Orthogonality

정보컴퓨터공학부 김민환 교수

Introduction

- Study useful properties of the scalar product of $\mathbf{R}^n$
- Study vector spaces that come equipped with an inner product, a generalization of the scalar product for abstract vectors
- Study usefulness of orthogonal (or orthonormal) bases
- Inner products are widely used from theoretical analysis to applied signal processing

Contents

Scalar Product in $\mathbf{R}^n$

To study the definition and useful properties of the scalar product in $\mathbf{R}^n$

Def. Scalar Product

Let $\mathbf{x}$ and $\mathbf{y}$ be n-vectors in $\mathbf{R}^n$. The product $\mathbf{x}^T\mathbf{y}$ is called the *scalar product* of $\mathbf{x}$ and $\mathbf{y}$. In particular, if $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, then

$$\mathbf{x}^T\mathbf{y} = x_1y_1 + x_2y_2 + \ldots + x_ny_n$$

- Notice that the product $\mathbf{x}^T\mathbf{y}$ is a real number, that is, a scalar

Def. Vector Length and Distance

Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in either $\mathbf{R}^2$ or $\mathbf{R}^3$. The *Euclidean length* of $\mathbf{x}$ is defined by $(\mathbf{x}^T\mathbf{x})^{1/2}$, denoted $\|\mathbf{x}\|$. The *distance* between $\mathbf{x}$ and $\mathbf{y}$ is defined to be the number $\|\mathbf{x}-\mathbf{y}\|$

- When $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$,

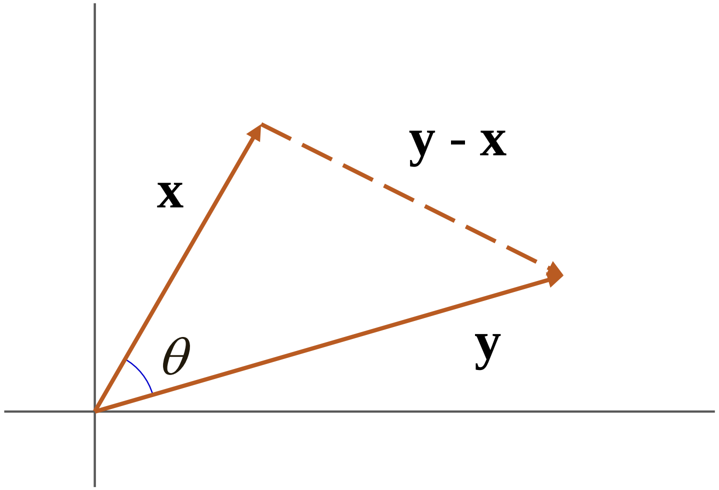
$$\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x} = x_1{}^2 + x_2{}^2 + \ldots + x_n{}^2$$

$$\|\mathbf{x}-\mathbf{y}\|^2 = (x_1 - y_1)^2 + \ldots + (x_n - y_n)^2$$

Th. 5.1.1

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

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



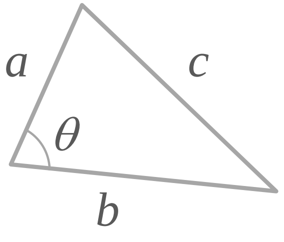
(*Proof*)

The vectors **x**, **y** and **y** - **x** may be used to form a triangle as shown. By the law of cosines, we have

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

and hence it follows that

$$\begin{aligned}\|\mathbf{x}\|\|\mathbf{y}\|\cos\theta &= (1/2)(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - \|\mathbf{y}-\mathbf{x}\|^2)\\ &= (1/2)(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - (\mathbf{y}-\mathbf{x})^T(\mathbf{y}-\mathbf{x}))\\ &= (1/2)(\|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - (\mathbf{y}^T\mathbf{y} - \mathbf{y}^T\mathbf{x} - \mathbf{x}^T\mathbf{y} + \mathbf{x}^T\mathbf{x}))\\ &= (1/2)(\mathbf{y}^T\mathbf{x} + \mathbf{x}^T\mathbf{y}) = \mathbf{x}^T\mathbf{y}\end{aligned}$$



$c^2 = a^2 + b^2 - 2ab\cos\theta$

- Corollary 5.1.2 *Cauchy-Schwartz Inequality*

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

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

  with equality holding if and only if one of the vectors is $\mathbf{0}$ or one vector is a multiple of the other

$$\left|\mathbf{x}^T\mathbf{y}\right| = \|\mathbf{x}\|\|\mathbf{y}\||\cos\theta| \leq \|\mathbf{x}\|\|\mathbf{y}\|$$

$$\text{If } \mathbf{x} = \mathbf{0} \text{ or } \mathbf{y} = \mathbf{0},\ \ \mathbf{x}^T\mathbf{y} = \|\mathbf{x}\|\|\mathbf{y}\| = 0$$

$$\text{If } \cos\theta = \pm 1,\ \ \mathbf{x} = \mathbf{y} \text{ or } \mathbf{x} = -\mathbf{y}$$

Def. Orthogonal

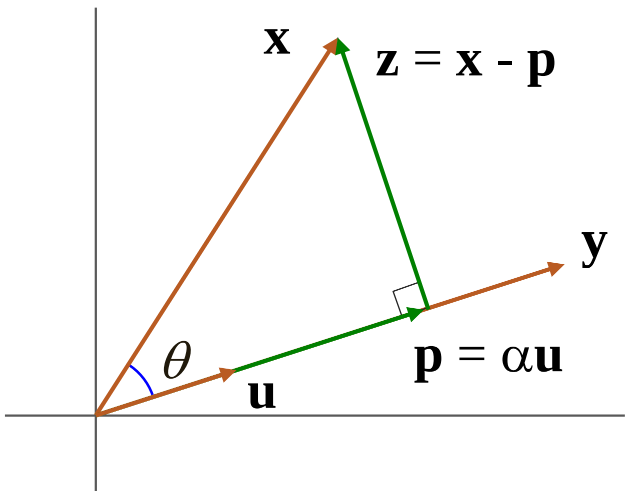
The vectors $\mathbf{x}$ and $\mathbf{y}$ in $\mathbf{R}^2$ (or $\mathbf{R}^3$), are said to be *orthogonal* if $\mathbf{x}^T\mathbf{y} = 0$

- The zero vector $\mathbf{0}$ is orthogonal to every vector in $\mathbf{R}^2$
- The vectors $(2,-3,1)$ and $(1,1,1)$ are orthogonal in $\mathbf{R}^3$

Scalar and Vector Projections

Scalar projection of $\mathbf{x}$ to $\mathbf{y}$: $\alpha = \mathbf{x}^T\mathbf{y} \,/\, \|\mathbf{y}\|$

Vector projection of $\mathbf{x}$ to $\mathbf{y}$: $\mathbf{p} = \alpha\,\mathbf{u} = \alpha\,\mathbf{y} \,/\, \|\mathbf{y}\| = (\mathbf{x}^T\mathbf{y} \,/\, \mathbf{y}^T\mathbf{y})\,\mathbf{y}$



$$\mathbf{x} = \mathbf{p} + \mathbf{z}$$

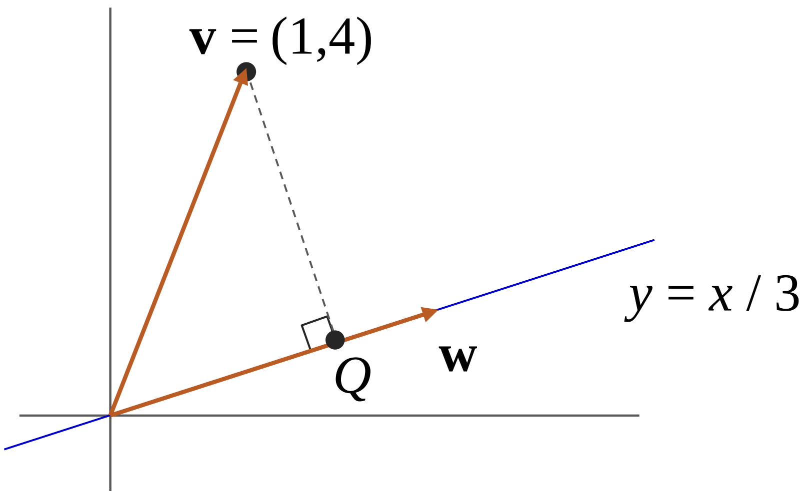
$$\begin{aligned}\mathbf{p}^T\mathbf{z} &= \mathbf{p}^T(\mathbf{x}-\mathbf{p}) = \mathbf{p}^T\mathbf{x} - \mathbf{p}^T\mathbf{p} \\ &= \alpha\mathbf{u}^T\mathbf{x} - \alpha^2 = 0\end{aligned}$$

$$\begin{aligned}\therefore \alpha &= \mathbf{u}^T\mathbf{x} = \|\mathbf{u}\|\|\mathbf{x}\|\cos\theta \\ &= \|\mathbf{x}\|\cos\theta = \|\mathbf{x}\|\|\mathbf{y}\|\cos\theta \,/\, \|\mathbf{y}\| \\ &= \mathbf{x}^T\mathbf{y} \,/\, \|\mathbf{y}\|\end{aligned}$$

(*Ex*.1) Determine the coordinate of the point Q

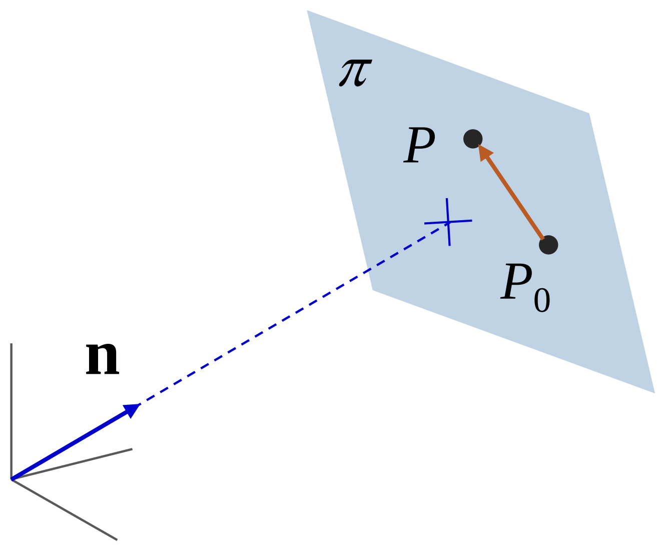
(*Solution*)

The vector $\mathbf{w} = (3,1)$ is a vector in the direction of the line $y = x / 3$. Q is the vector projection of $\mathbf{v} = (1,4)$ onto $\mathbf{w}$



Plane Equation

- 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 $\mathbf{n}$ is a nonzero vector and P_0 is a fixed point, the set of points P such that $\overrightarrow{P_0P}$ is orthogonal to $\mathbf{n}$ forms a plane π in 3-space that passes through P_0

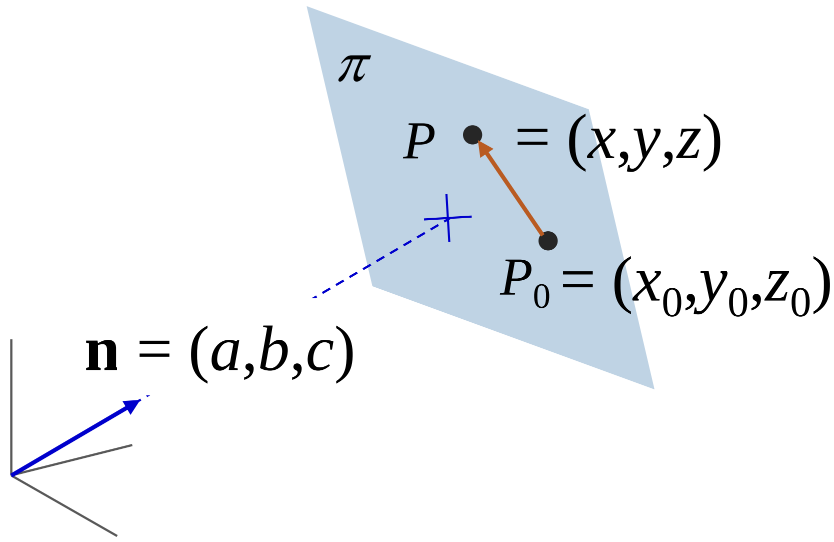


- A point $P = (x,y,z)$ will lie on the plane π if and only if

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

- If $\mathbf{n} = (a,b,c)$ and $P_0 = (x_0,y_0,z_0)$, this equation can be written in the form

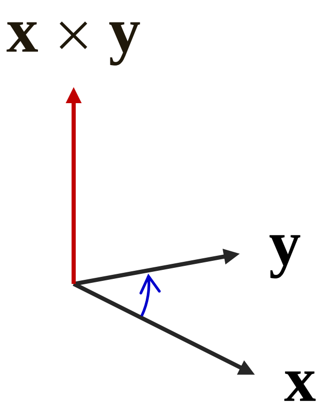
$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0$$

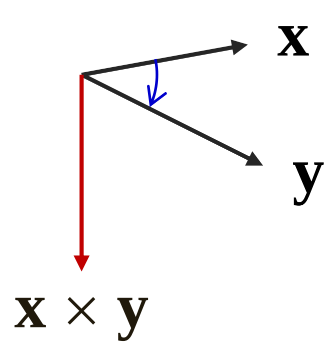


Def. Cross Product

Let $\mathbf{x}$ and $\mathbf{y}$ be vectors in $\mathbf{R}^3$. The *cross product* of $\mathbf{x}$ and $\mathbf{y}$, denoted $\mathbf{x} \times \mathbf{y}$, is a third vector defined by

$$\mathbf{x} \times \mathbf{y} = \begin{bmatrix} x_2 y_3 - y_2 x_3 \\ y_1 x_3 - x_1 y_3 \\ x_1 y_2 - y_1 x_2 \end{bmatrix}$$





예각 직각 둔각 평각 우각 ?

