Linear Algebra Thm Archive - for Final Exam

in exactly 100 pages!

by Gyeonggi Science High School for the Gifted 'Linear Algebra' Participants
Main Author : 14129 황동욱, 14041 박승원
Revised by : 14121 하석민, 14125 한지우
LaTeX Technician : 14041 박승원

Last Compilation Time : Tuesday 24th January, 2017 16:51

Contents

List of Theorems

Extra Theorems

LaTeX technician's note : Theorems from chapter 3 are omitted here, due to TeX-chnical issue.

Chapter 4

Eigenvalues and Eigenvectors

4.1 Introduction to Eigenvalues and Eigenspaces

Definition. Let A be an $n \times n$ matrix. A scalar λ is called an **eigenvalue** of A if there exists a nonzero vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Such a vector $\mathbf{x}$ is called an **eigenvector** of A corresponding to λ.

Definition. Let A be an $n \times n$ matrix and let λ be an eigenvalue of A. The set of all eigenvectors corresponding to λ, together with the zero vector, is called the **eigenspace** of λ and is denoted as E_λ.

4.2 Determinants

Definition. A_{ij} is the submatrix of a matrix A obtained by deleting ith row and jth column from A. The determinant of such matrix is called the (i, j)-minor of A.
The **(i, j)-cofactor** of A is defined as $C_{ij} = (-1)^{i+j} \det(A_{ij})$.

Definition. Let $A = [a_{ij}]$ be an $n \times n$ matrix, where $n \geq 2$. Then the **determinant** of A is the scalar

$$\det A = |A| = \sum_{j=1}^{n} (-1)^{j+1} a_{1j} \det A_{1j} = \sum_{j=1}^{n} a_{1j} C_{1j}$$

Lemma 4.13
Let A be $n \times n$ matrix. Then

$$\det A = \sum_{i=1}^{n} a_{i1} C_{i1} = \sum_{j=1}^{n} a_{1j} C_{1j}$$

Proof. We prove this lemma with mathematical induction.

(i) $n = 2$

$$\det A = a_{11}C_{11} + a_{12}C_{12} = a_{11}a_{22} + a_{12}(-a_{21}) = a_{11}C_{11} + a_{21}C_{21}$$

(ii) Suppose that the proposition is true for all $(n-1) \times (n-1)$ matrices.
For $n \times n$ matrix A,

$$\det A = \sum_{i=1}^{n} a_{i1} C_{i1} = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det A_{i1}$$

Since A_{i1} is an $(n-1)\times(n-1)$ matrix,

$$\begin{aligned}
&\sum_{i=1}^{n}(-1)^{i+1}a_{i1}\det A_{i1}\\
=&a_{11}C_{11}+\sum_{i=2}^{n}[(-1)^{i+1}a_{i1}(\sum_{j=2}^{n}(-1)^{j+1}a_{1j}\det A_{i1,1j})]\\
=&a_{11}C_{11}+\sum_{j=2}^{n}[(-1)^{j+1}a_{1j}(\sum_{i=2}^{n}(-1)^{i+1}a_{i1}\det A_{i1,1j})]\\
=&\sum_{j=1}^{n}(-1)^{j+1}a_{1j}\det A_{1j}=\sum_{j=1}^{n}a_{1j}C_{1j}
\end{aligned}$$

where $A_{ij,kl}$ stands for the minor matrix of A obtained by deleting i,kth rows and j,lth columns from A.

By (i) and (ii), The cofactor expansion along the first row and column of A is equal to $\det A$. □

Theorem 4.1 : The Laplace Expansion Theorem
The determinant of an $n\times n$ matrix $A=[a_{ij}]$, where $n\geq 2$, can be computed as

$$\det A=\sum_{j=1}^{n}a_{ij}C_{ij}$$

(which is the **cofactor expansion along the ith row**) and also as

$$\det A=\sum_{i=1}^{n}a_{ij}C_{ij}$$

(which is the **cofactor expansion along the jth column**)

Proof. Let $A=[a_{ij}]$ be an $n\times n$ matrix. Let B be the matrix obtained by moving ith row of A to the top row using $i-1$ interchanges of adjacent rows. By Theorem 4.3(b), $\det A=(-1)^{i-1}\det B$. Thus,

$$\begin{aligned}
\det A=(-1)^{i-1}\det B&=(-1)^{i-1}\sum_{j=1}^{n}(-1)^{1+j}b_{1j}\det B_{1j}\\
&=(-1)^{i-1}\sum_{j=1}^{n}(-1)^{1+j}a_{ij}\det A_{ij}=\sum_{j=1}^{n}(-1)^{i+j}a_{ij}\det A_{ij}
\end{aligned}$$

Therefore, the cofactor expansion along the rows of A are equal to the determinant of A.

(Exercise 4.2 68) By Lemma 4.13, the cofactor expansion along the first column is equal to $\det A$. With similar procedure, the cofactor expansion along the columns of A are equal to the determinant of A. □

Theorem 4.2
The determinant of a triangular matrix is the product of the entries on its main diagonal. Specifically, if $A = [a_{ij}]$ is an $n \times n$ triangular matrix, then

$$\det A = a_{11}a_{22}\cdots a_{nn}$$

Proof. **(Exercise 4.2 21)** We prove the theorem with mathematical induction.

(i) $n = 2$
For upper triangular 2×2 matrix $U = \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}$, $\det U = u_{11}u_{22}$.
For lower triangular 2×2 matrix $L = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix}$, $\det L = l_{11}l_{22}$.

(ii) Suppose that the proposition is true for all $(n-1) \times (n-1)$ triangular matrices.
For $n \times n$ upper triangular matrix $U = [u_{ij}]$,

$$\det U = \sum_{i=1}^{n} u_{i1}C_{i1} = u_{11}C_{11} = u_{11}u_{22}\cdots u_{nn}$$

For $n \times n$ lower triangular matrix $L = [l_{ij}]$,

$$\det L = \sum_{j=1}^{n} l_{1j}C_{1j} = l_{11}C_{11} = l_{11}l_{22}\cdots l_{nn}$$

By (i) and (ii), the determinant of a triangular matrix in any size is equal to the product of its diagonal entries. $\square$

Theorem 4.3 : Properties of Determinants
Let $A = [a_{ij}]$, $B = [b_{ij}]$, and $C = [c_{ij}]$ be a square matrices.

a. If A has a zero row (column), then $\det A = 0$.

b. **(Lemma 4.14)** If B is obtained by interchanging two rows (columns) of A, then $\det B = -\det A$.

c. If A has two identical rows (columns), then $\det A = 0$.

d. If B is obtained by multiplying a row (column) of A by k, then $\det B = k \det A$.

e. If A, B, and C are identical except that the ith row (column) of C is the sum of the ith rows (columns) of A and B, then $\det C = \det A + \det B$.

f. If B is obtained by adding a multiple of one row (column) of A to another row (column), then $\det B = \det A$.

Proof. Let $A = [a_{ij}]$ be a $n \times n$ matrix.

a. **(Exercise 4.2 41)** Suppose that ith row of A is a zero row. Then by Laplace Expansion Theorem,

$$\det A = \sum_{j=1}^{n} a_{ij}C_{ij} = \sum_{j=1}^{n} 0C_{ij} = 0$$

b. We first prove the case when two adjacent rows are interchanged by mathematical induction.

(i) $n = 2$
For 2×2 matrix $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, consider the determinant of $B = \begin{bmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{bmatrix}$, which is obtained by interchanging first and second rows of A. Then

$$\det A = a_{11}a_{22} - a_{12}a_{21} = -(a_{12}a_{21} - a_{11}a_{22}) = -\det B$$

(ii) Suppose that the proposition is true for all $(n-1) \times (n-1)$ matrices.
Consider the determinant of B which is obtained by interchanging rth and $(r+1)$th row of A, where $r \in \{1, 2, \cdots, n-1\}$. By Laplace Expansion Theorem,

$$\begin{aligned} \det B &= \sum_{i \neq r, r+1} (-1)^{i+1} b_{i1} \det B_{i1} + (-1)^{r+1} b_{r1} \det B_{r1} + (-1)^{(r+1)+1} b_{(r+1)1} \det B_{(r+1)1} \\ &= \sum_{i \neq r, r+1} (-1)^{i+1} a_{i1} (-\det A_{i1}) + (-1)^{r+1} a_{(r+1)1} \det A_{(r+1)1} + (-1)^{(r+1)+1} a_{r1} \det A_{r1} \\ &= -\sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det A_{i1} = -\det A \end{aligned}$$

By (i) and (ii), for square matrix A in any size, if B is obtained by interchanging two adjacent rows of A, $\det B = -\det A$.

