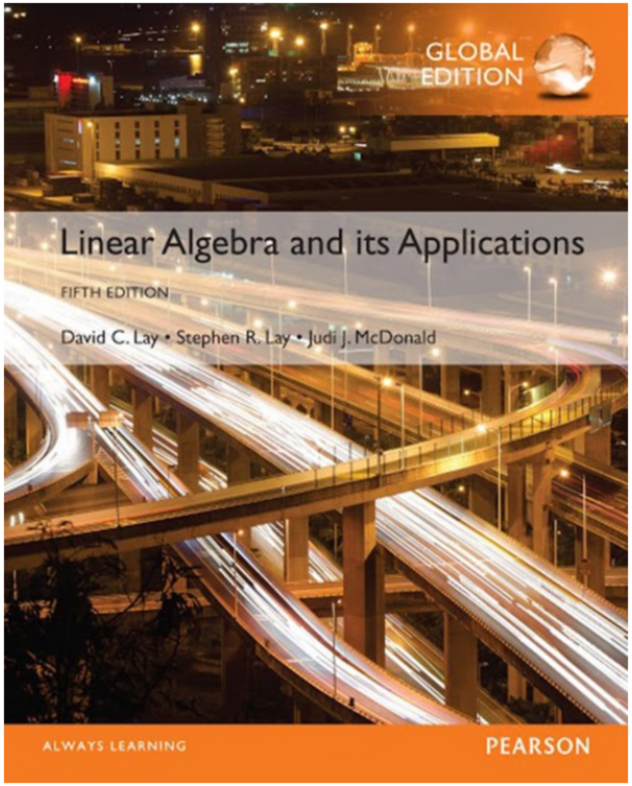
5

Eigenvalues and Eigenvectors

5.4

EIGENVECTORS AND LINEAR TRANSFORMATIONS



$$\begin{tikzcd}
p \in V = P_1 = \{a_1x + a_0\} \arrow[r, "T"] \arrow[d] & W \ni T(p),\ W = P_1 \arrow[d] \\
{[p]_B} \in R^2 \arrow[r, "A"] & A[p]_B = [T(p)]_C \in R^2
\end{tikzcd}$$

$B = \{1, x\}$, $C = \{1-x,\ 1+x\}$

개념 $T(p) = p' = (a_1x + a_0)' = a_1$

$$A = \begin{array}{cc} b_1 & b_2 \\ \end{array}$$

$$A = \left[\, [T(1)]_C \quad [T(x)]_C \,\right]$$

$$= \left[\, [0]_C \quad [1]_C \,\right]$$

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

$p = a_1x + a_0$

$$[p]_B = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$

$$\begin{tikzcd}
p \in V \arrow[r, "T"] \arrow[d] & W \ni T(p) \arrow[d] \\
{[p]_B} \in R^n \arrow[r, "A"] & A[p]_B = [T(p)]_C \in R^n
\end{tikzcd}$$

$$1 = \frac{(1-x) + (1+x)}{2}$$

$$\begin{tikzcd}
p \in V = P_1 = \{a_1x + a_0\} \arrow[r, "T"] \arrow[d] & W \ni T(p),\ W = P_1 \arrow[d] \\
{[p]_B} \in R^2 \arrow[r, "A"] & A[p]_B = [T(p)]_C \in R^2
\end{tikzcd}$$

$B = \{1, x\}$, $C = \{1-x,\ 1+x\}$

$T(p) = p' = (a_1x + a_0)' = a_1$

$$A = \left[\, [T(1)]_C \quad [T(x)]_C \,\right]$$

$$= \left[\, [0]_C \quad [1]_C \,\right]$$

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

$p = a_1x + a_0$

$$[p]_B = \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}$$

$$1 = \frac{(1-x) + (1+x)}{2}$$

ex

$p = 2 + 3x$

$$[p]_B = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

$$A[p]_B = \begin{pmatrix} 0 & 1/2 \\ 0 & 1/2 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 3/2 \\ 3/2 \end{pmatrix}$$

$$3/2\,(1-x) + 3/2\,(1+x) = 3$$

$$
\begin{tikzcd}
V \arrow[r, "T"] \arrow[d, "{[\ ]_B}"'] & W \arrow[d, shift right, "{[\ ]_C}"'] \\
\mathbb{R}^n \arrow[r, "A"] & \mathbb{R}^m \arrow[u, shift right, "{[\ ]_C^{-1}}"']
\end{tikzcd}
$$

왜 이런 Theorem이 있나?

Sometimes, you want to consider T operations several times

$$T(\cdots T(T(T(x)))) = T^n(x)$$

what if $A = PDP^{-1}$

$$A^n \longrightarrow A^n = (PDP^{-1})(PDP^{-1})\cdots(PDP^{-1})$$

$$P^{-1}P = I$$

$$= PD^nP^{-1} = P\begin{pmatrix} \lambda_1^n & & & \\ & \lambda_2^n & & \\ & & \ddots & \\ & & & \lambda_n^n \end{pmatrix}P^{-1}$$

$$D = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}$$

$$DD = \begin{pmatrix} \lambda_1^2 & & & \\ & \lambda_2^2 & & \\ & & \ddots & \\ & & & \lambda_n^2 \end{pmatrix}$$

T^n은 오래걸리니

A로 바꾸고 $A = PDP^{-1}$ 이용해서

빠르게 계산

THE MATRIX OF A LINEAR TRANSFORMATION

- Given any x in V, the coordinate vector $[\mathrm{x}]_{\beta}$ is in $\mathbb{R}^n$ and the coordinate vector of its image, $[\mathrm{T(x)}]_C$, is in $\mathbb{R}^m$, as shown in Fig. 1 below.

1-1 onto
linear T

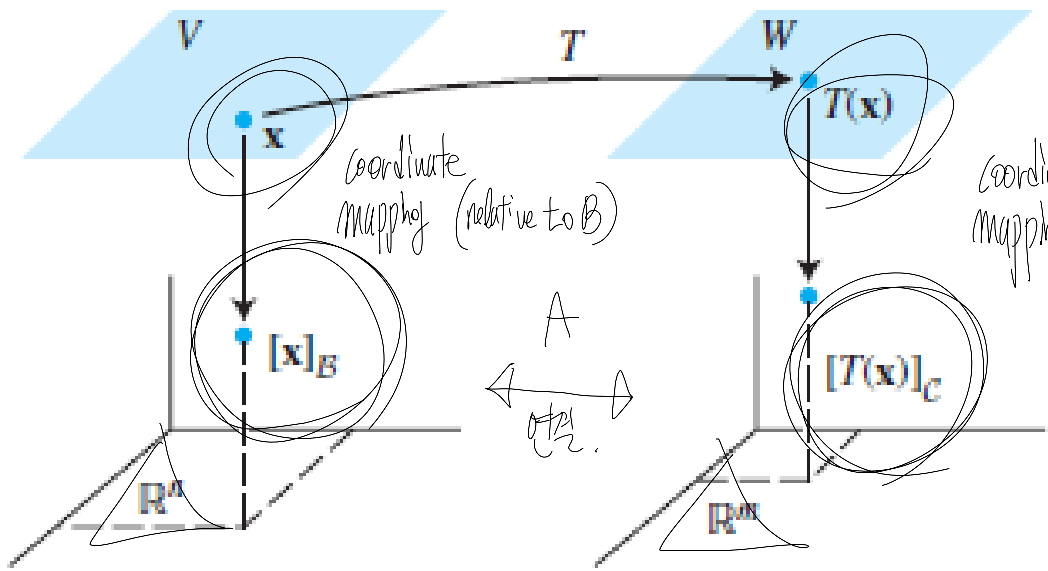


FIGURE 1 A linear transformation from V to W.

THE MATRIX OF A LINEAR TRANSFORMATION

- The connection between $[x]_\beta$ and $[T(x)]_C$ is easy to find. Let $\{b_1, \dots, b_2\}$ be the basis β for V. If $x = r_1b_1 + \dots + r_nb_n$, then,

$$[x]_\beta = \begin{bmatrix} r_1 \\ \cdot \\ \cdot \\ \cdot \\ r_n \end{bmatrix}$$

- And

$$T(x) = T(r_1b_1 + \cdots + r_nb_n) = r_1T(b_1) + \cdots + r_nT(b_n) \quad (1)$$

because T is linear.

THE MATRIX OF A LINEAR TRANSFORMATION

- Now, since the coordinate mapping from W to $\mathbb{R}^m$ is linear, equation (1) leads to

$$= [r_1 T(b_1) + r_2 T(b_2) + \cdots + r_n T(b_n)]$$

$$[T(x)]_C = r_1[T(b_1)]_C + \cdots + r_n[T(b_n)]_C \qquad (2)$$

$$= \left[[T(b_1)]_C + \cdots + [T(b_n)]_C \right] \begin{bmatrix} r_1 \\ \vdots \\ r_n \end{bmatrix}$$

- Since C-coordinate vectors are in $\mathbb{R}^m$, the vector equation (2) can be written as a matrix equation, namely,

Coordinate of image T(x) in the codomain

Coordinate of x in domain

$$[T(x)]_C = M[x]_\beta \qquad (3)$$

where

$$M = [[T(b_1)]_C [T(b_2)]_C \ \ldots \ [T(b_n)]_C] \qquad (4)$$

THE MATRIX OF A LINEAR TRANSFORMATION

- The matrix M is a matrix representation of T, called the matrix for T relative to the bases β and C. See Fig. 2 below:

$$\begin{tikzcd}
\mathbf{x} \arrow[rr, "T"] \arrow[d] & & T(\mathbf{x}) \arrow[d] \\
{[\mathbf{x}]_{\mathcal{B}}} \arrow[rr, "\text{Multiplication}", "\text{by } M"'] & & {[T(\mathbf{x})]_{\mathcal{C}}}
\end{tikzcd}$$

FIGURE 2

Example of theory

계산은 T 행렬을 결론

$$V = P_2 = \{a_2x^2 + a_1x + a_0\} \xrightarrow{T} W = P_1 = \{a_1x + a_0\}$$

$$T(p) = \frac{p(x) - p(0)}{x} \qquad \text{ex)}\ T(x^2+1) = \frac{(x^2+1) - (0+1)}{x} = \frac{x^2}{x} = \underline{\underline{x}}$$

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ \downarrow & & \uparrow \\ \mathbb{R}^3 & \xrightarrow{A} & \mathbb{R}^2 \end{array}$$

$B = \{1, 1+x, 1+x+x^2\}$ $\qquad$ $C = \{1, x\}$

$$\left([T(b_1)]_C \quad [T(b_2)]_C \quad [T(b_3)]_C\right) = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

$$T(b_1) = T(1) = \frac{1-1}{x} = 0 \qquad [T(b_1)]_C = [0]_C = 0 \cdot 1 + 0 \cdot x$$

$$T(b_2) = T(1+x) = \frac{(1+x)-1}{x} = 1 \qquad [T(b_2)]_C = [1]_C = \underline{1} \cdot 1 + \underline{0}\, x$$

$$T(b_3) = T(1+x+x^2) = \frac{(1+x+x^2)-1}{x} = 1+x \qquad [T(b_3)]_C = 1 \cdot 1 + 1 \cdot x$$

Let $p(x) = \underline{\underline{3+2x+x^2}}$ then $[p]_B = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

$$M[p]_B = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix} = [T(p)]_C \Rightarrow \underline{T(p) = 2 \cdot 1 + 1 \cdot x} = 2+x$$

Double Check: $T(3+2x+x^2) = \dfrac{3+2x+x^2-3}{x} = \dfrac{2x+x^2}{x} = 2+x$

THE MATRIX OF A LINEAR TRANSFORMATION

- **Example 1** Suppose $\beta = \{b_1, b_2\}$ is a basis for V and $C = \{c_1, c_2, c_3\}$ is a basis for W. Let $T: V \to W$ be a linear transformation with the property that

$$T(b_1) = 3c_1 - 2c_2 + 5c_3 \quad \text{and} \quad T(b_2) = 4c_1 + 7c_2 - c_3$$

Find the matrix M for T relative to β and C.

$$[T(b_1)]_C = \begin{bmatrix} 3 \\ -2 \\ 5 \end{bmatrix} \qquad [T(b_2)]_C = \begin{bmatrix} 4 \\ 7 \\ -1 \end{bmatrix}$$

THE MATRIX OF A LINEAR TRANSFORMATION

- **Solution** The C-coordinate vectors of the images of b_1 and b_2 are

$$[T(b_1)]_C = \begin{bmatrix} 3 \\ -2 \\ 5 \end{bmatrix} \text{ and } [T(b_2)]_C = \begin{bmatrix} 4 \\ 7 \\ -1 \end{bmatrix}$$

- Hence

$$M = \begin{bmatrix} 3 & 4 \\ -2 & 7 \\ 5 & -1 \end{bmatrix}$$

- If β and C are bases for the same space V and if T is the identity transformation T(x) = x for x in V, then matrix M in (4) is just a change-of-coordinates matrix.

LINEAR TRANSFORMATIONS FROM *V* INTO *V*

- In the common case where W is the same V and the basis C is the same as β, then the matrix M in (4) is called the **matrix for T relative to β**, or simply the **β-matrix for T**, and is denoted by $[T]_\beta$.
- See Fig. 3 below

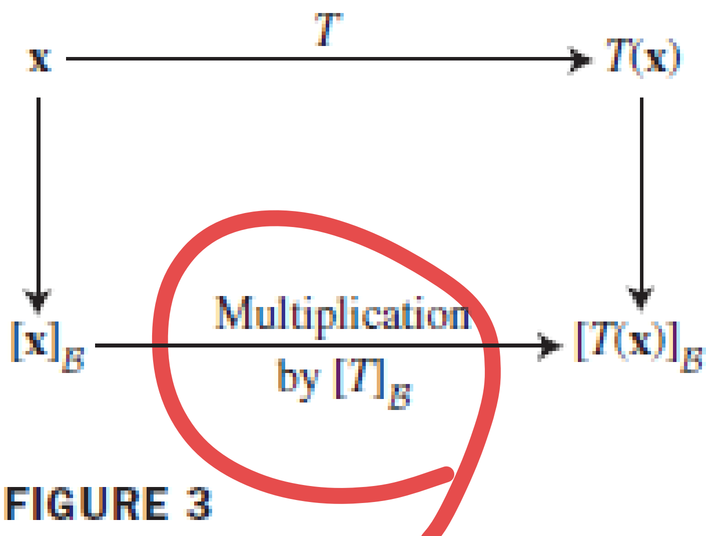


FIGURE 3

- **Theorem 8**: Suppose A = PDP^{-1}, where D is a diagonal $n \times n$ matrix. If β is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the β-matrix for the transformation $\mathrm{x} \mapsto A\mathrm{x}$.

columns of eigenvectors

Diagonal of eigenvalues

If β = an eigenvector basis

then D is β-matrix for the transformation $x \mapsto Ax$

(the matrix with eigenvalues)

LINEAR TRANSFORMATIONS ON $\mathbb{R}^n$

- **Theorem 8**: Suppose $A = PDP^{-1}$, where D is a diagonal $n \times n$ matrix. If β is the basis for $\mathbb{R}^n$ formed from the columns of P, then D is the β-matrix for the transformation $x \mapsto Ax$.