- If $\mathbf{A}$ is any matrix of the form, $\mathbf{A} = \begin{bmatrix} w_1 & w_2 & w_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{bmatrix}$

$$\mathbf{x} \times \mathbf{y} = \begin{bmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{bmatrix} = \begin{bmatrix} A_{11} \\ A_{12} \\ A_{13} \end{bmatrix} = A_{11}\mathbf{e}_1 + A_{12}\mathbf{e}_2 + A_{13}\mathbf{e}_3$$

$$\equiv \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$

where A_{11}, A_{12}, A_{13} are *cofactors* of $\mathbf{A}$

- On the one hand, we can see that

$$\det(\mathbf{A}) = w_1 A_{11} + w_2 A_{12} + w_3 A_{13} = \mathbf{w}^T (\mathbf{x} \times \mathbf{y})$$

and $\mathbf{x}^T (\mathbf{x} \times \mathbf{y}) = \mathbf{y}^T (\mathbf{x} \times \mathbf{y}) = 0$ Why ?

Therefore, $\mathbf{x} \times \mathbf{y}$ is *orthogonal* to both $\mathbf{x}$ and $\mathbf{y}$

- If $\mathbf{A}$ is any matrix of the form,

$$\mathbf{A} = \begin{bmatrix} w_1 & x_1 & y_1 \\ w_2 & x_2 & y_2 \\ w_3 & x_3 & y_3 \end{bmatrix}$$

$$\mathbf{x} \times \mathbf{y} = \begin{bmatrix} A_{11} \\ A_{21} \\ A_{31} \end{bmatrix} = \begin{bmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{bmatrix} \equiv \begin{vmatrix} \mathbf{e}_1 & x_1 & y_1 \\ \mathbf{e}_2 & x_2 & y_2 \\ \mathbf{e}_3 & x_3 & y_3 \end{vmatrix}$$

where A_{11}, A_{21}, and A_{31} are *cofactors* of $\mathbf{A}$

(*Ex*.2) Find the equation of the plane that passes through the points

$$P_1 = (1,1,2),\ \ P_2 = (2,3,3),\ \ P_3 = (3,-3,3)$$

(*Solution*)

$$\mathbf{x} = \overrightarrow{P_1P_2} = (2,3,3) - (1,1,2) = (1,2,1)$$

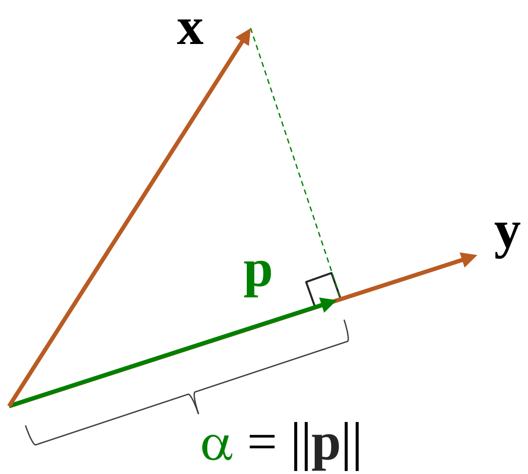
$$\mathbf{y} = \overrightarrow{P_1P_3} = (3,-3,3) - (1,1,2) = (2,-4,1)$$

$$\mathbf{n} = \mathbf{x} \times \mathbf{y} = (6,1,-8)$$

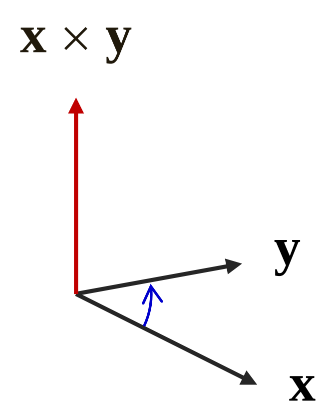
$$\therefore 6(x-1) + (y-1) - 8(z-2) = 0$$

Summary 1

1. Def. of Scalar Product $\mathbf{x}^T\mathbf{y}$
2. Vector Length, Distance, C.S. Inequality
3. Orthogonal Vector $\mathbf{x}^T\mathbf{y} = 0$
4. Scalar and Vector Projections
5. Cross Product



$$\mathbf{x}\times\mathbf{y} \equiv \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix}$$



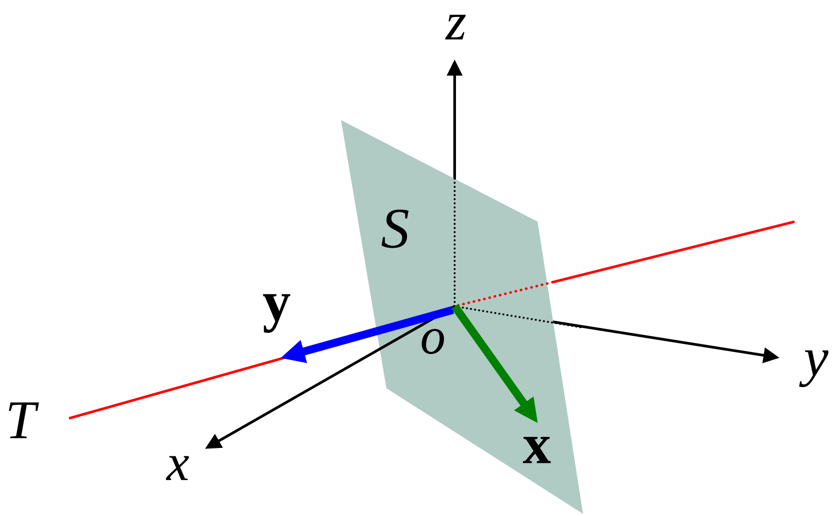
Orthogonal Subspaces

To define orthogonality between two subspaces

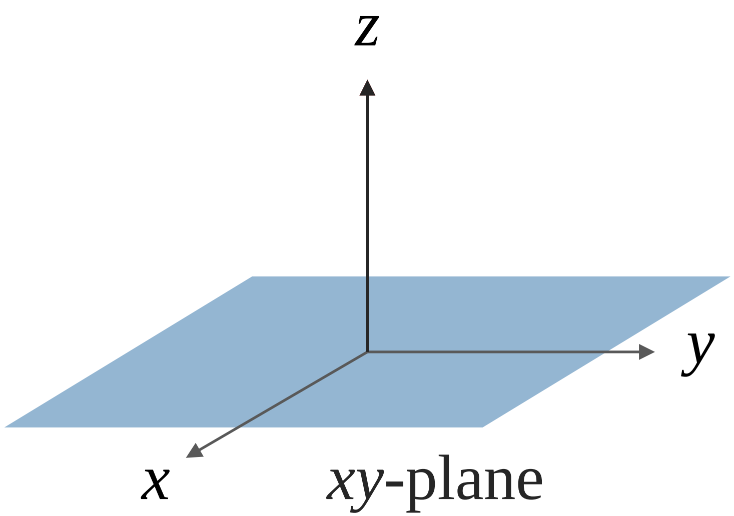
To understand subspaces of $\mathbf{R}^n$, $\mathrm{Null}(\mathbf{A})$, $\mathrm{Col}(\mathbf{A})$, and their relationships

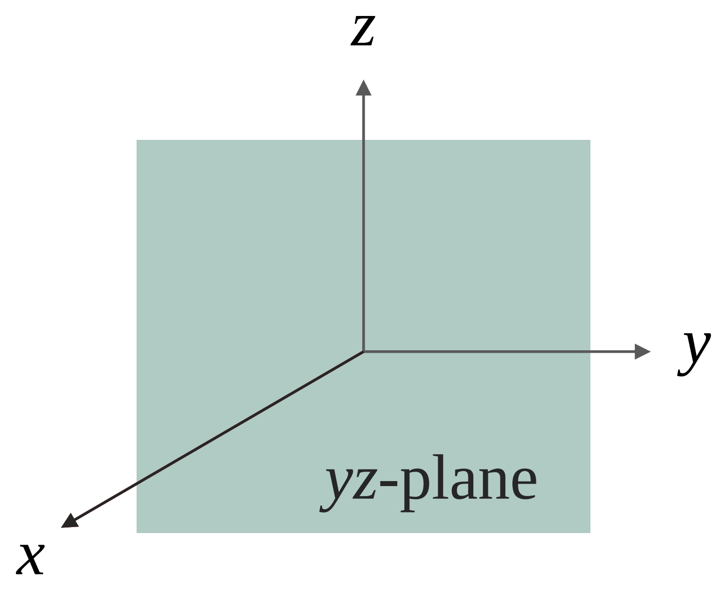
Def. Orthogonal Subspace

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



(*Q*) Orthogonal ? Span($\mathbf{e}_1$,$\mathbf{e}_2$) and Span($\mathbf{e}_2$,$\mathbf{e}_3$)





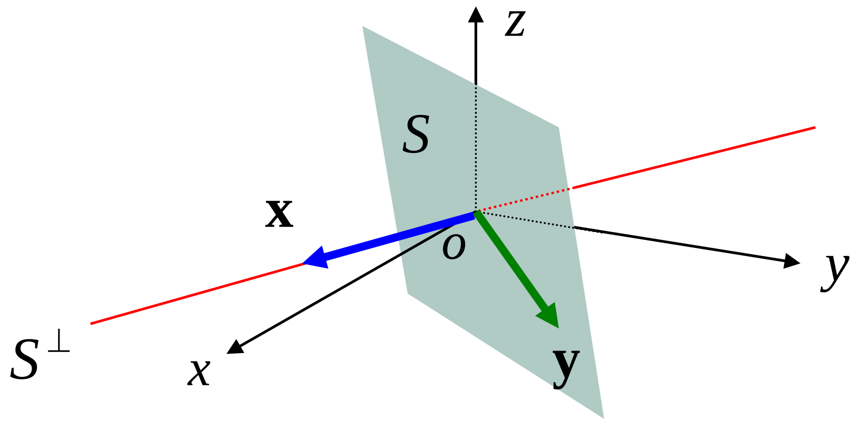
Def. Orthogonal Complement

Let S be a subspace of $\mathbf{R}^n$. The set of all vectors in $\mathbf{R}^n$ that are orthogonal to every vector in S will be denoted $S^\perp$.

$$S^\perp = \{ \mathbf{x} \in \mathbf{R}^n \mid \mathbf{x}^T\mathbf{y} = 0 \text{ for every } \mathbf{y} \in S \}$$

The set $S^\perp$ is called the *orthogonal complement* of S

$S^\perp$: read "S perp."



Remark

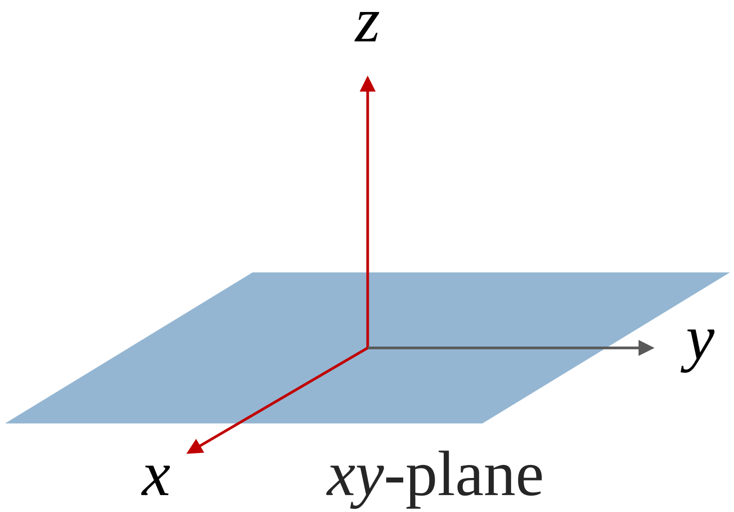
- If S and T are orthogonal subspaces of $\mathbf{R}^n$, then $S \cap T = \{\mathbf{0}\}$
  - If $\mathbf{x} \in S \cap T$ and $S \perp T$, then $\|\mathbf{x}\|^2 = \mathbf{x}^T\mathbf{x} = 0$ and hence $\mathbf{x} = \mathbf{0}$

- If S is a subspace of $\mathbf{R}^n$, then $S^{\perp}$ is also its subspace
  - For any $\mathbf{u}, \mathbf{v} \in S^{\perp}$, $\mathbf{y} \in S$, and a scalar c,

$$(c\mathbf{u})^T \mathbf{y} = c(\mathbf{u}^T \mathbf{y}) = c0 = 0$$

$$(\mathbf{u} + \mathbf{v})^T \mathbf{y} = \mathbf{u}^T \mathbf{y} + \mathbf{v}^T \mathbf{y} = 0 + 0 = 0$$

(Q) Orthogonal Complement ? $\mathrm{Span}(\mathbf{e}_3)^{\perp}$ *vs*. $\mathrm{Span}(\mathbf{e}_1)$



Fundamental Subspaces

Let $\mathbf{A}$ be an $m \times n$ matrix. A vector $\mathbf{b} \in \mathbf{R}^m$ is in $\mathrm{Col}(\mathbf{A})$ if and only if $\mathbf{b} = \mathbf{A}\mathbf{x}$ for some $\mathbf{x} \in \mathbf{R}^n$

- If we think of $\mathbf{A}$ as a linear transformation mapping $\mathbf{R}^n$ into $\mathbf{R}^m$, then $\mathrm{Col}(\mathbf{A})$ is the same as the range of $\mathbf{A}$, denoted $R(\mathbf{A})$

$$R(\mathbf{A}) = \{\ \mathbf{b} \in \mathbf{R}^m \ |\ \mathbf{b} = \mathbf{A}\mathbf{x} \text{ for some } \mathbf{x} \in \mathbf{R}^n\ \} = \mathrm{Col}(\mathbf{A})$$