(Exercise 4.2 67) Suppose that B is obtained by interchanging rth row and sth row ($r < s$) of A. This operation is equivalent to following sequence of operations:

$$R_r \leftrightarrow R_{r+1}, R_{r+1} \leftrightarrow R_{r+2}, \cdots, R_{s-1} \leftrightarrow R_s,$$
$$R_{s-2} \leftrightarrow R_{s-1}, \cdots, R_r \leftrightarrow R_{r+1}$$

which is $2(s-r)-1$ times of interchange between adjacent rows. Therefore,

$$\det B = (-1)^{2(s-r)-1} \det A = -\det A$$

c. Let B be the matrix obtained by swapping two identical rows of A. Since $A = B$, $\det A = \det B$. By Theorem 4.3(b), $\det A = -\det B$. Therefore, $\det A = 0$.

d. Let B be the matrix obtained by multiplying scalar k to ith row of A, then $B_{ij} = A_{ij}$ for j from 1 to n. By Laplace Expansion Theorem,

$$\begin{aligned} \det B &= \sum_{j=1}^{n} (-1)^{i+j} b_{ij} \det B_{ij} = \sum_{j=1}^{n} (-1)^{i+j} (ka_{ij}) \det A_{ij} \\ &= k \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} = k \det A \end{aligned}$$

e. Since A, B, and C are identical except for ith row, $A_{ij} = B_{ij} = C_{ij}$ for j from 1 to n. By Laplace Expansion Theorem,

$$\begin{aligned} \det C &= \sum_{j=1}^{n} (-1)^{i+j} c_{ij} \det C_{ij} = \sum_{j=1}^{n} (-1)^{i+j} (a_{ij} + b_{ij}) \det C_{ij} \\ &= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det A_{ij} + \sum_{j=1}^{n} (-1)^{i+j} b_{ij} \det B_{ij} = \det A + \det B \end{aligned}$$

f. **(Exercise 4.2 42)** Let $A = \begin{bmatrix} \vdots \\ \mathbf{A}_r \\ \vdots \\ \mathbf{A}_s \\ \vdots \end{bmatrix}$, and let $B = \begin{bmatrix} \vdots \\ \mathbf{A}_r \\ \vdots \\ \mathbf{A}_s + k\mathbf{A}_r \\ \vdots \end{bmatrix}$, where k is a scalar. By Theorem 4.3(e),

$$\det B = \det A + \begin{vmatrix} \vdots \\ \mathbf{A}_r \\ \vdots \\ k\mathbf{A}_r \\ \vdots \end{vmatrix}$$

By Theorem 4.3(c) and Theorem 4.3(d),

$$\begin{vmatrix} \vdots \\ \mathbf{A}_r \\ \vdots \\ k\mathbf{A}_r \\ \vdots \end{vmatrix} = k \begin{vmatrix} \vdots \\ \mathbf{A}_r \\ \vdots \\ \mathbf{A}_r \\ \vdots \end{vmatrix} = k0 = 0$$

Therefore, $\det B = \det A$.

Note that by Lemma 4.13, the proof above is also valid for columns of A. □

Theorem 4.4
Let E be an $n \times n$ elementary matrix.

a. If E results from interchanging two rows of I_n, then $\det E = -1$.

b. If E results from multiplying one row of I_n by k, then $\det E = k$.

c. If E results from adding a multiple of one row of I_n to another row, then $\det E = 1$.

Note. Theorem 4.4의 a, b, c 유형에 해당하는 elementary matrix를 구별하여 아래와 같이 쓰기로 하자. 새로운 정의니까, 이 notation을 쓰고 싶다면 증명하기 전에 꼭 언급하고 시작할 것.

a. E_{ij} is the elementary matrix resulting from interchanging ith row and jth row of I. ($i \neq j$)

b. $E_i(k)$ is the elementary matrix resulting from multiplying ith row of I by a nonzero scalar k.

c. $E_{ij}(k)$ is the elementary matrix resulting from adding a multiple of jth row with k to ith row of I.

That is,

$$I \xrightarrow{R_i \leftrightarrow R_j} E_{ij}$$
$$I \xrightarrow{kR_i} E_i(k)$$
$$I \xrightarrow{R_i + kR_j} E_{ij}(k)$$

Proof. Let E be an $n \times n$ elementary matrix.

a. By Theorem 4.3(b),
$$\det E_{ij} = -\det I_n = -1$$

b. By Theorem 4.3(d),
$$\det E_i(k) = k \det I_n = k$$

c. By Theorem 4.3(f),
$$\det E_{ij}(k) = \det I_n = 1$$

□

Lemma 4.5
Let B be an $n \times n$ matrix and let E be an $n \times n$ elementary matrix. Then

$$\det(EB) = (\det E)\,(\det B)$$

Proof. **(Exercise 4.2 43)** Let B be an $n \times n$ matrix and let E be an $n \times n$ elementary matrix. Consider the determinant of EB with the type of elementary matrix. By Theorem 3.10, the elementary row operation which converts I_n to E also converts B to EB.

(i) If $E = E_{ij}$ for some i and j, $\det E = -1$ by Theorem 4.4(a). By Theorem 4.3(b), $\det(EB) = -\det B = (\det E)(\det B)$.

(ii) If $E = E_i(k)$ for some i and k, $\det E = k$ by Theorem 4.4(b). By Theorem 4.3(d), $\det(EB) = k \det B = (\det E)(\det B)$.

(iii) If $E = E_{ij}(k)$ for some i, j, and k, $\det E = 1$ by Theorem 4.4(c). By Theorem 4.3(f), $\det(EB) = \det B = (\det E)(\det B)$. □

Theorem 4.6
A square matrix A is invertible if and only if $\det A \neq 0$.

Proof. Let R be RREF of A. Then there exist elementary matrices $E_1, \cdots, E_k$ such that

$$R = (E_k \cdots E_2 E_1)A$$

By Lemma 4.5,

$$\det R = (\det E_k) \cdots (\det E_2)(\det E_1)(\det A)$$

Note that Theorem 4.4 states the determinant of elementary matrices cannot be zero. Thus, $\det R \neq 0$ if and only if $\det A \neq 0$. Since $\det R \neq 0$ if and only if $R = I$, $\det A \neq 0$ if and only if A is invertible by F.T.I.M. □

Theorem 4.7
If A is an $n \times n$ matrix, then

$$\det(kA) = k^n \det A$$

.

Proof. **(Exercise 4.2 44)** Let $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ be a $n \times n$ matrix and k be a scalar. By Theorem 4.3(d),

$$\begin{aligned} \det kA &= \begin{vmatrix} k\mathbf{a}_1 & k\mathbf{a}_2 & \cdots & k\mathbf{a}_n \end{vmatrix} \\ &= k \begin{vmatrix} \mathbf{a}_1 & k\mathbf{a}_2 & \cdots & k\mathbf{a}_n \end{vmatrix} \\ &\vdots \\ &= k^n \begin{vmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{vmatrix} = k^n \det A \end{aligned}$$

□

Theorem 4.8
If A and B are $n \times n$ matrices, then

$$\det(AB) = (\det A)(\det B)$$

Proof. Let A and B be $n \times n$ matrices.

(i) If A is invertible, then A is a product of elementary matrices, by F.T.I.M. So there exist elementary matrices $E_1, E_2, \cdots, E_k$ such that $A = E_1 E_2 \cdots E_k$. By Lemma 4.5,

$$\begin{aligned} \det(AB) &= \det(E_1 E_2 \cdots E_k B) = (\det E_1)(\det E_2) \cdots (\det E_k)(\det B) \\ &= (\det(E_1 E_2 \cdots E_k))(\det B) = (\det A)(\det B) \end{aligned}$$

(ii) If A is not invertible, then $\det A = 0$ by Theorem 4.6. By Exercise 3.3 47 (For square matrices A and B, if AB is invertible, then A and B are both invertible.) AB is also not invertible, so $\det(AB) = 0$ by Theorem 4.6. Therefore, $\det(AB) = (\det A)(\det B) = 0$. □

Theorem 4.9
If A is invertible, then

$$\det A^{-1} = \frac{1}{\det A}$$

Proof. Let A be a invertible matrix. By Theorem 4.8, $(\det A)(\det A^{-1}) = \det I = 1$ since $AA^{-1} = I$. By Theorem 4.6, A is invertible so $\det A \neq 0$. Therefore,

$$\det A^{-1} = \frac{1}{\det A}$$

□

Lemma : Theorem 4.10 for Elementary Matrices
If E is an elementary matrix, then

$$\det E = \det E^T$$

Proof. (i) Since $E_{ij}^T = E_{ij}$, $\det E_{ij} = \det E_{ij}^T$.

(ii) Since $(E_i(k))^T = E_i(k)$e, $\det E_i(k) = \det (E_i(k))^T$.

(iii) Since $(E_{ij}(k))^T = E_{ji}(k)$, $\det (E_{ij}(k))^T = \det E_{ji}(k) = 1 = \det E_{ij}(k)$. □

Theorem 4.10
For any square matrix A,

$$\det A = \det A^T$$

Proof. (i) Suppose that A is an invertible matrix. By F.T.I.M, A is a product of elementary matrices. Let $E_1, E_2, \cdots, E_k$ be the elementary matrices whose product is A. Then by the lemma above,

$$\begin{aligned}\det A^T &= \det ({E_k}^T \cdots {E_2}^T {E_1}^T) = (\det E_k^T) \cdots (\det E_2^T)(\det E_1^T) \\ &= (\det E_k) \cdots (\det E_2)(\det E_1) = \det (E_1 E_2 \cdots E_k) = \det A\end{aligned}$$

(ii) If A is not invertible, then A^T is also not invertible. Therefore, by Theorem 4.6,

$$\det A = 0 = \det A^T$$

□

We denote the matrix obtained by replacing ith column of A by $\mathbf{x}$ as $A_i(\mathbf{x})$.