- **Proof** Denote the columns of P by $b_1, \dots, b_n$, so that $\beta = \{b_1, \dots, b_n\}$ and $P = [b_1 \ \dots \ b_n]$. In this case, P is the change-of-coordinates matrix P_β discussed in Section 4.4, where

$$\underline{P[x]_\beta = x \ \text{ and } [x]_\beta = P^{-1}x}$$

LINEAR TRANSFORMATIONS ON $\mathbb{R}^n$

- If $T(x) = Ax$ for x in $\mathbb{R}^n$, then

$$\begin{aligned}
[T]_\beta &= [[T(b_1)]_\beta \quad \cdots \quad [T(b_n)]_\beta] && \text{Definition of } [T]_\beta \\
&= [[Ab_1]_\beta \quad \cdots \quad [Ab_n]_\beta] && \text{Since } T(x) = Ax \\
&= [P^{-1}Ab_1 \quad \cdots \quad P^{-1}Ab_n] && \text{Change of coordinates} \\
&= P^{-1}A[b_1 \quad \cdots \quad b_n] && \text{Matrix multiplication} \\
&= P^{-1}AP \quad = P^{-1}(PDP^{-1})P = D
\end{aligned}$$

$[x]_\beta = P^{-1}x$

$[Ab_1]_\beta = P^{-1}Ab_1$

$\vdots$

$[Ab_n]_\beta = P^{-1}Ab_n$

- Since $A = PDP^{-1}$, we have $[T]_\beta = P^{-1}AP = D$.

이렇게 된다.

LINEAR TRANSFORMATIONS ON $\mathbb{R}^n$

- **Example 3** Define T: $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ by T(x)=Ax, where $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Find a basis β for $\mathbb{R}^2$ with the property that the β–matrix for T is a diagonal matrix.

Find an eigenvector basis i.e. find (two linearly independent) eigenvectors of A

$$0 = \begin{vmatrix} 7-\lambda & 2 \\ -4 & 1-\lambda \end{vmatrix} = (\lambda-3)(\lambda-5)$$

$$\lambda = 3 \;\; \Rightarrow \begin{pmatrix} 1 \\ -2 \end{pmatrix} \qquad \lambda = 5 \;\; \Rightarrow \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

LINEAR TRANSFORMATIONS ON $\mathbb{R}^n$

- **Example 3** Define T: $\mathbb{R}^2 \rightarrow \mathbb{R}^2$ by $T(x)=Ax$, where $A = \begin{bmatrix} 7 & 2 \\ -4 & 1 \end{bmatrix}$. Find a basis β for $\mathbb{R}^2$ with the property that the β–matrix for T is a diagonal matrix.
- **Solution** From Example 2 in Section 5.3 we know that $A = PDP^{-1}$, where

$$P = \begin{bmatrix} 1 & 1 \\ -1 & -2 \end{bmatrix} \text{ and } D = \begin{bmatrix} 5 & 0 \\ 0 & 3 \end{bmatrix}$$

- The columns of P, call them b_1 and b_2, are eigenvectors of A. By Theorem 8, D is the β-matrix for T when $\beta = \{b_1, b_2\}$. The mappings $x \mapsto Ax$ and $u \mapsto Du$ describe the same linear transformation, relative to different bases.

1. Let $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ and $\mathcal{D} = \{\mathbf{d}_1, \mathbf{d}_2\}$ be bases for vector spaces V and W, respectively. Let $T : V \to W$ be a linear transformation with the property that

$$T(\mathbf{b}_1) = 3\mathbf{d}_1 - 5\mathbf{d}_2, \quad T(\mathbf{b}_2) = -\mathbf{d}_1 + 6\mathbf{d}_2, \quad T(\mathbf{b}_3) = 4\mathbf{d}_2$$

Find the matrix for T relative to $\mathcal{B}$ and $\mathcal{D}$.

$$[T(b_1)]_D = \begin{pmatrix} 3 \\ -5 \end{pmatrix} \qquad [T(b_2)]_D = \begin{pmatrix} -1 \\ 6 \end{pmatrix} \qquad [T(b_3)]_D = \begin{pmatrix} 0 \\ 4 \end{pmatrix}$$

$$M = \begin{pmatrix} 3 & -1 & 0 \\ -5 & 6 & 4 \end{pmatrix}$$

In Exercises 11 and 12, find the $\mathcal{B}$-matrix for the transformation $\mathbf{x} \mapsto A\mathbf{x}$, when $\mathcal{B} = \{\mathbf{b}_1, \mathbf{b}_2\}$.

11. $A = \begin{bmatrix} 3 & 4 \\ -1 & -1 \end{bmatrix}, \mathbf{b}_1 = \begin{bmatrix} 2 \\ -1 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$$P = [b_1 \ b_2] = \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix}$$

$$P^{-1} = \frac{1}{5}\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}$$

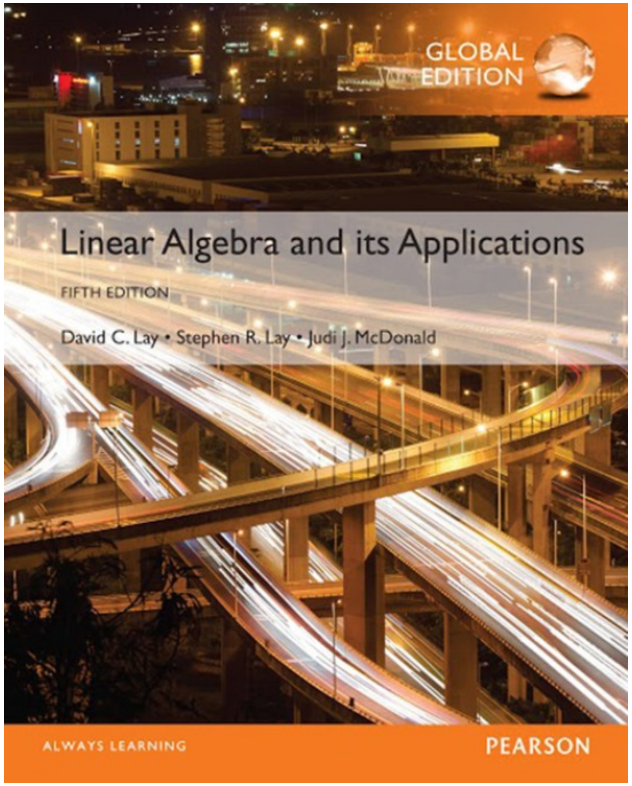
$$P^{-1}AP = \frac{1}{5}\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 3 & 4 \\ -1 & -1 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$$

$$= \frac{1}{5}\begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} 2 & 11 \\ -1 & -3 \end{pmatrix} = \frac{1}{5}\begin{pmatrix} 5 & 25 \\ 0 & 5 \end{pmatrix}$$

5 Eigenvalues and Eigenvectors

5.5

COMPLEX EIGENVALUES



PEARSON

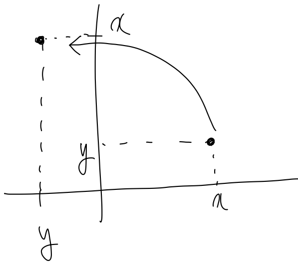
COMPLEX EIGENVALUES

- **Note.** Even for a matrix with real entries, the eigenvalues and eigenvectors may be complex-valued.

- **Example 1** If $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, then the linear transformation $\mathbf{x} \mapsto A\mathbf{x}$ on $\mathbb{R}^2$ rotates the plane counterclockwise through a quarter-turn.
- In fact, the characteristic equation of A is

$$\lambda^2 + 1 = 0$$

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -y \\ x \end{pmatrix}$$



COMPLEX EIGENVALUES

- The only roots are complex: $\lambda = i$ and $\lambda = -i$. However, if we permit A to act on $\mathbb{C}^2$, then

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} i \\ 1 \end{bmatrix} = +i \begin{bmatrix} 1 \\ i \end{bmatrix}$$
$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = \begin{bmatrix} -i \\ 1 \end{bmatrix} = -i \begin{bmatrix} 1 \\ i \end{bmatrix}$$

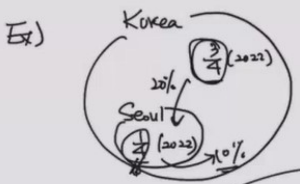
- Thus i and $-i$ are eigenvalues, with $\begin{bmatrix} 1 \\ -i \end{bmatrix}$ and $\begin{bmatrix} 1 \\ i \end{bmatrix}$ as corresponding eigenvectors.

$\lambda = i \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad A - iI \Rightarrow \begin{pmatrix} -i & -1 & 0 \\ 1 & -i & 0 \end{pmatrix} \to \begin{pmatrix} 1 & -i & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad x_1 - i x_2 = 0 \quad \therefore \begin{pmatrix} i \\ 1 \end{pmatrix} \overset{\times(-i)}{化} \begin{pmatrix} 1 \\ -i \end{pmatrix}$

$\lambda = -i \quad A + iI \Rightarrow \begin{pmatrix} 1 & i & 0 \\ i & -1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & i & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad x_1 = -i x_2 \quad \therefore \begin{pmatrix} -i \\ 1 \end{pmatrix} \xrightarrow{\times i} \begin{pmatrix} 1 \\ i \end{pmatrix}$

Given a real value matrix → eigen value & eigenvectors may be complex-valued.

Ex)



Suppose that 10% of the residents in Seoul move out every year and 20% of the people living outside Seoul move in.

$$\begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} 1/4 \\ 3/4 \end{pmatrix}$$

$$\begin{pmatrix} x_n \\ y_n \end{pmatrix} = \begin{pmatrix} \text{the population in Seoul in year } 2022+n \\ \text{—— outside —} \end{pmatrix}$$

In 2023, $\begin{cases} 0.9 \times \frac{1}{4} + 0.2 \times \frac{3}{4} & \text{live in Seoul} \\ 0.1 \times \frac{1}{4} + 0.8 \times \frac{3}{4} & \text{outside} \end{cases}$

In 2024 ⋮

$$\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} 0.9 \times \frac{1}{4} + 0.2 \times \frac{3}{4} \\ 0.1 \times \frac{1}{4} + 0.8 \times \frac{3}{4} \end{pmatrix} = \underbrace{\begin{pmatrix} 0.9 & 0.2 \\ 0.1 & 0.8 \end{pmatrix}}_{A} \begin{pmatrix} 1/4 \\ 3/4 \end{pmatrix}$$

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = A \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = A\left(A\begin{pmatrix} x_0 \\ y_0 \end{pmatrix}\right) = A^2 \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

In 2100 ?

$$A^{78} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix}$$

$$|A - \lambda I| = \left| \begin{pmatrix} 0.9-\lambda & 0.2 \\ 0.1 & 0.8-\lambda \end{pmatrix} \right| = \lambda^2 - 1.7\lambda + 0.72 - 0.02$$
$$= \lambda^2 - 1.7\lambda + 0.7$$
$$= (\lambda - 1)(\lambda - 0.7) = 0$$

$\lambda = 1$

0.7

$\lambda = 1 \rightarrow \begin{pmatrix} -0.1 & 0.2 & 0 \\ 0.1 & -0.2 & 0 \end{pmatrix} \sim \begin{pmatrix} -0.1 & 0.2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $\quad 0.1 x_1 = 0.2 x_2$, $x_1 = 2x_2$ $\quad \boxed{\begin{pmatrix} 2 \\ 1 \end{pmatrix}}$

$0.7 \rightarrow \begin{pmatrix} 0.2 & 0.2 & 0 \\ 0.1 & 0.1 & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ $\quad x_1 = -x_2$ $\quad \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$$A = \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0.7 \end{pmatrix} \frac{1}{-3} \begin{pmatrix} -1 & -1 \\ -1 & 2 \end{pmatrix} = \boxed{\frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0.7 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix}}$$

In 2100 ?

$$A^{78} \begin{pmatrix} x_0 \\ y_0 \end{pmatrix} = \boxed{\frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1^{78} & 0 \\ 0 & 0.7^{78} \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1/4 \\ 3/4 \end{pmatrix}}$$

ex) (Fibonacci sequence)

$a_1 = 1$

$a_2 = 1$

$a_3 = a_1 + a_2 = 2$

$a_4 = a_2 + a_3 = 3$

$a_5 = a_3 + a_4 = 5$

$$\boxed{a_{n+1} = a_n + a_{n-1}}$$

$\vdots$

$a_{10} = ?$

$a_{1000} = ?$

$a_n = a_n$

$a_{n+1} = a_n + a_{n-1}$

$$\begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} a_n \\ a_n + a_{n-1} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}}_{A} \begin{pmatrix} a_{n-1} \\ a_n \end{pmatrix}$$

$$\begin{pmatrix} a_2 \\ a_3 \end{pmatrix} = A \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

$$\begin{pmatrix} a_3 \\ a_4 \end{pmatrix} = A \begin{pmatrix} a_2 \\ a_3 \end{pmatrix} = A\left(A \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}\right) = A^2 \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

$\vdots$

$$\begin{pmatrix} a_{999} \\ \boxed{a_{1000}} \end{pmatrix} = \underline{A}^{998} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = A^{998} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

ex) (Fibonacci sequence)

$$|A - \lambda I| = \begin{vmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = \underline{\lambda^2 - \lambda - 1 = 0}$$

$$\lambda = \frac{1 \pm \sqrt{5}}{2}$$

$A - \lambda I \sim$ singular (i.e. not invertible)

$$\left(\begin{array}{cc|c} -\lambda & 1 & 0 \\ 1 & 1-\lambda & 0 \end{array}\right) \sim \left(\begin{array}{cc|c} -\lambda & 1 & 0 \\ 0 & 0 & 0 \end{array}\right)$$

$\Rightarrow$ Two ~~eqn~~ rows are dependent

$\Rightarrow -\lambda x_1 + x_2 = 0 \quad x_2 = \lambda x_1 \quad \Rightarrow \begin{pmatrix} 1 \\ \lambda \end{pmatrix} \sim$ eigenvector

$$\lambda_1 = \frac{1+\sqrt{5}}{2}$$

$$\lambda_2 = \frac{1-\sqrt{5}}{2}$$

$$\begin{pmatrix} a_n \\ a_{n+1} \end{pmatrix} = \begin{pmatrix} a_n \\ a_n + a_{n-1} \end{pmatrix} = \underbrace{\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}}_{A} \begin{pmatrix} a_{n-1} \\ a_n \end{pmatrix}$$

$$A = \underline{\begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix}} \underline{\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}} \underline{\frac{1}{\lambda_2 - \lambda_1} \begin{pmatrix} \lambda_2 & -1 \\ -\lambda_1 & 1 \end{pmatrix}}$$

PDP^{-1}

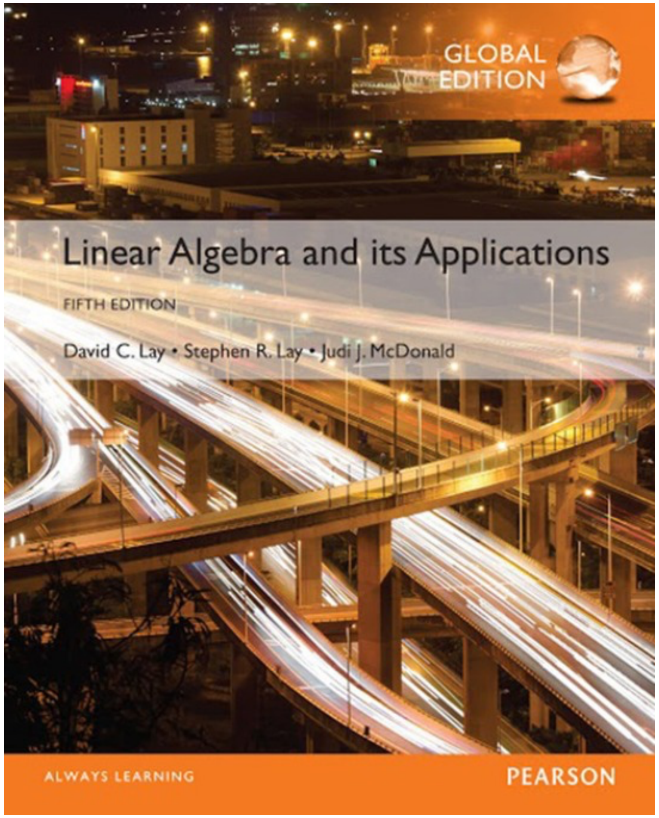
$$\begin{pmatrix} a_{999} \\ \boxed{a_{1000}} \end{pmatrix} = \underline{A}^{998} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = A^{998} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 \\ \lambda_1 & \lambda_2 \end{pmatrix} \begin{pmatrix} \lambda_1^{998} & 0 \\ 0 & \lambda_2^{998} \end{pmatrix} \frac{1}{\lambda_2 - \lambda_1} \begin{pmatrix} \lambda_2 & 1 \\ -\lambda_1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

6 Orthogonality and Least Squares

6.1

INNER PRODUCT, LENGTH, AND ORTHOGONALITY



INNER PRODUCT

- If **u** and **v** are vectors in $\mathbb{R}^n$, then we regard **u** and **v** as $n \times 1$ matrices.