- The column space of $\mathbf{A}^T$, $R(\mathbf{A}^T)$, is a subspace of $\mathbf{R}^n$

$$R(\mathbf{A}^T) = \{\ \mathbf{y} \in \mathbf{R}^n \ |\ \mathbf{y} = \mathbf{A}^T\mathbf{x} \text{ for some } \mathbf{x} \in \mathbf{R}^m\ \}$$

□ Th. 5.2.1 *Fundamental Subspaces Theorem*

If $\mathbf{A}$ be an $m \times n$ matrix, then $\mathrm{Null}(\mathbf{A}) = R(\mathbf{A}^T)^{\perp}$ and

$\mathrm{Null}(\mathbf{A}^T) = R(\mathbf{A})^{\perp}$

(*Proof*)

An n-vector $\mathbf{x} \in \mathrm{Null}(\mathbf{A}) \;\Rightarrow\; \mathbf{Ax} = \mathbf{0}$

$\Rightarrow\; \mathbf{c}^T\mathbf{x} = 0$ for each column vector $\mathbf{c}$ of $\mathbf{A}^T$

$\Rightarrow\; \mathbf{x} \in \mathrm{Col}(\mathbf{A}^T)^{\perp}$

$\Rightarrow\; \mathrm{Null}(\mathbf{A}) \subseteq \mathrm{Col}(\mathbf{A}^T)^{\perp}$

$\Rightarrow\; \mathrm{Null}(\mathbf{A}) \subseteq R(\mathbf{A}^T)^{\perp}$

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

Conversely, an n-vector $\mathbf{x} \in R(\mathbf{A}^T)^\perp$

$\Rightarrow$ $\mathbf{x}$ is orthogonal to each column vector of $\mathbf{A}^T$

$\Rightarrow$ $\mathbf{Ax} = \mathbf{0}$

$\Rightarrow$ $\mathbf{x} \in \mathrm{Null}(\mathbf{A})$

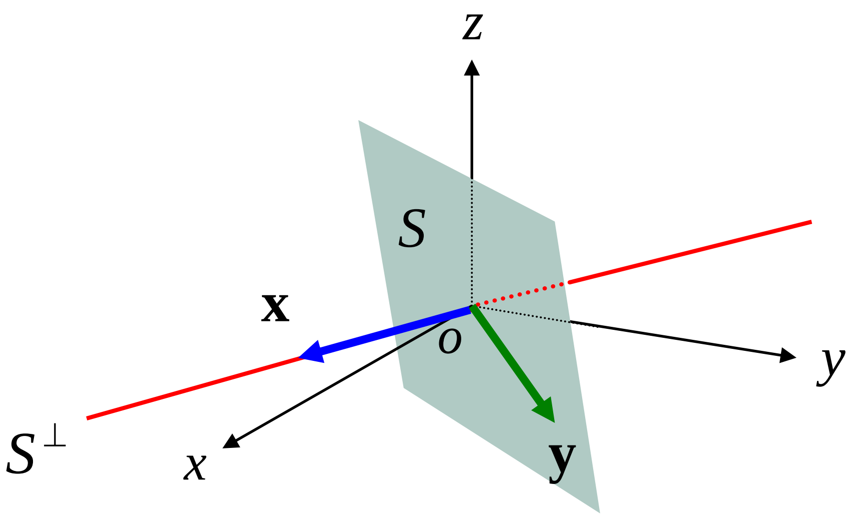
$\Rightarrow$ $R(\mathbf{A}^T)^\perp \subseteq \mathrm{Null}(\mathbf{A})$

Therefore, $\mathrm{Null}(\mathbf{A}) = R(\mathbf{A}^T)^\perp$

Since the above proof is not dependent on the dimension of $\mathbf{A}$, the result hold for the matrix $\mathbf{B} = \mathbf{A}^T$. That is, $\mathrm{Null}(\mathbf{B}) = R(\mathbf{B}^T)^\perp$. Therefore, $\mathrm{Null}(\mathbf{A}^T) = R(\mathbf{A})^\perp$

- Th. 5.2.2

  If S is a subspace of $\mathbf{R}^n$, then $\dim(S) + \dim(S^\perp) = n$.
  Furthermore, if $\{\mathbf{u}_1,\ldots,\mathbf{u}_r\}$ is a basis for S and $\{\mathbf{v}_1,\ldots,\mathbf{v}_{n\text{-}r}\}$ is a basis for $S^\perp$, then $\{\mathbf{u}_1,\ldots,\mathbf{u}_r ,\mathbf{v}_1,\ldots,\mathbf{v}_{n\text{-}r}\}$ is a basis for $\mathbf{R}^n$



(*Ex*.3) Find the bases for $\mathrm{Null}(\mathbf{A})$, $R(\mathbf{A}^T)$, $\mathrm{Null}(\mathbf{A}^T)$, and $R(\mathbf{A})$

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 1 & 3 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 2 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{U}$$

(*Solution*)

The first two column vectors of $\mathbf{A}$ are pivot columns, hence $\{(1,0,1), (1,1,3)\}$ is a basis of $R(\mathbf{A}) = \mathrm{Col}(\mathbf{A})$

The nonzero rows in $\mathbf{U}$ form the row space of $\mathbf{A}$, $\mathrm{Row}(\mathbf{A})$. Thus $\{(1,0,1), (0,1,1)\}$ is a basis of $R(\mathbf{A}^T) = \mathrm{Col}(\mathbf{A}^T)$

From $\mathbf{U}$, we can determine that $\mathrm{Null}(\mathbf{A}) = \mathrm{Span}((-1,-1,1))$. Thus $\{(-1,-1,1)\}$ is a basis of $\mathrm{Null}(\mathbf{A})$

$$\mathbf{A}^T = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 3 \\ 2 & 1 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 1 & 2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \end{bmatrix} = \mathbf{T}$$

From $\mathbf{T}$, we can determine that $\mathrm{Null}(\mathbf{A}^T) = \mathrm{Span}((-1,-2,1))$. Thus $\{(-1,-2,1)\}$ is a basis of $\mathrm{Null}(\mathbf{A}^T)$

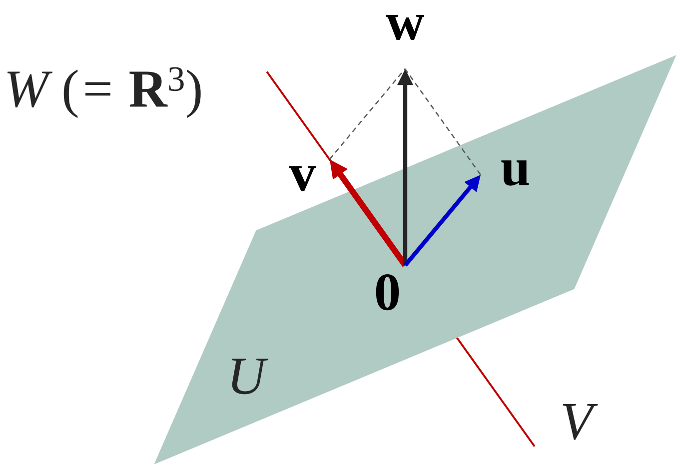
On the one hand, $\mathrm{Null}(\mathbf{A}^T)$ $(= R(\mathbf{A})^\perp)$ is a set of all the vectors that are orthogonal to each vector of the basis of $R(\mathbf{A})$.

$$\mathbf{Bx} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 3 \end{bmatrix} \mathbf{x} = \mathbf{0} \qquad \mathbf{B} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 3 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

- Note that $\{(1,0,1), (1,1,3), (-1,-2,1)\}$ is a basis of $\mathbf{R}^3$

Def. Direct Sum of Subspaces

If U and V are subspaces of a vector space W and each $\mathbf{w} \in W$ can be written uniquely as a sum of $\mathbf{u} + \mathbf{v}$, where $\mathbf{u} \in U$ and $\mathbf{v} \in V$, then we say that W is a *direct sum* of U and V, and we write $W = U \oplus V$

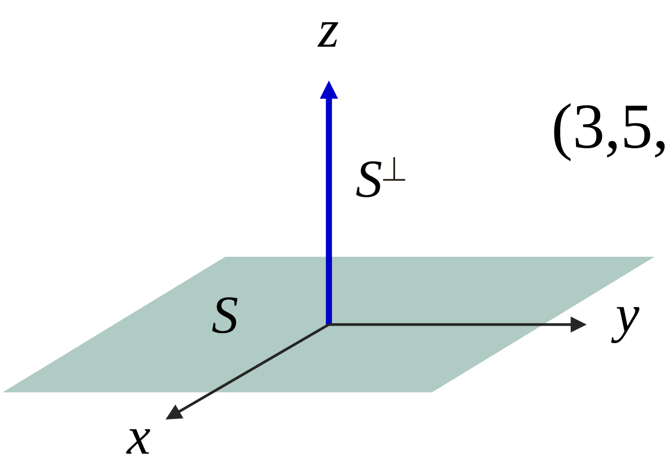


- Th. 5.2.3

  If S is a subspace of $\mathbf{R}^n$, then $\mathbf{R}^n = S \oplus S^{\perp}$

- Th. 5.2.4

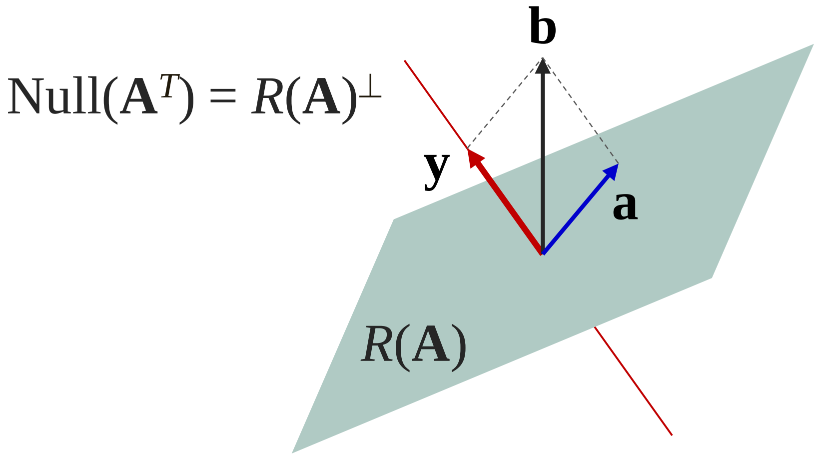
  If S is a subspace of $\mathbf{R}^n$, then $(S^{\perp})^{\perp} = S$



$$(3,5,7) = (3,5,0) + (0,0,7)$$

Corollary 5.2.5

If $\mathbf{A}$ is an $m\times n$ matrix and $\mathbf{b} \in \mathbf{R}^m$, then either there is a vector $\mathbf{x} \in \mathbf{R}^n$ such that $\mathbf{Ax} = \mathbf{b}$ or there is a vector $\mathbf{y} \in \mathbf{R}^m$ such that $\mathbf{A}^T\mathbf{y} = \mathbf{0}$ and $\mathbf{y}^T\mathbf{b} \neq 0$



$\mathbf{R}^m = R(\mathbf{A}) \oplus R(\mathbf{A})^\perp$

$\mathbf{b} = \mathbf{a} + \mathbf{y}$

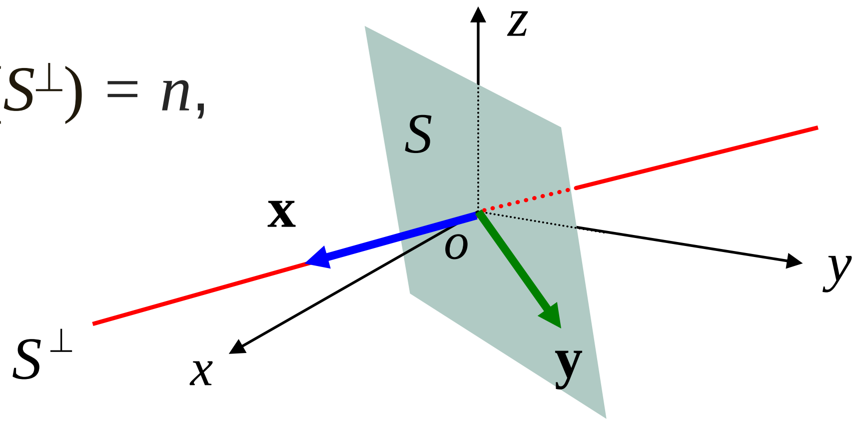
Summary 2

1. Def. of Orthogonal Subspaces $S \perp T$
2. Orthogonal Complement $S^{\perp}$
3. Fundamental Subspaces

   $\mathrm{Col}(\mathbf{A}) \equiv R(\mathbf{A}), \mathrm{Null}(\mathbf{A}) = R(\mathbf{A}^T)^{\perp}, \mathrm{Null}(\mathbf{A}^T) = R(\mathbf{A})^{\perp}$
4. Direct Sum $W = U \oplus V$

   $\mathbf{R}^n = S \oplus S^{\perp}, \dim(S) + \dim(S^{\perp}) = n,$

   $(S^{\perp})^{\perp} = S$



Least Squares Problems

To know how to solve basic least squares problems

1. Scalar Product in $\mathbf{R}^n$
2. Orthogonal Subspaces
3. Least Squares Problem
4. Inner Product Spaces
5. Orthonormal Sets
6. Gram-Schmidt Orthogonalization Process

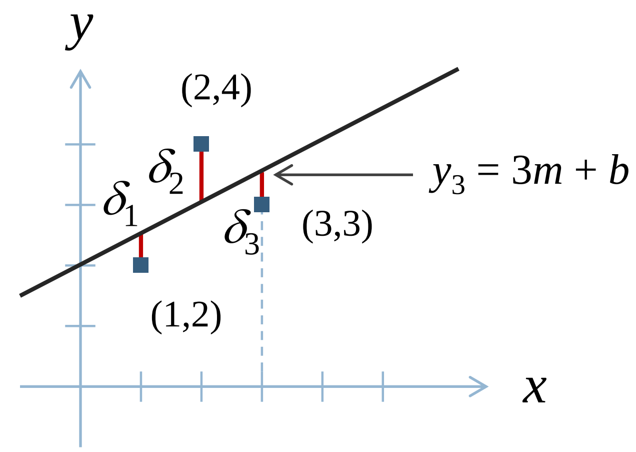
Determine a line passing through three points

Line eq. : $y = mx + b$

$$2 = m + b$$
$$4 = 2m + b$$
$$3 = 3m + b$$

$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix}$$

Inconsistent !!!