Theorem 4.11 : Cramer's Rule
Let A be an invertible $n \times n$ matrix and let $\mathbf{b}$ be a vector in $\mathbb{R}^n$. Then the unique solution $\mathbf{x}$ of the system $A\mathbf{x} = \mathbf{b}$ is given by

$$x_i = \frac{\det(A_i(\mathbf{b}))}{\det A} \text{ for } i = 1, \cdots, n$$

Proof. Let A be an invertible $n \times n$ matrix, and $\mathbf{b}$ be a vector in $\mathbb{R}^n$.

Let $\mathbf{x}$ be a vector in $\mathbb{R}^n$ such that $A\mathbf{x} = A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \mathbf{b}$. For $i = 1, 2, \cdots, n$,

$$\begin{aligned}AI_i(\mathbf{x}) = A \begin{bmatrix} \mathbf{e}_1 & \cdots & \mathbf{x} & \cdots & \mathbf{e}_n \end{bmatrix} &= \begin{bmatrix} A\mathbf{e}_1 & \cdots & A\mathbf{x} & \cdots & A\mathbf{e}_n \end{bmatrix} \\ &= \begin{bmatrix} A\mathbf{e}_1 & \cdots & \mathbf{b} & \cdots & A\mathbf{e}_n \end{bmatrix} = A_i(\mathbf{b})\end{aligned}$$

Then, by Laplace Expansion Theorem,

$$\det A_i(\mathbf{b}) = (\det A)(\det I_i(\mathbf{x})) = (\det A) \begin{vmatrix} 1 & 0 & \cdots & x_1 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & x_2 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & x_i & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & x_{n-1} & \cdots & 1 & 0 \\ 0 & 0 & \cdots & x_n & \cdots & 0 & 1 \end{vmatrix}$$

$$= (\det A)(x_i \det I_{n-1}) = x_i \det A$$

$$\therefore x_i = \frac{\det (A_i(\mathbf{b})))}{\det A}$$

□

Definition. The **adjoint** of a square matrix A is denoted by $\mathbf{adj(A)}$ and defined as

$$\operatorname{adj} A = [C_{ji}] = \begin{bmatrix} C_{11} & \cdots & C_{n1} \\ \vdots & & \vdots \\ C_{1n} & \cdots & C_{nn} \end{bmatrix}$$

Theorem 4.12
Let A be an invertible $n \times n$ matrix. Then

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A$$

Proof. Since $AA^{-1} = I_n$, the ith column of A^{-1} satisfies the equation $A\mathbf{x} = \mathbf{e}_i$. By Cramer's Rule, $(A^{-1})_i^C$ the ith column of A^{-1}, is given as

$$(A^{-1})_i^C = \begin{bmatrix} \frac{\det A_1(\mathbf{e}_i)}{\det A} \\ \vdots \\ \frac{\det A_n(\mathbf{e}_i)}{\det A} \end{bmatrix}$$

Since $\det A_j(\mathbf{e}_i) = C_{ij}$,

$$(A^{-1})_i^C = \frac{1}{\det A} \begin{bmatrix} C_{i1} \\ \vdots \\ C_{in} \end{bmatrix}$$

Therefore,

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix} = \frac{1}{\det A} \operatorname{adj} A$$

□

Extra Theorem
Let A be an $n \times n$ matrix. Then

$$(\operatorname{adj} A)A = A(\operatorname{adj} A) = (\det A)I_n$$

Proof. Let $A = [a_{ij}]$ be an $n \times n$ matrix. Consider each entries of $A(\operatorname{adj} A)$ and $(\operatorname{adj} A)A$.

First we compute the diagonal entries of $A(\operatorname{adj} A)$ and $(\operatorname{adj} A)A$. The ith diagonal entry of $A(\operatorname{adj} A)$ is given as

$$[A(\operatorname{adj} A)]_{ii} = \sum_{k=1}^{n} a_{ik}(\operatorname{adj} A)_{ki} = \sum_{k=1}^{n} a_{ik}C_{ik}$$

which is a cofactor expansion along the ith row. Also, the ith diagonal entry of $(\operatorname{adj} A)A$ is given as

$$[(\operatorname{adj} A)A]_{ii} = \sum_{k=1}^{n} (\operatorname{adj} A)_{ik}a_{ki} = \sum_{k=1}^{n} a_{ki}C_{ki}$$

which is a cofactor expansion along the ith column. Therefore, $[(\operatorname{adj} A)A]_{ii} = [A(\operatorname{adj} A)]_{ii} = \det A$ by the Laplace Expansion Theorem.

Now consider the non-diagonal entries of $A(\operatorname{adj} A)$ and $(\operatorname{adj} A)A$. The (i, j) entry of $A(\operatorname{adj} A)$ is

$$[A(\operatorname{adj} A)]_{ij} = \sum_{k=1}^{n} a_{ik}(\operatorname{adj} A)_{kj} = \sum_{k=1}^{n} a_{ik}C_{jk}, i \neq j$$

This is the cofactor expansion along the jth row of the matrix obtained by replacing the jth row by the ith row of A. The (i, j) entry of $(\operatorname{adj} A)A$ is

$$[(\operatorname{adj} A)A]_{ij} = \sum_{k=1}^{n} (\operatorname{adj} A)_{ik}a_{kj} = \sum_{k=1}^{n} a_{kj}C_{ki}$$

This is the cofactor expansion along the ith column of the matrix obtained by replacing the ith column by the jth column of A. Therefore, $[A(\operatorname{adj} A)]_{ij} = [(\operatorname{adj} A)A]_{ij} = 0$, since such matrices has two identical rows.

Consequently, we have

$$[(\operatorname{adj} A)A]_{ij} = [A(\operatorname{adj} A)]_{ij} = \begin{cases} \det A \ , \ i = j \\ 0 \ , \ i \neq j \end{cases}$$

thus $(\operatorname{adj} A)A = A(\operatorname{adj} A) = (\det A)I_n$. □

A가 invertible하지 않아도 성립한다는 점에서 Theorem 4.12의 확장이라고 할 수 있다.

4.3 Eigenvalues and Eigenvectors of $n \times n$ Matrices

Definition. The **characteristic polynomial** of A is defined as $\det(A - \lambda I)$, which is a polynomial in λ. The **characteristic equation** of A is defined as $\det(A - \lambda I) = 0$. The solution of characteristic equation of a matrix is the eigenvalue of the matrix.

Definition. The **algebraic multiplicity** of an eigenvalue is the multiplicity of it as a root of the characteristic equation. In other words, the algebraic multiplicity of an eigenvalue λ_0 is the degree of $(\lambda - \lambda_0)$ in the characteristic equation.

Definition. The **geometric multiplicity** of an eigenvalue is the dimension of eigenspace corresponding to it ($\dim E_\lambda$).

Theorem 4.15
The eigenvalues of a triangular matrix are the entries on its main diagonal.

Proof. Let A be a $n \times n$ triangular matrix. Since $A - \lambda I$ is also a triangular matrix, by Theorem 4.2, the characteristic equation of A is

$$\det(A - \lambda I) = (a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda) = 0$$

Therefore, the diagonal entries $a_{11}, a_{22}, \cdots, a_{nn}$ are eigenvalues of A, since they are solutions of the characteristic equation. $\square$

Theorem 4.16
A square matrix A is invertible if and only if 0 is *not* an eigenvalue of A.

Proof. By Theorem 4.6, A is not invertible if and only if $\det A = \det(A - 0I) = 0$. Since the eigenvalues of A are solutions of the characteristic equation $\det(A - \lambda I) = 0$, $\det(A - 0I) = 0$ if and only if 0 is an eigenvalue of A. $\square$

Theorem 4.17 : The Fundamental Theorem of Invertible Matrices : Version 3
Let A be an $n \times n$ matrix. The following propositions are equivalent:

a. A is invertible.

b. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^n$.

c. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

d. The RREF of A is I_n.

e. A is a product of elementary matrices.

f. $\text{rank}(A) = n$

g. $\text{nullity}(A) = 0$

h. The columns of A are linearly independent.

i. The columns of A span $\mathbb{R}^n$.

j. The columns of A form a basis for $\mathbb{R}^n$.

k. The rows of A are linearly independent.

l. The rows of A span $\mathbb{R}^n$.

m. The rows of A form a basis for $\mathbb{R}^n$.

n. $\det A \neq 0$

o. 0 is not an eigenvalue of A.

Proof. (a $\Leftrightarrow$ n) Theorem 4.6.
(a $\Leftrightarrow$ o) Theorem 4.16. □

Theorem 4.18
Let A be a square matrix with eigenvalue λ and corresponding eigenvector $\mathbf{x}$.

a. For any positive integer n, λ^n is an eigenvalue of A^n with corresponding eigenvector $\mathbf{x}$.

b. If A is invertible, then $1/\lambda$ is an eigenvalue of A^{-1} with corresponding eigenvector $\mathbf{x}$.

c. If A is invertible, then for any integer n, λ^n is an eigenvalue of A^n with corresponding eigenvector $\mathbf{x}$.