- The transpose $\mathbf{u}^T$ is a $1 \times n$ matrix, and the matrix product $\mathbf{u}^T\mathbf{v}$ is a 1×1 matrix, which we write as a single real number (a scalar) without brackets.

- The number $\mathbf{u}^T\mathbf{v}$ is called the **inner product** of **u** and **v**, and it is written as $u \cdot v$.

  scalar / dot product

- This inner product is also referred to as a **dot product**.

INNER PRODUCT

- If $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$,

ex) $u = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ $v = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$

$u \cdot v = u^T v = 1 \cdot 1 + 2 \cdot 1 + 3 \cdot 1 = 6$

X: $v u^T = = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (1\ 2\ 3)$

$= \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}$

then the inner product of **u** and **v** is

$$\boxed{\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.}$$

INNER PRODUCT

- **Theorem 1:** Let **u**, **v**, and **w** be vectors in $\mathbb{R}^n$, and let c be a scalar. Then
  a. $u \cdot v = v \cdot u$ (commutative)
  b. $(u + v) \cdot w = u \cdot w + v \cdot w$ (distributive)
  c. $(cu) \cdot v = c(u \cdot v) = u \cdot (cv)$
  d. $u \cdot u \geq 0$, and $u \cdot u = 0$ if and only if $\mathbf{u} = 0$

INNER PRODUCT

- **Theorem 1:** Let **u**, **v**, and **w** be vectors in $\mathbb{R}^n$, and let c be a scalar. Then
  - *a.* $u \cdot v = v \cdot u$
  - *b.* $(u + v) \cdot w = u \cdot w + v \cdot w$
  - *c.* $(cu) \cdot v = c(u \cdot v) = u \cdot (cv)$
  - *d.* $u \cdot u \geq 0$, and $u \cdot u = 0$if and only if $\mathrm{u} = 0$

- Properties (b) and (c) can be combined several times to produce the following useful rule
  $(c_1 u_1 + \ldots + c_p u_p) \cdot w = c_1(u_1 \cdot w) + \cdots + c_p(u_p \cdot w)$

THE LENGTH OF A VECTOR

- If **v** is in $\mathbb{R}^n$, with entries $v_1, \ldots, v_n$, then the square root of $v \cdot v$ is defined because $v \cdot v$ is nonnegative.

- **Definition:** The **length** (or **norm**) of **v** is the nonnegative scalar $\|\mathrm{v}\|$ defined by

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}, \text{ and } \|v\|^2 = v \cdot v^2$$

THE LENGTH OF A VECTOR

- If **v** is in $\mathbb{R}^n$, with entries $v_1, \ldots, v_n$, then the square root of $v \cdot v$ is defined because $v \cdot v$ is nonnegative.

- **Definition:** The **length** (or **norm**) of **v** is the nonnegative scalar $\|\mathrm{v}\|$ defined by

$$\|v\| = \sqrt{v \cdot v} = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}, \text{ and } \|v\|^2 = v \cdot v^2$$

- For any scalar c, the length $c\mathbf{v}$ is $|c|$ times the length of v. That is,

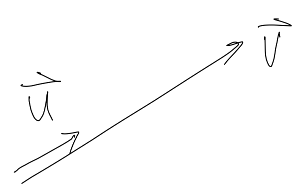
$$\|c\mathrm{v}\| = |c|\|\mathrm{v}\|$$

THE LENGTH OF A VECTOR

- A vector whose length is 1 is called a **unit vector**.

- If we *divide* a nonzero vector $\mathbf{v}$ by its length—that is, multiply by $1/\|\mathbf{v}\|$—we obtain a unit vector $\mathbf{u}$ because the length of $\mathbf{u}$ is $(1/\|\mathbf{v}\|)\|\mathbf{v}\|$. $=1$

- The process of creating $\mathbf{u}$ from $\mathbf{v}$ is sometimes called **normalizing v**, and we say that $\mathbf{u}$ is *in the same direction* as $\mathbf{v}$.



THE LENGTH OF A VECTOR

- **Example 2:** Let $v = (1, -2, 2, 0)$. Find a unit vector **u** in the same direction as **v**.

THE LENGTH OF A VECTOR

- **Example 2:** Let $v = (1, -2, 2, 0)$. Find a unit vector **u** in the same direction as **v**.
- **Solution:** First, compute the length of **v**:

$$\|v\|^2 = v \cdot v = (1)^2 + (-2)^2 + (2)^2 + (0)^2 = 9$$

$$\|v\| = \sqrt{9} = 3$$

- Then, multiply **v** by $1/\|v\|$ to obtain

$$u = \frac{1}{\|v\|}v = \frac{1}{3}v = \frac{1}{3}\begin{bmatrix} 1 \\ -2 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/3 \\ -2/3 \\ 2/3 \\ 0 \end{bmatrix}$$

DISTANCE IN $\mathbb{R}^n$

- To check that $\|\mathbf{u}\| = 1$, it suffices to show that $\|\mathbf{u}\|^2 = 1$.

$$\|\mathbf{u}\|^2 = u \cdot u = \left(\frac{1}{3}\right)^2 + \left(-\frac{2}{3}\right) + \left(\frac{2}{3}\right)^2 + (0)^2$$

$$= \frac{1}{9} + \frac{4}{9} + \frac{4}{9} + 0 = 1$$

DISTANCE IN $\mathbb{R}^n$

- **Definition:** For $\mathbf{u}$ and $\mathbf{v}$in $\mathbb{R}^n$, the **distance between u and v**, written as dist ($\mathbf{u}$, $\mathbf{v}$), is the length of the vector $\mathrm{u} - \mathrm{v}$. That is,

$$\text{dist}\,(\mathrm{u},\mathrm{v}) = \|\mathrm{u} - \mathrm{v}\|$$

DISTANCE IN $\mathbb{R}^n$

- **Example 4:** Compute the distance between the vectors $\mathbf{u} = (7,1)$ and $\mathbf{v} = (3,2)$.

- **Solution:** Calculate

$$\mathbf{u} - \mathbf{v} = \begin{bmatrix} 7 \\ 1 \end{bmatrix} - \begin{bmatrix} 3 \\ 2 \end{bmatrix} = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$$

$$\|\mathbf{u} - \mathbf{v}\| = \sqrt{4^2 + (-1)^2} = \sqrt{17}$$

ORTHOGONAL VECTORS

- **Definition:** Two vectors **u** and **v** in $\mathbb{R}^n$ are **orthogonal** (to each other) if $u \cdot v = 0$.

ex) $e_1 = \begin{pmatrix}1\\0\end{pmatrix}$ $e_2 = \begin{pmatrix}0\\1\end{pmatrix}$ $e_1 \cdot e_2 = 0$ $\to$ $e_1 \perp e_2$

ex) $v_1 = \begin{pmatrix}1\\1\end{pmatrix}$ $v_2 = \begin{pmatrix}1\\-1\end{pmatrix}$ $v_1 \cdot v_2 = 0$ $\to$ $v_1 \perp v_2$

ex) $v_3 \begin{pmatrix}0\\0\end{pmatrix}$ $v_1 \cdot v_3 = 0$, $v_2 \cdot v_3 = 0$ $\to$ $v_1 \perp v_3$, $v_2 \perp v_3$

ORTHOGONAL VECTORS

- **Definition:** Two vectors **u** and **v** in $\mathbb{R}^n$ are **orthogonal** (to each other) if $u \cdot v = 0$.

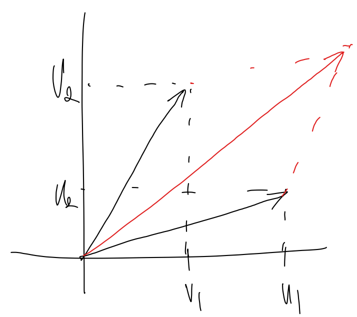
- The zero vector is orthogonal to every vector in $\mathbb{R}^n$ because $0^T \cdot \mathrm{v} = 0$ for all **v**.

주의!

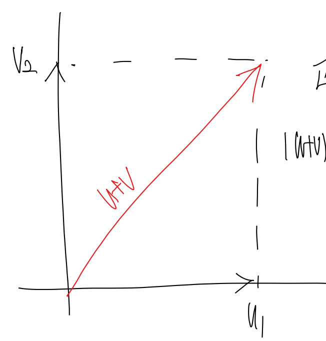
THE PYTHOGOREAN THEOREM

- **Theorem 2:** Two vectors **u** and **v** are orthogonal if and only if $\|\mathbf{u}+\mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$.

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



$$u+v = \begin{pmatrix} u_1+v_1 \\ u_2+v_2 \end{pmatrix}$$



고라스정리

$|(u+v)|^2 = |u|^2 + |v|^2$

THE PYTHOGOREAN THEOREM

Orthogonal Complements

정의

- If a vector $\mathbf{z}$ is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then $\mathbf{z}$ is said to be **orthogonal to** W.

$z \perp u$ if $z \cdot u = 0$

$z \perp W$ if $z \perp u$ for every $u \in W$

THE PYTHOGOREAN THEOREM

Orthogonal Complements

- If a vector **z** is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then **z** is said to be **orthogonal to** W.
- The set of all vectors **z** that are orthogonal to W is called the **orthogonal complement** of W and is denoted by $W^{\perp}$ (and read as "W perpendicular" or simply "Wperp").

$z \perp u$ if $z \cdot u = 0$

$z \perp W$ if $z \perp u$ for every $u \in W$

$W^{\perp} = \{ z \mid z \perp W \}$

ex) $0 \perp$ (every vector in W) $\rightarrow 0 \in W^{\perp}$

항상 포함.

Orthogonal Complements

- If a vector **z** is orthogonal to every vector in a subspace W of $\mathbb{R}^n$, then **z** is said to be **orthogonal to** W.
- The set of all vectors **z** that are orthogonal to W is called the **orthogonal complement** of W and is denoted by $W^{\perp}$ (and read as "W perpendicular" or simply "W perp").

1. A vector **x** is in $W^{\perp}$ if and only if **x** is orthogonal to every vector in a set that spans W.
2. $W^{\perp}$ is a subspace of $\mathbb{R}^n$.

generating set

ex) $x \in W^{\perp} \iff x \perp v_i$ where $\{v_1 \dots v_n\} \sim$ basis for W

$x \in W^\perp \iff \underline{x \perp W \iff x \cdot u = 0 \text{ for every}}\ u \in W$

$W = \mathrm{span}\{v_1, \cdots v_n\}$

If $x \in W^\perp$, then $x \perp v_i$ because $v_i \in W$

If $x \perp v_i$ for each i, then each $u \in W$ can be written as

$$u = c_1 v_1 + \cdots + c_n v_n$$

$$\Rightarrow x \cdot u = x \cdot (c_1 v_1 + \cdots + c_n v_n) = c_1 (\underline{x \cdot v_1}) + \cdots + c_n (\underline{x \cdot v_n}) = 0$$

$\underline{u, v \in W^\perp}$

$\hookrightarrow$ This implies $u, v \perp W \Rightarrow \begin{matrix} u \cdot z = 0 \\ v \cdot z = 0 \end{matrix}$ for any $z \in W$

For any $z \in W$ and $c \in \mathbb{R}$

$(u+v) \cdot z = \underline{u \cdot z} + \underline{v \cdot z} = 0 \quad \Rightarrow u + v \in W^\perp$

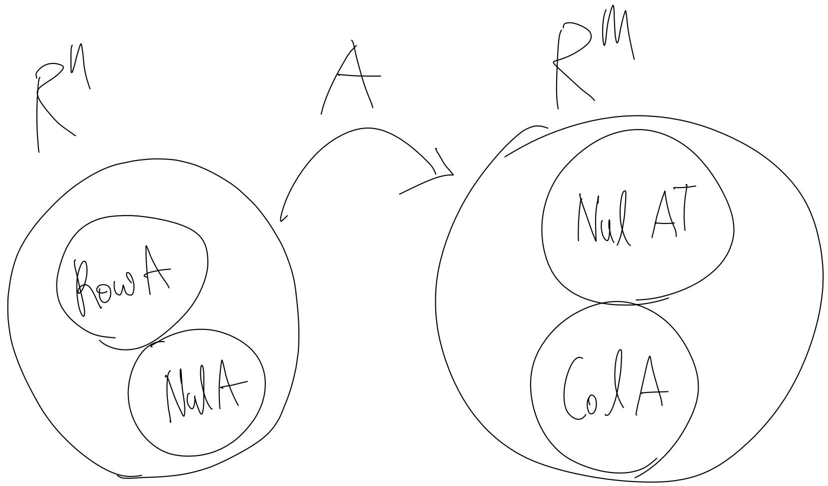
$(cu) \cdot z = c(u \cdot z) = 0 \qquad cu \in W^\perp$

ORTHOGONAL COMPLEMENTS

- **Theorem 3:** Let A be an $m \times n$ matrix. The orthogonal complement of the row space of A is the null space of A, and the orthogonal complement of the column space of A is the null space of A^T:

$\{x \mid Ax=0\}$

$$(\text{Row } A)^{\perp} = \text{Nul } A \text{ and } (\text{Col } A)^{\perp} = \text{Nul } A^T$$



$\{y \mid A^T y = 0\}$
$= \{y \mid y^T A = 0\}$ (Left Null Space)

ORTHOGONAL COMPLEMENTS

- **Proof:** The row-column rule for computing $A\mathbf{x}$ shows that if $\mathbf{x}$ is in Nul$A$, then $\mathbf{x}$ is orthogonal to each row of A (with the rows treated as vectors in $\mathbb{R}^n$).