$(\delta_1, \delta_2, \delta_3) = (2\text{-}y_1, 4\text{-}y_2, 3\text{-}y_3)$

Best fit line ?

The line that minimizes the sum of squares of y-directional error, $\|\Delta\|^2 = \|(\delta_1, \delta_2, \delta_3)\|^2 = \delta_1^{\,2} + \delta_2^{\,2} + \delta_3^{\,2}$

Least Squares Solutions

Suppose we have an inconsistent overdetermined linear system $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A}$ is an $m \times n$ matrix and $m > n$

We cannot expect in general to find $\mathbf{Ax} = \mathbf{b}$ for any n-vector $\mathbf{x}$, so we can form a *residual* $r(\mathbf{x}) = \mathbf{b} - \mathbf{Ax}$

/ rizídʒuəl /
오차, 나머지

The distance between $\mathbf{b}$ and $\mathbf{Ax}$ is given by

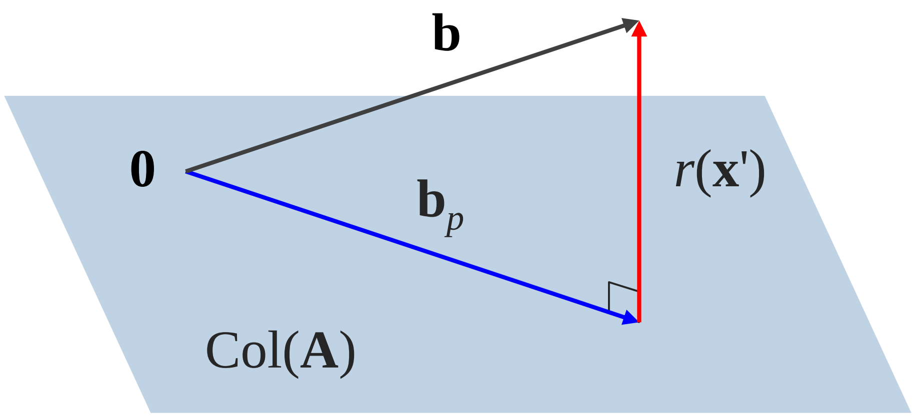
$$||r(\mathbf{x})|| = ||\mathbf{b} - \mathbf{Ax}||$$

Least Squares Problem

Find the solution $\mathbf{x}'$ such that $||\mathbf{b} - \mathbf{Ax}'||$ is minimum

$\|\mathbf{b} - \mathbf{A}\mathbf{x}'\|$ is minimum only when $\mathbf{A}\mathbf{x}'$ is the orthogonal projection of $\mathbf{b}$ onto $\mathrm{Col}(\mathbf{A})$, $\mathbf{b}_p$. Hence,

$$\|r(\mathbf{x}')\| \text{ is minimum} \iff \mathbf{A}\mathbf{x}' = \mathbf{b}_p \iff \mathbf{b} - \mathbf{A}\mathbf{x}' = r(\mathbf{x}')$$

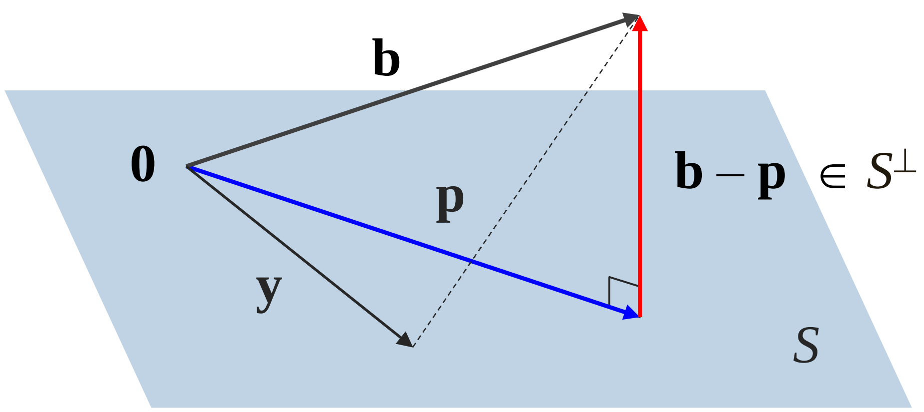


Th. 5.3.1

Let S is a subspace of $\mathbf{R}^m$. For each $\mathbf{b} \in \mathbf{R}^m$, there is a unique element $\mathbf{p}$ of S that is closest to $\mathbf{b}$, that is,

$$\|\mathbf{b}-\mathbf{y}\| > \|\mathbf{b}-\mathbf{p}\|$$

for any $\mathbf{y} \neq \mathbf{p}$ in S. Furthermore, $\mathbf{p}$ in S is closest to a given vector $\mathbf{b}$ if and only if $\mathbf{b}-\mathbf{p} \in S^{\perp}$



(*Proof*)

Since $\mathbf{R}^m = S \oplus S^\perp$, each element $\mathbf{b}$ in $\mathbf{R}^m$ can be expressed *uniquely* as a sum $\mathbf{b} = \mathbf{p} + \mathbf{z}$, where $\mathbf{p} \in S$ and $\mathbf{z} \in S^\perp$.

If $\mathbf{y}$ is any other element of S, then

$$\|\mathbf{b} - \mathbf{y}\|^2 = \|(\mathbf{b} - \mathbf{p}) + (\mathbf{p} - \mathbf{y})\|^2$$

Since $\mathbf{p} - \mathbf{y} \in S$, $\mathbf{b} - \mathbf{p} \in S^\perp$, and $(\mathbf{p} - \mathbf{y})^T(\mathbf{b} - \mathbf{p}) = 0$, it follows from the Pythagorean law that

$$\|\mathbf{b} - \mathbf{y}\|^2 = \|\mathbf{b} - \mathbf{p}\|^2 + \|\mathbf{p} - \mathbf{y}\|^2$$

Therefore, $\|\mathbf{b} - \mathbf{y}\| > \|\mathbf{b} - \mathbf{p}\|$. Thus, if $\mathbf{p} \in S$ and $\mathbf{b} - \mathbf{p} \in S^\perp$, $\mathbf{p}$ is closest to $\mathbf{b}$. Conversely, …

check

- How to find the least squares solution of $\mathbf{Ax} = \mathbf{b}$
  - A vector $\mathbf{x}'$ will be a solution of the least squares problem $\mathbf{Ax} = \mathbf{b}$ if and only if $\mathbf{p} = \mathbf{Ax}'$ is the vector in $\mathrm{Col}(\mathbf{A})$, or $R(\mathbf{A})$
  - Since the residual $r(\mathbf{x}') = \mathbf{b} - \mathbf{p} = \mathbf{b} - \mathbf{Ax}' \in R(\mathbf{A})^{\perp}$ and $R(\mathbf{A})^{\perp} = \mathrm{Null}(\mathbf{A}^T)$,

$$\mathbf{0} = \mathbf{A}^T r(\mathbf{x}') = \mathbf{A}^T(\mathbf{b} - \mathbf{Ax}')$$

  - Thus, to solve the least squares problem $\mathbf{Ax} = \mathbf{b}$, we must solve the *normal equations*

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

□ Th. 5.3.2

If $\mathbf{A}$ is an $m \times n$ matrix of *rank* n, the normal equations

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

have a *unique* solution

$$\mathbf{x}' = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

and $\mathbf{x}'$ is the unique least squares solution of the system $\mathbf{A}\mathbf{x} = \mathbf{b}$

- The projection vector $\mathbf{p} = \mathbf{A}\mathbf{x}' = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$ is in $R(\mathbf{A})$ and $\mathbf{P} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is called the *projection matrix*

(*Ex.*4)

Solve the least squares problem and compute the least squares error for the system $\mathbf{Ax} = \mathbf{b}$

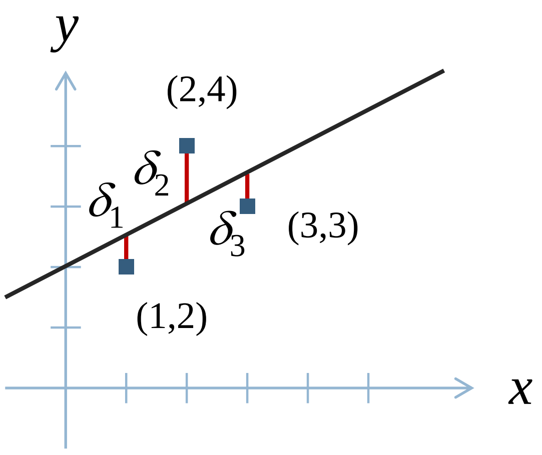
$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix}$$

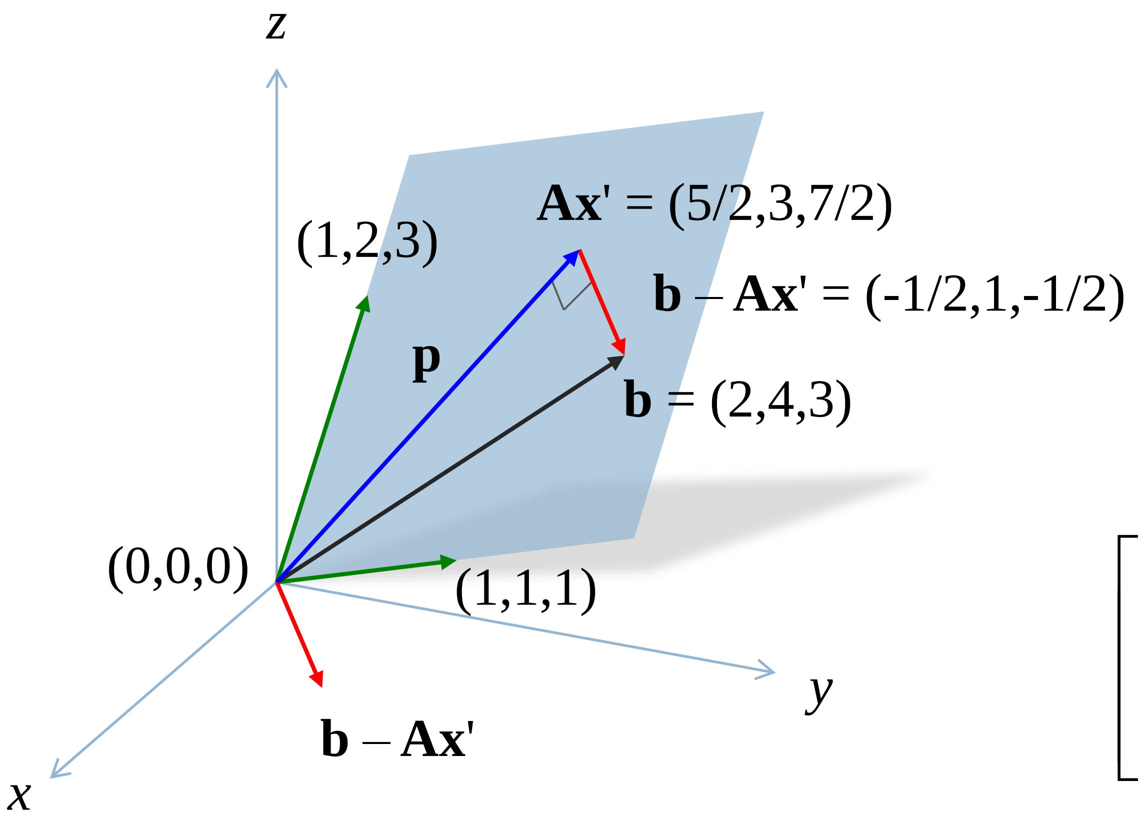
(*Solution*)

$$\begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix}$$

$$\begin{bmatrix} 14 & 6 \\ 6 & 3 \end{bmatrix} \mathbf{x} = \begin{bmatrix} 19 \\ 9 \end{bmatrix} \quad \therefore \mathbf{x}' = \begin{bmatrix} 1/2 \\ 2 \end{bmatrix}$$

$$\left\| \mathbf{b} - \mathbf{Ax}' \right\| = \frac{\sqrt{6}}{2}$$





$$\begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} m \\ b \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix}$$

Inner Product Spaces

To know the definition and basic properties of inner products

To find more properties of inner products analogous to those for dot products

4. Inner Product Spaces
5. Orthonoraml Sets
6. Gram-Schmidt Orthogonalization Process

Introduction

- Inner Product ?
  - An extension of the scalar product in $\mathbf{R}^n$ to an scalar mapping of vectors in a vector space V
  - Thus the scalar product is an inner product

Def. Inner Product

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 $<\mathbf{x},\mathbf{y}>$ satisfying the following conditions

1. $<\mathbf{x},\mathbf{x}> \ \geq \ 0$ with equality if and only if $\mathbf{x}=\mathbf{0}$ *positivity*
2. $<\mathbf{x},\mathbf{y}> \ = \ <\mathbf{y},\mathbf{x}>$ for all $\mathbf{x}$ and $\mathbf{y}$ in V *symmetry*
3. $<a\mathbf{x}+b\mathbf{y},\mathbf{z}> \ = \ a<\mathbf{x},\mathbf{z}>+\,b<\mathbf{y},\mathbf{z}>$ for any vectors $\mathbf{x},\mathbf{y},\mathbf{z}$ and scalars a and b *linearity*