Proof. Let A be a square matrix and let λ be an eigenvalue of A, and $\mathbf{x}$ be an eigenvector corresponding to λ. Then $A\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{x} \neq \mathbf{0}$.

a. We prove the proposition by mathematical induction.

   (i) $n = 1$
   Trivially, λ is an eigenvalue of A.

   (ii) Suppose that the proposition is true if $n = k$ ($k \in \mathbb{N}$). Since λ^k is an eigenvalue of A^k with corresponding eigenvector $\mathbf{x}$,

$$A^{k+1}\mathbf{x} = (AA^k)\mathbf{x} = A(A^k\mathbf{x}) = A(\lambda^k\mathbf{x}) = \lambda^k(A\mathbf{x}) = \lambda^{k+1}\mathbf{x}$$

   Thus, λ^{k+1} is an eigenvalue of A^{k+1} with corresponding eigenvector $\mathbf{x}$.

   By (i) and (ii), for all positive integer n, λ^n is an eigenvalue of A^n with corresponding eigenvector $\mathbf{x}$.

b. **(Exercise 4.3 13)** Suppose that A is an invertible matrix. Since $A\mathbf{x} = \lambda\mathbf{x}$,

$$\mathbf{x} = I\mathbf{x} = (A^{-1}A)\mathbf{x} = A^{-1}(A\mathbf{x}) = A^{-1}(\lambda\mathbf{x}) = \lambda(A^{-1}\mathbf{x})$$

   Since A is invertible, by F.T.I.M, $\lambda \neq 0$.

$$\therefore A^{-1}\mathbf{x} = \frac{1}{\lambda}\mathbf{x}$$

   Therefore, $1/\lambda$ is an eigenvalue of A^{-1} with corresponding eigenvector $\mathbf{x}$.

c. **(Exericse 4.3 14)**
(i) $n > 0$, Theorem 4.18(a) gives the proof.
(ii) $n = 0$, $A^0 = I$, so $I\mathbf{x} = \mathbf{x} = \lambda^0\mathbf{x}$.
(iii) $n < 0$, Since $A^n = (A^{-1})^{-n}$, Theorem 4.18(a) and Theorem 4.18(b) gives the proof.

□

Theorem 4.19
Suppose the $n \times n$ matrix A has eigenvectors $\mathbf{v}_1, \cdots, \mathbf{v}_m$ with $\lambda_1, \lambda_2, \cdots, \lambda_m$. If $\mathbf{x}$ is a vector in $\mathbb{R}^n$ that can be expressed as a linear combination of these eigenvectors so that scalars $c_1, c_2, \cdots, c_m$ exist such that

$$\mathbf{x} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_m\mathbf{v}_m$$

then, for any integer k,

$$A^k\mathbf{x} = c_1\lambda_1^k\mathbf{v}_1 + c_2\lambda_2^k\mathbf{v}_2 + \cdots + c_m\lambda_m^k\mathbf{v}_m$$

Proof. **(Exercise 4.3 42)** Let A be an $n \times n$ matrix and let $\lambda_1, \cdots, \lambda_m$ be eigenvalues of A with corresponding eigenvectors $\mathbf{v}_1, \cdots, \mathbf{v}_m$. Suppose that there exist scalars $c_1, \cdots, c_m$ such that

$$\mathbf{x} = c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m$$

For any integer k,

$$A^k\mathbf{x} = A^k(c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m) = c_1(A^k\mathbf{v}_1) + \cdots + c_m(A^k\mathbf{v}_m)$$

By Theorem 4.18(c), λ_i^k is an eigenvalue of A^k with corresponding eigenvector $\mathbf{v}_i$ for i from 1 to m. Therefore,

$$c_1(A^k\mathbf{v}_1) + \cdots + c_m(A^k\mathbf{v}_m) = c_1\lambda_1^k\mathbf{v}_1 + \cdots + c_m\lambda_m^k\mathbf{v}_m$$

□

Theorem 4.20
Let A be an $n \times n$ matrix and let $\lambda_1, \cdots, \lambda_m$ be distinct eigenvalues of A with corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m$. Then $\mathbf{v}_1, \cdots, \mathbf{v}_m$ are linearly independent.

Proof. Suppose that $\mathbf{v}_1, \cdots, \mathbf{v}_m$ are linearly dependent. Then one of the vectors can be expressed as a linear combination of the previous vectors. Let $\mathbf{v}_k$ be the first of such vectors. Then there exist scalars $c_1, c_2, \cdots, c_{k-1}$ such that

$$\mathbf{v}_k = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{k-1}\mathbf{v}_{k-1}$$

Since $\mathbf{v}_1, \cdots, \mathbf{v}_m$ are eigenvectors of A with corresponding eigenvalue $\lambda_1, \lambda_2, \cdots, \lambda_m$,

$$\begin{aligned}\lambda_k\mathbf{v}_k = A\mathbf{v}_k &= A(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_{k-1}\mathbf{v}_{k-1}) \\ &= c_1(A\mathbf{v}_1) + c_2(A\mathbf{v}_2) + \cdots + c_{k-1}(A\mathbf{v}_{k-1}) \\ &= c_1\lambda_1\mathbf{v}_1 + c_2\lambda_2\mathbf{v}_2 + \cdots + c_{k-1}\lambda_{k-1}\mathbf{v}_{k-1}\end{aligned}$$

Since $\lambda_k\mathbf{v}_k = c_1\lambda_k\mathbf{v}_1 + \cdots + c_{k-1}\lambda_{k-1}\mathbf{v}_{k-1}$,

$$c_1(\lambda_1 - \lambda_k)\mathbf{v}_1 + c_2(\lambda_2 - \lambda_k)\mathbf{v}_2 + \cdots + c_{k-1}(\lambda_{k-1} - \lambda_k)\mathbf{v}_{k-1} = \lambda_k\mathbf{v}_k - \lambda_k\mathbf{v}_k = \mathbf{0}$$

Since the vectors $\mathbf{v}_1, \cdots, \mathbf{v}_{k-1}$ are linearly independent (remind that $\mathbf{v}_k$ was the 'first' vector to be expressed as the linear combination of other vectors), $c_1(\lambda_1 - \lambda_k) = \cdots = c_{k-1}(\lambda_{k-1} - \lambda_k) = 0$. The eignevalues are distinct, thus $c_1 = c_2 = \cdots = c_{k-1} = 0$. Therefore, $\mathbf{v}_k = \mathbf{0}$.

It's a contradiction since eigenvectors cannot be a zero vector. Therefore, $\mathbf{v}_1, \cdots, \mathbf{v}_m$s are linearly independent. □

4.4 Similarity and Diagonalization

Definition. Let A and B be $n \times n$ matrices. We say that **A is similar to B** if there exists an invertible $n \times n$ matrix P such that $P^{-1}AP = B$. We denote this as $A \sim B$.

Theorem 4.21
Let A, B, and C be $n \times n$ matrices.

a. $A \sim A$ (Reflectivity, 반사율)

b. If $A \sim B$, then $B \sim A$. (Symmetricity, 대칭률)

c. If $A \sim B$ and $B \sim C$, then $A \sim C$. (Transitivity, 추이율)

Proof. Let A, B, and C be $n \times n$ matrices.

a. $A = {I_n}^{-1} A I_n$, therefore $A \sim A$.

b. If $A \sim B$, there exists an invertible $n \times n$ matrix P such that $B = P^{-1}AP$. Then $A = PBP^{-1} = (P^{-1})^{-1}BP^{-1}$, therefore $B \sim A$.

c. **(Exercise 4.4 30)** If $A \sim B$ and $B \sim C$, there exist invertible $n \times n$ matrices P_1, P_2 such that $B = {P_1}^{-1}AP_1$ and $C = {P_2}^{-1}BP_2$. Then $C = {P_2}^{-1}BP_2 = {P_2}^{-1}({P_1}^{-1}AP_1)P_2 = (P_1P_2)^{-1}A(P_1P_2)$. Therefore, $A \sim C$.

□

Theorem 4.22
Let A and B be $n \times n$ matrices with $A \sim B$. Then

a. $\det A = \det B$

b. A is invertible if and only if B is invertible.

c. A and B have the same rank.

d. A and B have the same characteristic polynomial.

e. A and B have the same eigenvalues.