$x \in \text{Nul} A \Rightarrow Ax = 0$

$$\underset{m \times n}{\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_m \end{bmatrix}} x = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \rightarrow \left.\begin{array}{l} A_1 x = 0 \\ A_2 x = 0 \\ \vdots \\ A_m x = 0 \end{array}\right\} \left.\begin{array}{l} A_1 \perp x \\ A_2 \perp x \\ \vdots \\ A_m \perp x \end{array}\right\} \rightarrow \text{perpendicular to all row}$$

- Since the rows of A span the row space, $\mathbf{x}$ is orthogonal to Row A.

the rows of A form a generating set.

- Conversely, if $\mathbf{x}$ is orthogonal to Row A, then $\mathbf{x}$ is certainly orthogonal to each row of A, and hence $A\mathbf{x} = 0$.

$$\begin{bmatrix} A_1 \\ A_2 \\ \\ \end{bmatrix} x = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \text{Row} A \perp \text{Nul} A \rightarrow \text{Row} A^{\perp} = \text{Nul} A$$

- This proves the first statement of the theorem.

ORTHOGONAL COMPLEMENTS

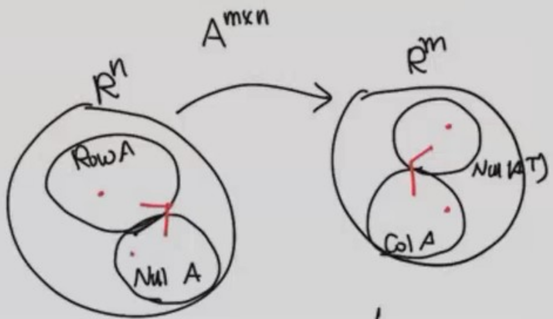
$\text{Row}A^{\perp} = \text{Nul}A$

- Since this statement is true for any matrix, it is true for A^T.

$\text{col}A^{\perp} = \text{Row}(A^T)^{\perp} = \text{Nul}A^T$ 이된다.

- That is, the orthogonal complement of the row space of A^T is the null space of A^T.

- This proves the second statement, because $\text{Row}\,A^T = \text{Col}\,A$.



~~Let $u \in W$ (≠0) and $v \in W^\perp$ (≠0)~~
~~then u and v~~ are
~~linearly~~ independent.

$\text{Row}\,A^\perp = \text{Nul}\,A$

$\text{Col}\,A^\perp = \text{Nul}\,A^T$

원래는 u,v 0일수있지만
이론을 성립하려면
u, v ≠0 추가해라.

Let $u \in W$ $v \in W^\perp$ (then $u \cdot v = 0$)

and, $\boxed{\text{u and v are nonzero}}$

Q. u and v are linearly independent   True ? False ?

Set $c_1 u + c_2 v = 0$

$0 = u\cdot(c_1u + c_2v) = c_1(u\cdot u) + c_2\cancel{(u\cdot v)}$ $c_1(u\cdot u) \overset{\neq 0}{} \Rightarrow \boxed{c_1 = 0}$ $u \neq 0$

$0 = v\cdot(c_1u + c_2v) = c_1\cancel{(v\cdot u)}^{0} + c_2(v\cdot v)$ $c_2(v\cdot v) = 0 \Rightarrow c_2 = 0$ $v \neq 0$

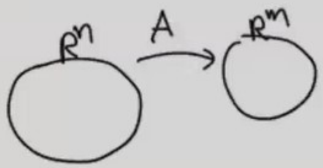
u & v are linearly independent.

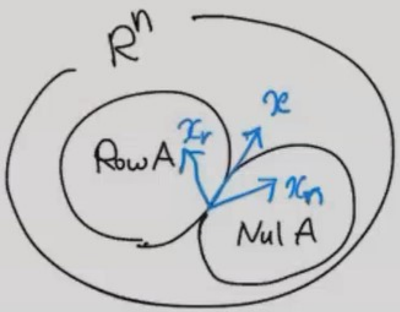
- dim Row A = rank = # pivot columns in A = # basic variables
  dim Nul A = n − rank = # free variables
  (n = # unknowns)

a basis of Row A = non-zero rows in an echelon form of A   → basis 구성

a basis of Nul A   Ax=0 이용   ← linearly independent vectors   dependent 한게 없음

$\boxed{\underset{\text{Independent vectors in } \mathbb{R}^n}{\overset{\text{rank}}{(\text{a basis of Row }A)}} \cup \underset{\text{independent vectors in } \mathbb{R}^n}{\overset{\text{dim Nul }A}{(\text{a basis of Nul }A)}} = \overset{n}{\text{linearly independent basis for } \mathbb{R}^n}}$





a basis for Row A $\cup$ a basis for $\underline{\text{Nul } A}$

= a basis for R^n

Let $\{v_1 \cdots v_r\}$ be a basis for Row A

$\{v_{r+1}, \cdots, v_n\}$ $\underline{\text{Nul } A}$

Then $\{v_1, \cdots, v_n\}$ is a basis for R^n

Each $x \in R^n$ can be written as $x = \overbrace{\boxed{c_1 v_1 + \cdots + c_r v_r}}^{x_r} + \overbrace{\boxed{c_{r+1} v_{r+1} + \cdots + c_n v_n}}^{x_n}$

$$Ax = A(c_1 v_1 + \cdots + c_n v_n) = c_1 A v_1 + \cdots + c_n A v_n$$
$$= c_1 A v_1 + \cdots + c_r A v_r + \underbrace{c_{r+1} A v_{r+1} + \cdots + c_n A v_n}_{0}$$
$$= A(\underline{c_1 v_1 + \cdots + c_r v_r})$$

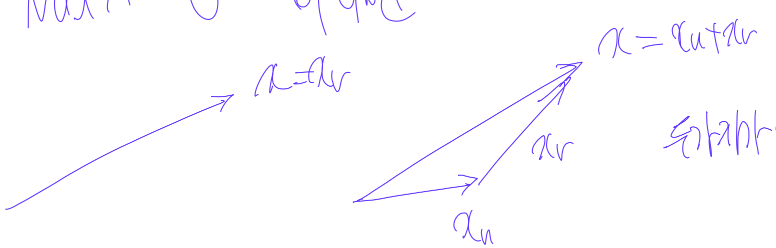
$A x_r$

$$\begin{cases} x = x_r + x_n \\ Ax = Ax_r \end{cases}$$

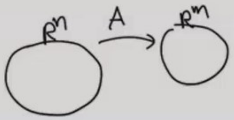
The mapping defined by Ax is 1-1 if $\underline{\underline{x_n = 0}}$ i.e. $\text{Nul } A = \{0\}$

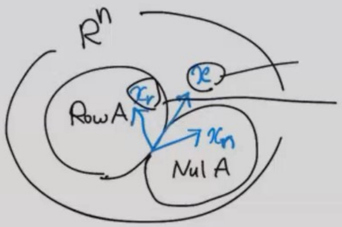
Why?

Nul A = 0 아니면



두개가 가능해진다.





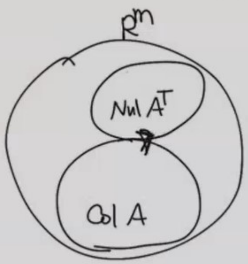
$$\mathrm{Row}\,A^{\perp} = \mathrm{Nul}\,A$$

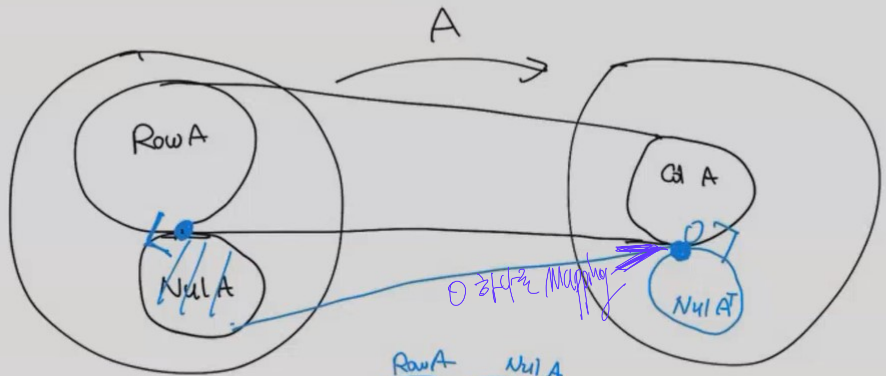
$$\downarrow A^T$$

$$(\mathrm{Row}\,A^T)^{\perp} = \mathrm{Nul}\,A^T$$

$$\|$$

$$\mathrm{Col}\,A^{\perp}$$





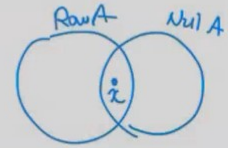
$$x \in \mathrm{Row}\,A \cap \mathrm{Nul}\,A$$

$$\Downarrow$$

$$x \cdot x = 0$$

$$\Downarrow$$

$$x = 0$$



$$\mathrm{Row}\,A \cap \mathrm{Nul}\,A = \{0\}$$

$$\mathrm{Col}\,A \cap \mathrm{Nul}\,A^T = \{0\}$$

10.

In Exercises 9–12, find a unit vector in the direction of the given vector.

10. $\begin{bmatrix} -6 \\ 4 \\ -3 \end{bmatrix}$

$\begin{bmatrix} -6 \\ 4 \\ -3 \end{bmatrix}$ $\|v\|^2 = 6^2 + 4^2 + 3^2 = 61$

24.

24. Verify the *parallelogram law* for vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$:

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

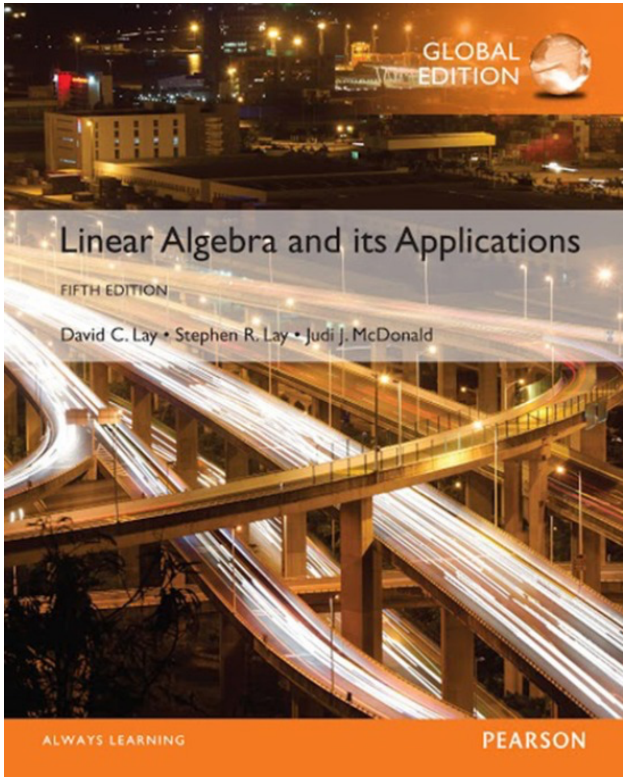
$|u+v|^2 + |u-v|^2 = (u+v)\cdot(u+v) + (u-v)\cdot(u-v)$

$= |u|^2 + \cancel{2u\cdot v} + |v|^2 + |u|^2 - \cancel{2u\cdot v} + |v|^2 = 2|u|^2 + 2|v|^2$

6 Orthogonality and Least Squares

6.2

ORTHOGONAL SETS



ORTHOGONAL SETS

- A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_p\}$ in $\mathbb{R}^n$ is said to be an **orthogonal set** if each pair of distinct vectors from the set is orthogonal, that is, if $u_i \cdot u_j = 0$ whenever $i \neq j$.

ex) $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$, $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$, $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$

→ 0. An orthogonal set is linearly independent

→ 아 ㄲ 0이 있을수 없어서 안된다고 ㄲ

ORTHOGONAL SETS

원래 거에서 가져와서

- **Theorem 4:** If $S = \{\mathrm{u}_1, \ldots, \mathrm{u}_p\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then S is linearly independent and hence is a basis for the subspace spanned by S.

ORTHOGONAL SETS

- **Proof:** If $0 = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$ for some scalars $c_1,\ldots,c_p$, then

$$0 = 0\bullet\mathbf{u}_1 = (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_p\mathbf{u}_p)\bullet\mathbf{u}_1$$
$$= (c_1\mathbf{u}_1)\bullet\mathbf{u}_1 + (c_2\mathbf{u}_2)\bullet\mathbf{u}_1 + \cdots + (c_p\mathbf{u}_p)\bullet\mathbf{u}_1$$
$$= c_1(\mathbf{u}_1\bullet\mathbf{u}_1) + c_2(\mathbf{u}_2\bullet\mathbf{u}_1) + \cdots + c_p(\mathbf{u}_p\bullet\mathbf{u}_1)$$
$$= c_1(\mathbf{u}_1\bullet\mathbf{u}_1)$$

$\neq 0$

because $\mathbf{u}_1$ is orthogonal to $\mathbf{u}_2,\ldots,\mathbf{u}_p$.

- Since $\mathbf{u}_1$ is nonzero, $u_1 \cdot u_1$ is not zero and so $c_1 = 0$
- Similarly, $c_2,\ldots,c_p$ must be zero and S is linearly independent

ORTHOGONAL SETS

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

ex) $\{e_1\ e_2\ e_n\}$ in $\mathbb{R}^3$

$\left\{ \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ in $\mathbb{R}^3$

$\frac{\ell}{0}$ ...

ORTHOGONAL SETS

cannot be zero (L.I.)

- **Theorem 5:** Let $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$. For each $\mathbf{y}$ in W, the weights in the linear combination $\mathbf{y} = c_1\mathbf{u}_1 + \cdots + c_p\mathbf{u}_p$

  are given by $c_j = \dfrac{\mathbf{y}\bullet\mathbf{u}_j}{\mathbf{u}_j\bullet\mathbf{u}_j} \quad (j = 1,\ldots,p)$

$\therefore\ y = \left(\frac{y\cdot u_1}{u_1\cdot u_1}\right)u_1 + \cdots + \left(\frac{y\cdot u_p}{u_p\cdot u_p}\right)u_p$

$y\cdot u_j = c_j(u_j\cdot u_j)$

$y\cdot u_j = (c_1u_1 + \cdots + c_pu_p)\cdot u_j$

$= c_1\cancel{(u_1\cdot u_j)}^{0} + \cdots + c_p\cancel{(u_p\cdot u_j)}^{0} = c_j(u_j\cdot u_j)$

ORTHOGONAL SETS

- **Proof:** The orthogonality of $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ shows that

$$y \cdot u_1 = (c_1\mathrm{u}_1 + c_2\mathrm{u}_2 + \cdots + c_p\mathrm{u}_p)\bullet\mathrm{u}_1 = c_1(\mathrm{u}_1\bullet\mathrm{u}_1)$$

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

- To find c_j for $j = 2,\ldots, p$, compute $y \cdot u_j$ and solve for c_j in a similar way.

AN ORTHOGONAL PROJECTION

- Given a nonzero vector **u** in $\mathbb{R}^n$, consider the problem of decomposing a vector **y** in $\mathbb{R}^n$ into the sum of two vectors, one a multiple of **u** and the other orthogonal to **u**.
- We wish to write

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

where $\hat{\mathbf{y}} = \alpha \mathbf{u}$ for some scalar α and **z** is some vector orthogonal to **u**.

