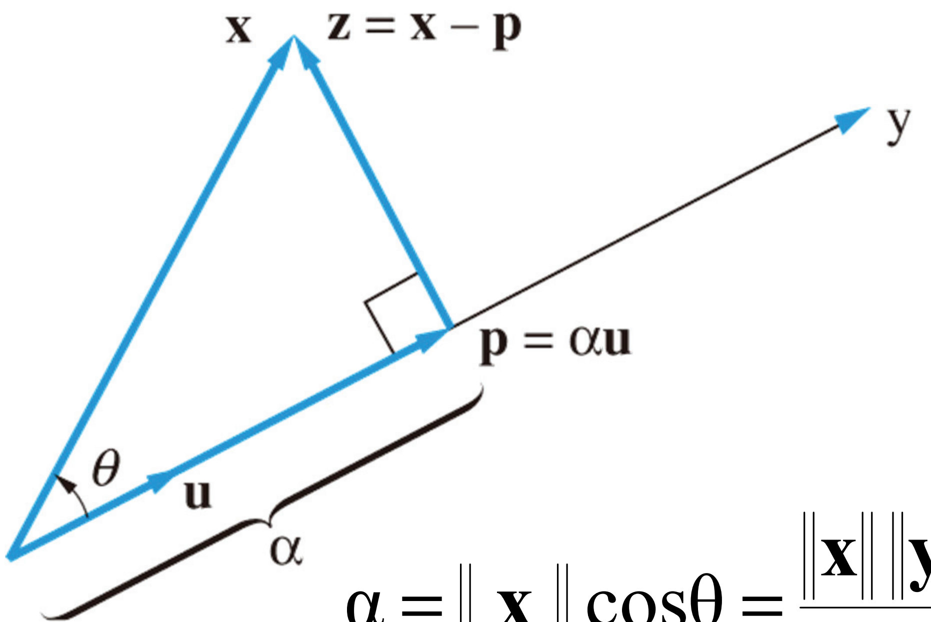
(a) The vector $\mathbf{0}$ is orthogonal to every vector in $\boldsymbol{R}^2$.

(b) The vectors $(3, 2)^T$ and $(-4, 6)^T$ are orthogonal in $\boldsymbol{R}^2$.

(c) The vectors $(2, -3, 1)^T$ and $(1, 1, 1)^T$ are orthogonal in $\boldsymbol{R}^3$.

Scalar and Vector Projections

- Let **x** and **y** be in either $\boldsymbol{R}^2$ or $\boldsymbol{R}^3$, then **x** can be represented as **p+z**, where **p** is in the direction of **y** and **z** is orthogonal to **p**.



$$\alpha = \| \mathbf{x} \| \cos\theta = \frac{\|\mathbf{x}\| \|\mathbf{y}\| \cos\theta}{\|\mathbf{y}\|} = \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{y}\|},$$

Scalar and Vector Projections

- Let $\mathbf{u} = (1/\|\mathbf{y}\|)\mathbf{y}$, thus $\mathbf{u}$ is a unit vector (length 1) in the direction of $\mathbf{y}$. We wish to find α such that $\mathbf{p} = \alpha\mathbf{u}$ and is orthogonal to $\mathbf{z} = \mathbf{x} - \alpha\mathbf{u}$. Thus

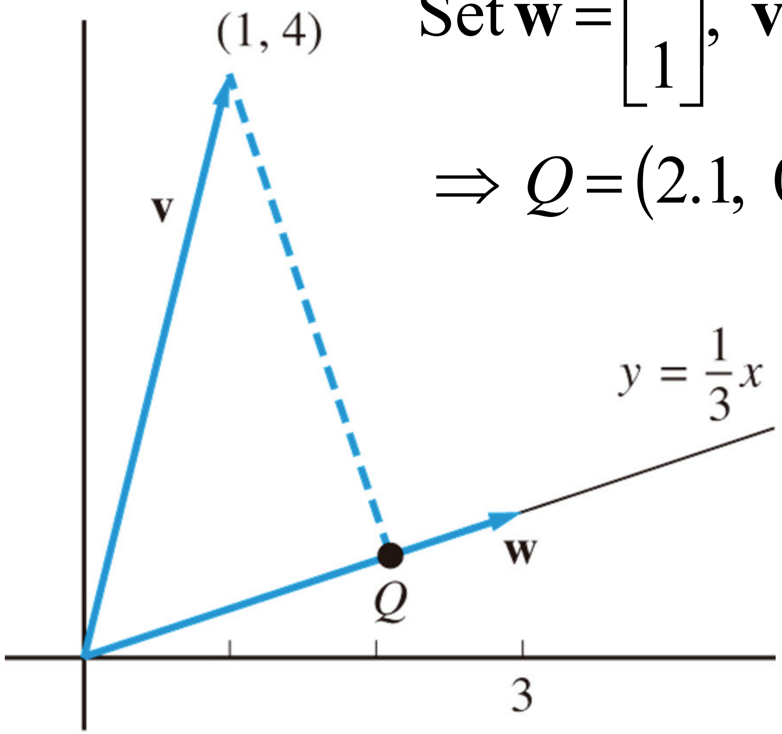
$$\alpha = \|\mathbf{x}\| \cos\theta = \frac{\|\mathbf{x}\|\,\|\mathbf{y}\|\cos\theta}{\|\mathbf{y}\|} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|},$$

α is called the **scalar projection** of $\mathbf{x}$ onto $\mathbf{y}$ and $\mathbf{p}$ is called the **vector projection** of $\mathbf{x}$ onto $\mathbf{y}$:

$$\mathbf{p} = \alpha\mathbf{u} = \alpha\frac{\mathbf{y}}{\|\mathbf{y}\|} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|}\frac{\mathbf{y}}{\|\mathbf{y}\|} = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{y}\|^2}\mathbf{y} = \frac{\mathbf{x}^T\mathbf{y}}{\mathbf{y}^T\mathbf{y}}\mathbf{y}$$

Example 5

- Find the point Q on the line $y = 1/3\ x$ that is closest to the point $(1, 4)$.



$$\text{Set } \mathbf{w} = \begin{bmatrix} 3 \\ 1 \end{bmatrix},\ \mathbf{v} = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \text{ then } Q = \frac{\mathbf{v}^T \mathbf{w}}{\|\mathbf{w}\|^2}\mathbf{w} = \frac{7}{10}\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2.1 \\ 0.7 \end{bmatrix}$$

$$\Rightarrow Q = (2.1,\ 0.7) \text{ is the closest point}$$

Notation

If P_1 and P_2 are two points in 3-space, we will denote the vector from P_1 to P_2 by $\overrightarrow{P_1P_2}$

- If **N** is a nonzero vector and P_0 is a fixed point, the set of points ***P*** such that $\overrightarrow{P_1P_2}$ is orthogonal to **N** forms a **plane π** in 3-space that passes through P_0. The vector **N** and the plane π are said to be **normal** to each other. A point $P = (x, y, z)$ will lie on π if and only if

$$(\overrightarrow{P_0P})^T \mathbf{N} = 0$$

If $\mathbf{N} = (a, b, c)^T$ and $P_0 = (x_0, y_0, z_0)$, the above equation can be written as

$$\boldsymbol{a(x - x_0) + b(y - y_0) + c(z - z_0) = 0}$$

Example 6

- Find the equation of the plane passing through the point (2, -1, 3) and normal to the vector $\mathbf{N} = (2, 3, 4)^T$.
- *Sol:*

$$\overrightarrow{P_0P} = (x-2, y+1, z-3)^T$$

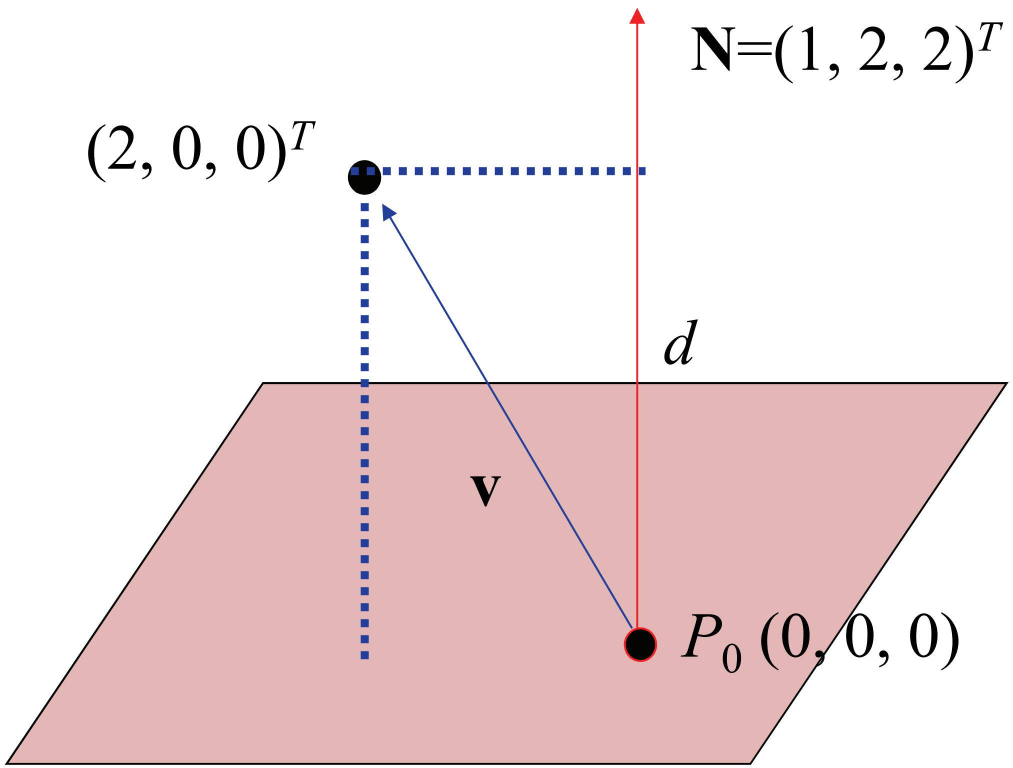
The equation is $(\overrightarrow{P_0P})^T \mathbf{N} = 0$

So, $2(x-2)+3(y+1)+4(z-3)=0$

Example 8

- Find the distance from the point (2, 0 ,0) to the plane $x + 2y + 2z = 0$.
- *Sol:*
- The plane equation: $x + 2y + 2z = 0$
  $\Rightarrow 1 \cdot (x - 0) + 2 \cdot (y - 0) + 2 \cdot (z - 0) = 0$
  So, $\mathbf{N} = (1, 2, 2)^T$ and $P_0 = (0, 0, 0)$
  Let $= \mathbf{v} = (2, 0, 0)^T$

$$\Rightarrow d = \frac{\mathbf{v}^T \mathbf{N}}{\|\mathbf{N}\|} = \frac{1}{\sqrt{1^2 + 2^2 + 2^2}} (2, 0, 0) \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} = \frac{2}{3}$$



$\mathbf{N}=(1, 2, 2)^T$
$(2, 0, 0)^T$
d
$\mathbf{v}$
$P_0\ (0, 0, 0)$

$\|\mathbf{x}\|\|\mathbf{y}\|\sin\theta = \|\mathbf{x}\times\mathbf{y}\|$

If $\mathbf{x}$ and $\mathbf{y}$ are nonzero vectors in $\mathbb{R}^3$ and θ is the angle between the vectors, then