Proof. Let A and B be $n \times n$ matrices such that $A \sim B$. Then there exists an invertible $n \times n$ matrix P such that $B = P^{-1}AP$.

a.
$$\det B = \det(P^{-1}AP) = \det P^{-1} \det A \det P = \frac{1}{\det P} \det A \det P = \det A$$

b. **(Exercise 4.4 31)** By Theorem 4.22(a), $\det A = \det B$. By F.T.I.M, A is invertible if and only if $\det A \neq 0$, and so as B. Therefore, A is invertible if and only if B is invertible.

c. **(Exercise 4.4 32)** Since $B = P^{-1}AP$, $\text{rank}(B) = \text{rank}(P^{-1}AP) \leq \text{rank}(A)$ ($\because$ Exercise 3.5 59, 60) Similarly, $\text{rank}(A) = \text{rank}(PBP^{-1}) \leq \text{rank}(B)$. Therefore, $\text{rank}(A) = \text{rank}(B)$.

d.

$$\begin{aligned}\det(B - \lambda I) &= \det(P^{-1}AP - \lambda I) \\ &= \det(P^{-1}AP - P^{-1}(\lambda I)P) = \det(P^{-1}(A - \lambda I)P) = \det(A - \lambda I)\end{aligned}$$

e. **(Exercise 4.4 33)** Since A and B have same characteristic polynomial and characteristic equation, the eigenvalues of A and B should have the same eigenvalues.

□

Note. $A \sim B$ 일때 $\text{tr}(A) = \text{tr}(B)$ 도 성립한다. Exercise 4.3 40 참고.

Definition. An $n \times n$ matrix A is **diagonalizable** if there exists a diagonal $n \times n$ matrix D such that $A \sim D$, that is there exists an invertible $n \times n$ matrix P which satisfies $P^{-1}AP = D$.

Theorem 4.23
Let A be $n \times n$ matrix. Then A is diagonalizable if and only if A has n linearly independent eigenvectors.
More precisely, there exist an invertible matrix P and diagonal matrix D such that $P^{-1}AP = D$ if and only if the columns of P are n linearly independent eigenvectors of A and the diagonal entries of D are the eigenvalues of A corresponding to the eigenvectors in P in the same order.

Proof. ($\Rightarrow$) Suppose that A is diagonalizable, then there exists a diagonal matrix $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$ and invertible matrix $P = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix}$ such that $P^{-1}AP = D$. Since $AP = PD$,

$$\begin{aligned} AP = A\begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix} &= \begin{bmatrix} A\mathbf{p}_1 & A\mathbf{p}_2 & \cdots & A\mathbf{p}_n \end{bmatrix} \\ = PD = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix} &= \begin{bmatrix} \lambda_1\mathbf{p}_1 & \lambda_2\mathbf{p}_2 & \cdots & \lambda_n\mathbf{p}_n \end{bmatrix} \end{aligned}$$

Thus $A\mathbf{p}_1 = \lambda_1\mathbf{p}_1, A\mathbf{p}_2 = \lambda_2\mathbf{p}_2, \cdots, A\mathbf{p}_n = \lambda_n\mathbf{p}_n$. Also, none of $\mathbf{p}_1, \cdots, \mathbf{p}_n$ is nonzero since P is an invertible matrix. Therefore, $\mathbf{p}_1, \cdots, \mathbf{p}_n$ are eigenvectors of A with corresponding eigenvalues $\lambda_1, \cdots, \lambda_n$, respectively.

($\Leftarrow$) Similarly, let $D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$ and let $P = \begin{bmatrix} \mathbf{p}_1 & \mathbf{p}_2 & \cdots & \mathbf{p}_n \end{bmatrix}$, where P is invertible. Since $\mathbf{p}_1, \cdots, \mathbf{p}_n$ are eigenvectors of A with corresponding eigenvalue $\lambda_1, \cdots, \lambda_n$, $AP = PD$. Also, P is invertible since columns of P are linearly independent. Therefore, $D = P^{-1}AP$. □

Theorem 4.24
Let A be an $n \times n$ matrix and let $\lambda_1, \cdots, \lambda_k$ be distinct eigenvalues of A. If $\mathcal{B}_i$ is a basis for the eigenspace E_{λ_i}, then $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \mathcal{B}_k$ (i.e., the total collection of basis vectors for all of the eigenspaces) is linearly independent.

Proof. Let $\mathcal{B}_i = \{\mathbf{v}_{i1}, \mathbf{v}_{i2}, \cdots, \mathbf{v}_{in_i}\}$, for i from 1 to k. Consider the equation

$$c_{11}\mathbf{v}_{11} + \cdots + c_{1n_1}\mathbf{v}_{1n_1} + \cdots + c_{k1}\mathbf{v}_{k1} + \cdots + c_{kn_k}\mathbf{v}_{kn_k} = \mathbf{0}$$

Let $\mathbf{x}_i = c_{i1}\mathbf{v}_{i1} + \cdots + c_{in_i}\mathbf{v}_{in_i}$, then $\mathbf{x}_1 + \cdots + \mathbf{x}_k = \mathbf{0}$. Since $\mathbf{x}_i \in \operatorname{span}(\mathcal{B}_i)$, $\mathbf{x}_i$ is either an eigenvector of A corresponding to λ_i or a zero vector. By Theorem 4.20, the eigenvectors corresponding to distinct eigenvalues should be linear independent. Thus, $\mathbf{x}_1 + \cdots + \mathbf{x}_k = \mathbf{0}$ gives $\mathbf{x}_1 = \cdots = \mathbf{x}_k = \mathbf{0}$. (If at least one of $\mathbf{x}_1, \cdots, \mathbf{x}_k$ is nonzero, then a nontrivial solution for the equation exists so that it contradicts with the linear independence of the eigenvectors.) Since the vectors in each $\mathcal{B}_i$ are linear independent, all c_{ij} should be zero. Therefore, $\mathcal{B}$ is a set of linearly independent vectors. □

Theorem 4.25
If A is an $n \times n$ matrix with n distinct eigenvalues, then A is diagonalizable.

Proof. By Theorem 4.20, n eigenvectors of A corresponding to n distinct eigenvalues should be linearly independent. Therefore, by Theorem 4.23, A is diagonalizable. □

Lemma 4.26
If A is $n \times n$ matrix, for each eigenvalues,

$$\text{(geometric multiplicity)} \leq \text{(algebraic multiplicity)}$$

Proof. Suppose that λ_0 is an eigenvalue of A with geometric multiplicity p. Then $\dim E_{\lambda_0} = p$, so let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_p\}$ be a basis for E_{λ_0}.

Let Q be a invertible $n \times n$ matrix which has $\mathbf{v}_1, \cdots, \mathbf{v}_p$ at its first p columns, so

$$Q = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_p & \mathbf{v}_{p+1} & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} U & V \end{bmatrix}$$

where U is an $n \times p$ matrix and V is an $n \times (n-p)$ matrix. Let $Q^{-1} = \begin{bmatrix} C \\ D \end{bmatrix}$, where C is a $p \times n$ matrix and D is an $(n-p) \times n$ matrix. Then

$$\begin{bmatrix} I_p & O \\ O & I_{n-p} \end{bmatrix} = I_n = Q^{-1}Q = \begin{bmatrix} C \\ D \end{bmatrix} \begin{bmatrix} U & V \end{bmatrix} = \begin{bmatrix} CU & CV \\ DU & DV \end{bmatrix}$$

Thus, $CU = I_p$, $CV = O$, $DU = O$, and $DV = I_{n-p}$.

Since $AU = A\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_p \end{bmatrix} = \begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_p \end{bmatrix} = \begin{bmatrix} \lambda_0\mathbf{v}_1 & \lambda_0\mathbf{v}_2 & \cdots & \lambda_0\mathbf{v}_p \end{bmatrix} = \lambda_0 U$,

$$Q^{-1}AQ = \begin{bmatrix} C \\ D \end{bmatrix} A \begin{bmatrix} U & V \end{bmatrix} = \begin{bmatrix} CAU & CAV \\ DAU & DAV \end{bmatrix} = \begin{bmatrix} \lambda_0 CU & CAV \\ \lambda_0 DU & DAV \end{bmatrix} = \begin{bmatrix} \lambda_0 I_p & CAV \\ O & DAV \end{bmatrix}$$

which is a block upper triangular matrix, so by Exercise 4.2 69 the characteristic polynomial of $Q^{-1}AQ$ is,

$$\begin{aligned}\det(A-\lambda_0 I) = \det(Q^{-1}AQ-\lambda_0 I) &= \begin{vmatrix}(\lambda_0-\lambda)I_p & CAV \\ O & DAV-\lambda I_{n-p}\end{vmatrix} \\ &= (\det(\lambda_0-\lambda)I_p)(\det(DAV-\lambda I_{n-p})) \\ &= (\lambda_0-\lambda)^p(\det(DAV-\lambda I_{n-p}))\end{aligned}$$

Therefore, algebraic multiplicity of λ_0 is at least p. □

Theorem 4.27 : The Diagonalization Theorem
Let A be an $n \times n$ matrix whose distinct eigenvalues are $\lambda_1, \lambda_2, \cdots, \lambda_k$. The following statements are equivalent :

a. A is diagonalizable.

b. The union $\mathcal{B}$ of the bases of the eigenspaces of A(as in Theorem 4.24) contains n vectors.

c. The algebraic multiplicity of each eigenvalue equals its geometric multiplicity.