AN ORTHOGONAL PROJECTION

- Given any scalar α, let $z = y - \alpha u$, so that (1) is satisfied.

given known

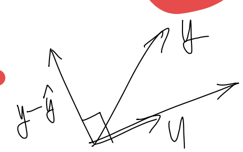
- Then $y - \hat{y}$ is orthogonal to **u** if an only if

$$0 = (y - \alpha u)\bullet u = y\bullet u - (\alpha u)\bullet u = y\bullet u - \alpha(u\bullet u)$$

- That is, (1) is satisfied with **z** orthogonal to **u** if and only if $\alpha = \dfrac{y\bullet u}{u\bullet u}$ and $\hat{y} = \dfrac{y\bullet u}{u\bullet u}u.$

← 구하는 대상 $\alpha, \hat{y}$

- The vector $\hat{y}$ is called the **orthogonal projection of y onto u**, and the vector **z** is called the **component of y orthogonal to u**.

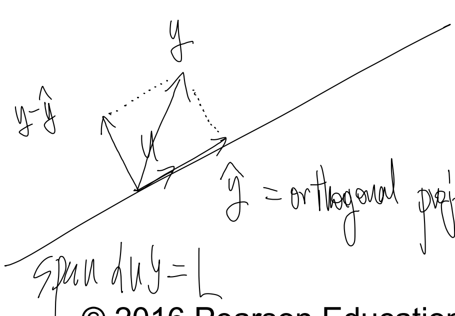


$\hat{y} = \left(\frac{y\cdot u}{u\cdot u}\right)u$

AN ORTHOGONAL PROJECTION

- If c is any nonzero scalar and if $\mathbf{u}$ is replaced by $c\mathbf{u}$ in the definition of $\hat{\mathbf{y}}$, then the orthogonal projection of $\mathbf{y}$ onto $c\mathbf{u}$ is exactly the same as the orthogonal projection of $\mathbf{y}$ onto $\mathbf{u}$.
- Sometimes $\hat{\mathbf{y}}$ is denoted by $\text{proj}_L\mathbf{y}$ and is called the **orthogonal projection of y onto** L.
- That is, $\text{proj}_u y =$

$$\hat{\mathbf{y}} = \text{proj}_L \mathbf{y} = \frac{\mathbf{y} \bullet \mathbf{u}}{\mathbf{u} \bullet \mathbf{u}}\mathbf{u} = \frac{y^T u}{u^T u} u \qquad (2)$$



$\hat{y}$ = orthogonal projection onto u

$$\hat{y} = \overset{\in \mathbb{R}}{\left(\frac{y \cdot u}{u \cdot u}\right)} u$$

$$= u \, \frac{y \cdot u}{u \cdot u}$$

$$= u \, \frac{u^T y}{u^T u}$$

$$= \boxed{\left(\frac{\overset{\text{matrix}}{u u^T}}{\underset{\text{number}}{u^T u}}\right)} \overset{\text{vector}}{y}$$

$$\underline{\text{matrix}} = P$$

Note that P is independent for y

AN ORTHOGONAL PROJECTION

- **Example 3:** Let $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Find the orthogonal projection of **y** onto **u**. Then write **y** as the sum of two orthogonal vectors, one in Span$\{\mathbf{u}\}$ and one orthogonal to **u**.

AN ORTHOGONAL PROJECTION

- **Example 3:** Let $\mathbf{y} = \begin{bmatrix} 7 \\ 6 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Find the orthogonal projection of **y** onto **u**. Then write **y** as the sum of two orthogonal vectors, one in Span$\{\mathbf{u}\}$ and one orthogonal to **u**.
- **Solution:** Compute

$$y \cdot u = \begin{bmatrix} 7 \\ 6 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 40$$

$$\mathrm{u} \cdot u = \begin{bmatrix} 4 \\ 2 \end{bmatrix} \cdot \begin{bmatrix} 4 \\ 2 \end{bmatrix} = 20$$

AN ORTHOGONAL PROJECTION

- The orthogonal projection of **y** onto **u** is

$$\hat{\mathrm{y}} = \frac{y \cdot u}{u \cdot u}\mathrm{u} = \frac{40}{20}\mathrm{u} = 2\begin{bmatrix} 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix}$$

and the component of **y** orthogonal to **u** is

$$\mathrm{y} - \hat{\mathrm{y}} = \begin{bmatrix} 7 \\ 6 \end{bmatrix} - \begin{bmatrix} 8 \\ 4 \end{bmatrix} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

- The sum of these two vectors is **y**.

AN ORTHOGONAL PROJECTION

- That is,

$$\begin{bmatrix} 7 \\ 6 \end{bmatrix} = \begin{bmatrix} 8 \\ 4 \end{bmatrix} + \begin{bmatrix} -1 \\ 2 \end{bmatrix}$$

$$\underset{\mathbf{y}}{\uparrow} \qquad\qquad \underset{\hat{\mathbf{y}}}{\uparrow} \qquad\qquad \underset{(\mathbf{y}-\hat{\mathbf{y}})}{\uparrow}$$

orthogonal : 서로직교
+ normalized : 길이=1

ORTHONORMAL SETS

- A set $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ is an **orthonormal set** if it is an orthogonal set of unit vectors.

ex) An orthonormal set is linearly independent (no zero vector)

- If W is the subspace spanned by such a set, then $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ is an **orthonormal basis** for W, since the set is automatically linearly independent, by Theorem 4.

- The simplest example of an orthonormal set is the standard basis $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ for $\mathbb{R}^n$.

- Any nonempty subset of $\{\mathbf{e}_1,\ldots,\mathbf{e}_n\}$ is orthonormal, too.

ORTHONORMAL SETS

- **Example 2:** Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of $\mathbb{R}^3$, where

$$\mathrm{v}_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \quad \mathrm{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \quad \mathrm{v}_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

ORTHONORMAL SETS

- **Example 2:** Show that $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal basis of $\mathbb{R}^3$, where

$$\mathrm{v}_1 = \begin{bmatrix} 3/\sqrt{11} \\ 1/\sqrt{11} \\ 1/\sqrt{11} \end{bmatrix}, \quad \mathrm{v}_2 = \begin{bmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{bmatrix}, \quad \mathrm{v}_3 = \begin{bmatrix} -1/\sqrt{66} \\ -4/\sqrt{66} \\ 7/\sqrt{66} \end{bmatrix}$$

- **Solution:** Compute

$$v_1 \cdot v_2 = -3/\sqrt{66} + 2/\sqrt{66} + 1/\sqrt{66} = 0$$

$$v_1 \cdot v_3 = -3/\sqrt{726} - 4/\sqrt{726} + 7/\sqrt{726} = 0$$

ORTHONORMAL SETS

$$v_2 \bullet v_3 = 1/\sqrt{396} - 8/\sqrt{396} + 7/\sqrt{396} = 0$$

- Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal set.
- Also, $v_1 \cdot v_1 = 9/11 + 1/11 + 1/11 = 0$

$$v_2 \cdot v_2 = 1/6 + 4/6 + 1/6 = 1$$

$$v_3 \cdot v_3 = 1/66 + 16/66 + 49/66 = 1$$

which shows that $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ are unit vectors.

- Thus $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthonormal set.

ORTHONORMAL SETS

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

ORTHONORMAL SETS

$\begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$

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

$u_i^T u_j = u_i \cdot u_j = 0 \quad (i \neq j)$

$u_i^T u_i = u_i \cdot u_i = 1$

- **Proof:** To simplify notation, we suppose that U has only three columns, each a vector in $\mathbb{R}^m$.
- Let $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix}$ and compute

$$U^T U = \begin{bmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T\mathbf{u}_1 & \mathbf{u}_1^T\mathbf{u}_2 & \mathbf{u}_1^T\mathbf{u}_3 \\ \mathbf{u}_2^T\mathbf{u}_1 & \mathbf{u}_2^T\mathbf{u}_2 & \mathbf{u}_2^T\mathbf{u}_3 \\ \mathbf{u}_3^T\mathbf{u}_1 & \mathbf{u}_3^T\mathbf{u}_2 & \mathbf{u}_3^T\mathbf{u}_3 \end{bmatrix} \tag{4}$$

ORTHONORMAL SETS

- The entries in the matrix at the right are inner products, using transpose notation.
- The columns of U are orthogonal if and only if

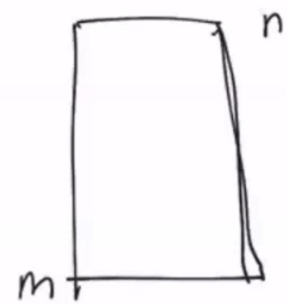
$$\mathbf{u}_1^T\mathbf{u}_2 = \mathbf{u}_2^T\mathbf{u}_1 = 0, \mathbf{u}_1^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_1 = 0, \mathbf{u}_2^T\mathbf{u}_3 = \mathbf{u}_3^T\mathbf{u}_2 = 0 \tag{5}$$

- The columns of U all have unit length if and only if

$$\mathbf{u}_1^T\mathbf{u}_1 = 1,\ \mathbf{u}_2^T\mathbf{u}_2 = 1,\ \mathbf{u}_3^T\mathbf{u}_3 = 1 \tag{6}$$

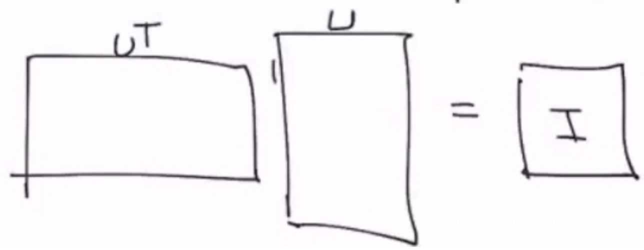
- The theorem follows immediately from (4)–(6).

An $m \times n$ matrix $\underline{U}$ has <u>orthonormal columns</u>



Then $\underline{\underline{m}} \geq n$

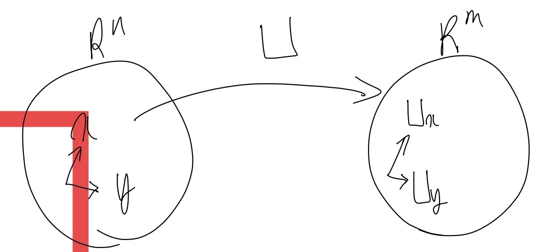
(Otherwise the columns are dependent)



orthonormal $\subset$ Linearly Independent

ORTHONORMAL SETS

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



a. $(Ux)\cdot(Uy) = x \cdot y$

b. $(Ux)\cdot(Uy) = 0$ if and only if $x \cdot y = 0$

c. $\|U\mathbf{x}\| = \|\mathbf{x}\|$

- Properties (b) and (c) say that the linear mapping $\mathbf{x} \mapsto U\mathbf{x}$ preserves lengths and orthogonality. → (b) (c)

(a) → Inner product is preserved

$(Ux)\cdot(Uy) = (Ux)^T(Uy) = x^T U^T U y = x^T y = x \cdot y$

$\|Ux\|^2 = (Ux)\cdot(Ux) = x \cdot x = \|x\|^2$

11.

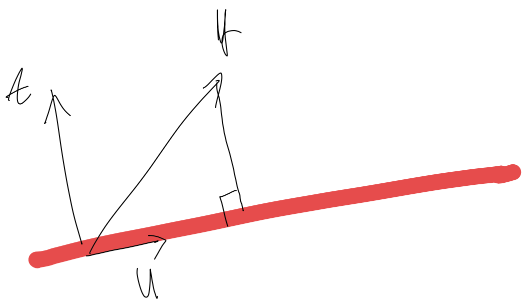
11. Compute the orthogonal projection of $\overset{\overset{y}{\shortparallel}}{\begin{bmatrix} 1 \\ 7 \end{bmatrix}}$ onto the line through $\underset{\underset{u}{\shortparallel}}{\begin{bmatrix} -4 \\ 2 \end{bmatrix}}$ and the origin.

$$\hat{y} = \left(\frac{y \cdot u}{u \cdot u}\right) \cdot u = \frac{-4+14}{16+4} \begin{pmatrix} -4 \\ 2 \end{pmatrix} = \frac{10}{20} \begin{pmatrix} -4 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

$$z = y - \hat{y} = \begin{pmatrix} 1 \\ 7 \end{pmatrix} - \begin{pmatrix} -2 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 6 \end{pmatrix} \quad \therefore y = \begin{pmatrix} \; \\ \; \end{pmatrix} + \begin{pmatrix} \; \\ \; \end{pmatrix}$$

15.

5. Let $\mathbf{y} = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$ and $\mathbf{u} = \begin{bmatrix} 8 \\ 6 \end{bmatrix}$. Compute the distance from $\mathbf{y}$ to the line through $\mathbf{u}$ and the origin.



$$\text{dist}(y, L) = \|z\| = \|y - \hat{y}\|$$

$$\hat{y} = \frac{24+6}{64+36}\begin{pmatrix} 8 \\ 6 \end{pmatrix}$$

$$= \frac{3}{5}\begin{pmatrix} 4 \\ 3 \end{pmatrix} = \begin{pmatrix} 12/5 \\ 9/5 \end{pmatrix}$$

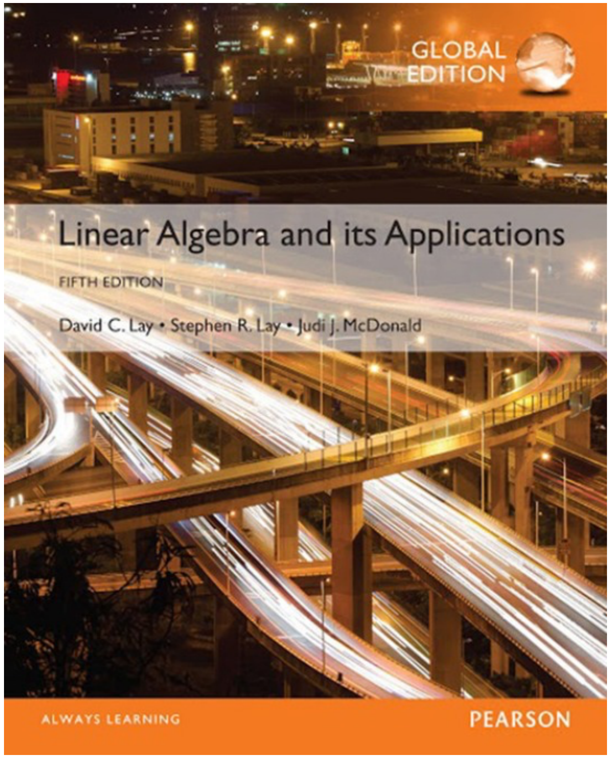
$$z = \begin{pmatrix} 3 \\ 1 \end{pmatrix} - \hat{y} = \begin{pmatrix} 3/5 \\ -4/5 \end{pmatrix}$$

$$\therefore \|z\| = \sqrt{\left(3/5\right)^2 + \left(4/5\right)^2}$$

6 Orthogonality and Least Squares

6.3

ORTHOGONAL PROJECTIONS

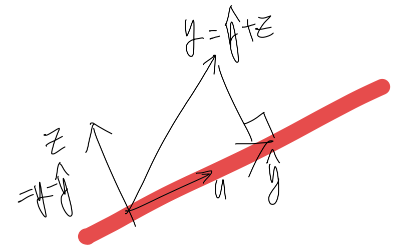


THE ORTHOGONAL DECOMPOSITION THEOREM

- **Theorem 8:** Let W be a subspace of $\mathbb{R}^n$. Then each $\mathbf{y}$ in $\mathbb{R}^n$ can be written uniquely in the form

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

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



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

(2) $$\hat{\mathrm{y}} = \frac{\mathrm{y} \bullet \mathrm{u}_1}{u_1 \cdot u_1} \mathrm{u}_1 + \cdots + \frac{y \cdot u_p}{u_p \cdot u_p} \mathrm{u}_p$$

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

THE ORTHOGONAL DECOMPOSITION THEOREM

- The vector $\hat{\mathbf{y}}$ in (1) is called the **orthogonal projection of y onto** W and often is written as $\text{proj}_W\mathbf{y}$.