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}$$

It then follows that

$$\sin\theta = \sqrt{1-\cos^2\theta} = \sqrt{1-\frac{(\mathbf{x}^T\mathbf{y})^2}{\|\mathbf{x}\|^2\|\mathbf{y}\|^2}} = \frac{\sqrt{\|\mathbf{x}\|^2\|\mathbf{y}\|^2-(\mathbf{x}^T\mathbf{y})^2}}{\|\mathbf{x}\|\,\|\mathbf{y}\|}$$

and hence

$$\begin{aligned}
\|\mathbf{x}\|\,\|\mathbf{y}\|\sin\theta &= \sqrt{\|\mathbf{x}\|^2\|\mathbf{y}\|^2-(\mathbf{x}^T\mathbf{y})^2} \\
&= \sqrt{(x_1^2+x_2^2+x_3^2)(y_1^2+y_2^2+y_3^2)-(x_1y_1+x_2y_2+x_3y_3)^2} \\
&= \sqrt{(x_2y_3-x_3y_2)^2+(x_3y_1-x_1y_3)^2+(x_1y_2-x_2y_1)^2} \\
&= \|\mathbf{x}\times\mathbf{y}\|
\end{aligned}$$

Thus, we have, for any nonzero vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^3$,

Orthogonality in $\boldsymbol{R^n}$

- If $\mathbf{x} \in \boldsymbol{R^n}$ then the **Euclidean length** of $\mathbf{x}$ is defined by

$$\| \mathbf{x} \| = (\mathbf{x}^T \mathbf{x})^{1/2} = (x_1^2 + x_2^2 + \cdots + x_n^2)^{1/2}$$

- If $\mathbf{x}$ and $\mathbf{y}$ are two vectors in $\boldsymbol{R^n}$, then the distance between $\mathbf{x}$ and $\mathbf{y}$ is $\|\mathbf{y} - \mathbf{x}\|$
- The Cauchy-Schwarz inequality holds in $\boldsymbol{R^n}$:

$$-1 \leq \frac{\mathbf{x}^T \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} \leq 1$$

- The angle θ between two vectors **x** and **y** in $\boldsymbol{R^n}$ is given by

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|}, 0 \le \theta \le \pi$$

- If **u** and **v** are the unit vectors of **x** and **y**:

$$\mathbf{u} = \frac{1}{\|\mathbf{x}\|}\mathbf{x} \quad \text{and } \mathbf{v} = \frac{1}{\|\mathbf{y}\|}\mathbf{y}$$

then the angle θ between **u** and **v** is the same as the angle between **x** and **y**:

$$\cos\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|\|\mathbf{y}\|} = \mathbf{u}^T\mathbf{v}$$

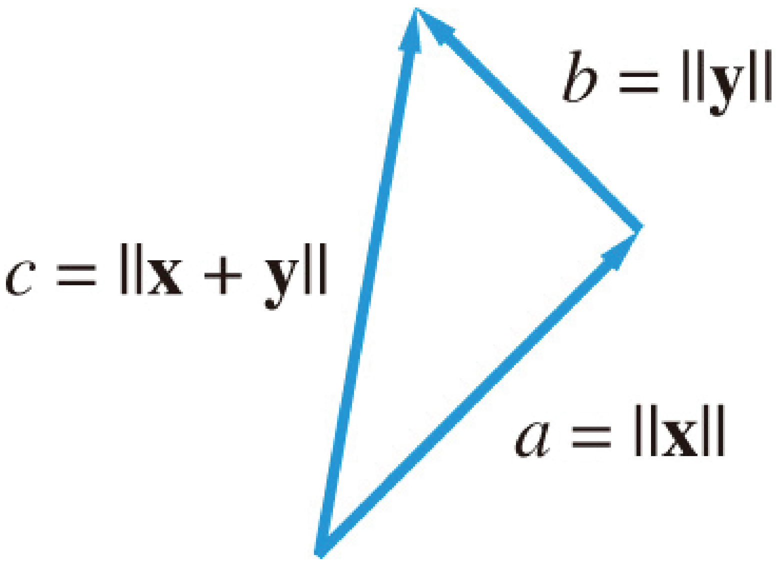
- *The cosine can be computed by simply taking the <u>scalar product</u> of the two <u>unit vectors</u>.*
- Two vectors **x** and **y** in $\boldsymbol{R}^n$ are said to be **orthogonal** if $\mathbf{x}^T\mathbf{y} = 0$ and often the symbol "$\perp$" is used to indicate orthogonality.
- If **x** and **y** are orthogonal, we will write $\mathbf{x} \perp \mathbf{y}$

- If $\mathbf{x}$ and $\mathbf{y}$ are vectors in $\boldsymbol{R}^n$, then

$$
\begin{aligned}
&\|\mathbf{x}+\mathbf{y}\|^2 \\
&= (\mathbf{x}+\mathbf{y})^T\,(\mathbf{x}+\mathbf{y}) \\
&= (\mathbf{x}^T+\mathbf{y}^T)\,(\mathbf{x}+\mathbf{y}) \\
&= \mathbf{x}^T\mathbf{x} + \mathbf{x}^T\mathbf{y} + \mathbf{y}^T\mathbf{x} + \mathbf{y}^T\mathbf{y} \\
&= \|\mathbf{x}\|^2 + 2\mathbf{x}^T\mathbf{y} + \|\mathbf{y}\|^2
\end{aligned}
$$

- If **x** and **y** are orthogonal, the above equation becomes the **Pythagorean Law**:

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$$



5.2 Orthogonal Subspaces

- Let A be an $m\times n$ matrix and let $\mathbf{x} \in N(A)$, the null space of A.
- $A\mathbf{x} = \mathbf{0}$, i.e.,

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n = 0, \text{ for } i = 1, \cdots, m$$

- $\mathbf{x}$ is orthogonal to the ith column vector of A^T for $i = 1, 2, \ldots, m$
- $\mathbf{x}$ is orthogonal to any linear combination of the column vector of A^T
- If $\mathbf{y}$ is any vector in the column space of A^T, then $\mathbf{x}^T\mathbf{y} = \mathbf{0}$
- Each vector in N(A) is orthogonal to every vector in the column space of A^T

Definition

Two subspaces X and Y of $\boldsymbol{R}^n$ are said to be **orthogonal** if $\mathbf{x}^T\mathbf{y} = 0$ for <u>every</u> $\mathbf{x} \in X$ and <u>every</u> $\mathbf{y} \in Y$. If X and Y are orthogonal, we write $X \perp Y$.

Example 1

- Let X be the subspace of $\boldsymbol{R}^3$ spanned by $\mathbf{e}_1$ and Y be the subspace of $\boldsymbol{R}^3$ spanned by $\mathbf{e}_2$, if $\mathbf{x} \in X$ and $\mathbf{y} \in Y$, then

$$\mathbf{x} = \alpha \mathbf{e}_1 = \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha \\ 0 \\ 0 \end{bmatrix} \text{ and } \mathbf{y} = \beta \mathbf{e}_2 = \beta \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ \beta \\ 0 \end{bmatrix}$$

$$\Rightarrow \mathbf{x}^T \mathbf{y} = \begin{bmatrix} \alpha & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ \beta \\ 0 \end{bmatrix} = 0 \Rightarrow X \perp Y$$

- The concept of orthogonal subspaces does not always agree with our intuitive idea of perpendicularity.
  - For example, the floor and wall of the classroom "look" orthogonal, but the *xy*-plane and the *yz*-plane are not orthogonal subspaces
  - Think of the vectors $\mathbf{x}_1 = (1, 1, 0)^T$ and $\mathbf{x}_2 = (0, 1, 1)^T$ lying in the *xy*-plane and the *yz*-plane, respectively, then

$$\mathbf{x}_1^T\mathbf{x}_2 = \begin{bmatrix} 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} = 1$$

  - These two subspaces are not orthogonal!

Example 2

- Let X be the subspace of $\boldsymbol{R}^3$ spanned by $\mathbf{e}_1$ and $\mathbf{e}_2$, and let Y be the subspace spanned by $\mathbf{e}_3$, if $\mathbf{x} \in X$ and $\mathbf{y} \in Y$, then

$$\mathbf{x} = \alpha\mathbf{e_1} + \beta\mathbf{e_2} = \alpha\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \alpha \\ \beta \\ 0 \end{bmatrix} \text{ and } \mathbf{y} = \gamma\mathbf{e_3} = \gamma\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \gamma \end{bmatrix}$$

$$\Rightarrow \mathbf{x}^T\mathbf{y} = \begin{bmatrix} \alpha & \beta & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ \gamma \end{bmatrix} = 0 \Rightarrow X \perp Y$$

- If $\mathbf{z} = (z_1, z_2, z_3)^T$ is any vector in $\boldsymbol{R}^3$ that is orthogonal to every vector in Y, then $\mathbf{z} \perp \mathbf{e_3}$, and hence

$$\mathbf{z}^T\mathbf{e}_3 = \begin{bmatrix} z_1 & z_2 & z_3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = z_3 = 0$$

- If $z_3 = 0$, then $\mathbf{z} \in X$
- X is the set of all vectors in $\boldsymbol{R}^3$ that is orthogonal to every vectors in Y

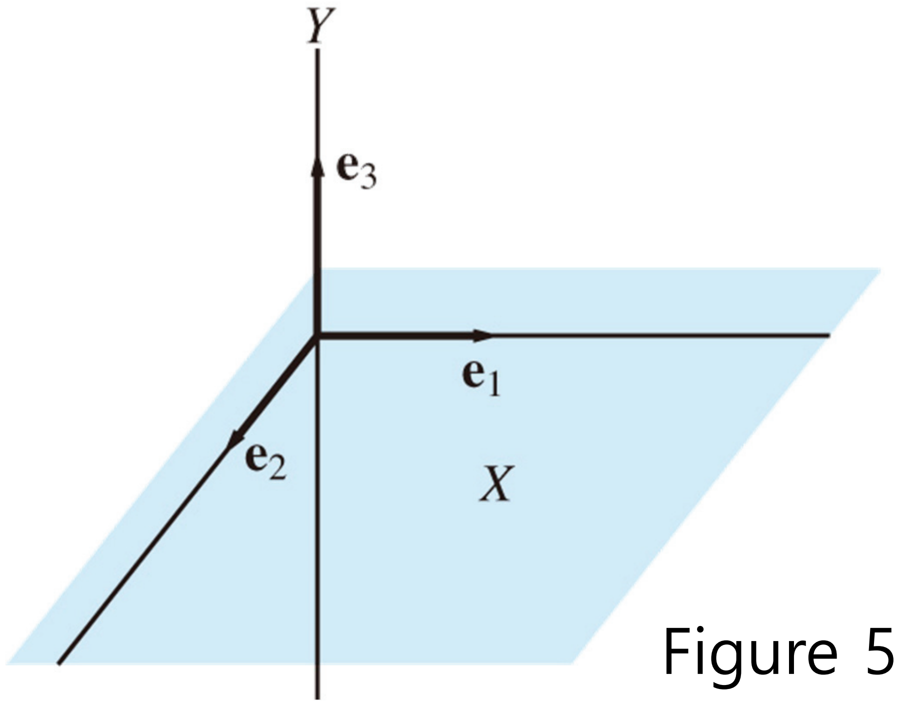


Figure 5.2.1

5.4 Inner Product Spaces
Definition

An **inner product** on a vector space V is an operation on V that assigns to each pair of vectors $\mathbf{x}$ and $\mathbf{y}$ in V a real number $\langle \mathbf{x}, \mathbf{y} \rangle$ satisfying the following conditions:

I. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ with equality if and only if $\mathbf{x} = \mathbf{0}$

II. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$ for all $\mathbf{x}$ and $\mathbf{y}$ in V

III. $\langle \alpha\mathbf{x}+\beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle$ for all $\mathbf{x}$, $\mathbf{y}$, $\mathbf{z}$ in V and all scalars α and β

- A vector space V with an inner product is called an **inner product space**.