Proof. (a $\Rightarrow$ b) Suppose that A is diagonalizable, then A has n linearly independent eigenvectors by Theorem 4.23. If n_i vectors of these n eigenvectors correspond to the eigenvalue λ_i, the basis $\mathcal{B}_i$ for the eigenspace E_{λ_i} should have at least n_i vectors. Thus, $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \cdots \cup \mathcal{B}_k$ contains at least $n_1 + n_2 + \cdots + n_k = n$ vectors. By Theorem 4.24, $\mathcal{B}$ is a linearly independent set of vectors in $\mathbb{R}^n$, so it has at most n vectors. Therefore, $\mathcal{B}$ has exactly n vectors.

(b $\Rightarrow$ a) By Theorem 4.24, the union of the bases for the eigenspaces is a linearly independent set of vectors. Since $\mathcal{B}$ has exactly n vectors, A has n linearly independent eigenvectors. Therefore, A is diagonalizable by Theorem 4.23.

(b $\Rightarrow$ c) Let the algebraic multiplicity of $\lambda_1, \lambda_2, \cdots, \lambda_k$ be $m_1, m_2, \cdots, m_k$. By Theorem 4.26, the geometric multiplicity of each eigenvalue is lesser or equal to the algebraic multiplicity. Thus,

$$\begin{gathered} n = \dim E_{\lambda_1} + \dim E_{\lambda_2} + \cdots + \dim E_{\lambda_k} \le m_1 + m_2 + \cdots + m_k = n \\ \therefore (m_1 - \dim E_{\lambda_1}) + \cdots + (m_k - \dim E_{\lambda_k}) = 0 \end{gathered}$$

Therefore, the algebraic multiplicity and the geometric multiplicity (which is the dimension of eigenspace corresponding to the eigenvalue) should be the same.

(c $\Rightarrow$ b) Let the algebraic multiplicity of $\lambda_1, \lambda_2, \cdots, \lambda_k$ be $m_1, m_2, \cdots, m_k$. Then $\mathcal{B}$ has $\dim E_{\lambda_2} + \cdots + \dim E_{\lambda_k} = m_1 + m_2 + \cdots + m_k = n$ vectors. □

Chapter 5

Orthogonality

5.1 Orthogonality in $\mathbb{R}^n$

Definition. A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$ in $\mathbb{R}^n$ is an **orthogonal set** if all pairs of distinct vectors in the set are orthogonal, that is

$$\mathbf{v}_i \cdot \mathbf{v}_j = 0 \text{ for } i \neq j, 1 \leq i, j \leq k$$

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

Definition. A set of vectors in $\mathbb{R}^n$ is an **orthonormal set** if it is an orthogonal set of unit vectors. Similarly, an **orthonormal basis** for a subspace W of $\mathbb{R}^n$ is a basis of W that is an orthonormal set.

Definition. An **orthogonal matrix** is a square matrix whose columns form an *orthonormal* set.

Theorem 5.1
If $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is an orthogonal set of nonzero vectors in $\mathbb{R}^n$, then these vectors are linearly independent.

Proof. Consider the equation $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$. For any $1 \leq i \leq k$,

$$\begin{aligned}(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) \cdot \mathbf{v}_i &= c_1(\mathbf{v}_1 \cdot \mathbf{v}_i) + \cdots + c_k(\mathbf{v}_k \cdot \mathbf{v}_i) \\ &= c_i(\mathbf{v}_i \cdot \mathbf{v}_i) = \mathbf{0} \cdot \mathbf{v} = 0\end{aligned}$$

Since $\mathbf{v}_i$ is a nonzero vector, $c_i = 0$. Thus $c_1 = c_2 = \cdots = c_k = 0$. Therefore, $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k\}$is a linearly independent set of vectors. □

Theorem 5.2
Let $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ be an orthogonal basis for a subspace W of $\mathbb{R}^n$ and let $\mathbf{w}$ be any vector in W. Then the unique scalars $c_1, \cdots, c_k$ such that

$$\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$$

are given by

$$c_i = \frac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \text{ for } i = 1, \cdots, k$$

Proof. Since $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is a basis for W, the scalars $c_1, \cdots, c_k$ which satisfies $\mathbf{w} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$ are unique. (Remind Theorem 3.29.) For any $1 \leq i \leq k$,

$$\begin{aligned}
\mathbf{w} \cdot \mathbf{v}_i &= (c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) \cdot \mathbf{v}_i \\
&= c_1(\mathbf{v}_1 \cdot \mathbf{v}_i) + \cdots + c_k(\mathbf{v}_k \cdot \mathbf{v}_i) \\
&= c_i(\mathbf{v}_i \cdot \mathbf{v}_i) \\
\therefore c_i &= \frac{\mathbf{w} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i} \text{ for } i = 1, 2, \cdots, k
\end{aligned}$$

□

Theorem 5.3
Let $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ be an orthonormal basis for a subspace W of $\mathbb{R}^n$, and let $\mathbf{w}$ be any vector in W. Then $\mathbf{w}$ can be represented as the linear combination of vectors in such basis as

$$\mathbf{w} = (\mathbf{w} \cdot \mathbf{q}_1)\mathbf{q}_1 + (\mathbf{w} \cdot \mathbf{q}_2)\mathbf{q}_2 + \cdots + (\mathbf{w} \cdot \mathbf{q}_k)\mathbf{q}_k$$

and the representation is unique.

Proof. By Theorem 5.2, the representation of $\mathbf{w}$ as the linear combination of $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ is unique, and is expressed as

$$\mathbf{w} = \left(\frac{\mathbf{w} \cdot \mathbf{q}_1}{\mathbf{q}_1 \cdot \mathbf{q}_1}\right)\mathbf{q}_1 + \left(\frac{\mathbf{w} \cdot \mathbf{q}_2}{\mathbf{q}_2 \cdot \mathbf{q}_2}\right)\mathbf{q}_2 + \cdots + \left(\frac{\mathbf{w} \cdot \mathbf{q}_k}{\mathbf{q}_k \cdot \mathbf{q}_k}\right)\mathbf{q}_k$$

Since $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ is an orthonormal set, $\mathbf{q}_i \cdot \mathbf{q}_i = 1$ for all $1 \leq i \leq k$. Therefore,

$$\mathbf{w} = (\mathbf{w} \cdot \mathbf{q}_1)\mathbf{q}_1 + \cdots + (\mathbf{w} \cdot \mathbf{q}_k)\mathbf{q}_k$$

□

Theorem 5.4
The columns of an $m \times n$ matrix Q form an orthonormal set if and only if $Q^TQ = I_n$.

Proof. Let $Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}$, then

$$(Q^TQ)_{ij} = \mathbf{q}_i \cdot \mathbf{q}_j$$

Therefore, $\{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_k\}$ is an orthonormal set if and only if

$$\mathbf{q}_i \cdot \mathbf{q}_j = (Q^TQ)_{ij} = \begin{cases} 0 \text{ if } i \neq j \\ 1 \text{ if } i = j \end{cases}$$

which holds if and only if $Q^TQ = I_n$. □

Theorem 5.5
A square matrix Q is orthogonal if and only if $Q^{-1} = Q^T$.

Proof. By Theorem 5.4, Q is orthogonal if and only if $Q^TQ = I_n$, which holds if and only if $Q^T = Q^{-1}$. □

Theorem 5.6
Let Q be an $n \times n$ matrix. The following propositions are equivalent:

a. Q is orthogonal.

b. $\|Q\mathbf{x}\| = \|\mathbf{x}\|$

c. $Q\mathbf{x} \cdot Q\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$ for any $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^n$.

Proof. (a $\Rightarrow$ c) Suppose that Q is an orthogonal matrix. By Theorem 5.4, $Q^T Q = I_n$. Therefore, for any $\mathbf{x} \in \mathbb{R}^n$and $\mathbf{y} \in \mathbb{R}^n$,

$$Q\mathbf{x} \cdot Q\mathbf{y} = (Q\mathbf{x})^T Q\mathbf{y} = \mathbf{x}^T Q^T Q\mathbf{y} = \mathbf{x}^T \mathbf{y} = \mathbf{x} \cdot \mathbf{y}$$

(c $\Rightarrow$ a) Let $Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}$, then $\mathbf{q}_i = Q\mathbf{e}_i$. Then,

$$(Q^T Q)_{ij} = \mathbf{q}_i \cdot \mathbf{q}_j = (Q\mathbf{e}_i) \cdot (Q\mathbf{e}_j) = \mathbf{e}_i \cdot \mathbf{e}_j = \begin{cases} 0 \text{ if } i \neq j \\ 1 \text{ if } i = j \end{cases}$$

which shows that Q is an orthogonal matrix since $Q^T Q = I_n$.

(b $\Rightarrow$ c)

$$\begin{aligned} \mathbf{x} \cdot \mathbf{y} &= \frac{1}{4}(\|\mathbf{x} + \mathbf{y}\|^2 - \|\mathbf{x} - \mathbf{y}\|^2) \\ &= \frac{1}{4}(\|Q(\mathbf{x} + \mathbf{y})\|^2 - \|Q(\mathbf{x} - \mathbf{y})\|^2) \\ &= \frac{1}{4}(\|Q\mathbf{x} + Q\mathbf{y}\|^2 - \|Q\mathbf{x} - Q\mathbf{y}\|^2) \\ &= (Q\mathbf{x}) \cdot (Q\mathbf{y}) \end{aligned}$$

(c $\Rightarrow$ b)

$$\|Q\mathbf{x}\| = \sqrt{(Q\mathbf{x}) \cdot (Q\mathbf{x})} = \sqrt{\mathbf{x} \cdot \mathbf{x}} = \|\mathbf{x}\|$$

□

Theorem 5.7
If Q is an orthogonal matrix, then its rows form an orthonormal set.