$y = \overset{W}{\hat{y}} + z$

$\hat{y} = c_1u_1 + \cdots + c_pu_p \Rightarrow \hat{y}\cdot u_j = c_j(u_j \cdot u_j)$

- **Proof:** Let $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ be any orthogonal basis for W, and define $\hat{\mathbf{y}}$ by (2).
- Then $\hat{\mathbf{y}}$ is in W because $\hat{\mathbf{y}}$ is a linear combination of the basis $\mathbf{u}_1,\ldots,\mathbf{u}_p$.

THE ORTHOGONAL DECOMPOSITION THEOREM

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

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

$c_j u_j \cdot u_1 = 0$

$$= y \cdot u_1 \text{-} y \cdot u_1 = 0$$

- Thus $\mathbf{z}$ is orthogonal to $\mathbf{u}_1$.
- Similarly, $\mathbf{z}$ is orthogonal to each $\mathbf{u}_j$ in the basis for W.
- Hence $\mathbf{z}$ is orthogonal to every vector in W.
- That is, $\mathbf{z}$ is in $W^{\perp}$.

THE ORTHOGONAL DECOMPOSITION THEOREM

- To show that the decomposition in (1) is unique, suppose **y** can also be written as $\mathrm{y} = \hat{\mathrm{y}}_1 + \mathrm{z}_1$, with $\hat{\mathrm{y}}_1$ in W and $\mathbf{z}_1$ in $W^{\perp}$.
- Then $\hat{\mathrm{y}} + \mathrm{z} = \hat{\mathrm{y}}_1 + \mathrm{z}_1$ (since both sides equal **y**), and so

$$\hat{\mathrm{y}} - \hat{\mathrm{y}}_1 = \mathrm{z}_1 - \mathrm{z}$$

W

$W^{\perp}$

- This equality shows that the vector $\mathrm{v} = \hat{\mathrm{y}} - \hat{\mathrm{y}}_1$ is in W and in $W^{\perp}$ (because $\mathbf{z}_1$ and $\mathbf{z}$ are both in $W^{\perp}$, and W is a subspace).
- Hence $v \cdot v = 0$, which shows that $\mathrm{v} = 0$.
- This proves that $\hat{\mathrm{y}} = \hat{\mathrm{y}}_1$ and also $\mathrm{z}_1 = \mathrm{z}$.

THE ORTHOGONAL DECOMPOSITION THEOREM

- The uniqueness of the decomposition (1) shows that the orthogonal projection $\hat{\mathbf{y}}$ depends only on W and not on the particular basis used in (2).

THE ORTHOGONAL DECOMPOSITION THEOREM

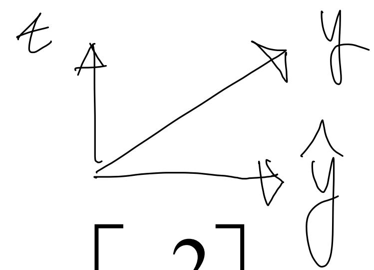
- **Example 1:** Let $\mathbf{u}_1 = \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

  Observe that $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Write **y** as the sum of a vector in W and a vector orthogonal to W.

THE ORTHOGONAL DECOMPOSITION THEOREM

- **Solution:** The orthogonal projection of **y** onto W is

$$\hat{\mathrm{y}} = \frac{y \cdot u_1}{u_1 \cdot u_1}\mathrm{u}_1 + \frac{y \cdot u_2}{u_2 \cdot u_2}\mathrm{u}_2$$



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

- Also

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

THE ORTHOGONAL DECOMPOSITION THEOREM

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

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

PROPERTIES OF ORTHOGONAL PROJECTIONS

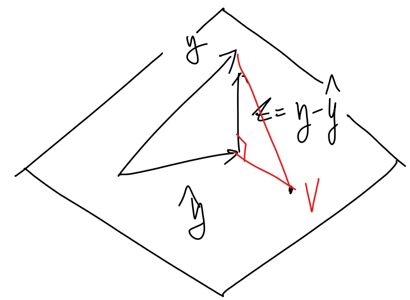
- If $\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$ is an orthogonal basis for W and if $\mathbf{y}$ happens to be in W, then the formula for $\text{proj}_W\mathbf{y}$ is exactly the same as the representation of $\mathbf{y}$ given in Theorem 5 in Section 6.2.

- In this case, $\text{proj}_W \mathbf{y} = \mathbf{y}$. $\rightarrow$ $\dfrac{y \cdot u_1}{u_1 \cdot u_1} u_1 + \cdots + \dfrac{y \cdot u_p}{u_p \cdot u_p} u_p$

- If $\mathbf{y}$ is in $W = \text{Span}\{\mathbf{u}_1,\ldots,\mathbf{u}_p\}$, then $\text{proj}_W \mathbf{y} = \mathbf{y}$.

THE BEST APPROXIMATION THEOREM

- **Theorem 9:** Let W be a subspace of $\mathbb{R}^n$, let $\mathbf{y}$ be any vector in $\mathbb{R}^n$, and let $\hat{\mathrm{y}}$ be the orthogonal projection of $\mathbf{y}$ onto W. Then $\hat{\mathrm{y}}$ is the closest point in W to $\mathbf{y}$, in the sense that



$$\|\mathrm{y}-\hat{\mathrm{y}}\| < \|\mathrm{y}-\mathrm{v}\| \tag{3}$$

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

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

THE BEST APPROXIMATION THEOREM

- If a different orthogonal basis for W were used to construct an orthogonal projection of $\mathbf{y}$, then this projection would also be the closest point in W to $\mathbf{y}$, namely, $\hat{\mathrm{y}}$.

THE BEST APPROXIMATION THEOREM

- **Proof:** Take $\mathbf{v}$ in W distinct from $\hat{\mathbf{y}}$.
- Then $\hat{\mathbf{y}} - \mathbf{v}$ is in W.
- By the Orthogonal Decomposition Theorem, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to W.
- In particular, $\mathbf{y} - \hat{\mathbf{y}}$ is orthogonal to $\hat{\mathbf{y}} - \mathbf{v}$ (which is in W).

THE BEST APPROXIMATION THEOREM

- Since

$$\mathbf{y}-\mathbf{v}=(\mathbf{y}-\hat{\mathbf{y}})+(\hat{\mathbf{y}}-\mathbf{v})$$

the Pythagorean Theorem gives

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

- (See the colored right triangle in the figure on the previous slide. The length of each side is labeled.)

- Now $\|\hat{\mathbf{y}}-\mathbf{v}\|^2>0$ because $\hat{\mathbf{y}}-\mathbf{v}\neq 0$, and so inequality (3) follows immediately.

PROPERTIES OF ORTHOGONAL PROJECTIONS

- **Example 4:** The distance from a point **y** in $\mathbb{R}^n$ to a subspace W is defined as the distance from **y** to the nearest point in W. Find the distance from **y** to $W = \text{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$, where

$$\mathbf{y} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix}, \mathbf{u}_1 = \begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$$

- **Solution:** By the Best Approximation Theorem, the distance from **y** to W is $\|\mathbf{y} - \hat{\mathbf{y}}\|$, where $\hat{\mathbf{y}} = \text{proj}_W \mathbf{y}$.

PROPERTIES OF ORTHOGONAL PROJECTIONS

- Since $\{\mathbf{u}_1, \mathbf{u}_2\}$ is an orthogonal basis for W,

$\left(\frac{y \cdot u_1}{u_1 \cdot u_1} \right.$ 이용 $\left. \right)$

$$\hat{\mathrm{y}} = \frac{15}{30}\mathrm{u}_1 + \frac{-21}{6}\mathrm{u}_2 = \frac{1}{2}\begin{bmatrix} 5 \\ -2 \\ 1 \end{bmatrix} - \frac{7}{2}\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix}$$

$$z = \mathrm{y} - \hat{\mathrm{y}} = \begin{bmatrix} -1 \\ -5 \\ 10 \end{bmatrix} - \begin{bmatrix} -1 \\ -8 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 6 \end{bmatrix}$$

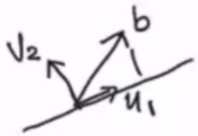
$$\|\mathrm{y} - \hat{\mathrm{y}}\|^2 = 3^2 + 6^2 = 45$$

- The distance from $\mathbf{y}$ to W is $\sqrt{45} = 3\sqrt{5}$.

Suppose that $\{a, b, c\}$ ~ linearly independent.

Let $v_1 = a$ and $u_1 = \frac{1}{\|v_1\|} v_1$ ($\|v_1\| \neq 0$ because $\{a, b, c\}$ ~ linearly independent)

$\mathrm{span}\{a\} = \mathrm{span}\{v_1\} = \mathrm{span}\{u_1\}$



$$\mathrm{proj}_{u_1} b = \frac{b \cdot u_1}{u_1 \cdot u_1} u_1 = (b \cdot u_1) u_1 \qquad u_1 \cdot u_1 = 1$$

Let $v_2 = b - \mathrm{proj}_{u_1} b$ then $v_2 \perp u_1$ ($v_2 \perp a$)

※ b는 $v_1 v_2$의 linear Comb 이므로 등식 성립

Let $u_2 = \frac{1}{\|v_2\|} v_2$ ($\|v_2\| \neq 0$ because $\|v_2\| = 0$ implies $b = \mathrm{proj}_{v_1} b$ i.e. $b \in \mathrm{span}\{u_1\} = \mathrm{span}\{a\}$)

$\mathrm{span}\{u_1, u_2\} = \mathrm{span}\{v_1, v_2\} = \mathrm{span}\{a, b\}$

$$\mathrm{proj}_{\mathrm{span}\{u_1, u_2\}} c = \left(\frac{c \cdot u_1}{u_1 \cdot u_1}\right) u_1 + \left(\frac{c \cdot u_2}{u_2 \cdot u_2}\right) u_2 = (c \cdot u_1) u_1 + (c \cdot u_2) u_2$$

Let $v_3 = c - \mathrm{proj}_{\mathrm{span}\{u_1, u_2\}} c = c - (c \cdot u_1) u_1 - (c \cdot u_2) u_2$

Let $u_3 = \frac{1}{\|v_3\|} v_3$ ($\|v_3\| \neq 0$)

$\mathrm{span}\{u_1, u_2, u_3\} = \mathrm{span}\{v_1, v_2, v_3\} = \mathrm{span}\{a, b, c\}$

Note $\{v_1, v_2, v_3\}$ ~ orthogonal

$\{u_1, u_2, u_3\}$ ~ orthonormal

이들의 linear comb로

a, b, c 가 만들어지니까

6 Orthogonality and Least Squares

6.4

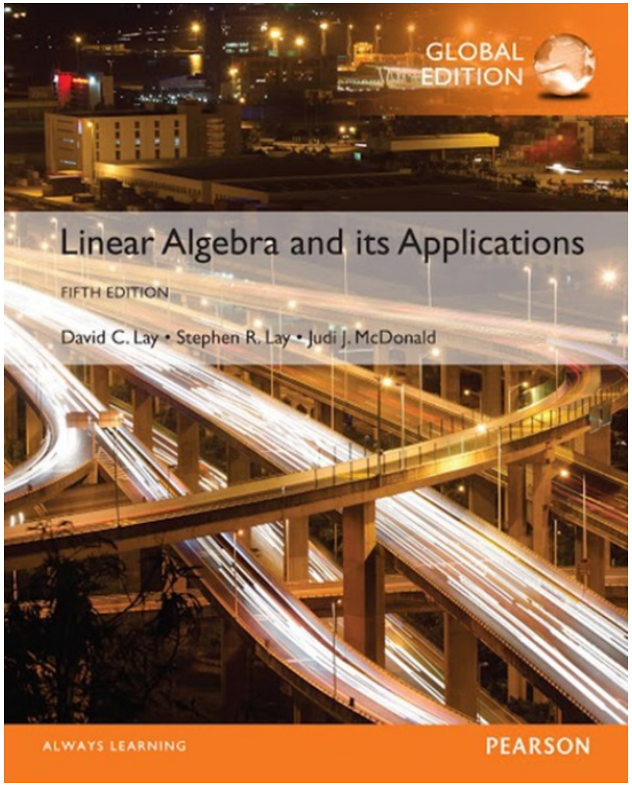
THE GRAM-SCHMIDT PROCESS

In Ch 1 & 2 → Given a set vectors

Linearly Independent vectors could be identified

previous, we consider orthogonal projection, given an orthogonal basis

→ Orthogonal basis 找]



THE GRAM-SCHMIDT PROCESS

- **Theorem 11: The Gram-Schmidt Process**
- Given a basis $\{x_1, \ldots, x_p\}$ for a nonzero subspace W of $\mathbb{R}^n$, define

linearly independent

$$v_1 = x_1$$

$$v_2 = x_2 - \frac{x_2 \cdot v_1}{v_1 \cdot v_1} v_1$$

$\mathrm{proj}_{u_1} x_2$

$$v_3 = x_3 - \frac{x_3 \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_3 \cdot v_2}{v_2 \cdot v_2} v_2$$

$\mathrm{proj}_{\mathrm{span}\{u_1, u_2\}} x_3$

$$\vdots$$

$$v_p = xp - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \frac{x_p \cdot v_2}{v_2 \cdot v_2} v_2 - \ldots - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1}$$

$\mathrm{proj}_{\mathrm{span}\{u_1, \ldots, u_p\}} x_p$

- Then $\{v_1, \ldots, v_p\}$ is an orthogonal basis for W. In addition

$$\mathrm{Span}\{v_1, \ldots, v_k\} = \mathrm{Span}\{x_1, \ldots, x_k\} \quad \text{for } 1 \le k \le p \qquad (1)$$

THE GRAM-SCHMIDT PROCESS

- **Proof** For $1 \leq k \leq p$, let $W_k = \text{Span}\{x_1, \ldots, x_k\}$. Set $v_1 = x_1$, so that $\text{Span}\{v_1\} = \text{Span}\ \{x_1\}$. Suppose, for some $k<p$, we have constructed $v_1, \ldots, v_k$ so that $\{v_1, \ldots, v_k\}$ is an orthogonal basis for W_k. Define

$$v_k+1 = x_{k+1} - \text{proj}_{W_k} x_{k+1} \qquad (2)$$

- By the Orthogonal Decomposition Theorem, v_{k+1} is orthogonal to W_k. Furthermore, $v_{k+1} \neq 0$ because x_{k+1} is not in $W_k = \text{Span}\{x_1, \ldots, x_k\}$
- Hence $\{v_1, \ldots, v_k\}$ is an orthogonal set of nonzero vectors in the $(k + 1)$-dimensional space W_{k+1}. By the Basis Theorem in Section 4.5, this set is an orthogonal basis for W_{k+1}. Hence $W_{k+1} = \text{Span}\{v_1, \ldots, v_{k+1}\}$. When $k + 1 = p$, the process stops.