Case 1: The vector space R^n

- The standard inner product for R^n is the scalar product

$$<\mathbf{x}, \mathbf{y}> = \mathbf{x}^T\mathbf{y}$$

Given a vector **w** with positive entries

$$< \mathbf{x}, \mathbf{y} >= \sum_{i=1}^{n} x_i y_i w_i \qquad (1)$$

where w_i are referred to as ***weights***

Case 2: The vector space $R^{m \times n}$

- Given A and B in $R^{m \times n}$

$$< A, B > = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij} \qquad (2)$$

Case 3: The vector space *C*[*a*, *b*]

- In $C[a, b]$ we may define an inner product by

$$< f , g >= \int_a^b f(x)g(x)dx \qquad (3)$$

- If $w(x)$ is a positive continuous function on $C[a, b]$, then

$$< f , g >= \int_a^b f(x)g(x)w(x)dx \qquad (4)$$

- The function $w(x)$ is called a **weight function**

Basic Properties of Inner Product Spaces

- If **v** is a vector in an inner product space V, the **length** or **norm** of **v** is given by

$$\|\mathbf{v}\| = \sqrt{<\mathbf{v}, \mathbf{v}>}$$

- Two vectors **u** and **v** are said to be **orthogonal** if $<\mathbf{u}, \mathbf{v}> = 0$

Theorem 5.4.1

(The Pythagorean Law)

If $\mathbf{u}$ and $\mathbf{v}$ are orthogonal vectors in an inner product space V, then

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

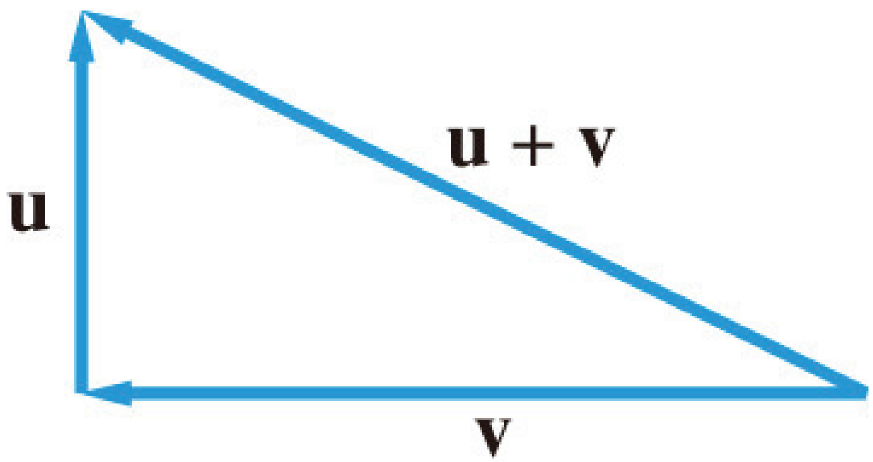


Figure 5.4.1

Theorem 5.4.1

Pf: $||\mathbf{u}+\mathbf{v}||^2 = \langle\mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle$

$= \langle\mathbf{u}, \mathbf{u}+\mathbf{v}\rangle + \langle\mathbf{v}, \mathbf{u}+\mathbf{v}\rangle$

$= \langle\mathbf{u}, \mathbf{u}\rangle + \langle\mathbf{u}, \mathbf{v}\rangle + \langle\mathbf{v}, \mathbf{u}\rangle + \langle\mathbf{v}, \mathbf{v}\rangle$

$= ||\mathbf{u}||^2 + 2\langle\mathbf{u}, \mathbf{v}\rangle + ||\mathbf{v}||^2$ (Note $\langle\mathbf{u}, \mathbf{v}\rangle = 0$)

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

Example 1

- Consider the vector space $C[-1, 1]$ with inner product defined by (3).
- The vector 1 and x are orthogonal since

$$< 1, x > = \int_{-1}^{1} 1 \cdot x \, dx = \frac{1}{2} x^2 \Big|_{-1}^{1} = \frac{1}{2} 1^2 - \frac{1}{2} (-1)^2 = 0$$

- To determine the length of each vector, we compute

$$< 1, 1 > = \int_{-1}^{1} 1 \cdot 1 \, dx = x \Big|_{-1}^{1} = 1 - (-1) = 2$$

$$< x, x > = \int_{-1}^{1} x \cdot x \, dx = \frac{1}{3} x^3 \Big|_{-1}^{1} = \frac{1}{3} 1^3 - \frac{1}{3} (-1)^3 = \frac{2}{3}$$

Example 1

$$\| 1 \| = (<1,1>)^{1/2} = \sqrt{2}$$

$$\| x \| = (<x,x>)^{1/2} = \sqrt{2/3} = \sqrt{6}/3$$

$$\| 1+x \|^2 = \| 1 \|^2 + \| x \|^2 = 2 + 2/3 = 8/3$$

- Verification:

$$\| 1+x \|^2 = <1+x, 1+x> = \int_{-1}^{1} (1+x)\cdot(1+x)\, dx = \int_{-1}^{1} (1+x)^2\, dx$$

$$= \frac{1}{3}(1+x)^3 \Big|_{-1}^{1} = \frac{1}{3}(1+1)^3 - \frac{1}{3}(1+(-1))^3 = \frac{1}{3}2^3 - \frac{1}{3}0^3 = \frac{8}{3}$$

Example 2

- For the vector space $C[-\pi, \pi]$, if we use a constant weight function $w(x) = 1/\pi$ to define an inner product:

$$< f, g > = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\, dx$$

- Then

$$< \cos x, \sin x > = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos x \sin x \, dx = 0$$

$$< \cos x, \cos x > = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos x \cos x \, dx = 1$$

$$< \sin x, \sin x > = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin x \sin x \, dx = 1$$

Example 2

- Thus cos$x$ and sin$x$ are orthogonal unit vectors with respect to this inner product.
- From the

$$\|\cos x + \sin x\| = \sqrt{\|\cos x\|^2 + \|\sin x\|^2} = \sqrt{1+1} = \sqrt{2}$$

Example 2

- Verification:

$$\| \cos x + \sin x \|^2 = \langle \cos x + \sin x, \cos x + \sin x \rangle$$

$$= \frac{1}{\pi}\int_{-\pi}^{\pi} (\cos x + \sin x) \cdot (\cos x + \sin x)\, dx$$

$$= \frac{1}{\pi}\int_{-\pi}^{\pi} [(\cos x)^2 + 2\sin x \cos x + (\sin x)^2]\, dx$$

$$= \frac{1}{\pi}\int_{-\pi}^{\pi} (\cos x)^2\, dx + 2\frac{1}{\pi}\int_{-\pi}^{\pi} \sin x \cos x\, dx + \frac{1}{\pi}\int_{-\pi}^{\pi} (\sin x)^2\, dx$$

$$= 1 + 2 \cdot 0 + 1 = 2$$

- For the vector space $R^{m\times n}$ the norm derived from the inner product is called the **Frobenius norm** and is denoted by $\|\bullet\|_F$. Thus if $A \in R^{m\times n}$, then

$$\|A\|_F = \sqrt{< A, A >} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}$$

Example 3

- If $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 3 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} -1 & 1 \\ 3 & 0 \\ -3 & 4 \end{bmatrix}$, then
- $<A, B>$
  $= 1\times(-1) + 1\times1 + 1\times3 + 2\times0 + 3\times(-3) + 3\times4$
  $= 6$
-

  $\|A\|_F = (1^2 + 1^2 + 1^2 + 2^2 + 3^2 + 3^2)^{1/2} = 5$
  $\|B\|_F = [(-1)^2+ 1^2+ 3^2+ 0^2 +(-3)^2 +4^2]^{1/2} = 6$

Example 4

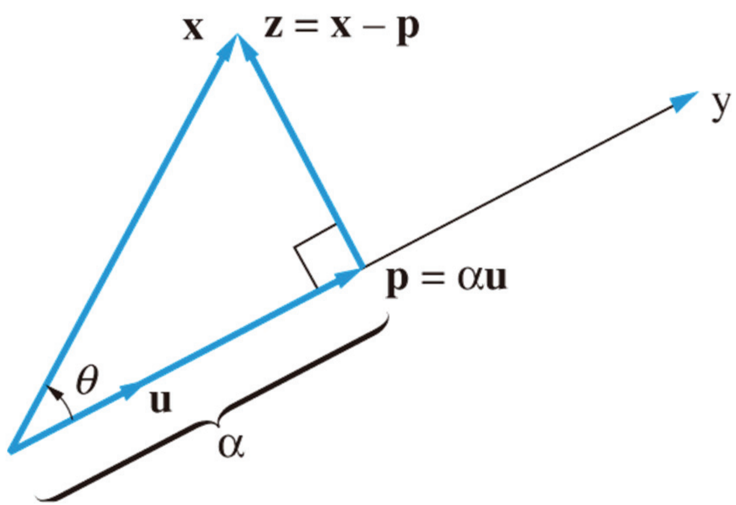
- In P_5, define an inner product by (5) with $x_i = (i-1)/4$ for $i = 1, 2, \ldots, 5$. The length of the function $p(x) = 4x$ is given by

$$\|4x\| = \left(\langle 4x, 4x \rangle\right)^{1/2} = \left(\sum_{i=1}^{5} 16x^2\right)^{1/2} = \left(\sum_{i=1}^{5} (i-1)^2\right)^{1/2} = \sqrt{30}$$

Definition

If **u** and **v** are vectors in an inner product space V and **v** $\neq$ **0**, then the **scalar projection** α and **vector projection** **p** of **u** onto **v** are given by

$$\alpha = \frac{<\mathbf{u}, \mathbf{v}>}{\|\mathbf{v}\|} \quad \text{and} \quad \mathbf{p} = \alpha \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{<\mathbf{u}, \mathbf{v}>}{\|\mathbf{v}\|} \frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{<\mathbf{u}, \mathbf{v}>}{\|\mathbf{v}\|^2} \mathbf{v} = \frac{<\mathbf{u}, \mathbf{v}>}{<\mathbf{v}, \mathbf{v}>} \mathbf{v}$$