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

(*Examples*)

Weighted Dot Product of n - vectors

$$\mathbf{u} = (u_1, \ldots, u_n), \mathbf{v} = (v_1, \ldots, v_n), \text{ positive } c_1, \ldots, c_n \in \mathbf{R}$$

$$< \mathbf{u}, \mathbf{v} > = \mathbf{u}^T \mathbf{W} \mathbf{v} = c_1 u_1 v_1 + \ldots + c_n u_n v_n$$

where $\mathbf{W} = \begin{bmatrix} c_1 & .. & 0 \\ : & .. & : \\ 0 & .. & c_n \end{bmatrix}$

$$p(x) = a_0 + a_1 x + \ldots + a_n x^n, \quad q(x) = b_0 + b_1 x + \ldots + b_n x^n$$

$$< p, q > = a_0 b_0 + a_1 b_1 + \ldots + a_n b_n$$

(*Examples*)

$$\mathbf{A}=\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix},\ \ \mathbf{B}=\begin{bmatrix} b_1 & b_2 \\ b_3 & b_4 \end{bmatrix}$$

$$<\mathbf{A},\mathbf{B}> = a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4$$

$$\|\mathbf{A}\|_F = \sqrt{<\mathbf{A},\mathbf{A}>} = \sqrt{a_1^2+a_2^2+a_3^2+a_4^2}$$ *Frobenius norm*

$$p(x) = a_0 + a_1x + \ldots + a_nx^n, \quad q(x) = b_0 + b_1x + \ldots + b_nx^n$$

Let $r_0, r_1, \ldots, r_n$ be $n+1$ distinct real numbers.

$$< p,q > = p(r_0)q(r_0) + p(r_1)q(r_1) + \ldots + p(r_n)q(r_n)$$

(Another *Example*)

Let $f(x)$ and $g(x)$ be in $C[a,b]$, the vector space of the continuous real - valued functions defined on $[a,b]$.

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

$$Positivity\ :\ < f, f > = \int_a^b (f(x))^2 dx \ \geq 0$$

$$Symmetry\ :\ < f, g > = \int_a^b f(x)g(x)dx = \int_a^b g(x)f(x)dx = < g, f >$$

$$Linearity\ :\ < cf + dg, h > = \int_a^b \{cf(x) + dg(x)\}h(x)dx$$

$$= c\int_a^b f(x)h(x)dx + d\int_a^b g(x)h(x)dx$$

$$= c < f, h > + d < g, h >$$

Def. Orthogonal, Norm, Distance

Let V be an inner product space. Two vectors $\mathbf{u}$ and $\mathbf{v}$ are called *orthogonal* if their inner product is zero

$$\mathbf{u} \text{ and } \mathbf{v} \text{ are orthogonal if } <\mathbf{u},\mathbf{v}> = 0$$

The *length* (or *magnitude*, or *norm*) of $\mathbf{v}$ is the nonnegative number $||\mathbf{v}||$ defined by

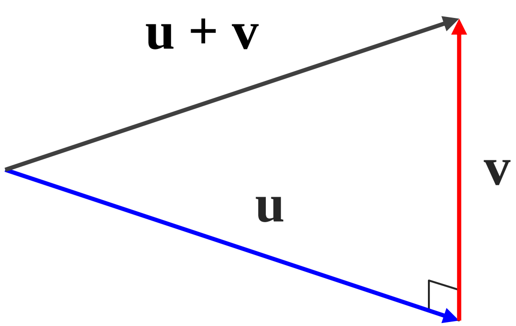
$$||\mathbf{v}|| = \sqrt{<\mathbf{v},\mathbf{v}>}$$

The *distance* between two vectors $\mathbf{v}$ and $\mathbf{w}$ is defined to be the number $||\mathbf{v}-\mathbf{w}||$

- Th. 5.4.1 *Pythagorean Law*

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

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



/ piθǣgərí:ən / , / piθǽgərəs /

(*Ex*.5) For the vector space $C[-\pi,\pi]$ of real-valued functions defined on $[-\pi,\pi]$, the inner product is defined by

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

Compute $\|\cos x + \sin x\|$

(*Solution*)

$$< \cos x, \sin x > = (1/\pi)\int_{-\pi}^{\pi} \cos x \sin x dx = (1/2\pi)\int_{-\pi}^{\pi} \sin 2x dx = 0$$

$$\therefore \cos x, \sin x \ \textit{are orthogonal}$$

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

$$= (1/\pi)\int_{-\pi}^{\pi} 1 dx = 2 \qquad \therefore \|\cos x + \sin x\| = \sqrt{2}$$

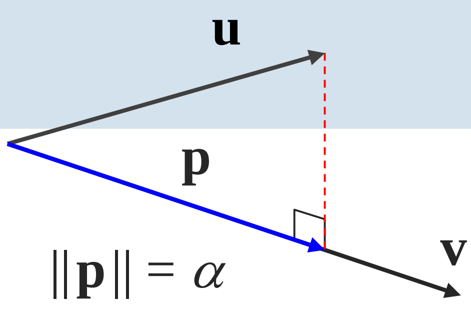
Def. Scalar and Vector Projection

If $\mathbf{u}$ and $\mathbf{v}$ are vectors in an inner product space V and $\mathbf{v} \neq \mathbf{0}$, then the scalar projection of $\mathbf{u}$ onto $\mathbf{v}$ is defined by

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

and the vector projection of $\mathbf{u}$ onto $\mathbf{v}$ is defined by

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



- Observations

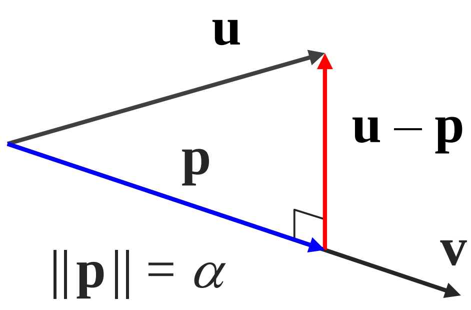
If $\mathbf{v} \neq \mathbf{0}$ and $\mathbf{p}$ is the vector projection of $\mathbf{u}$ onto $\mathbf{v}$, then

1. $\mathbf{u} - \mathbf{p}$ and $\mathbf{p}$ are orthogonal
2. $\mathbf{u} = \mathbf{p}$ if and only if $\mathbf{u}$ is a scalar multiple of $\mathbf{v}$

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

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

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



□ Th. 5.4.2 *Cauchy-Schwartz Inequality*

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

$$|<\mathbf{u},\mathbf{v}>| \le \|\mathbf{u}\|\,\|\mathbf{v}\|$$

Equality holds if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent

(*Proof*)

If $\mathbf{v} = \mathbf{0}$, then $| < \mathbf{u},\mathbf{v} > | = 0 = \|\mathbf{u}\|\|\mathbf{v}\|$

If $\mathbf{v} \neq \mathbf{0}$, then let $\mathbf{p}$ be the vector projection of $\mathbf{u}$ onto $\mathbf{v}$.

From the Pythagorean law, $\|\mathbf{p}\|^2 + \|\mathbf{u}-\mathbf{p}\|^2 = \|\mathbf{u}\|^2$

Thus, $$\frac{(<\mathbf{u},\mathbf{v}>)^2}{\|\mathbf{v}\|^2} = \|\mathbf{p}\|^2 = \|\mathbf{u}\|^2 - \|\mathbf{u}-\mathbf{p}\|^2$$

and hence

$$(<\mathbf{u},\mathbf{v}>)^2 = \|\mathbf{u}\|^2\|\mathbf{v}\|^2 - \|\mathbf{u}-\mathbf{p}\|^2\|\mathbf{v}\|^2 \le \|\mathbf{u}\|^2\|\mathbf{v}\|^2$$

Therefore, $|<\mathbf{u},\mathbf{v}>| \le \|\mathbf{u}\|\|\mathbf{v}\|$

Equality holds in the given inequality if and only if $\mathbf{u} = \mathbf{p}$ or $\mathbf{v} = \mathbf{0}$. More simply stated, equality will hold if and only if $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent

Def. Normed Linear Space

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

1. $\|\mathbf{v}\| \geq 0$ with equality if an only if $\mathbf{v} = \mathbf{0}$; Positivity
2. $\|\alpha\mathbf{v}\| = |\alpha|\|\mathbf{v}\|$ for any scalar α ; Scalability
3. $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ for all $\mathbf{v}, \mathbf{w} \in V$; Triangle Inequality

- Th. 5.4.3

If V is an inner product space, then the equation

$$\|\mathbf{v}\| = \sqrt{<\mathbf{v}, \mathbf{v}>} \quad \text{for all } \mathbf{v} \in V$$

defines a norm on V

(*Proof*)

The first condition is satisfied from the positivity condition in inner product space

For any scalar α, $\|\alpha\mathbf{v}\|^2 = <\alpha\mathbf{v},\alpha\mathbf{v}> = \alpha <\mathbf{v},\alpha\mathbf{v}> = \alpha^2 <\mathbf{v},\mathbf{v}>$. Thus $\|\alpha\mathbf{v}\| = |\alpha|\|\mathbf{v}\|$ and the second condition is satisfied

$$
\begin{aligned}
\|\mathbf{v} + \mathbf{w}\|^2 &= <\mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w}> \\
&= <\mathbf{v},\mathbf{v}> + 2<\mathbf{v},\mathbf{w}> + <\mathbf{w},\mathbf{w}> \\
&\le \|\mathbf{v}\|^2 + 2\,\|\mathbf{v}\|\|\mathbf{w}\| + \|\mathbf{w}\|^2 \\
&= (\,\|\mathbf{v}\| + \|\mathbf{w}\|\,)^2
\end{aligned}
$$

Thus, the triangle inequality condition, $\|\mathbf{v} + \mathbf{w}\| \le \|\mathbf{v}\| + \|\mathbf{w}\|$, is also satisfied

(Note)

- It is possible to define many different norms on a given vector space.

  For example, for a n-vector $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{R}^n$,

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i| \quad \text{or} \quad \|\mathbf{x}\|_\infty = \max_{1 \le i \le n} |x_i|$$

?

- However, in a case of a norm that is not derived from an inner product, the Pythagorean law will not hold

  For orthogoanl vectors, $\mathbf{x}_1 = (1,2)$ and $\mathbf{x}_2 = (-4,2)$,

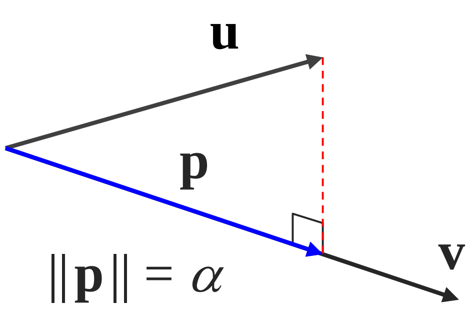
$$\|\mathbf{x}_1\|_1^2 + \|\mathbf{x}_2\|_1^2 = 9 + 36 \neq 49 = \|\mathbf{x}_1 + \mathbf{x}_2\|_1^2$$

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

Summary 3

1. Inner Product Space: *Positivity*, *Symmetry*, *Linearity*
2. Orthogonal, Norm, Distance in Inner P.S.
3. Pythagorean Law, Scalar and Vector Projection, C-S Inequality
4. A Normed Linear Space:

   Positivity, *Scalability*, *Triangle Inequality*



Orthonormal Sets

To know the definitions and basic properties of orthogonal and orthonormal sets

To show usefulness of a basis of mutually orthogonal unit vectors

4. Inner Product Spaces
5. Orthonormal Sets
6. Gram-Schmidt Orthogonalization Process

Def. Orthogonal Set

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

An orthogonal set of unit vectors is said to be an *orthonormal set*

$$<\mathbf{x}, \mathbf{y}> = \mathbf{x}^T \mathbf{W} \mathbf{y} = (1/6)\mathbf{x}^T \mathbf{y} \text{ in } \mathbf{R}^4$$

$\{(1,1,2,0), (0,2,-1,1), (-2,0,1,1)\}$ a orthonormal set ?