ORTHONORMAL BASES

- **Example 3** Example 1 constructed the orthogonal basis

$$v_1 = \begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix},\ v_2 = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$$

- An orthonormal basis is

$$u_1 = \frac{1}{\|v_1\|}\mathrm{v}_1 = \frac{1}{\sqrt{45}}\begin{bmatrix} 3 \\ 6 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \\ 0 \end{bmatrix}$$

$$u_2 = \frac{1}{\|v_2\|}\mathrm{v}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Given linearly independent $\{a, b, c\}$

$v_1 = a \qquad u_1 = \frac{1}{\|v_1\|} v_1 \qquad a = v_1 = \underbrace{\|v_1\|}_{x} u_1 \Rightarrow a = x_1 u_1 \qquad a \cdot u_1 = (x_1 u_1)\cdot u_1 = x_1 (u_1 \cdot u_1) = x_1$

i.e. $\boxed{a = (a\cdot u_1) u_1}$

$v_2 = b - (b\cdot u_1) u_1 \qquad u_2 = \frac{1}{\|v_2\|} v_2 \qquad b = v_2 + (b\cdot u_1)u_1 = \underbrace{\|v_2\|}_{= x_2} u_2 + (b\cdot u_1) u_1$

$b \cdot u_2 = \left(x_2 u_2 + (b\cdot u_1) u_1 \right) \cdot u_2$

$= x_2 (u_2 \cdot u_2) + (b\cdot u_1)(\cancel{u_1 \cdot u_2})^{0} = x_2$

$\Rightarrow b = (b\cdot u_2) u_2 + (b \cdot u_1) u_1$

$\boxed{b = (b\cdot u_1) u_1 + (b\cdot u_2) u_2}$

$v_3 = c - (c\cdot u_1)u_1 - (c\cdot u_2) u_2 \qquad u_3 = \frac{1}{\|v_3\|} v_3$

$v_3 = \|v_3\| u_3$

$c = (c\cdot u_1) u_1 + (c\cdot u_2) u_2 + v_3$

$= (c\cdot u_1) u_1 + (c\cdot u_2) u_2 + \underbrace{\|v_3\|}_{x_3} u_3$

$c \cdot u_3 = (c\cdot u_1)(\cancel{u_1 \cdot u_3}) + (c\cdot u_2)(\cancel{u_2\cdot u_3}) + x_3 (u_3 \cdot u_3) = x_3$

$\boxed{c = (c\cdot u_1) u_1 + (c\cdot u_2) u_2 + (c\cdot u_3) u_3}$

$\boxed{a = (a\cdot u_1) u_1}$ $\qquad (a\cdot u_1) \neq 0$

$b = (b\cdot u_1) u_1 + (b\cdot u_2) u_2$ $\qquad (b\cdot u_2) \neq 0$

$c = (c\cdot u_1) u_1 + (c\cdot u_2) u_2 + (c\cdot u_3) u_3$ $\qquad (c\cdot u_3) \neq 0$

matrix with linearly independent columns)

$$[a \;\; b \;\; c] = [\, (a\cdot u_1)u_1 \quad (b\cdot u_1)u_1 + (b\cdot u_2)u_2 \quad (c\cdot u_1)u_1 + (c\cdot u_2)u_2 + (c\cdot u_3)u_3 \,]$$

$$= [u_1 \;\; u_2 \;\; u_3] \begin{pmatrix} a\cdot u_1 & b\cdot u_1 & c\cdot u_1 \\ 0 & b\cdot u_2 & c\cdot u_2 \\ 0 & 0 & c\cdot u_3 \end{pmatrix}$$

$\neq 0$

$= $ (a matrix with orthonormal columns) $\times$ (an upper-triangular matrix) $= Q \times R$

invertable

QR FACTORIZATION OF MATRICES

$A = LU$

$A = PDP^{-1}$

$A = QR$

- **Theorem 12: The QR Factorization**
- If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

↳ 각 length 크기1...

Gram-Schmit → QR decomposition

QR FACTORIZATION OF MATRICES

- **Theorem 12: The QR Factorization**
- If A is an $m \times n$ matrix with linearly independent columns, then A can be factored as $A = QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal basis for Col A and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

- **Proof** The columns of A form a basis $\{x_1, \ldots, x_n\}$ for Col A. Construct an orthonormal basis $\{u_1, \ldots, u_n\}$ for W = Col A with property (1) in Theorem 11. This basis may be constructed by the Gram-Schmidt process or some other means.

QR FACTORIZATION OF MATRICES

- Let

$$Q = [\mathrm{u}_1 \quad \mathrm{u}_2 \quad \dots \quad \mathrm{u}_\mathrm{n}]$$

- For $k = 1, \dots, \mathrm{x}_k$ is in $\mathrm{Span}\{\mathrm{x}_1, \dots, \mathrm{x}_k\} = \mathrm{Span}\{\mathrm{u}_1, \dots, \mathrm{u}_k\}$. So there are constants, $\mathrm{r}_{1k}, \dots, \mathrm{r}_{kk}$, such that

$$x_k = r_{1k}u_1 + \cdots + r_{kk}u_k + 0 \cdot u_{k+1} + \cdots + 0 \cdot u_n$$

- We may assume that $\mathrm{r}_{kk} \geq 0$. This shows that x_k is a linear combination of the columns of Q using as weights the entries in the vector

QR FACTORIZATION OF MATRICES

$$r_k = \begin{bmatrix} r_{1k} \\ \cdot \\ \cdot \\ \cdot \\ r_{kk} \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

- That is, $x_k = Qr_k$ for $k = 1, \ldots, n$. Let $R = [r_1 \ldots r_n]$. Then
$$A = [x_1 \ldots x_n] = [Qr_1 \ldots Qr_n] = QR$$
- The fact that R is invertible follows easily from the fact that the columns of A are linearly independent. Since R is clearly upper triangular, its nonnegative diagonal entries must be positive.

QR FACTORIZATION OF MATRICES

- **Example 4** Find a QR factorization of $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.

$V_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ $\|V_1\| = 2$ $u_1 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ $\begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} \cdot u_1 = \frac{3}{2}$

$V_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{2} \cdot \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$ $u_2 = \frac{1}{\sqrt{12}}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ $\left(V_2^* = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right)$ $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} \cdot u_1 = \frac{1}{2} \cdot 2 = 1$, $\begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} u_2 = \frac{2}{\sqrt{12}} = \frac{1}{\sqrt{3}}$

$V_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{3}\frac{1}{2\sqrt{3}}\begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1/2 + 1/2 \\ -1/2 - 1/6 \\ 1/2 - 1/6 \\ 1/2 - 1/6 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}$ $\therefore u_3 = \frac{1}{\sqrt{6}}\begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}$

QR FACTORIZATION OF MATRICES

- **Example 4** Find a QR factorization of $A = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$.
- **Solution** The columns of A are the vectors $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ in Example 2. An orthogonal basis for Col A = Span$\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ was found in that example:

$$v_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad v_2 = \begin{bmatrix} -3 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad v_3 = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix}$$

QR FACTORIZATION OF MATRICES

- To simplify the arithmetic that follows, scale v_3 by letting $v_3 = 3v_3$. Then normalize the three vectors to obtain u_1, u_2, and u_3, and use these vectors as the columns of Q:

$$Q = \begin{bmatrix} 1/2 & -3/\sqrt{12} & 0 \\ 1/2 & 1/\sqrt{12} & -2/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \\ 1/2 & 1/\sqrt{12} & 1/\sqrt{6} \end{bmatrix}.$$

- By construction, the first k columns of Q are an orthonormal basis of Span$\{x_1, \ldots, x_k\}$.

QR FACTORIZATION OF MATRICES

- From the proof of Theorem 12, $A = QR$ for some R. To find R, observe that $Q^TQ = I$, because the columns of Q are orthonormal. Hence

$$Q^TA = Q^T(QR) = IR = R$$

- and

$$R = \begin{bmatrix} 1/2 & 1/2 & 1/2 & 1/2 \\ -3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 0 & -2/\sqrt{6} & 1/\sqrt{12} & 1/\sqrt{6} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & 3/2 & 1 \\ 0 & 3/\sqrt{12} & 2/\sqrt{12} \\ 0 & 0 & 3/\sqrt{12} \end{bmatrix}$$

$$Q^T \cdot A = R$$

$$\begin{pmatrix} a \cdot u_1 & b \cdot u_1 & c \cdot u_1 \\ 0 & b \cdot u_2 & c \cdot u_2 \\ 0 & 0 & c \cdot u_3 \end{pmatrix}$$

내적을 통해
이렇게 해도 된다.

$$A = QR$$

Suppose that A and Q are known.

What is the property of Q? Q has orthonormal columns

$$Q^T A = Q^T Q R = \begin{pmatrix} u_1^T \\ \vdots \\ u_n^T \end{pmatrix} \begin{pmatrix} u_1 & \cdots & u_n \end{pmatrix} R$$

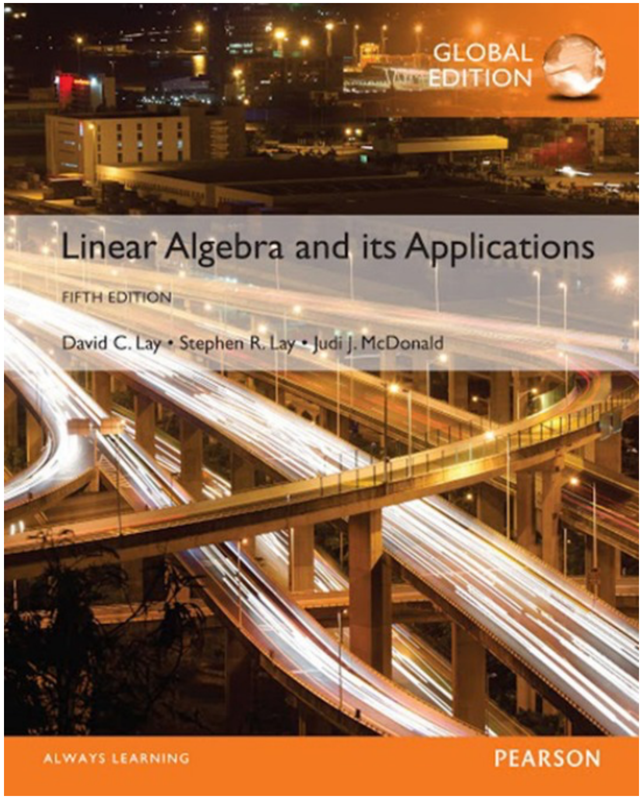
$$= \begin{pmatrix} u_1^T u_1 & u_1^T u_2 & \cdots & u_1^T u_n \\ u_2^T u_1 & u_2^T u_2 & \cdots & u_2^T u_n \\ & \vdots & & \end{pmatrix} R$$

$$= I R = R$$

6 Orthogonality and Least Squares

6.5

LEAST-SQUARES PROBLEMS



LEAST-SQUARES PROBLEMS

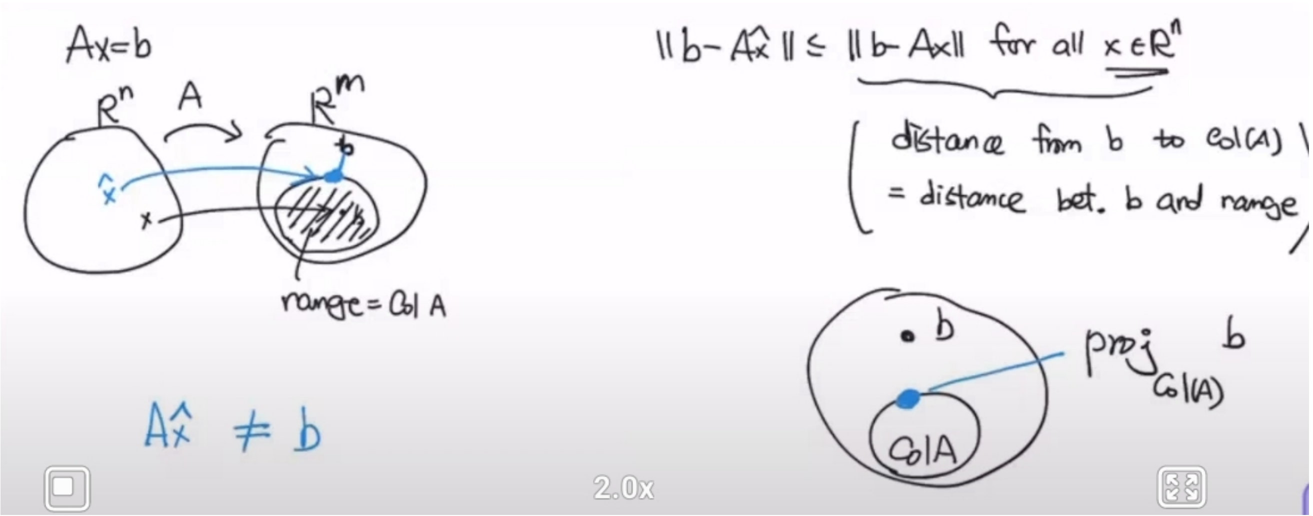
- **Definition:** If A is $m \times n$ and **b** is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathrm{x} = \mathrm{b}$ is an $\hat{\mathrm{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathrm{b} - A\hat{\mathrm{x}}\| \le \|\mathrm{b} - A\mathrm{x}\|$$

for all **x** in $\mathbb{R}^n$.

- The most important aspect of the least-squares problem is that no matter what **x** we select, the vector A**x** will necessarily be in the column space, Col A.

- So we seek an **x** that makes A**x** the closest point in Col A to **b**.



error 가 최소가 되는 $\hat{x}$ 찾기 ($\hat{x}$는 range로 mapping 됨)

+ $\hat{x}$: orthogonal projection of preimage

LEAST-SQUARES PROBLEMS

- **Definition:** If A is $m \times n$ and **b** is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

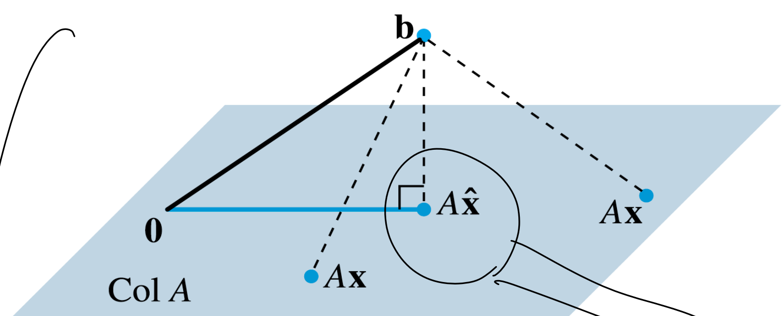
$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \leq \|\mathbf{b} - A\mathbf{x}\|$$ ← the error between b and Ax

for all **x** in $\mathbb{R}^n$.

$\{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\}$
$=$ range $=$ Col A

- So we seek an **x** that makes A**x** the closest point in Col A to **b**.

$\mathbb{R}^m$



$\text{proj}_{\text{Col}A}\,\mathbf{b}$

The vector **b** is closer to $A\hat{\mathbf{x}}$ than to $A\mathbf{x}$ for other **x**.