Observation

If $\mathbf{v} \neq \mathbf{0}$ and $\mathbf{p}$ is the vector projection of $\mathbf{u}$ onto $\mathbf{v}$, then

I. $\mathbf{u} - \mathbf{p}$ and $\mathbf{p}$ are orthogonal

II. $\mathbf{u} = \mathbf{p}$ if and only if $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$

Pf:

I. Since $<\mathbf{p}, \mathbf{p}> = <\frac{\alpha}{\|\mathbf{v}\|}\mathbf{v}, \frac{\alpha}{\|\mathbf{v}\|}\mathbf{v}> = (\frac{\alpha}{\|\mathbf{v}\|})^2 <\mathbf{v}, \mathbf{v}> = \alpha^2$ and

$$<\mathbf{u}, \mathbf{p}> = <\mathbf{u}, \frac{<\mathbf{u}, \mathbf{v}>}{<\mathbf{v}, \mathbf{v}>}\mathbf{v}> = \frac{<\mathbf{u}, \mathbf{v}>}{<\mathbf{v}, \mathbf{v}>}<\mathbf{u}, \mathbf{v}> = \frac{<\mathbf{u}, \mathbf{v}>^2}{\|\mathbf{v}\|^2} = \alpha^2$$

$$\Rightarrow <\mathbf{u} - \mathbf{p}, \mathbf{p}> = <\mathbf{u}, \mathbf{p}> - \mathbf{2} <\mathbf{p}, \mathbf{p}> = \alpha^2 - \alpha^2 = 0$$

$\Rightarrow \mathbf{u} - \mathbf{p}$ and $\mathbf{p}$ are orthogonal

Observation

II. If $\mathbf{u} = \beta\mathbf{v}$, then the vector projection of $\mathbf{u}$ onto $\mathbf{v}$ is given by

$$\mathbf{p} = \frac{<\mathbf{u}, \mathbf{v}>}{<\mathbf{v}, \mathbf{v}>}\mathbf{v} = \frac{<\beta\mathbf{v}, \mathbf{v}>}{<\mathbf{v}, \mathbf{v}>}\mathbf{v} = \beta\mathbf{v} = \mathbf{u}$$

If $\mathbf{u} = \mathbf{p} \Rightarrow$ $\mathbf{u} = \mathbf{p} = \alpha\frac{\mathbf{v}}{\|\mathbf{v}\|} = \frac{\alpha}{\|\mathbf{v}\|}\mathbf{v} = \beta\mathbf{v}$, with $\beta = \frac{\alpha}{\|\mathbf{v}\|}$

Theorem 5.4.2

(The Cauchy-Schwarz Inequality)

If **u** and **v** are any two vectors in an inner product space V, then

$$| \langle \mathbf{u}, \mathbf{v} \rangle | \leq \|\mathbf{u}\| \, \|\mathbf{v}\|$$

Equality holds if and only if **u** and **v** are **linearly dependent**.

- From the above theorem, if **u** and **v** are nonzero vectors, then

$$-1 \leq \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|} \leq 1$$

and hence there is a unique angle $\theta \in [0, \pi]$ such that

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\| \, \|\mathbf{v}\|}$$

this equation can be used to define the angle θ between two nonzero vectors **u** and **v**

Definition

A vector space V is said to be a **normed linear space** if to each vector $\mathbf{v} \in V$ there is associated a real number $\|\mathbf{v}\|$ called the **norm** of $\mathbf{v}$, satisfying

I. $\|\mathbf{v}\| \geq 0$ with equality if and only if $\mathbf{v} = \mathbf{0}$

II. $\|\alpha\mathbf{v}\| = |\alpha|\ \|\mathbf{v}\|$ for any scalar α

III. $\|\mathbf{v}+\mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ for all $\mathbf{v}, \mathbf{w} \in V$

- The third condition is called the **triangle inequality**.

Theorem 5.4.3

If V is an **inner product space**, then the equation

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}, \text{ for all } \mathbf{v} \in \mathbf{V}$$

defines a **norm** on V

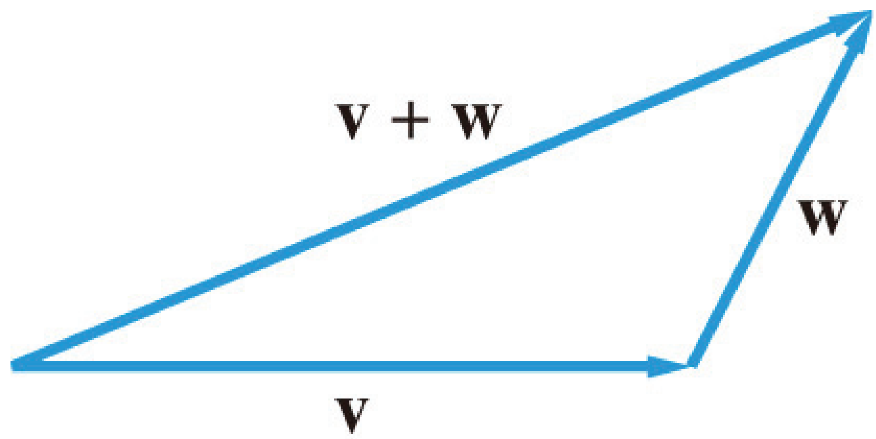


Figure 5.4.2

Theorem 5.4.3

- *Pf:* It is easily seen that conditions **I** and **II** are satisfied. We have to show that condition **III** is satisfied:

  $\|\mathbf{u}+\mathbf{v}\|^2 = \langle \mathbf{u}+\mathbf{v}, \mathbf{u}+\mathbf{v} \rangle$

  $= \langle \mathbf{u}, \mathbf{u} \rangle + 2 \underline{\langle \mathbf{u}, \mathbf{v} \rangle} + \langle \mathbf{v}, \mathbf{v} \rangle$

  $\leq \|\mathbf{u}\|^2 + 2 \underline{\|\mathbf{u}\| \, \|\mathbf{v}\|} + \|\mathbf{v}\|^2$

  (from **The Cauchy-Schwarz Inequality**)

  $= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2$

  Thus

  $$\|\mathbf{u}+\mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

Definition

- For every vector $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$ in R^n

1-norm: $\|\mathbf{x}\|_1 = |x_1| + |x_2| + \ldots + |x_n|$

2-norm: $\|\mathbf{x}\|_2 = (|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2)^{1/2} =$

$$\left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2} = \sqrt{<\mathbf{x}, \mathbf{x}>}$$

***p*-norm:** $\|\mathbf{x}\|_p = (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{1/p} = \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}$

∞-norm (uniform norm, infinity norm):

$\|\mathbf{x}\|_\infty = \mathbf{max}(|x_1|, |x_2|, \ldots, |x_n|)$

- If $p \neq 2$, $\|\bullet\|_p$ does not correspond to any inner product, thus the Pythogorean Law will not hold.
- For example, $\mathbf{x}_1 = [1 , 2]^T$ and $\mathbf{x}_2 = [-4 , 2]^T$ are orthogonal. However,

$$\mathbf{x}_1 + \mathbf{x}_2 = [1 , 2]^T + [-4 , 2]^T = [-3 , 4]^T$$

$$(\|\mathbf{x}_1\|_\infty)^2 + (\|\mathbf{x}_2\|_\infty)^2 = 2^2 + 4^2 = 4 + 16 = 20$$

$$(\|\mathbf{x}_1 + \mathbf{x}_2\|_\infty)^2 = 4^2 = 16$$

$$(\|\mathbf{x}_1\|_2)^2 + (\|\mathbf{x}_2\|_2)^2 = (1^2 + 2^2) + ((-4)^2 + 2^2) = 5 + 20 = 25$$

$$(\|\mathbf{x}_1 + \mathbf{x}_2\|_2)^2 = (-3)^2 + 4^2 = 9 + 16 = 25$$

Example 5

- Let $\mathbf{x} = (4, -5, 3)^T$ in R^3, then

  $\|\mathbf{x}\|_1 = |4| + |-5| + |3| = 12$
  $\|\mathbf{x}\|_2 = (|4|^2 + |-5|^2 + |3|^2)^{1/2} = (50)^{1/2}$
  $\|\mathbf{x}\|_\infty = \max(|4|, |-5|, |3|) = 5$

- A norm provides a way of measuring <u>the distance between two vectors</u>

Definition

Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in a normed linear space. The **distance** between $\mathbf{x}$ and $\mathbf{y}$ is defined to be the number $\|\mathbf{y} - \mathbf{x}\|$.

5.5 Orthonormal Sets

- In R^2, the elements of the standard basis $\{\mathbf{e}_1, \mathbf{e}_2\}$ are orthogonal unit vectors
- In working with an inner product space V, it is generally desirable to have a basis of mutually orthogonal unit vectors
- Convenient in finding coordinates of vectors and solving least square problems

Definition

Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be nonzero vectors in an inner product space V. If $<\mathbf{v}_i, \mathbf{v}_j> = 0$ whenever $i \neq j$, then $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is said to be an **orthogonal set** of vectors.

Example 1

- The set $\{[1, 1, 1]^T, [2, 1, -3]^T, [4, -5, 1]^T\}$ is an orthogonal set in R^3.
- Since

$$[1, 1, 1]^T [2, 1, -3] = 0$$
$$[1, 1, 1]^T [4, -5, 1] = 0$$
$$[2, 1, -3]^T [4, -5, 1] = 0$$

Theorem 5.5.1

If $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ is an **orthogonal set** of nonzero vectors in an inner product space V, then $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are **linearly independent**.

- *Pf*:

  Let $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n = \mathbf{0}$ (1)

  $(\mathbf{v}_j)^T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \ldots + c_n\mathbf{v}_n) = 0$ for all j

  $\Rightarrow c_1\langle\mathbf{v}_j, \mathbf{v}_1\rangle + c_2\langle\mathbf{v}_j, \mathbf{v}_2\rangle + \ldots + c_n\langle\mathbf{v}_j, \mathbf{v}_n\rangle = 0$

  $\Rightarrow c_j\|\mathbf{v}_j\|^2 = 0 \Rightarrow c_j = 0$ for all j

  $\Rightarrow \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ are linearly independent.

Definition: An **orthonormal set** of vectors is an **orthogonal** set of **unit vectors**.