□ Th. 5.5.1

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

(*Proof*)

Suppose that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \Lambda + c_n\mathbf{v}_n = \mathbf{0}$

If $1 \le j \le n$, then, taking the inner product of $\mathbf{v}_j$ with both sides of the above equation, we see that

$$c_1 < \mathbf{v}_1, \mathbf{v}_j > + \Lambda + c_n < \mathbf{v}_n, \mathbf{v}_j > = c_j \| \mathbf{v}_j \|^2 = 0$$

Since all scalars are 0, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent

(*Ex*.6) Show that in $C[-\pi,\pi]$ with inner product

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

the set $\{1, \cos x, \cos 2x, \ldots, \cos nx\}$ is an orthogonal set

(*Solution*)

For positive integers j and k,

$$< 1, \cos kx > = (1/\pi)\int_{-\pi}^{\pi} \cos kx\, dx = (1/k\pi)\sin kx \Big|_{-\pi}^{\pi} = 0 \quad k = 1, \ldots, n$$

$$< \cos jx, \cos kx > = (1/\pi)\int_{-\pi}^{\pi} \cos jx \cos kx\, dx$$

$$= (1/2\pi)\int_{-\pi}^{\pi} \{\cos(j+k)x + \cos(j-k)x\}dx = 0 \quad (j \neq k)$$

$\therefore \{1, \cos x, \cos 2x, \ldots, \cos nx\}$ is orthogonal

On the one hand, $\cos x$, $\cos 2x$, …, $\cos nx$ are already unit vectors, since for $k = 1, 2, \ldots, n$

$$< \cos kx, \cos kx > = (1/\pi)\int_{-\pi}^{\pi} \cos^2 kx\,dx = (1/2\pi)\int_{-\pi}^{\pi} \{1 + \cos 2kx\}dx$$

$$= (1/2\pi)\,\{x + (1/2k)\sin 2kx\}\Big|_{-\pi}^{\pi} = (1/2\pi)\,\{2\pi + 0\} = 1$$

To form an orthonormal set, we need only find a unit vector in the direction of the constant function 1,

$$\|1\|^2 = <1, 1> = (1/\pi)\int_{-\pi}^{\pi} 1\,dx = 2$$

Thus, $\{1/\sqrt{2}, \cos x, \cos 2x, \ldots, \cos nx\}$ is an orthonormal set

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

Th. 5.5.2

If $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ is an orthonormal basis for an inner product space V. If $\mathbf{v} = c_1\mathbf{u}_1 + \ldots + c_n\mathbf{u}_n$, then $c_i = <\mathbf{v}, \mathbf{u}_i>$

(*Proof*)

$$<\mathbf{v},\mathbf{u}_i> = <\sum_{j=1}^{n} c_j\mathbf{u}_j, \mathbf{u}_i> = \sum_{j=1}^{n} c_j <\mathbf{u}_j, \mathbf{u}_i> = \sum_{j=1}^{n} c_j \delta_{ji} = c_i$$

Corollary 5.5.3

If $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for an inner product space V. If $\mathbf{v} = a_1\mathbf{u}_1 + \ldots + a_n\mathbf{u}_n$ and $\mathbf{w} = b_1\mathbf{u}_1 + \ldots + b_n\mathbf{u}_n$, then $<\mathbf{v}, \mathbf{w}> = a_1 b_1 + \ldots + a_n b_n$

(*Proof*)

From Th. 5.5.2, $<\mathbf{w}, \mathbf{u}_i> = b_i$ for $i = 1, 2, \ldots, n$

Then, $$<\mathbf{v}, \mathbf{w}> = <\sum_{i=1}^{n} a_i \mathbf{u}_i, \mathbf{w}> = \sum_{i=1}^{n} a_i <\mathbf{u}_i, \mathbf{w}>$$

$$= \sum_{i=1}^{n} a_i <\mathbf{w}, \mathbf{u}_i> = \sum_{i=1}^{n} a_i b_i$$

- Corollary 5.5.4 *Parseval's Formula*

  If $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ is an orthonormal basis for an inner product space V and $\mathbf{v} = c_1\mathbf{u}_1 + \ldots + c_n\mathbf{u}_n$, then

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

(*Proof*)

From Th. 5.5.3, $\left\|\mathbf{v}\right\|^2 = <\mathbf{v},\mathbf{v}> = \sum_{i=1}^{n} c_i^2$

Def. Orthogonal Matrix

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

- Th. 5.5.5

  An $n \times n$ matrix $\mathbf{Q}$ is orthogonal if and only if $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$

(Note) $\mathbf{Q}$ is invertible and $\mathbf{Q}^{-1} = \mathbf{Q}^T$

Properties of Orthogonal Matrices

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

a. The column vectors of $\mathbf{Q}$ form an orthonormal basis for $\mathbf{R}^n$

b. $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}$

c. $\mathbf{Q}^T = \mathbf{Q}^{-1}$

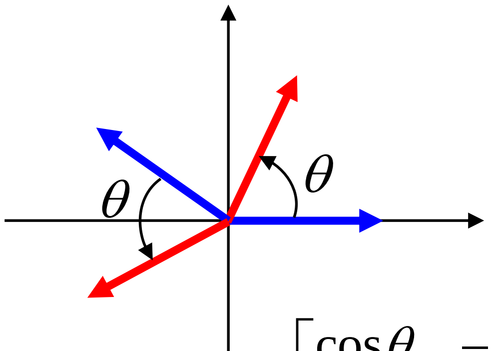
d. $< \mathbf{Qx}, \mathbf{Qy} > = < \mathbf{x}, \mathbf{y} >$ preservation of inner products

e. $\|\mathbf{Qx}\| = \|\mathbf{x}\|$ preservation of norms

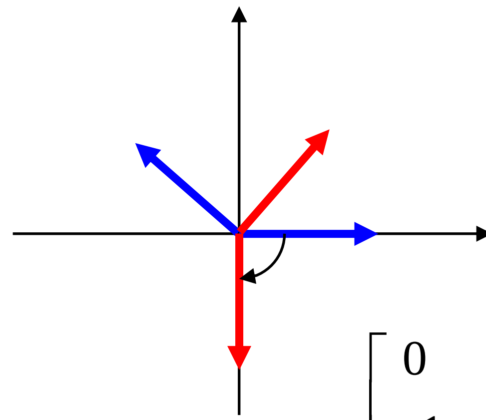
$$< \mathbf{Qx}, \mathbf{Qy} > = (\mathbf{Qx})^T(\mathbf{Qy}) = \mathbf{x}^T\mathbf{Q}^T\mathbf{Qy} = \mathbf{x}^T\mathbf{y} = < \mathbf{x}, \mathbf{y} >$$

$$\|\mathbf{Qx}\|^2 = < \mathbf{Qx}, \mathbf{Qx} > = < \mathbf{x}, \mathbf{x} > = \|\mathbf{x}\|^2$$

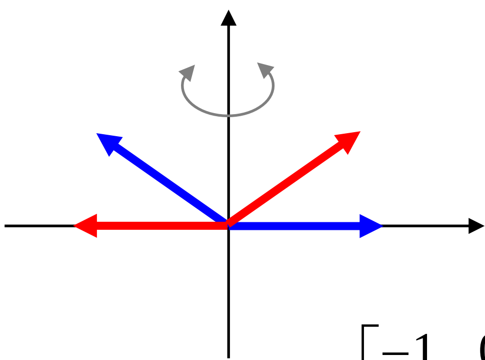
Generalization ?



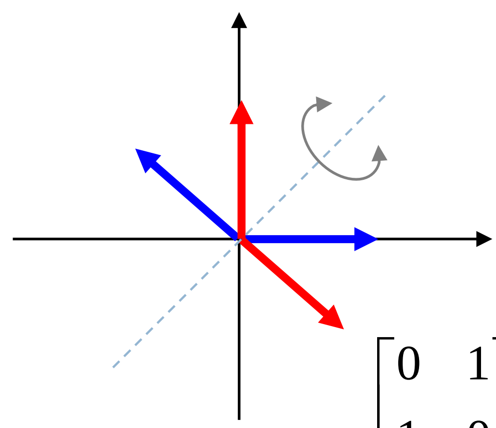
$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$



$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$



$$\begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}$$



$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

(Example) **Permutation Matrices**

- A permutation matrix is a matrix formed from the identity matrix by reordering its columns
- The permutation matrix formed by reordering $\mathbf{I}$ in the order $(k_1, \ldots, k_n)$ is $\mathbf{P} = (\mathbf{e}_{k1}, \ldots, \mathbf{e}_{kn})$

$$\mathbf{P} = (\mathbf{e}_3, \mathbf{e}_1, \mathbf{e}_2) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

- Clearly, permutation matrices are *orthogonal matrices*. Then,

$$\mathbf{P}^{-1} = \mathbf{P}^T = \begin{bmatrix} \mathbf{e}_3^T \\ \mathbf{e}_1^T \\ \mathbf{e}_2^T \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} = \mathbf{Q}$$

- If $\mathbf{A}$ is an $m \times n$ matrix, then

$$\mathbf{AP} = (\mathbf{Ae}_{k1}, \ldots, \mathbf{Ae}_{kn}) = (\mathbf{a}_{k1}, \ldots, \mathbf{a}_{kn})$$

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}, \quad \mathbf{AP} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 1 & 2 \\ 6 & 4 & 5 \end{bmatrix}$$

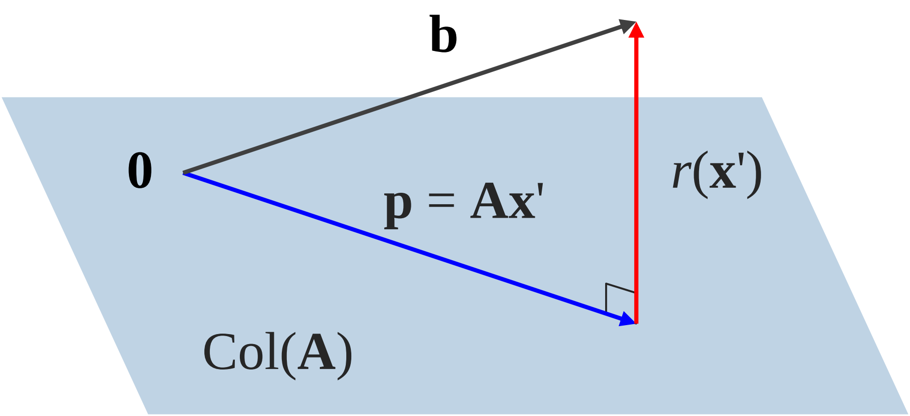
- Similarly, row-wise reordering can be also formed

$$\mathbf{P} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \quad \mathbf{PB} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 5 \\ 3 & 6 \\ 1 & 4 \end{bmatrix}$$

Orthonormal Set & Least Squares

- Th. 5.5.6

  If the column vectors of an $m \times n$ matrix $\mathbf{A}$ form an orthonormal set of vectors in $\mathbf{R}^m$, then $\mathbf{A}^T\mathbf{A} = \mathbf{I}$ and the solution to the least squares problem $\mathbf{A}\mathbf{x} = \mathbf{b}$ is $\mathbf{x}' = \mathbf{A}^T\mathbf{b}$



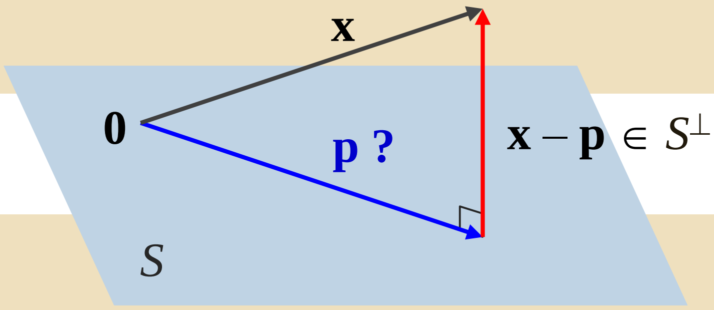
Normal eq.

$\mathbf{A}^T\mathbf{A}\mathbf{x}' = \mathbf{A}^T\mathbf{b}$

Th. 5.5.7

Let S be a subspace of an inner product space V and let $\mathbf{x} \in V$. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ be an orthonormal basis for S.

If $\mathbf{p} = c_1\mathbf{u}_1 + \ldots + c_n\mathbf{u}_k$ where $c_i = <\mathbf{x}, \mathbf{u}_i>$ for each i, then $\mathbf{x} - \mathbf{p} \in S^{\perp}$



Th. 5.5.8

Under the hypothesis of Th. 5.5.7, $\mathbf{p}$ is the element of S that is closest to $\mathbf{x}$, that is, for any $\mathbf{y} \neq \mathbf{p}$ in S,

$$\|\mathbf{x} - \mathbf{y}\| > \|\mathbf{x} - \mathbf{p}\|$$

□ Corollary 5.5.9

Let S be a nonzero subspace of $\mathbf{R}^n$ and let $\mathbf{b} \in \mathbf{R}^n$.

If $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ is an orthonormal basis for S and $\mathbf{U} = [\, \mathbf{u}_1 \; \mathbf{u}_2 \; \ldots \; \mathbf{u}_k \,]$, then the projection $\mathbf{p}$ of $\mathbf{b}$ onto S is given by

$$\mathbf{p} = \mathbf{U}\mathbf{U}^T\mathbf{b}$$

(Proof) From Th. 5.5.7,

$$\mathbf{p} = c_1\mathbf{u}_1 + \ldots + c_k\mathbf{u}_k = \mathbf{U}\mathbf{c} \quad \text{where} \quad \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{b} \\ \mathbf{u}_2^T\mathbf{b} \\ \vdots \\ \mathbf{u}_k^T\mathbf{b} \end{bmatrix} = \mathbf{U}^T\mathbf{b}$$

Therefore, $\mathbf{p} = \mathbf{U}\mathbf{U}^T\mathbf{b}$

(*Ex*.7) Let S be the set of all vectors in $\mathbf{R}^3$ of the form $(x,y,0)$. Find the vector $\mathbf{p}$ in S that is closest to $\mathbf{w} = (5,3,4)$

(*Solution*)

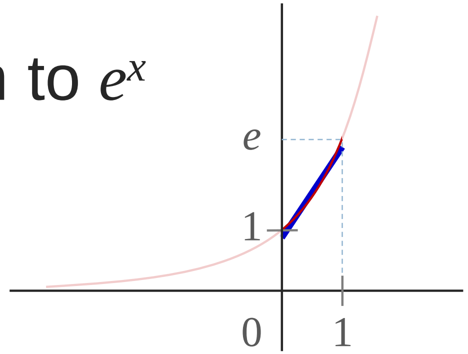
Let $\mathbf{u}_1 = (1,0,0)$ and $\mathbf{u}_2 = (0,1,0)$. Clearly, $\mathbf{u}_1$ and $\mathbf{u}_2$ form an orthonormal basis for S. Now $c_1 = \mathbf{w}^T\mathbf{u}_1 = 5$ and $c_2 = \mathbf{w}^T\mathbf{u}_2 = 3$. Thus $\mathbf{p} = 5\mathbf{u}_1 + 3\mathbf{u}_2 = (5,3,0)$

Alternatively, $\mathbf{p}$ could have been calculated using the projection matrix $\mathbf{U}\mathbf{U}^T$:

$$\mathbf{p} = \mathbf{U}\mathbf{U}^T\mathbf{w} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 0 \end{bmatrix}$$

Approximation of Functions

Find the best least squares approximation to e^x on the interval $[0,1]$ by a linear function $mx + b$



(*Solution*)

Let S be the subspace of all linear functions in $[0,1]$. Although the functions 1 and x span S, they are not orthogonal.