LEAST-SQUARES PROBLEMS

- **Solution of the General Least-Squares Problem**

- Given A and $\mathbf{b}$, apply the Best Approximation Theorem to the subspace Col A.

- Let $$\hat{\mathbf{b}} = \text{proj}_{\text{Col}\,A}\,\mathbf{b}$$

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- Because $\hat{\mathrm{b}}$ is in the column space A, the equation $A\mathrm{x} = \hat{\mathrm{b}}$ *is* consistent, and there is an $\hat{\mathrm{x}}$ in $\mathbb{R}^n$ such that

(1) $$A\hat{\mathrm{x}} = \hat{\mathrm{b}}$$ 이건 consistent b에 가장 가까운

- Since $\hat{\mathrm{b}}$ is the closest point in Col A to $\mathbf{b}$, a vector $\hat{\mathrm{x}}$ is a least-squares solution of $A\mathrm{x} = \mathrm{b}$ if and only if $\hat{\mathrm{x}}$ satisfies (1). i.e. $A\hat{x} = \mathrm{proj}_{\mathrm{col}\,A}\, b$ 을 찾는다.

- Such an $\hat{\mathrm{x}}$ in $\mathbb{R}^n$ is a list of weights that will build $\hat{\mathrm{b}}$ out of the columns of A.

If $Ax=b$ has a solution (consistent), ← 근이 있는 경우

then $\hat{b} = \text{proj}_{\text{Col}A}\, b = b$

$\|b - Ax\| = 0$ for solutions x

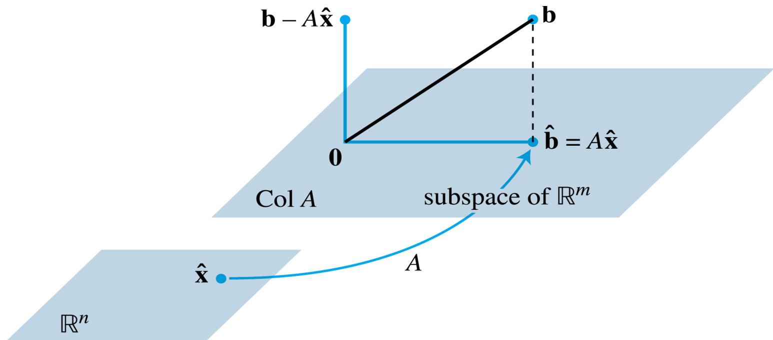
우리는 근 없는 경우를 따질 것.

So, if $v \in \mathbb{R}^n$,

then $\|b - Ax\| = 0 \le \|b - Av\|$

i.e. the solutions x are least-squares solutions in that case.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM



The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

If $\hat{x}$ is a least squares solution of $Ax=b$

$\rightarrow A\hat{x} = \hat{b} = \text{proj}_{\text{Col}A}\, b$

- Suppose $\hat{x}$ satisfies $A\hat{x} = \hat{b}$.
- By the Orthogonal Decomposition Theorem, the projection $\hat{b}$ has the property that $b - \hat{b}$ is orthogonal to Col A, so $b - A\hat{x}$ is orthogonal to each column of A.
- If $\mathbf{a}_j$ is any column of A, then $a_j \cdot (b - A\hat{x}) = 0$,and $a_j^T(b - A\hat{x}) = 0$.

$$a_j^T(b - A\hat{x}) = 0$$

$$a_1^T(b - A\hat{x}) = 0$$
$$a_2^T(b - A\hat{x}) = 0$$
$$\vdots$$
$$a_n^T(b - A\hat{x}) = 0$$

$$A^T(b - A\hat{x}) = \begin{bmatrix} a_1^T \\ \vdots \\ a_n^T \end{bmatrix}(b - A\hat{x}) = 0$$

If $\hat{b}$ $\left(= \mathrm{proj}_{\mathrm{Col}A} b\right)$ is known, $\hat{x}$ can be found by solving

$$A\hat{x} = \hat{b}$$

$\hat{x}$ can be also found by solving the normal equation:

$$A^TA\hat{x} = A^Tb$$

[illegible]

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- Since each $\mathbf{a}_j^T$ is a row of A^T,

$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = 0 \qquad (2)$$

- Thus

$$A^T\mathbf{b} - A^T A\hat{\mathbf{x}} = 0$$

得到 → $$A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$$ A^T的row = A的col

- These calculations show that each least-squares solution of $A\mathbf{x} = \mathbf{b}$ satisfies the equation

$$A^T A\mathbf{x} = A^T\mathbf{b} \qquad (3)$$

- The matrix equation (3) represents a system of equations called the **normal equations** for $A\mathbf{x} = \mathbf{b}$.
- A solution of (3) is often denoted by $\hat{\mathbf{x}}$.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- **Theorem 13:** The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equation $A^T A\mathbf{x} = A^T\mathbf{b}$.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- **Theorem 13:** The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equation $A^T A\mathbf{x} = A^T\mathbf{b}$.
- **Proof:** The set of least-squares solutions is nonempty and each least-squares solution $\hat{\mathbf{x}}$ satisfies the normal equations.

  $A^T(A\hat{x} - b) = 0$
- Conversely, suppose $\hat{\mathbf{x}}$ satisfies $A^T A\hat{\mathbf{x}} = A^T\mathbf{b}$.
- Then $\hat{\mathbf{x}}$ satisfies (2), which shows that $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to the rows of A^T and hence is orthogonal to the columns of A.
- Since the columns of A span Col A, the vector $\mathbf{b} - A\hat{\mathbf{x}}$ is orthogonal to all of Col A.

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- Hence the equation

$$\mathrm{b} = \underbrace{A\hat{\mathrm{x}}}_{\text{Col}\,A} + \underbrace{(\mathrm{b} - A\hat{\mathrm{x}})}_{\perp\,\text{Col}\,A}$$

  is a decomposition of **b** into the sum of a vector in Col A and a vector orthogonal to Col A.

- By the uniqueness of the orthogonal decomposition, $A\hat{\mathrm{x}}$ must be the orthogonal projection of **b** onto Col A.

$A\hat{x} = \mathrm{proj}_{\mathrm{Col}A}\, b = \hat{b}$

- That is, $A\hat{\mathrm{x}} = \hat{\mathrm{b}}$ and $\hat{\mathrm{x}}$ is a least-squares solution.

$\hat{x}$ is a least-squares solution

$\Leftrightarrow \|b - A\hat{x}\| \le \|b - Ax\|$ for "all" $x \in \mathbb{R}^n$

$\Leftrightarrow \quad \underline{A\hat{x} = \text{proj}_{\text{Col}A}\, b} = \hat{b}$ ✓

$\Leftrightarrow$ <u>normal equation</u> $\quad \underline{A^TA\hat{x} = A^Tb}$

정리

<u>Note</u>. $\cancel{x}$ $\hat{x}$ (a least-squares solution) may not be <u>unique</u>

i.e. $\underline{A^TA}$ may be a singular matrix

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- **Example 1:** Find a least-squares solution of the inconsistent system $A\mathrm{x} = \mathrm{b}$ for

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

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- **Example 1:** Find a least-squares solution of the inconsistent system $A\mathrm{x} = \mathrm{b}$ for

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

- **Solution:** To use normal equations (3), compute:

$$A^T A = \begin{bmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}$$

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

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

- Then the equation $A^T A\mathbf{x} = A^T\mathbf{b}$ becomes

$$\begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

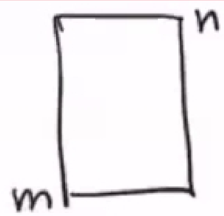
ERO

or

Inverse

① $A \sim m\times n$ matrix $(m \geq n)$

(If $m<n$, the columns of A cannot be independent)



② Claim : $\mathrm{Nul}(A) = \mathrm{Nul}(A^TA)$

If $x \in \mathrm{Nul}\, A$, then $Ax=0 \Rightarrow A^T(Ax) = A^T 0 = 0 \Rightarrow (A^TA)x = 0$

$\Rightarrow x \in \mathrm{Nul}(A^TA)$

i.e. $\boxed{\mathrm{Nul}\, A \subset \mathrm{Nul}(A^TA)}$

If $x \in \mathrm{Nul}(A^TA)$, then $A^TAx=0$

$$\left.\begin{array}{l} x^T A^T A x = x^T 0 = 0 \\ \quad \parallel \\ (Ax)^T(Ax) = \|Ax\|^2 \end{array}\right\} \; Ax = 0$$

$\Downarrow$

$x \in \mathrm{Nul}\, A$

$\Downarrow$

$\therefore$ $\boxed{\mathrm{Nul}(A^TA) \subset \mathrm{Nul}\, A}$

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM

- Row operations can be used to solve the system on the previous slide, but since A^TA is invertible and 2×2, it is probably faster to compute

$$(A^TA)^{-1} = \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}$$

and then solve $A^TA\mathrm{x} = A^T\mathrm{b}$ as

$$\hat{\mathrm{x}} = (A^TA)^{-1}A^T\mathrm{b}$$

$$= \frac{1}{84}\begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix}\begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84}\begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

SOLUTION OF THE GENREAL LEAST-SQUARES PROBLEM when Unique $\hat{x}$?

- **Theorem 14**: Let A be an $m \times n$ matrix. The following statements are logically equivalent:
  a. The equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution for each **b** in $\mathbb{R}^m$.
  b. The columns of A are linearly independent.
  c. The matrix $A^T A$ is invertible. $\Leftrightarrow$ $\mathrm{Nul}(A^TA) = \{0\} \to \mathrm{Nul}\,A = \{0\}$
  $\to$ Cols are Linearly Independent

  When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b} \tag{4}$$

ex. Col A dependent $\to$ "many" least-square solutions. $\to \begin{bmatrix} A^TA & A^Tb \end{bmatrix}$ 하기

ALTERNATIVE CALCULATIONS OF LEAST-SQUARES SOLUTIONS

- **Example 4:** Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

ALTERNATIVE CALCULATIONS OF LEAST-SQUARES SOLUTIONS

- **Example 4:** Find a least-squares solution of $A\mathbf{x} = \mathbf{b}$ for

$$A = \begin{bmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} -1 \\ 2 \\ 1 \\ 6 \end{bmatrix}$$

- **Solution:** Because the columns $\mathbf{a}_1$ and $\mathbf{a}_2$ of A are orthogonal, the orthogonal projection of $\mathbf{b}$ onto Col A is given by

$$\hat{b} = \frac{b \cdot a_1}{a_1 \cdot a_1} a_1 + \frac{b \cdot a_2}{a_2 \cdot a_2} a_2 = \frac{8}{4} a_1 + \frac{45}{90} a_2 \qquad (5)$$

ALTERNATIVE CALCULATIONS OF LEAST-SQUARES SOLUTIONS

$$= \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} + \begin{bmatrix} -3 \\ -1 \\ 1/2 \\ 7/2 \end{bmatrix} = \begin{bmatrix} -1 \\ 1 \\ 5/2 \\ 11/2 \end{bmatrix}$$

x is given by

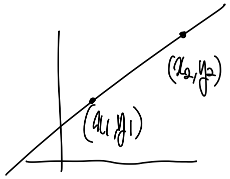
$$\hat{\mathbf{x}} = (A^T A)^{-1} A^T \mathbf{b}$$

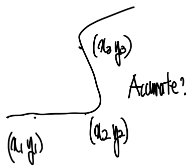
이렇게 계산할수도

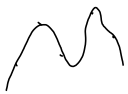
혹은

- Now that $\hat{\mathbf{b}}$ is known, we can solve $A\hat{\mathbf{x}} = \hat{\mathbf{b}}$.
- But this is trivial, since we already know weights to place on the columns of A to produce $\hat{\mathbf{b}}$.
- It is clear from (5) that

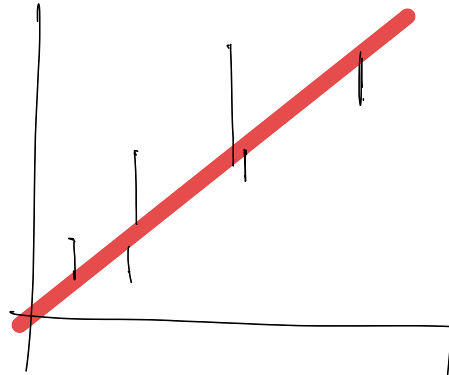
$$\hat{x} = \begin{bmatrix} 8/4 \\ 45/90 \end{bmatrix} = \begin{bmatrix} 2 \\ 1/2 \end{bmatrix}$$







1차 1개 → points all accurate



$$A^TAx = A^Tb$$

$$\underbrace{\begin{pmatrix} x_1 & \cdots & x_6 \\ 1 & \cdots & 1 \end{pmatrix}\begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_6 & 1 \end{pmatrix}}\begin{pmatrix} m \\ n \end{pmatrix} = \begin{pmatrix} x_1 & \cdots & x_6 \\ 1 & \cdots & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_6 \end{pmatrix}$$

$$\begin{pmatrix} \sum_{i=1}^{6} x_i^2 & \sum_{i=1}^{6} x_i \\ \sum_{i=1}^{6} & 6 \end{pmatrix}\begin{pmatrix} m \\ n \end{pmatrix} = \begin{pmatrix} \sum_{1}^{6} x_i y_i \\ \sum_{i=1}^{6} y_i \end{pmatrix}$$

$$\begin{pmatrix} m \\ n \end{pmatrix} = \frac{1}{6\sum x_i^2 - (\sum x_i)^2}\begin{pmatrix} 6 & -\sum x_i \\ -\sum x_i & \sum x_i^2 \end{pmatrix}\begin{pmatrix} \sum x_i y_i \\ \sum y_i \end{pmatrix}$$

a least-squares solution
(linear regression)

Least square solution
linear regression

$$y = mx + n$$

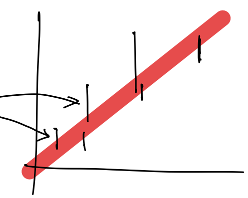
$$\downarrow$$

$$\begin{cases} mx_1 + n = y_1 \\ mx_2 + n = y_2 \\ \vdots \\ mx_6 + n = y_6 \end{cases}$$

$$\downarrow$$

$$\begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \\ x_6 & 1 \end{pmatrix}\begin{pmatrix} m \\ n \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_6 \end{pmatrix}$$

$|b - Ax|$



(m, n) minimizes $\|b - Ax\|^2$

$$= \sum (y_i - m_i x + n)^2$$

Error minimized

Given $(x_1, y_1) \cdots (x_n, y_n)$ find a problem that best fits the given data

$y = ax^2 + bx + c \rightarrow \sum_{i=1}^{n} (ax_i^2 + bx_i + c - y_i)^2$ is minimized

Need to find a least-squared solutions of

$$\begin{pmatrix} ax_1^2 + bx_1 + c = y_1 \\ ax_2^2 + bx_2 + c = y_2 \\ \vdots \\ ax_n^2 + bx_n + c = y_n \end{pmatrix} = \underbrace{\begin{pmatrix} x_1^2 & x_1 & 1 \\ & \vdots & \\ x_n^2 & x_n & 1 \end{pmatrix}}_{A} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}}_{b} \rightarrow A^T A \hat{x} = A^T b$$

$[A^T A \quad A^T b]$ 만들기

always consistent — Unique or Many

Orthogonal Projection to Col A 이기에