- The set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ will be orthonormal iff

$$\langle \mathbf{u}_i, \mathbf{u}_j \rangle = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

- Given any orthogonal set of nonzero vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, it is possible to form an **orthonormal set** by defining

$$\mathbf{u}_i = \frac{1}{\|\mathbf{v}_i\|}\mathbf{v}_i, \quad \text{for } i = 1, 2, \cdots, n$$

Then $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ will be an orthonormal set

Example 2

- To get an orthonormal set from the set $\{[1, 1, 1]^T, [2, 1, -3]^T, [4, -5, 1]^T\}$

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{3}}[1,1,1]^T$$

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2\|}\mathbf{v}_2 = \frac{1}{\sqrt{14}}[2,1,-3]^T$$

$$\mathbf{u}_3 = \frac{1}{\|\mathbf{v}_3\|}\mathbf{v}_3 = \frac{1}{\sqrt{42}}[4,-5,1]^T$$

Example 3

- In $C[-\pi, \pi]$ with inner product:

$$< f, g > = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)\,dx \tag{2}$$

The set $\{1, \cos x, \cos 2x, \ldots, \cos nx\}$ is an orthogonal set of vectors

- *Sol:*

(1) For any positive integers j and k

$$<1, \cos kx> = \frac{1}{\pi}\int_{-\pi}^{\pi} 1\cdot\cos kx\,dx = 0 = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos kx\,dx = 0$$

$$<\cos jx, \cos kx> = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos jx \cos kx\,dx = 0$$

Example 3

(2) The functions cos x, cos $2x$, …, cos nx are unit vectors since

$$< \cos kx, \cos kx > = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos kx \cos kx \, dx = \frac{1}{\pi}\int_{-\pi}^{\pi} \cos^2 kx \, dx = 1$$

$$< 1, 1 > = \frac{1}{\pi}\int_{-\pi}^{\pi} 1 \cdot 1 \, dx = \frac{1}{\pi}\int_{-\pi}^{\pi} 1 \, dx = 2$$

Thus, $1/\sqrt{2}$ is a unit vector

$\Rightarrow \{1/\sqrt{2}$, cos x, cos $2x$, …, cos $nx\}$ is an orthonormal set of vectors

- From **Theorem 5.5.1**, if $B = \{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is an orthonormal set in an inner product space V
  $\Rightarrow$ $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ are **linearly independent**
  $\Rightarrow$ B is a **basis** for a subspace $S = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ of V
  $\Rightarrow$ B is an **orthonormal basis** for S

Theorem 5.5.2

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be an **orthonormal basis** for an inner product space V. If , then $c_i = <\mathbf{v}, \mathbf{u}_i>$ **(Note: the scalar projection of v onto $\mathbf{u}_i$)**

- *Pf:*

$$\langle \mathbf{v}, \mathbf{u}_i \rangle = \left\langle \sum_{j=1}^{n} c_j \mathbf{u}_j, \mathbf{u}_i \right\rangle = \sum_{j=1}^{n} c_j \langle \mathbf{u}_i, \mathbf{u}_j \rangle = \sum_{j=1}^{n} c_j \, \delta_{ij} = c_i$$

Corollary 5.5.3

Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for an inner product space V. If

$$\mathbf{u} = \sum_{i=1}^{n} a_i \mathbf{u}_i \text{ and } \mathbf{v} = \sum_{i=1}^{n} b_i \mathbf{u}_i$$

then

$$\langle \mathbf{u}, \mathbf{v} \rangle = \sum_{i=1}^{n} a_i\, b_i$$

(Note: the inner product of two vectors is equivalent to the inner product of their coordinate vectors)

- *Pf:* From **Theorem 5.5.2**, $<\mathbf{v}, \mathbf{u}_i> = b_i$, $i = 1, 2, \ldots, n$

$$\Rightarrow \langle \mathbf{u}, \mathbf{v} \rangle = \left\langle \sum_{i=1}^{n} a_i \mathbf{u}_i, \mathbf{v} \right\rangle = \sum_{i=1}^{n} a_i \langle \mathbf{u}_i, \mathbf{v} \rangle = \sum_{i=1}^{n} a_i b_i$$

Corollary 5.5.4

(Parseval's Formula)

If $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for an inner product space V and

$$\mathbf{v} = \sum_{i=1}^{n} c_i \mathbf{u}_i$$

then

$$\|\mathbf{v}\|^2 = \sum_{i=1}^{n} c_i^2$$

- *Pf:* From **Corollary 5.5.3**

$$\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle = \sum_{i=1}^{n} c_i c_i = \sum_{i=1}^{n} c_i^2$$

Example 4

- The vectors $\mathbf{u}_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})^T$ and $\mathbf{u}_2 = (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})^T$

  form an orthonormal basis for R^2.
- *Sol:* If $\mathbf{x} = (x_1, x_2)^T \in R^2$, then

$$c_1 = \mathbf{x}^T\mathbf{u}_1 = \frac{x_1 + x_2}{\sqrt{2}} \text{ and } c_2 = \mathbf{x}^T\mathbf{u}_2 = \frac{x_1 - x_2}{\sqrt{2}}$$

$$\Rightarrow \mathbf{x} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 = \frac{x_1 + x_2}{\sqrt{2}}\mathbf{u}_1 + \frac{x_1 - x_2}{\sqrt{2}}\mathbf{u}_2$$

$$\Rightarrow \|\mathbf{x}\|^2 = \sum_{i=1}^{2} c_i^2 = \left(\frac{x_1 + x_2}{\sqrt{2}}\right)^2 + \left(\frac{x_1 - x_2}{\sqrt{2}}\right)^2 = x_1^2 + x_2^2$$

Example 5

- Given that $\{ 1/\sqrt{2}, \cos 2x \}$ is an orthonormal set in $C[-\pi, \pi]$ (with inner product as in Ex 3), determine the value of $\int_{-\pi}^{\pi} \sin^4 x \, dx$ without computing antiderivatives.
- *Sol:*
  Since $\sin^2 x = \dfrac{1-\cos 2x}{2} = \dfrac{1}{\sqrt{2}}\dfrac{1}{\sqrt{2}} + (-\dfrac{1}{2})\cos 2x$
  From Parseval's Formula, we can get

$$\| \sin^2 x \|^2 = < \sin^2 x, \sin^2 x > = \frac{1}{\pi}\int_{-\pi}^{\pi} \sin^4 x \, dx = \left(\frac{1}{\sqrt{2}}\right)^2 + \left(-\frac{1}{2}\right)^2 = \frac{3}{4}$$

$$\Rightarrow \int_{-\pi}^{\pi} \sin^4 x \, dx = \frac{3}{4}\pi$$

Orthogonal Matrices
Definition

An $n \times n$ matrix Q is said to be an **orthogonal matrix** if the **column vectors** of Q form an **orthonormal set** in R^n.

Theorem 5.5.5

An $n \times n$ matrix Q is orthogonal if and only if $Q^T Q = I$.
($Q^{-1} = Q^T$)

- *Pf:*
  An $n \times n$ matrix Q is orthogonal if and only if its column vectors satisfy
  $$\mathbf{q}_i^T \mathbf{q}_j = \delta_{ij}$$
- However, is the (i, j) entry of the matrix $Q^T Q$
  $\Rightarrow Q$ is orthogonal if and only if $Q^T Q = I$

Note

$$Q^T Q = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n]^T [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n]$$

$$= \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n]$$

$$= \begin{bmatrix} \mathbf{q}_1^T\mathbf{q}_1 & \mathbf{q}_1^T\mathbf{q}_2 & \cdots & \mathbf{q}_1^T\mathbf{q}_n \\ \mathbf{q}_2^T\mathbf{q}_1 & \mathbf{q}_2^T\mathbf{q}_2 & \cdots & \mathbf{q}_2^T\mathbf{q}_n \\ \vdots & \vdots & & \vdots \\ \mathbf{q}_n^T\mathbf{q}_1 & \mathbf{q}_n^T\mathbf{q}_2 & \cdots & \mathbf{q}_n^T\mathbf{q}_n \end{bmatrix}$$

- If Q is an orthogonal matrix then Q is invertible and $Q^{-1} = Q^T$

Example 6

$$Q^TQ = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$
$$= \begin{bmatrix} \cos\theta\cos\theta + \sin\theta\sin\theta = 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$Q = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is an orthogonal matrix.

Sol:

$Q = (\mathbf{q}_1, \mathbf{q}_2)$ with $\mathbf{q}_1 = (\cos\theta, \sin\theta)^T$, $\mathbf{q}_2 = (-\sin\theta, \cos\theta)^T$

$\Rightarrow \mathbf{q}_1{}^T\mathbf{q}_1 = (\cos\theta, \sin\theta)(\cos\theta, \sin\theta)^T = \cos^2\theta + \sin^2\theta = 1$

$\Rightarrow \mathbf{q}_2{}^T\mathbf{q}_2 = (-\sin\theta, \cos\theta)(-\sin\theta, \cos\theta)^T$

$= \sin^2\theta + \cos^2\theta = 1$

$\Rightarrow \mathbf{q}_1{}^T\mathbf{q}_2 = (\cos\theta, \sin\theta)(-\sin\theta, \cos\theta)^T$

$= -\sin\theta\cos\theta + \sin\theta\cos\theta = 0$

So, Q is an orthogonal matrix.

Verification

$$Q^{-1} = Q^T = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

$$QQ^{-1} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

- The matrix Q can be thought of as a linear transformation from R^2 to R^2 that has the effect of rotating each vector by an angle θ while leaving the length of the vector unchanged.
- Q^{-1} can be thought of as a rotation by the angle -θ

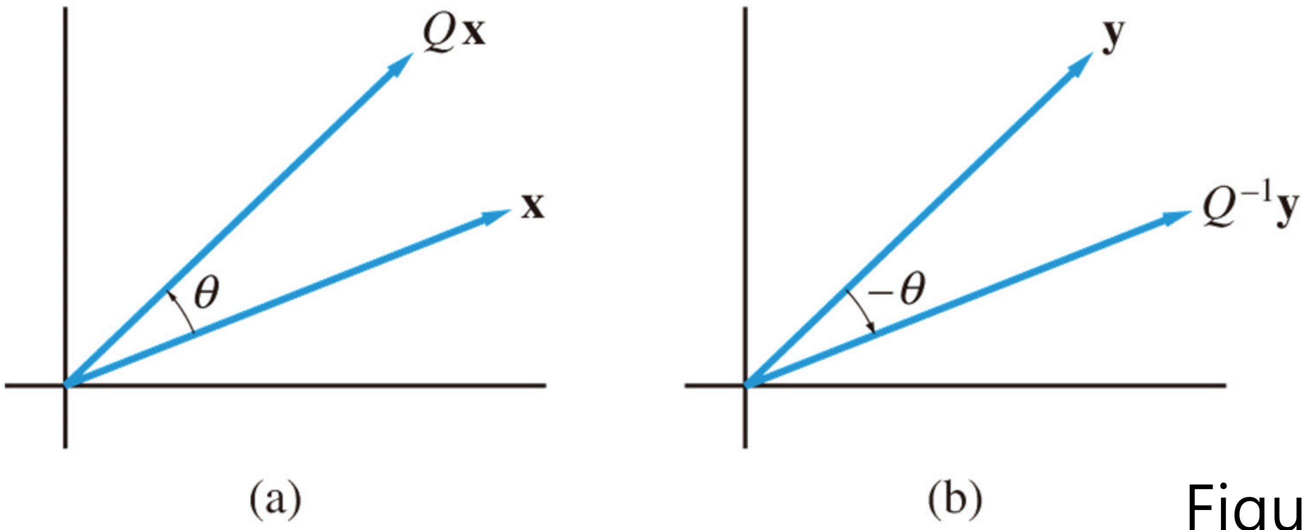


Figure 5.5.1

- In general, *inner product are preserved under multiplication by an orthogonal matrix* (i.e. $< \mathbf{x}, \mathbf{y} > = < Q\mathbf{x}, Q\mathbf{y} >$):
  $< Q\mathbf{x}, Q\mathbf{y} > = (Q\mathbf{x})^T Q\mathbf{y} = \mathbf{x}^T\underline{Q^T Q}\mathbf{y} = \mathbf{x}^T\mathbf{y} = < \mathbf{x}, \mathbf{y} >$
- If $\mathbf{x} = \mathbf{y}$ (i.e. $< \mathbf{x}, \mathbf{x} > = < Q\mathbf{x}, Q\mathbf{x} >$), then $\|Q\mathbf{x}\|^2 = \|\mathbf{x}\|^2$ and hence $\| Q\mathbf{x} \| = \| \mathbf{x} \|$
- *Multiplication by an orthogonal matrix preserves the lengths of vectors*

Properties of Orthogonal Matrices

If Q is an $n \times n$ orthogonal matrix, then

(1) The column vectors of Q form an orthonormal basis for R^n.