Proof. By Theorem 5.5, $Q^T = Q^{-1}$. Since $(Q^T)^{-1} = (Q^{-1})^{-1} = Q = (Q^T)^T$, Q^T is also an orthogonal matrix by Theorem 5.4. Therefore, the columns of Q^T, which are rows of Q, form an orthonormal set. □

Theorem 5.8
Let Q be an orthogonal matrix.

a. Q^{-1} is orthogonal.

b. $|\det Q| = 1$

c. If λ is an eigenvalue of Q, then $|\lambda| = 1$.

d. If Q_1 and Q_2 are orthogonal $n \times n$ matrices, then $Q_1 Q_2$ is also an orthogonal matrix.

Proof. Let Q be an orthogonal matrix.

a. **(Exercise 5.1 22)** Since Q is an orthogonal matrix, $Q^T = Q^{-1}$. So

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

By Theorem 5.4, Q^{-1} is also orthogonal.

b. **(Exercise 5.1 23)** Since $Q^T Q = I_n$,

$$\det I_n = \det(Q^T Q) = (\det Q^T)(\det Q) = (\det Q)^2 = 1$$

Therefore, $|\det Q| = 1$.

c. Let λ be a eigenvalue of Q with corresponding eigenvector $\mathbf{v}$. By Theorem 5.6(b),

$$\|\mathbf{v}\| = \|Q\mathbf{v}\| = \|\lambda\mathbf{v}\| = |\lambda|\|\mathbf{v}\|$$

Therefore $|\lambda| = 1$.

d. **(Exercise 5.1 24)** Let Q_1 and Q_2 be orthogonal matrices of the same size. Since ${Q_1}^{-1} = Q_1^T$ and ${Q_2}^{-1} = Q_2^T$.

$$(Q_1 Q_2)^{-1} = {Q_2}^{-1}{Q_1}^{-1} = Q_2^T Q_1^T = (Q_1 Q_2)^T$$

Therefore $Q_1 Q_2$ is also an orthogonal matrix.

□

5.2 Orthogonal Complements and Orthogonal Projections

Definition. Let W be a subspace of $\mathbb{R}^n$. A vector **v in** $\mathbb{R}^n$ **is orthogonal to** W if $\mathbf{v}$ is orthogonal to every vector in W. The set of all vectors that are orthogonal to W is the **orthogonal complement** of W, denoted by $W^\perp$.

Theorem 5.9
Let W be a subspace of $\mathbb{R}^n$. The following propositions hold:

a. $W^\perp$ is a subspace of $\mathbb{R}^n$.

b. $(W^\perp)^\perp = W$

c. $W \cap W^\perp = \{\mathbf{0}\}$

d. If $W = \text{span}(\mathbf{w}_1, \cdots, \mathbf{w}_k)$, then $\mathbf{v} \in W^\perp$ if and only if $\mathbf{v} \cdot \mathbf{w}_i = 0$ for all i from 1 to k.

Proof. Let W be a subspace of $\mathbb{R}^n$.

a. (i) $\mathbf{0} \cdot \mathbf{w} = 0$ for every $\mathbf{w} \in W$, so $\mathbf{0} \in W^\perp$.

Let $\mathbf{u} \in W$ and $\mathbf{v} \in W$, and c a scalar. Then for any $\mathbf{w} \in W$, $\mathbf{u} \cdot \mathbf{w} = \mathbf{v} \cdot \mathbf{w} = 0$.
(ii) $(\mathbf{u} + \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot \mathbf{w} + \mathbf{v} \cdot \mathbf{w} = 0$

(iii) $(c\mathbf{u}) \cdot \mathbf{w} = c(\mathbf{u} \cdot \mathbf{w}) = 0$

b. (i) Let $\mathbf{w}$ be a vector in W, then for every vector $\mathbf{w}^\perp$ in $W^\perp$, $\mathbf{w} \cdot \mathbf{w}^\perp = 0$. This implies that $\mathbf{w} \in (W^\perp)^\perp$, so $W \subseteq (W^\perp)^\perp$.

(ii) Now let $\mathbf{v}$ be a vector in $(W^\perp)^\perp$. By Orthogonal Decomposition Theorem (Theorem 5.11), there are unique vectors $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$ such that $\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$. Since $\mathbf{v} \in (W^\perp)^\perp$. $\mathbf{v}$ and $\mathbf{w}^\perp$ are orthogonal, so

$$\mathbf{v} \cdot \mathbf{w}^\perp = (\mathbf{w} + \mathbf{w}^\perp) \cdot \mathbf{w}^\perp = \mathbf{w} \cdot \mathbf{w}^\perp + \mathbf{w}^\perp \cdot \mathbf{w}^\perp = \mathbf{w}^\perp \cdot \mathbf{w}^\perp = 0$$

Therefore $\mathbf{w}^\perp = \mathbf{0}$, which indicates $\mathbf{v} = \mathbf{w} \in W$. Thus, $(W^\perp)^\perp \subseteq W$.

By (i) and (ii), $W = (W^\perp)^\perp$.

c. **(Exercise 5.2 23)** Suppose that $\mathbf{v} \in W \cap W^\perp$. Then $\mathbf{v} \cdot \mathbf{v} = 0$, which is true if and only if $\mathbf{v} = \mathbf{0}$. Therefore, $W \cap W^\perp = \{\mathbf{0}\}$.

d. **(Exercise 5.2 24)** ($\Rightarrow$) Since $\mathbf{v} \in W^\perp$, $\mathbf{v}$ is orthogonal with every vector in W, including the vectors $\mathbf{w}_1, \cdots, \mathbf{w}_k$.

($\Leftarrow$) For any vector $\mathbf{w} \in W$, there exist scalars $c_1, \cdots, c_k$ such that

$$\mathbf{w} = c_1\mathbf{w}_1 + \cdots + c_k\mathbf{w}_k$$

Since $\mathbf{v}$ is orthogonal to $\mathbf{w}_1, \cdots, \mathbf{w}_k$,

$$\mathbf{w} \cdot \mathbf{v} = (c_1\mathbf{w}_1 + \cdots + c_k\mathbf{w}_k) \cdot \mathbf{v} = c_1(\mathbf{w}_1 \cdot \mathbf{v}) + \cdots + c_k(\mathbf{w}_k \cdot \mathbf{v}) = 0$$

so $\mathbf{w}$ is orthogonal to $\mathbf{v}$. Therefore, $\mathbf{v} \in W^\perp$.

$\square$

Theorem 5.10
Let A be an $m \times n$ matrix. Then 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. That is,

$$(\text{row}(A))^\perp = \text{null}(A) \text{ and } (\text{col}(A))^\perp = \text{null}(A^T)$$

Proof. Let $\mathbf{v}$ be a vector in $\mathbb{R}^n$. Then $\mathbf{v}$ satisfies the equation $A\mathbf{x} = \mathbf{0}$ (which means $\mathbf{v} \in \text{null}(A)$) if and only if $\mathbf{v}$ is orthogonal to every rows of A. Since the rows of A span the row space of A, by Theorem 5.9(d), this is true if and only if $\mathbf{v} \in (\text{row}(A))^\perp$.

Let $\mathbf{v}$ be a vector in $\mathbb{R}^m$. Then $\mathbf{v}$ satisfies the equation $A^T\mathbf{x} = \mathbf{0}$ (which means $\mathbf{v} \in \text{null}(A^T)$) if and only if $\mathbf{v}$ is orthogonal to every rows of A^T, which are columns of A. Since the columns of A span the column space of A, by Theorem 5.9(d), this is true if and only if $\mathbf{v} \in (\text{col}(A))^\perp$.

Also note that
$$(\text{col}(A))^\perp = (\text{row}(A^T))^\perp = \text{null}(A^T)$$

□

The subspaces $\text{row}(A)$, $\text{col}(A)$, $\text{null}(A)$, $\text{null}(A^T)$ are **fundamental subspaces** of the matrix A.

Definition. Let W be a subspace of $\mathbb{R}^n$and let $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ be an orthogonal basis for W. For any vector $\mathbf{v} \in \mathbb{R}^n$, the orthogonal projection of $\mathbf{v}$ onto W is defined as

$$\text{proj}_W(\mathbf{v}) = \left(\frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)\mathbf{u}_1 + \cdots + \left(\frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k}\right)\mathbf{u}_k$$

and the component of $\mathbf{v}$ orthogonal to W is defined as

$$\text{perp}_W\mathbf{v} = \mathbf{v} - \text{proj}_W(\mathbf{v})$$

Note that $\text{proj}_W(\mathbf{v}) = \text{proj}_{\mathbf{u}_1}(\mathbf{v}) + \cdots + \text{proj}_{\mathbf{u}_k}(\mathbf{v})$.