We seek a function of form $x - a$ that is orthogonal to 1.

$$<1, x-a> = \int_0^1 (x-a)\,dx = 1/2 - a = 0$$

Thus, $a = 1/2$.

Since $\|x - 1/2\| = 1/\sqrt{12}$, it follows that

$$u_1(x) = 1, \quad u_2(x) = \sqrt{12}(x - 1/2)$$

form an orthonormal basis for S

Let $c_1 = \int_0^1 u_1(x) e^x \, dx = e - 1, \; c_2 = \int_0^1 u_2(x) e^x \, dx = \sqrt{3}(3 - e)$

The projection

$$p(x) = c_1 u_1(x) + c_2 u_2(x) = (e-1)\cdot 1 + \sqrt{3}(3-e)[\sqrt{12}(x - 1/2)]$$
$$= (4e - 10) + 6(3 - e)x$$

is the best linear least squares approximation to e^x

$$\|x - 1/2\|^2 = \int_0^1 (x - 1/2)^2 \, dx = x^3/3 - x^2/2 + x/4 \Big|_0^1 = 1/12$$

$$\int_0^1 x e^x \, dx = \int_0^1 \{(xe^x)' - e^x\} \, dx = (xe^x - e^x) \Big|_0^1 = 1$$

(*Another Solution*)

Let the squared sum error between e^x and $ax + b$ be Δ^2.

$$\Delta^2 = \int_0^1 \{e^x - (ax+b)\}^2\,dx = \int_0^1 \{e^{2x} - 2(ax+b)e^x + (ax+b)^2\}\,dx$$

$$= (e^2 - 1)/2 - 2a - 2b(e-1) + a^2/3 + ab + b^2$$

$$\frac{\partial \Delta^2}{\partial a} = -2 + 2a/3 + b = 0 \quad \therefore b = 2 - 2a/3$$

$$\frac{\partial \Delta^2}{\partial b} = -2(e-1) + a + 2b = -2(e-1) + a + 2(2 - 2a/3)$$

$$= (6 - 2e) - a/3 = 0 \quad \therefore a = 6(3-e) \quad \therefore b = 4e - 10$$

Therefore, the best approximation is $6(3-e)x + (4e-10)$

Summary 4

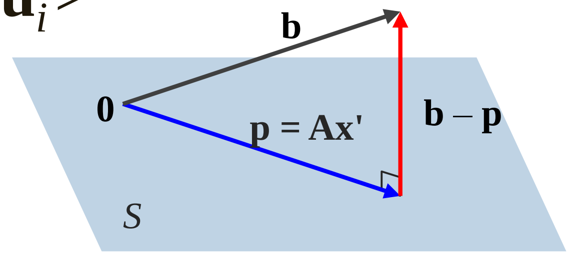
1. Orthogonal Set, Orthonormal Set
2. Vectors in Orthogonal Set $\Rightarrow$ Linearly Independent
3. Orthonormal Basis $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$

$$\Rightarrow\ \mathbf{v} = c_1\mathbf{u}_1 + \ldots + c_n\mathbf{u}_n,\ \ c_i = <\mathbf{v},\mathbf{u}_i>\ ,\ \ \|\mathbf{v}\|^2 = \sum_{i=1}^{n} c_i^2$$

$$\Rightarrow\ <\mathbf{v},\mathbf{w}> = a_1b_1 + \ldots + a_nb_n$$

$$\Rightarrow\ \mathbf{p} = c_1\mathbf{u}_1 + \ldots + c_n\mathbf{u}_n = \mathbf{U}\mathbf{U}^T\mathbf{b},\ \ c_i = <\mathbf{b},\mathbf{u}_i>$$

$$\Rightarrow\ \mathbf{A}^T\mathbf{A}\mathbf{x}' = \mathbf{x}' = \mathbf{A}^T\mathbf{b}$$



4. Orthogonal Matrix $\mathbf{Q}$

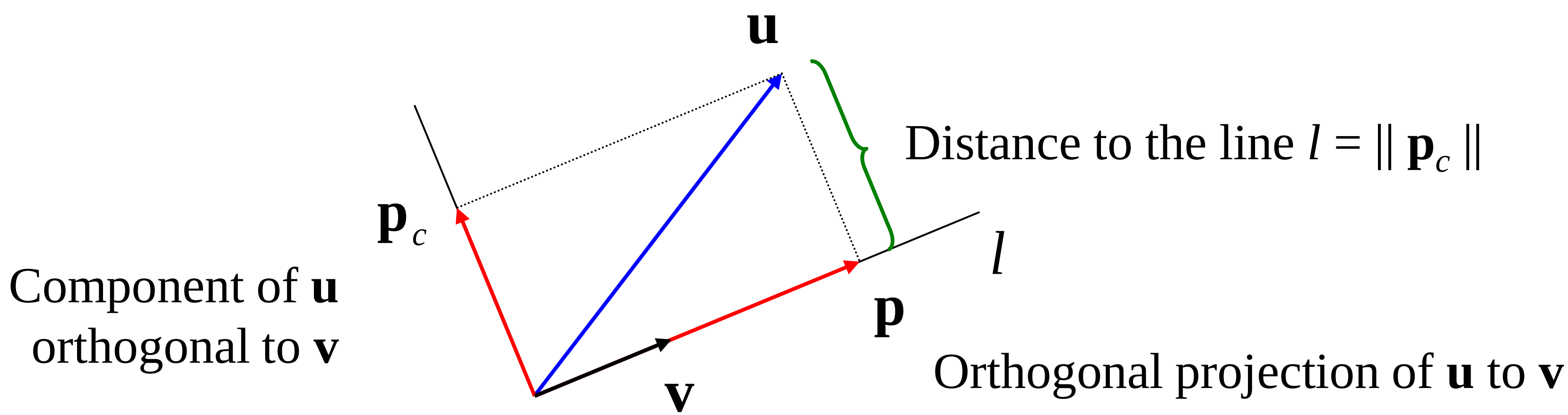
$$\Rightarrow\ \mathbf{Q}^T = \mathbf{Q}^{-1},\ \ <\mathbf{Q}\mathbf{x}, \mathbf{Q}\mathbf{y}> = <\mathbf{x}, \mathbf{y}>,\ \|\mathbf{Q}\mathbf{x}\| = \|\mathbf{x}\|$$

Gram-Schmidt Orthogonalization Process

To know how to construct an orthonormal basis for an n-dimensional inner product space V

4. Inner Product Spaces
5. Orthonoraml Sets
6. Gram-Schmidt Orthogonalization Process

Review: Orthogonal Projections



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

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

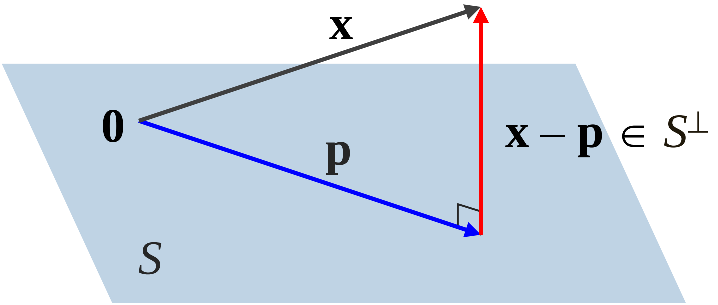
- (Review) Th. 5.5.7 and 5.5.8

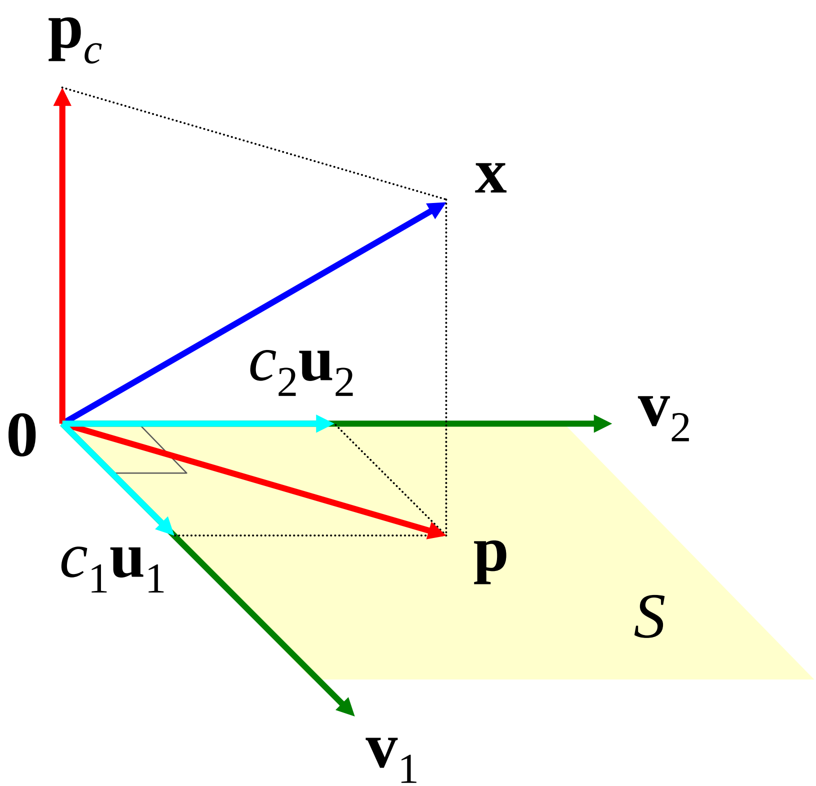
Let S be a subspace of an inner product space V and let $\mathbf{x} \in V$. Let $\{\mathbf{u}_1,\ldots,\mathbf{u}_n\}$ be an orthonormal basis for S.

If $\mathbf{p} = c_1\mathbf{u}_1 + \ldots + c_n\mathbf{u}_n$ where $c_i = \langle \mathbf{x}, \mathbf{u}_i \rangle$ for each i, then $\mathbf{x} - \mathbf{p} \in S^{\perp}$

The projection vector $\mathbf{p}$ onto S is the vector closest to $\mathbf{x}$, that is, for any $\mathbf{y} \neq \mathbf{p}$ in S,

$$\|\mathbf{x} - \mathbf{y}\| > \|\mathbf{x} - \mathbf{p}\|$$





$$S = \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$$

$$< \mathbf{v}_1, \mathbf{v}_2 > = 0$$

$$\mathbf{u}_1 = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|}, \;\; \mathbf{u}_2 = \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|}$$

$$c_1 = < \mathbf{x}, \mathbf{u}_1 >, \;\; c_2 = < \mathbf{x}, \mathbf{u}_2 >$$

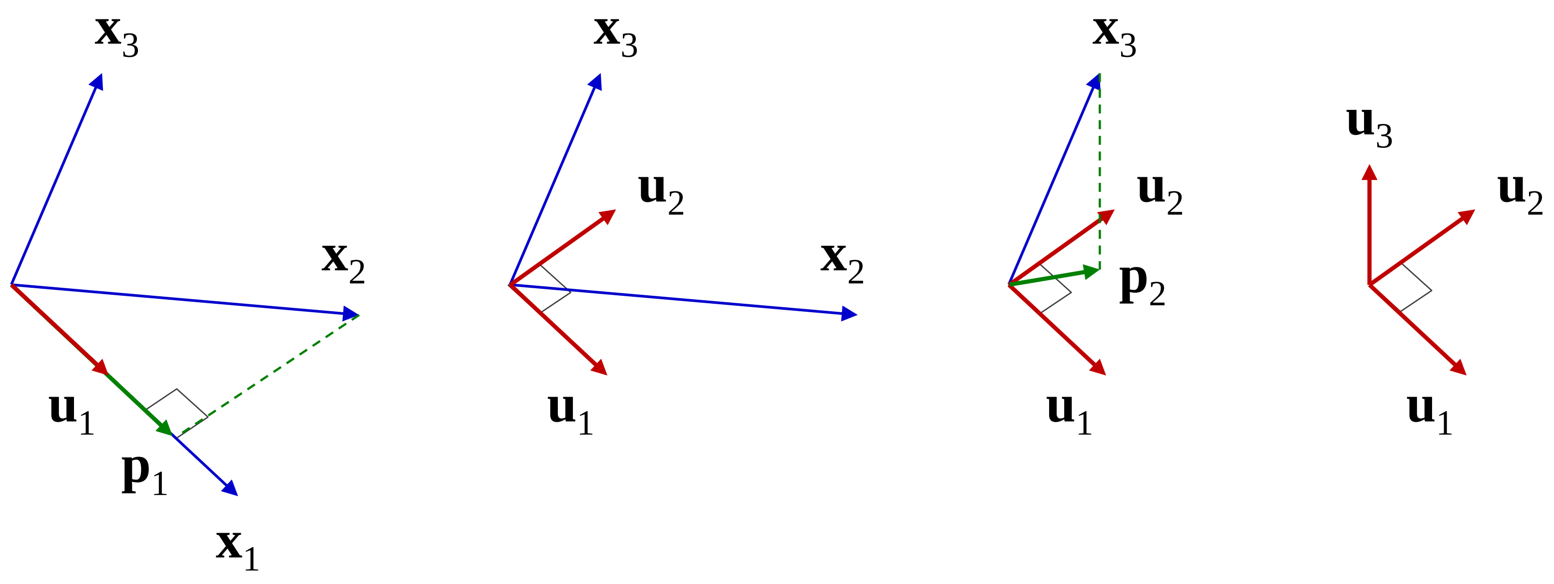
$$\mathbf{p} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 \in S$$

$$\mathbf{p}_c = \mathbf{x} - \mathbf{p} \in S^{\perp}$$

$\mathbf{p}$ is the closest point in S to $\mathbf{x}$, or

the *best approximation* of $\mathbf{x}$ by vectors of S

- Transformation of an ordinary basis $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ into an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$



$$\mathbf{u}_1 = \frac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$$

$$\mathbf{p}_1 = <\mathbf{x}_2, \mathbf{u}_1> \mathbf{u}_1$$

□ Th. 5.6.1 *Gram-Schmidt Process*

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

Let $\mathbf{u}_1 = \dfrac{\mathbf{x}_1}{\|\mathbf{x}_1\|}$ and define $\mathbf{u}_2, \ldots, \mathbf{u}_n$ recursively by

$$\mathbf{u}_{k+1} = \frac{\mathbf{x}_{k+1} - \mathbf{p}_k}{\|\mathbf{x}_{k+1} - \mathbf{p}_k\|} \qquad \mathit{for} \;\; k = 1, \ldots, n-1$$

where $\mathbf{p}_k = <\mathbf{x}_{k+1}, \mathbf{u}_1> \mathbf{u}_1 + \;\Lambda\; + <\mathbf{x}_{k+1}, \mathbf{u}_k> \mathbf{u}_k$

is the projection of $\mathbf{x}_{k+1}$ onto $\mathrm{Span}(\mathbf{u}_1, \ldots, \mathbf{u}_k)$.