(2) $Q^T Q = I$

(3) $Q^T = Q^{-1}$

(4) $\langle Q\mathbf{x}, Q\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle$

(5) $\|Q\mathbf{x}\|_2 = \|\mathbf{x}\|_2$

Permutation Matrices

- A **permutation matrix** is a matrix formed from the identity matrix by **reordering its columns**.
- A permutation matrix is an **orthogonal matrix**.
- If P is the permutation matrix formed by reordering the columns of I in the order $(k_1, k_2, \ldots, k_n)$, then $P = (\mathbf{e}_{k1}, \mathbf{e}_{k2}, \ldots, \mathbf{e}_{kn})$. If A is an $m \times n$ matrix, then
$$AP = (A\mathbf{e}_{k1}, A\mathbf{e}_{k2}, \ldots, A\mathbf{e}_{kn}) = (\mathbf{a}_{k1}, \mathbf{a}_{k2}, \ldots, \mathbf{a}_{kn})$$
- Post multiplication of A by P reorders the columns of A in the order $(k_1, k_2, \ldots, k_n)$

Example

$$A = \begin{bmatrix}1 & 2 & 3\\ 1 & 2 & 3\end{bmatrix} \text{ and } P = \begin{array}{c}\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix}\begin{matrix}2\\ 3\\ 1\end{matrix}\\ \begin{matrix}3 & 1 & 2\end{matrix}\end{array}$$

Then $AP = \begin{bmatrix}1 & 2 & 3\\ 1 & 2 & 3\end{bmatrix}\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix} = \begin{bmatrix}3 & 1 & 2\\ 3 & 1 & 2\end{bmatrix}$, and

$$P^{-1} = P^T = \begin{bmatrix}0 & 0 & 1\\ 1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}$$

- Since $P = (\mathbf{e}_{k1}, \mathbf{e}_{k2}, ..., \mathbf{e}_{kn})$ is orthogonal, then

$$P^{-1} = P^T = \begin{bmatrix} e_{k_1}^T \\ e_{k_2}^T \\ \vdots \\ e_{k_n}^T \end{bmatrix}$$

- The k_1 column of P^T is $\mathbf{e}_1$, the k_2 column of P^T is $\mathbf{e}_2$, and so on.
- P^T is also a permutation matrix

- P^T is formed from ***I*** by **reordering the rows** of I in the order $(k_1, k_2, ..., k_\text{n})$
- If Q is the permutation matrix formed by reordering the rows of I in the order $(k_1, k_2, ..., k_n)$ and B is an $n \times r$ matrix, then

$$QB = \begin{bmatrix} e_{k_1}^T \\ e_{k_2}^T \\ \vdots \\ e_{k_n}^T \end{bmatrix} B = \begin{bmatrix} e_{k_1}^T B \\ e_{k_2}^T B \\ \vdots \\ e_{k_n}^T B \end{bmatrix} = \begin{bmatrix} \vec{\mathbf{b}}_{k_1} \\ \vec{\mathbf{b}}_{k_2} \\ \vdots \\ \vec{\mathbf{b}}_{k_n} \end{bmatrix}$$

$\Rightarrow$ QB is formed by <u>reordering the rows of B</u> in the order $(k_1, k_2, ..., k_n)$.

Example

$$Q = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{matrix} 3 \\ 1 \\ 2 \end{matrix} \text{ and } B = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \end{bmatrix}$$

$$\Rightarrow QB = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 1 & 1 \\ 2 & 2 \end{bmatrix}$$

5.6 The Gram-Schmidt Orthogonalization Process

- Learn a process for constructing an orthonormal basis for an n-dimensional inner product space V.
- Using projections to transform an ordinary basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ into an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$.

- Construction process – construct the $\mathbf{u}_i$'s so that

$$\text{Span}(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k) = \text{Span}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)$$

$$\text{for } k = 1, \ldots, n$$

  - **Step 1:** $\mathbf{u}_1 = \left(\dfrac{1}{\|\mathbf{x}_1\|}\right)\mathbf{x}_1 \quad \Rightarrow \text{Span}(\mathbf{u}_1) = \text{Span}(\mathbf{x}_1)$

– **Step 2:** Let $\mathbf{p}_1$ denote the projection of $\mathbf{x}_2$ onto Span($\mathbf{x}_1$) = Span($\mathbf{u}_1$):

$$\mathbf{p}_1 = \langle \mathbf{x}_2, \mathbf{u}_1 \rangle \mathbf{u}_1$$

From Theorem 5.5.7: $(\mathbf{x}_2 - \mathbf{p}_1) \perp \mathbf{u}_1$

Note that $\mathbf{x}_2 - \mathbf{p}_1 \neq \mathbf{0}$ since

$$\mathbf{x}_2 - \mathbf{p}_1 = \mathbf{x}_2 - \langle \mathbf{x}_2, \mathbf{u}_1 \rangle \mathbf{u}_1 = \mathbf{x}_2 - \frac{\langle \mathbf{x}_2, \mathbf{u}_1 \rangle}{\| \mathbf{x}_1 \|} \mathbf{x}_1$$

and $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent

$$\Rightarrow \mathbf{u}_2 = \frac{1}{\| \mathbf{x}_2 - \mathbf{p}_1 \|} (\mathbf{x}_2 - \mathbf{p}_1)$$

$\Rightarrow$ $\mathbf{u}_2$ is a unit vector orthogonal to $\mathbf{u}_1$

$\Rightarrow$ Span($\mathbf{u}_1$, $\mathbf{u}_2$) = Span($\mathbf{x}_1$, $\mathbf{x}_2$)

– **Step 3:** Let $\mathbf{p}_2$ be the projection of $\mathbf{x}_3$ onto Span($\mathbf{x}_1$, $\mathbf{x}_2$) = Span($\mathbf{u}_1$, $\mathbf{u}_2$):

$$\mathbf{p}_2 = \langle \mathbf{x}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{x}_3, \mathbf{u}_2 \rangle \mathbf{u}_2$$

$$\Rightarrow \qquad \mathbf{u}_3 = \frac{1}{\| \mathbf{x}_3 - \mathbf{p}_2 \|}(\mathbf{x}_3 - \mathbf{p}_2)$$

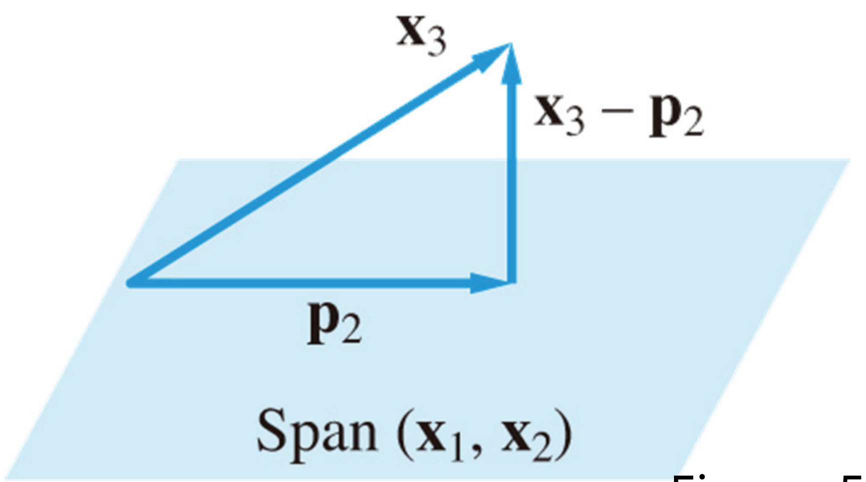


Figure 5.6.1

Theorem 5.6.1

(The Gram-Schmidt Process)

Let $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ be a basis for the inner product space V. Let

$$\mathbf{u}_1 = \left(\frac{1}{\|\mathbf{x}_1\|}\right)\mathbf{x}_1$$

and define $\mathbf{u}_2, \ldots, \mathbf{u}_n$ recursively by

$$\mathbf{u}_{k+1} = \frac{1}{\|\mathbf{x}_{k+1} - \mathbf{p}_k\|}(\mathbf{x}_{k+1} - \mathbf{p}_k)$$

where $\mathbf{p}_k = \langle \mathbf{x}_{k+1}, \mathbf{u}_1 \rangle \mathbf{u}_1 + \langle \mathbf{x}_{k+1}, \mathbf{u}_2 \rangle \mathbf{u}_2 + \cdots + \langle \mathbf{x}_{k+1}, \mathbf{u}_k \rangle \mathbf{u}_k$ is the projection of $\mathbf{x}_{k+1}$ onto Span$\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$. The set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an **orthonormal basis** for V.

Example 1

- Find an orthonormal basis for P_3 if the inner product on P_3 is defined by

$$\langle p, q \rangle = \sum_{i=1}^{3} p(x_i)q(x_i)$$

where x_1 = -1, x_2 = 0, and x_3 = 1.

- *Sol:*

Use the Gram-Schmidt Process to generate an orthonormal basis.

$$\left\| 1^2 \right\| = \langle 1, 1 \rangle = 3$$

So

$$\mathbf{u}_1 = \left(\frac{1}{\|1\|} \right) 1 = \frac{1}{\sqrt{3}}$$

Example 1

- Set
$$p_1 = \left\langle x, \frac{1}{\sqrt{3}} \right\rangle \frac{1}{\sqrt{3}} = \left(-1 \cdot \frac{1}{\sqrt{3}} + 0 \cdot \frac{1}{\sqrt{3}} + 1 \cdot \frac{1}{\sqrt{3}} \right) \frac{1}{\sqrt{3}} = 0$$
- Therefore,
$$x - p_1 = x \quad \text{and} \quad \left\| x - p_1 \right\|^2 = \langle x, x \rangle = 2$$
- Hence
$$\mathbf{u}_2 = \frac{1}{\sqrt{2}} x$$

Example 1

- Finally, $$p_2 = \left\langle x^2, \frac{1}{\sqrt{3}} \right\rangle \frac{1}{\sqrt{3}} + \left\langle x^2, \frac{1}{\sqrt{2}} \right\rangle \frac{1}{\sqrt{2}} x = \frac{2}{3}$$

$$\left\| x^2 - p_2 \right\|^2 = \left\langle x^2 - \frac{2}{3}, x^2 - \frac{2}{3} \right\rangle = \frac{2}{3}$$

  and hence

$$u_3 = \frac{\sqrt{6}}{2}\left(x^2 - \frac{2}{3} \right)$$

Example 2

- Let
$$A = \begin{bmatrix} 1 & -1 & 4 \\ 1 & 4 & -2 \\ 1 & 4 & 2 \\ 1 & -1 & 0 \end{bmatrix}$$
Find an orthonormal basis for the column space of A.
- ***Sol:***
The column vectors of A are linearly independent **(why?)** and hence form a basis for a 3-dimensional subspace of R^4. The Gram-Schmidt process used to construct an orthonormal basis as follows.

Example 2

- **Step 1:** $\mathbf{q_1} = \left(\dfrac{1}{\|\mathbf{a}_1\|}\right)\mathbf{a}_1$

$$r_{11} = \|\,\mathbf{a}_1\,\| = \sqrt{1^2 + 1^2 + 1^2 + 1^2} = 2$$

$$\mathbf{q_1} = \frac{1}{r_{11}}\mathbf{a}_1 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

Example 2

- **<u>Step 2:</u>** $\mathbf{q}_2 = \left(\dfrac{1}{\|\mathbf{a}_2 - \mathbf{p}_1\|}\right)(\mathbf{a}_2 - \mathbf{p}_1)$, where $\mathbf{p}_1 = \langle \mathbf{a}_2, \mathbf{q}_1 \rangle \mathbf{q}_1$

$$r_{12} = \langle \mathbf{a}_2, \mathbf{q}_1 \rangle = \mathbf{a}_2^T \mathbf{q}_1 = \begin{bmatrix} -1 & 4 & 4 & -1 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = 3$$

$$\mathbf{p}_1 = \langle \mathbf{a}_2, \mathbf{q}_1 \rangle \mathbf{q}_1 = r_{12}\mathbf{q}_1 = 3\mathbf{q}_1 = \begin{bmatrix} 3/2 \\ 3/2 \\ 3/2 \\ 3/2 \end{bmatrix}$$

Examp1e 2

$$\mathbf{a}_2 - \mathbf{p}_1 = \begin{bmatrix} -1 \\ 4 \\ 4 \\ -1 \end{bmatrix} - \begin{bmatrix} 3/2 \\ 3/2 \\ 3/2 \\ 3/2 \end{bmatrix} = \begin{bmatrix} -5/2 \\ 5/2 \\ 5/2 \\ -5/2 \end{bmatrix}$$

$$r_{22} = \left\| \mathbf{a}_2 - \mathbf{p}_1 \right\| = \sqrt{(-5/2)^2 + (5/2)^2 (5/2)^2 (-5/2)^2} = 5$$

$$\mathbf{q_2} = \frac{1}{r_{22}} (\mathbf{a}_2 - \mathbf{p}_1) = \frac{1}{5} \begin{bmatrix} -5/2 \\ 5/2 \\ 5/2 \\ -5/2 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 1/2 \\ 1/2 \\ -1/2 \end{bmatrix}$$

Example 2

- **Step 3:**

$$\mathbf{q_3} = \left(\frac{1}{\left\|\mathbf{a}_3 - \mathbf{p}_2\right\|}\right)(\mathbf{a}_3 - \mathbf{p}_2), \text{where } \mathbf{p}_2 = \left\langle \mathbf{a}_3, \mathbf{q}_1 \right\rangle \mathbf{q}_1 + \left\langle \mathbf{a}_3, \mathbf{q}_2 \right\rangle \mathbf{q}_2$$

$$r_{13} = \left\langle \mathbf{a}_3, \mathbf{q}_1 \right\rangle = \mathbf{a}_3^T \mathbf{q}_1 = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \\ 2 \\ 0 \end{bmatrix} = 2$$

$$r_{23} = \left\langle \mathbf{a}_3, \mathbf{q}_2 \right\rangle = \mathbf{q}_2^T \mathbf{a}_3 = \begin{bmatrix} -1/2 & 1/2 & 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \\ 2 \\ 0 \end{bmatrix} = -2$$

Example 2

$$\mathbf{p}_2 = \langle \mathbf{a}_3, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{a}_3, \mathbf{q}_2 \rangle \mathbf{q}_2 = r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 = 2\begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} - 2\begin{bmatrix} -1/2 \\ 1/2 \\ 1/2 \\ -1/2 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 0 \\ 2 \end{bmatrix}$$

$$\mathbf{a}_3 - \mathbf{p}_2 = \begin{bmatrix} 4 \\ -2 \\ 2 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 2 \\ -2 \end{bmatrix}$$

$$r_{33} = \|\mathbf{a}_3 - \mathbf{p}_2\| = 4$$

$$\mathbf{q}_3 = \left(\frac{1}{\|\mathbf{a}_3 - \mathbf{p}_2\|} \right)(\mathbf{a}_3 - \mathbf{p}_2) = \frac{1}{r_{33}}(\mathbf{a}_3 - \mathbf{p}_2) = \frac{1}{4}\begin{bmatrix} 2 \\ -2 \\ 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix}$$

Note

$$\textbf{(1)}\ \mathbf{q_1} = \frac{1}{r_{11}}\mathbf{a}_1 \Rightarrow \mathbf{a}_1 = r_{11}\mathbf{q_1}$$

$$\textbf{(2)}\ \mathbf{q_2} = \frac{1}{r_{22}}(\mathbf{a}_2 - \mathbf{p}_1)$$

$$\Rightarrow \mathbf{a}_2 = r_{22}\mathbf{q_2} + \mathbf{p}_1 = r_{22}\mathbf{q_2} + r_{12}\mathbf{q_1} \quad (\because \mathbf{p}_1 = r_{12}\mathbf{q_1})$$

$$\textbf{(3)}\ \mathbf{q_3} = \frac{1}{r_{33}}(\mathbf{a}_3 - \mathbf{p}_2)$$

$$\Rightarrow \mathbf{a}_3 = r_{33}\mathbf{q_3} + \mathbf{p}_2 = r_{22}\mathbf{q_2} + r_{13}\mathbf{q_1} + r_{23}\mathbf{q_2} \quad (\because \mathbf{p}_2 = r_{13}\mathbf{q_1} + r_{23}\mathbf{q_2})$$

- If the r_{ij}'s are used to form a matrix $R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix}$

  and we set $Q = [\mathbf{q}_1\ \mathbf{q}_2\ \mathbf{q}_3]$
  Then,

$$
\begin{aligned}
QR &= [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \mathbf{q}_3] \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix} \\
&= [r_{11}\mathbf{q}_1 \quad r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2 \quad r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 + r_{33}\mathbf{q}_3] \\
&= [\mathbf{a}_1 \quad \mathbf{a}_2 \quad \mathbf{a}_3] \\
&= A
\end{aligned}
$$

Theorem 5.6.2

(Gram-Schmidt *QR* Factorization)

If A is an $m\times n$ matrix of rank n, then A can be factored into a product QR, where Q is an $m\times n$ matrix with orthonormal columns and R is an $n\times n$ matrix whose diagonal entries are all positive [Note: R must be nonsingular since $\det(R) > 0$]

Note

$$Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \quad \text{and} \quad R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & & & \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix}$$

Example 3

- Compute the Gram-Schmidt QR factorization of the matrix

$$A = \begin{bmatrix} 1 & -2 & -1 \\ 2 & 0 & 1 \\ 2 & -4 & 2 \\ 4 & 0 & 0 \end{bmatrix}$$

- *Sol:*
- **Step 1:** $\mathbf{q}_1 = \dfrac{1}{r_{11}} \mathbf{a}_1 = \dfrac{1}{\|\mathbf{a}_1\|} \mathbf{a}_1$

$$r_{11} = \| \mathbf{a}_1 \| = \sqrt{1^2 + 2^2 + 2^2 + 4^2} = 5 \qquad \mathbf{q}_1 = \frac{1}{r_{11}} \mathbf{a}_1 = \frac{1}{5} \begin{bmatrix} 1 \\ 2 \\ 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 1/5 \\ 2/5 \\ 2/5 \\ 4/5 \end{bmatrix}$$

Example 3

- **<u>Step 2:</u>**

$$\mathbf{q_2} = \frac{1}{\left\|\mathbf{a}_2 - \mathbf{p}_1\right\|}(\mathbf{a}_2 - \mathbf{p}_1) = \frac{1}{r_{22}}(\mathbf{a}_2 - \mathbf{p}_1), \text{ where } \mathbf{p}_1 = \left\langle \mathbf{a}_2, \mathbf{q}_1 \right\rangle \mathbf{q}_1 = r_{12}\mathbf{q}_1$$

$$r_{12} = \left\langle \mathbf{a}_2, \mathbf{q}_1 \right\rangle = \mathbf{a}_2^T \mathbf{q}_1 = \begin{bmatrix} -2 & 0 & -4 & 0 \end{bmatrix} \begin{bmatrix} 1/5 \\ 2/5 \\ 2/5 \\ 4/5 \end{bmatrix} = -2$$

$$\mathbf{p_1} = \left\langle \mathbf{a}_2, \mathbf{q}_1 \right\rangle \mathbf{q}_1 = r_{12}\mathbf{q}_1 = -2\mathbf{q}_1 = \begin{bmatrix} -2/5 \\ -4/5 \\ -4/5 \\ -8/5 \end{bmatrix}$$

Example 3

$$\mathbf{a}_2 - \mathbf{p_1} = \begin{bmatrix} -2 \\ 0 \\ -4 \\ 0 \end{bmatrix} - \begin{bmatrix} -2/5 \\ -4/5 \\ -4/5 \\ -8/5 \end{bmatrix} = \begin{bmatrix} -8/5 \\ 4/5 \\ -16/5 \\ 8/5 \end{bmatrix}$$

$$r_{22} = \|\mathbf{a}_2 - \mathbf{p}_1\| = \sqrt{(-8/5)^2 + (4/5)^2 + (-16/5)^2 + (8/5)^2} = 4$$

$$\mathbf{q_2} = \frac{1}{r_{22}}(\mathbf{a}_2 - \mathbf{p}_1) = \frac{1}{4}\begin{bmatrix} -8/5 \\ 4/5 \\ -16/5 \\ 8/5 \end{bmatrix} = \begin{bmatrix} -2/5 \\ 1/5 \\ -4/5 \\ 2/5 \end{bmatrix}$$

Example 3

- **Step 3:**

$$\mathbf{q_3} = \frac{1}{\left\|\mathbf{a}_3 - \mathbf{p}_2\right\|}(\mathbf{a}_3 - \mathbf{p}_2) = \frac{1}{r_{33}}(\mathbf{a}_3 - \mathbf{p}_2),$$

where $\mathbf{p}_2 = \langle \mathbf{a}_3,\ \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{a}_3,\ \mathbf{q}_2 \rangle \mathbf{q}_2 = r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2$

$$r_{13} = \langle \mathbf{a}_3, \mathbf{q}_1 \rangle = \mathbf{a}_3^T \mathbf{q}_1 = \begin{bmatrix} -1 & 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} 1/5 \\ 2/5 \\ 2/5 \\ 4/5 \end{bmatrix} = 1$$