Then the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ is an orthonormal basis for V

(*Ex*.8) Find an orthonormal basis for $\mathbf{R}^3$ if inner product is defined by $< \mathbf{x}, \mathbf{y} > = x_1y_1 + 2\,x_2y_2 + x_3y_3$

(*Solution*)

$\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{x} = (1,1,1)\}$ is a *basis* for $\mathbf{R}^3$.

Then $\mathbf{u}_1 = \mathbf{e}_1 / \|\mathbf{e}_1\| = \mathbf{e}_1$.

The projection of $\mathbf{e}_2$ onto $\mathrm{Span}(\mathbf{u}_1)$ is $\mathbf{p}_1 = < \mathbf{e}_2, \mathbf{u}_1 > \mathbf{u}_1 = \mathbf{0}$.
Thus $\mathbf{u}_2 = (\mathbf{e}_2 - \mathbf{p}_1) / \|\mathbf{e}_2 - \mathbf{p}_1\| = \mathbf{e}_2 / \|\mathbf{e}_2\| = \mathbf{e}_2 / \sqrt{2}$

The projection of $\mathbf{x}$ onto $\mathrm{Span}(\mathbf{u}_1,\mathbf{u}_2)$ is $\mathbf{p}_2 = < \mathbf{x}, \mathbf{u}_1 > \mathbf{u}_1 + < \mathbf{x}, \mathbf{u}_2 > \mathbf{u}_2 = \mathbf{u}_1 + \sqrt{2}\,\mathbf{u}_2 = (1,1,0)$. Thus $\mathbf{u}_3 = (\mathbf{x} - \mathbf{p}_2) / \|\mathbf{x} - \mathbf{p}_2\| = (0,0,1) / \|(0,0,1)\| = \mathbf{e}_3$

Therefore, $\{\mathbf{e}_1, \mathbf{e}_2 / \sqrt{2}, \mathbf{e}_3\}$ is an orthonormal basis for $\mathbf{R}^3$

(*Ex*.9) Find an orthonormal basis for P_3 if inner product on P_3 is defined, for $x_1 = -1$, $x_2 = 0$, $x_3 = 1$, by

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

(*Solution*)

$\{1, x, x^2\}$ is a basis for P_3. $\mathbf{u}_1 = 1 / \|1\| = 1/\sqrt{3}$.

The projection of x onto $\mathrm{Span}(\mathbf{u}_1)$ is $\mathbf{p}_1 = < x, \mathbf{u}_1 > \mathbf{u}_1 =$ $< x, 1/\sqrt{3} > (1/\sqrt{3}) = (-1+0+1)/\sqrt{3} \cdot (1/\sqrt{3}) = 0$. Then $\mathbf{u}_2 =$ $(x - \mathbf{p}_1) / \|x - \mathbf{p}_1\| = x / \|x\| = x/\sqrt{2}$

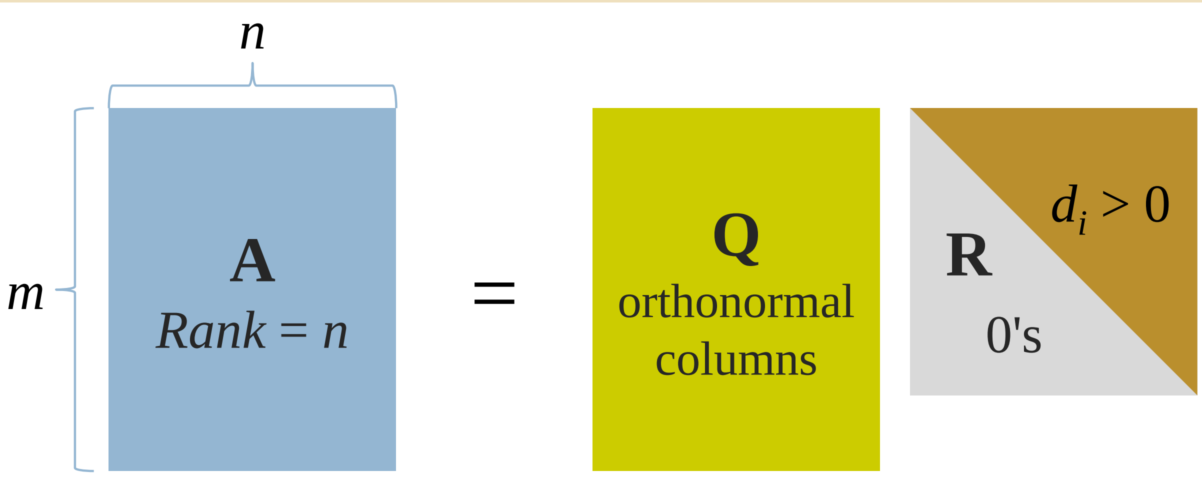
The projection of x^2 onto $\mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2)$ is $\mathbf{p}_2 = < x^2, 1/\sqrt{3} >$ $1/\sqrt{3} + < x^2, x/\sqrt{2} > x/\sqrt{2} = 2/3$. Then $\mathbf{u}_3 = (x^2 - 2/3) / \|x^2 - 2/3\| = (\sqrt{6}/2)(x^2 - 2/3)$

Thus $\{1/\sqrt{3}, x/\sqrt{2}, (\sqrt{6}/2)(x^2-2/3)\}$ is an orthonormal basis

□ Th. 5.6.2 *Gram-Schmidt QR Factorization*

If $\mathbf{A}$ is an $m \times n$ matrix of *rank* n (hence, $m \geq n$), then $\mathbf{A}$ can be factored into a product $\mathbf{QR}$, where $\mathbf{Q}$ is an $m \times n$ matrix with orthonormal column vectors and $\mathbf{R}$ is an upper triangular $n \times n$ matrix whose *diagonal entries* are all *positive*.

[Note: $\mathbf{R}$ must be nonsingular, since $\det(\mathbf{R}) > 0$.]



(*Proof*)

Let $\mathbf{p}_1, \ldots, \mathbf{p}_{n\text{-}1}$ be the projection vectors and let $\{\mathbf{q}_1, \ldots, \mathbf{q}_n\}$ be the orthonormal basis of $R(\mathbf{A})$ derived from the Gram-Schmidt process for the column vectors of $\mathbf{A}$, $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$

Define $r_{11} = \|\mathbf{a}_1\|$, $r_{kk} = \|\mathbf{a}_k - \mathbf{p}_{k\text{-}1}\|$ for $k = 2, \ldots, n$ and $r_{ik} = \mathbf{q}_i^T \mathbf{a}_k$ for $i = 1, \ldots, k\text{-}1$ and $k = 2, \ldots, n$

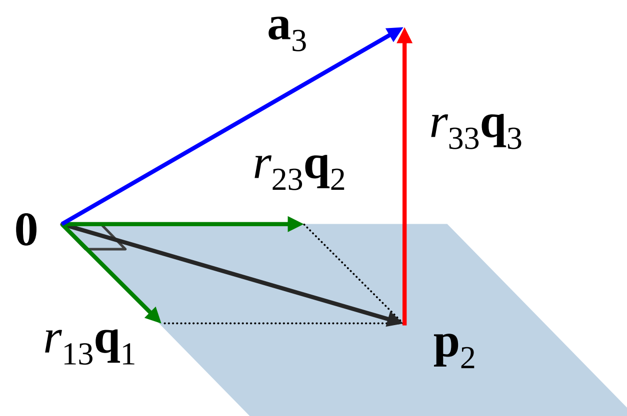
By the Gram-Schmidt process,

$$\mathbf{a}_1 = r_{11}\mathbf{q}_1$$

$$\mathbf{a}_2 = r_{12}\mathbf{q}_1 + r_{22}\mathbf{q}_2$$

$$\vdots$$

$$\mathbf{a}_n = r_{1n}\mathbf{q}_1 + \Lambda + r_{nn}\mathbf{q}_n$$



If we set $\mathbf{Q} = [\ \mathbf{q}_1 \ \ldots \ \mathbf{q}_n\ ]$ and define $\mathbf{R}$ to be the upper triangular matrix

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \Lambda & r_{1n} \\ 0 & r_{22} & \Lambda & r_{2n} \\ : & : & \mathrm{O} & : \\ 0 & 0 & \Lambda & r_{nn} \end{bmatrix}$$

then the jth column of the product $\mathbf{QR}$ will be

$$\mathbf{Qr}_j = r_{1j}\mathbf{q}_1 + \Lambda + r_{jj}\mathbf{q}_n = \mathbf{a}_j$$

for $j = 1,\ldots,n$. Therefore,

$$\mathbf{QR} = [\,\mathbf{a}_1 \ \mathbf{a}_2 \, \Lambda \ \mathbf{a}_n\,] = \mathbf{A}$$

- Th. 5.6.3

If $\mathbf{A}$ is an $m \times n$ matrix of rank n, then the least squares solution of $\mathbf{Ax} = \mathbf{b}$ is given by

$$\mathbf{x}' = \mathbf{R}^{-1}\mathbf{Q}^T\mathbf{b}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are the matrices obtained from the QR factorization, that is, $\mathbf{A} = \mathbf{QR}$

The solution $\mathbf{x}'$ may be obtained by using back substitution to solve $\mathbf{Rx} = \mathbf{Q}^T\mathbf{b}$

$$\mathbf{A}^T\mathbf{A}\mathbf{x}' = \mathbf{A}^T\mathbf{b} \quad \Rightarrow \quad (\mathbf{QR})^T(\mathbf{QR})\mathbf{x}' = (\mathbf{QR})^T\mathbf{b}$$

$$\Rightarrow \quad \mathbf{R}^T(\mathbf{Q}^T\mathbf{Q})\mathbf{R}\mathbf{x}' = \mathbf{R}^T\mathbf{Q}^T\mathbf{b} \quad \Rightarrow \quad \mathbf{R}^T\mathbf{R}\mathbf{x}' = \mathbf{R}^T\mathbf{Q}^T\mathbf{b}$$

Since $\mathbf{R}$ is invertible, $\mathbf{x}' = \mathbf{R}^{-1}\mathbf{Q}^T\mathbf{b}$

(*Ex*.10) Find the QR factorization of $\mathbf{A}$

$$\mathbf{A}=\begin{bmatrix}1 & 1 & 0\\ 1 & -1 & 0\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix}\sim\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix}$$

(*Solution*)

By the Gram-Schmidt process we have

$$\mathbf{q}_1=\mathbf{a}_1/\|\mathbf{a}_1\|=(1,1,1,1)/2$$

$$\mathbf{p}_1=<\mathbf{a}_2,\mathbf{q}_1>\mathbf{q}_1=<(1,-1,1,1),\mathbf{q}_1>\mathbf{q}_1=\mathbf{q}_1$$

$$\mathbf{q}_2=(\mathbf{a}_2-\mathbf{p}_1)/\|\mathbf{a}_2-\mathbf{p}_1\|=(1,-3,1,1)/2\sqrt{3}$$

$$\mathbf{p}_2=<\mathbf{a}_3,\mathbf{q}_1>\mathbf{q}_1+<\mathbf{a}_3,\mathbf{q}_2>\mathbf{q}_2=\mathbf{q}_1+\mathbf{q}_2/\sqrt{3}=(2,0,2,2)/3$$

$$\mathbf{q}_3=(\mathbf{a}_3-\mathbf{p}_2)/\|\mathbf{a}_3-\mathbf{p}_2\|=(-2,0,1,1)/\sqrt{6}$$

Thus the orthonormal matrix is

$$\mathbf{Q} = \begin{bmatrix} 1/2 & \sqrt{3}/6 & -\sqrt{6}/3 \\ 1/2 & -\sqrt{3}/2 & 0 \\ 1/2 & \sqrt{3}/6 & \sqrt{6}/6 \\ 1/2 & \sqrt{3}/6 & \sqrt{6}/6 \end{bmatrix}$$

Since $\mathbf{Q}^T\mathbf{A} = \mathbf{Q}^T\mathbf{QR} = \mathbf{IR} = \mathbf{R}$, we have

$$\mathbf{R} = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ \sqrt{3}/6 & -\sqrt{3}/2 & \sqrt{3}/6 & \sqrt{3}/6 \\ -\sqrt{6}/3 & 0 & \sqrt{6}/6 & \sqrt{6}/6 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 1 & 1 \\ 0 & \sqrt{3} & \sqrt{3}/3 \\ 0 & 0 & \sqrt{6}/3 \end{bmatrix}$$