Example 3

$$r_{23} = \langle \mathbf{a}_3, \mathbf{q}_2 \rangle = \mathbf{a}_3^T \mathbf{q}_2 = \begin{bmatrix} -1 & 1 & 2 & 0 \end{bmatrix} \begin{bmatrix} -2/5 \\ 1/5 \\ -4/5 \\ 2/5 \end{bmatrix} = -1$$

$$\mathbf{p}_2 = \langle \mathbf{a}_3, \mathbf{q}_1 \rangle \mathbf{q}_1 + \langle \mathbf{a}_3, \mathbf{q}_2 \rangle \mathbf{q}_2 = r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2$$

$$= \begin{bmatrix} 1/5 \\ 2/5 \\ 2/5 \\ 4/5 \end{bmatrix} - \begin{bmatrix} -2/5 \\ 1/5 \\ -4/5 \\ 2/5 \end{bmatrix} = \begin{bmatrix} 3/5 \\ 1/5 \\ 6/5 \\ 2/5 \end{bmatrix}$$

Example 3

$$\mathbf{a}_3 - \mathbf{p}_2 = \begin{bmatrix} -1 \\ 1 \\ 2 \\ 0 \end{bmatrix} - \begin{bmatrix} 3/5 \\ 1/5 \\ 6/5 \\ 2/5 \end{bmatrix} = \begin{bmatrix} -8/5 \\ 4/5 \\ 4/5 \\ -2/5 \end{bmatrix}$$

$$r_{33} = \left\|\mathbf{a}_3 - \mathbf{p}_2\right\| = \sqrt{(-8/5)^2 + (4/5)^2 + (4/5)^2 + (-2/5)^2} = 2$$

$$\mathbf{q_3} = \frac{1}{r_{33}}(\mathbf{a}_3 - \mathbf{p}_2) = \frac{1}{2}\begin{bmatrix} -8/5 \\ 4/5 \\ 4/5 \\ -2/5 \end{bmatrix} = \begin{bmatrix} -4/5 \\ 2/5 \\ 2/5 \\ -1/5 \end{bmatrix}$$

Example 3

- **Step 4:**

$$A = QR = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ 0 & r_{22} & r_{23} \\ 0 & 0 & r_{33} \end{bmatrix}$$

$$= \begin{bmatrix} 1/5 & -2/5 & -4/5 \\ 2/5 & 1/5 & 2/5 \\ 2/5 & -4/5 & 2/5 \\ 4/5 & 2/5 & -1/5 \end{bmatrix} \begin{bmatrix} 5 & -2 & 1 \\ 0 & 4 & -1 \\ 0 & 0 & 2 \end{bmatrix}$$

Example 4

- The system $R\mathbf{x} = Q^T\mathbf{b}$ is easily solved by back substitution: $\left[\begin{array}{ccc|c} 5 & -2 & 1 & -1 \\ 0 & 4 & -1 & -1 \\ 0 & 0 & 2 & 2 \end{array}\right]$

$\Rightarrow$ The solution is $x = \begin{bmatrix} -2/5 \\ 0 \\ 1 \end{bmatrix}$

Example 3

EXAMPLE 3. Compute the Gram–Schmidt QR factorization of the matrix

$$A = \begin{bmatrix} 1 & -2 & -1 \\ 2 & 0 & 1 \\ 2 & -4 & 2 \\ 4 & 0 & 0 \end{bmatrix}$$

SOLUTION.

Step 1. Set

$$r_{11} = \|\mathbf{a}_1\| = 5$$

$$\mathbf{q}_1 = \frac{1}{r_{11}}\mathbf{a}_1 = \left(\frac{1}{5}, \frac{2}{5}, \frac{2}{5}, \frac{4}{5}\right)^T$$

Step 2. Set

$$r_{12} = \mathbf{q}_1^T\mathbf{a}_2 = -2$$

$$\mathbf{p}_1 = r_{12}\mathbf{q}_1 = -2\mathbf{q}_1$$

$$\mathbf{a}_2 - \mathbf{p}_1 = \left(-\frac{8}{5}, \frac{4}{5}, -\frac{16}{5}, \frac{8}{5}\right)^T$$

$$r_{22} = \|\mathbf{a}_2 - \mathbf{p}_1\| = 4$$

$$\mathbf{q}_2 = \frac{1}{r_{22}}(\mathbf{a}_2 - \mathbf{p}_1) = \left(-\frac{2}{5}, \frac{1}{5}, -\frac{4}{5}, \frac{2}{5}\right)^T$$

Step 3. Set

$$r_{13} = \mathbf{q}_1^T \mathbf{a}_3 = 1, \qquad r_{23} = \mathbf{q}_2^T \mathbf{a}_3 = -1$$

$$\mathbf{p}_2 = r_{13}\mathbf{q}_1 + r_{23}\mathbf{q}_2 = \mathbf{q}_1 - \mathbf{q}_2 = \left(\frac{3}{5}, \frac{1}{5}, \frac{6}{5}, \frac{2}{5}\right)^T$$

$$\mathbf{a}_3 - \mathbf{p}_2 = \left(-\frac{8}{5}, \frac{4}{5}, \frac{4}{5}, -\frac{2}{5}\right)^T$$

$$r_{33} = \|\mathbf{a}_3 - \mathbf{p}_2\| = 2$$

$$\mathbf{q}_3 = \frac{1}{r_{33}}(\mathbf{a}_3 - \mathbf{p}_2) = \left(-\frac{4}{5}, \frac{2}{5}, \frac{2}{5}, -\frac{1}{5}\right)^T$$

At each step we have determined a column of Q and a column of R. The factorization is given by

$$A = QR = \begin{pmatrix} \frac{1}{5} & -\frac{2}{5} & -\frac{4}{5} \\ \frac{2}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & -\frac{4}{5} & \frac{2}{5} \\ \frac{4}{5} & \frac{2}{5} & -\frac{1}{5} \end{pmatrix} \begin{pmatrix} 5 & -2 & 1 \\ 0 & 4 & -1 \\ 0 & 0 & 2 \end{pmatrix}$$

◀

Theorem 5.6.3

▶ **THEOREM 5.6.3** *If A is an $m \times n$ matrix of rank n, then the solution to the least squares problem $A\mathbf{x} = \mathbf{b}$ is given by $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$, where Q and R are the matrices obtained from the factorization given in Theorem 5.6.2. The solution $\hat{\mathbf{x}}$ may be obtained by using back substitution to solve $R\mathbf{x} = Q^T\mathbf{b}$.*

▶***Proof.*** Let $\hat{\mathbf{x}}$ be the solution to the least squares problem $A\mathbf{x} = \mathbf{b}$ guaranteed by Theorem 5.3.2. Thus $\hat{\mathbf{x}}$ is the solution to the normal equations

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

If A is factored into a product QR, these equations become

$$(QR)^T QR\mathbf{x} = (QR)^T\mathbf{b}$$

or

$$R^T(Q^TQ)R\mathbf{x} = R^TQ^T\mathbf{b}$$

Since Q has orthonormal columns, it follows that $Q^TQ = I$ and hence

$$R^TR\mathbf{x} = R^TQ^T\mathbf{b}$$

Since R^T is invertible, this simplifies to

$$R\mathbf{x} = Q^T\mathbf{b} \qquad \text{or} \qquad \mathbf{x} = R^{-1}Q^T\mathbf{b}$$ ◀

EXAMPLE 4. Find the least squares solution to

$$\begin{bmatrix} 1 & -2 & -1 \\ 2 & 0 & 1 \\ 2 & -4 & 2 \\ 4 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 1 \\ -2 \end{bmatrix}$$

SOLUTION. The coefficient matrix of this system was factored in Example 3. Using that factorization, we have

$$Q^T\mathbf{b} = \begin{bmatrix} \frac{1}{5} & \frac{2}{5} & \frac{2}{5} & \frac{4}{5} \\ -\frac{2}{5} & \frac{1}{5} & -\frac{4}{5} & \frac{2}{5} \\ -\frac{4}{5} & \frac{2}{5} & \frac{2}{5} & -\frac{1}{5} \end{bmatrix} \begin{bmatrix} -1 \\ 1 \\ 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 2 \end{bmatrix}$$

The system $R\mathbf{x} = Q^T\mathbf{b}$ is easily solved by back substitution:

$$\left[\begin{array}{rrr|r} 5 & -2 & 1 & -1 \\ 0 & 4 & -1 & -1 \\ 0 & 0 & 2 & 2 \end{array}\right]$$

The solution is $\mathbf{x} = \left(-\frac{2}{5}, 0, 1\right)^T$. ◀

The Modified Gram–Schmidt Process

In Chapter 7 we will consider computer methods for solving least squares problems. The QR method of Example 4 does not in general produce accurate results when carried out with finite-precision arithmetic. In practice, there may be a loss of orthogonality due to roundoff error in computing $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$. We can achieve better numerical accuracy using a modified version of the Gram–Schmidt method. In the modified version the vector $\mathbf{q}_1$ is constructed as before:

$$\mathbf{q}_1 = \frac{1}{\|\mathbf{a}_1\|}\mathbf{a}_1$$

However, the remaining vectors $\mathbf{a}_2, \ldots, \mathbf{a}_n$ are then modified so as to be orthogonal to $\mathbf{q}_1$. This can be done by subtracting from each vector $\mathbf{a}_k$ the projection of $\mathbf{a}_k$ onto $\mathbf{q}_1$.

$$\mathbf{a}_k^{(1)} = \mathbf{a}_k - (\mathbf{q}_1^T\mathbf{a}_k)\mathbf{q}_1 \qquad k = 2, \ldots, n$$

At the second step we take

$$\mathbf{q}_2 = \frac{1}{\|\mathbf{a}_2^{(1)}\|}\mathbf{a}_2^{(1)}$$

The vector $\mathbf{q}_2$ is already orthogonal to $\mathbf{q}_1$. We then modify the remaining vectors to make them orthogonal to $\mathbf{q}_2$.

$$\mathbf{a}_k^{(2)} = \mathbf{a}_k^{(1)} - (\mathbf{q}_2^T\mathbf{a}_k^{(1)})\mathbf{q}_2 \qquad k = 3, \ldots, n$$

In a similar manner $\mathbf{q}_3, \mathbf{q}_4, \ldots, \mathbf{q}_n$ are successively determined. At the last step we need only set

$$\mathbf{q}_n = \frac{1}{\|\mathbf{a}_n^{(n-1)}\|}\mathbf{a}_n^{(n-1)}$$

to achieve an orthonormal set $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$. The following algorithm summarizes the process.

Algorithm 5.6.4

▶ **ALGORITHM 5.6.4** *[The Modified Gram–Schmidt Process]*

For $k = 1, 2, \ldots, n$ *set*

$$r_{kk} = \|\mathbf{a}_k\|$$

$$\mathbf{q}_k = \frac{1}{r_{kk}}\mathbf{a}_k$$

For $j = k+1, k+2, \ldots, n$, *set*

$$r_{kj} = \mathbf{q}_k^T \mathbf{a}_j$$
$$\mathbf{a}_j = \mathbf{a}_j - r_{kj}\mathbf{q}_k$$

End for loop

End for loop ◀

If the modified Gram–Schmidt process is applied to the column vectors of an $m \times n$ matrix A having rank n, then, as before, we can obtain a QR factorization of A. This factorization may then be used computationally to determine the least squares solution to $A\mathbf{x} = \mathbf{b}$.