Theorem 5.11 : The Orthogonal Decomposition Theorem
Let W be a subspace of $\mathbb{R}^n$ and let $\mathbf{v}$ be a vector in $\mathbb{R}^n$. Then there are unique vectors $\mathbf{w} \in W$ and $\mathbf{w}^\perp \in W^\perp$ such that
$$\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp$$
which is an orthogonal decomposition of $\mathbf{v}$ with respect to W.

Proof. Existence. Choose an orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ for W, and let $\mathbf{w} = \text{proj}_W(\mathbf{v})$ and $\mathbf{w}^\perp = \text{perp}_W(\mathbf{v})$. Then
$$\mathbf{w} + \mathbf{w}^\perp = \text{proj}_W(\mathbf{v}) + (\mathbf{v} - \text{proj}_W(\mathbf{v})) = \mathbf{v}$$
Since $\mathbf{w}$ is a linear combination of $\mathbf{u}_1, \cdots, \mathbf{u}_k$, $\mathbf{w} \in W$. Also,

$$\begin{aligned}
\mathbf{u}_i \cdot \mathbf{w}^\perp &= \mathbf{u}_i \cdot \text{perp}_W(\mathbf{v}) \\
&= \mathbf{u}_i \cdot \left(\mathbf{v} - \left(\frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)\mathbf{u}_1 - \cdots - \left(\frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k}\right)\mathbf{u}_k\right) \\
&= \mathbf{u}_i \cdot \mathbf{v} - \left(\frac{\mathbf{u}_1 \cdot \mathbf{v}}{\mathbf{u}_1 \cdot \mathbf{u}_1}\right)(\mathbf{u}_1 \cdot \mathbf{u}_i) - \cdots - \left(\frac{\mathbf{u}_k \cdot \mathbf{v}}{\mathbf{u}_k \cdot \mathbf{u}_k}\right)(\mathbf{u}_k \cdot \mathbf{u}_i) \\
&= \mathbf{u}_i \cdot \mathbf{v} - \left(\frac{\mathbf{u}_i \cdot \mathbf{v}}{\mathbf{u}_i \cdot \mathbf{u}_i}\right)(\mathbf{u}_i \cdot \mathbf{u}_i) \\
&= \mathbf{u}_i \cdot \mathbf{v} - \mathbf{u}_i \cdot \mathbf{v} = 0
\end{aligned}$$

By Theorem 5.9(d), $\mathbf{w}^\perp \in W^\perp$. Therefore, $\mathbf{w}$ and $\mathbf{w}^\perp$ satisfies the condition.

Uniqueness. Suppose that there are another vectors $\mathbf{w}_1$ and $\mathbf{w}_1^\perp$ which satisfy the condition: so $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_1^\perp$, $\mathbf{w}_1 \in W$ and $\mathbf{w}_1^\perp \in W^\perp$. Then $\mathbf{w} + \mathbf{w}^\perp = \mathbf{w}_1 + \mathbf{w}_1^\perp = \mathbf{v}$, so

$$\mathbf{w} - \mathbf{w}_1 = \mathbf{w}_1^\perp - \mathbf{w}^\perp$$

Since $\mathbf{w} - \mathbf{w}_1 \in W$ and $\mathbf{w}_1^{\perp} - \mathbf{w}^{\perp} \in W^{\perp}$, it should be $\mathbf{0}$ since the only common vector of W and $W^{\perp}$ is the zero vector. Therefore, $\mathbf{w} = \mathbf{w}_1$ and $\mathbf{w}^{\perp} = \mathbf{w}_1^{\perp}$. □

Note. Orthogonal Decomposition Theorem은 직교성분분해가 존재하고 유일하다는 정리이지만, 그 직교성분분해가 $\mathbf{v} = \text{proj}_W(\mathbf{v}) + \text{perp}_W(\mathbf{v})$ 인 것으로 확장시켜도 문제 없다. 또한, orthogonal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ 가 변해도 orthogonal decomposition은 변화 없다.

Corollary 5.12
If W is a subspace of $\mathbb{R}^n$, then

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

Proof. Corollary 5.12 is equal to Theorem 5.9(b). □

Theorem 5.13
If W is a subspace of $\mathbb{R}^n$, then

$$\dim W + \dim W^{\perp} = n$$

Proof. Let $\mathcal{B}_W = \{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ be an orthogonal basis for W and let $\mathcal{B}_{W^{\perp}} = \{\mathbf{v}_1, \cdots, \mathbf{v}_l\}$ be an orthogonal basis for $W^{\perp}$. Let $\mathcal{B} = \mathcal{B}_W \cup \mathcal{B}_{W^{\perp}}$. Since $\mathbf{u}_i \cdot \mathbf{v}_j = 0$ for every i and j, $\mathcal{B}$ is an orthogonal set of nonzero vectors, which is also a linearly independent set. (Theorem 5.1) By Theorem 5.11, for any $\mathbf{v} \in \mathbb{R}^n$, there exist $\mathbf{w} \in W$ and $\mathbf{w}^{\perp} \in W^{\perp}$ such that $\mathbf{v} = \mathbf{w} + \mathbf{w}^{\perp}$, which indicates that $\mathbf{v} \in \text{span}(\mathcal{B})$. Thus, $\text{span}(\mathcal{B}) = \mathbb{R}^n$, so $\mathcal{B}$ is an orthogonal basis for $\mathbb{R}^n$. Therefore,

$$\dim \mathbb{R}^n = n = k + l = \dim W + \dim W^{\perp}$$

□

Corollary 5.14
If A is an $m \times n$ matrix,

$$\text{rank}(A) + \text{nullity}(A) = n$$

Proof. By Theorem 5.13,

$$\dim(\text{row}(A)) + \dim(\text{row}(A))^{\perp} = \dim(\text{row}(A)) + \dim(\text{null}(A)) = \text{rank}(A) + \text{nullity}(A)$$

□

5.3 The QR Factorization

Theorem 5.15 : The Gram-Schmidt Process
Let $\{\mathbf{x}_1, \cdots, \mathbf{x}_k\}$ be a basis for a subspace W of $\mathbb{R}^n$. Define $\mathbf{v}_1, \cdots, \mathbf{v}_k$ and $W_1, \cdots, W_k$ as following:

$$\mathbf{v}_1 = \mathbf{x}_1 \qquad\qquad W_1 = \mathrm{span}(\mathbf{x}_1)$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \left(\tfrac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 = \mathbf{x}_2 - \mathrm{proj}_{W_1}(\mathbf{x}_2) \qquad\qquad W_2 = \mathrm{span}(\mathbf{x}_1, \mathbf{x}_2)$$

$$\vdots \qquad\qquad \vdots$$

$$\mathbf{v}_k = \mathbf{x}_k - \left(\tfrac{\mathbf{x}_k \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \cdots - \left(\tfrac{\mathbf{x}_k \cdot \mathbf{v}_{k-1}}{\mathbf{v}_{k-1} \cdot \mathbf{v}_{k-1}}\right)\mathbf{v}_{k-1} = \mathbf{x}_k - \mathrm{proj}_{W_{k-1}}(\mathbf{x}_k) \quad W_k = \mathrm{span}(\mathbf{x}_1, \cdots, \mathbf{x}_k)$$

Then $\{\mathbf{v}_1, \cdots, \mathbf{v}_i\}$ is an orthogonal basis for W_i.

Proof. We prove the theorem using the mathematical induction.

(i) Since $\mathbf{v}_1 = \mathbf{x}_1$, $\{\mathbf{v}_1\}$ is an orthogonal basis for W_1.

(ii) Suppose that $\{\mathbf{v}_1, \cdots, \mathbf{v}_i\}$ is an orthogonal basis for $W_i = \mathrm{span}(\mathbf{x}_1, \cdots, \mathbf{x}_i)$. Let

$$\mathbf{v}_{i+1} = \mathbf{x}_{i+1} - \left(\frac{\mathbf{x}_{i+1} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\right)\mathbf{v}_1 - \cdots - \left(\frac{\mathbf{x}_{i+1} \cdot \mathbf{v}_i}{\mathbf{v}_i \cdot \mathbf{v}_i}\right)\mathbf{v}_i = \mathbf{x}_{i+1} - \mathrm{proj}_{W_i}(\mathbf{x}_{i+1}) = \mathrm{perp}_{W_i}(\mathbf{x}_{i+1})$$

which is orthogonal to W_i. (Orthogonal Decomposition Theorem 증명 과정에서 보였다.)
Also, $\mathbf{v}_{i+1} \neq \mathbf{0}$, because if $\mathbf{v}_{i+1} = \mathbf{0}$, $\mathbf{x}_{i+1} = \mathrm{proj}_{W_i}(\mathbf{x}_{i+1}) \in W_i = \mathrm{span}(\mathbf{x}_1, \cdots, \mathbf{x}_i)$, which is impossible since $\{\mathbf{x}_1, \cdots, \mathbf{x}_{i+1}\}$ is a linearly independent set. ($\{\mathbf{x}_1, \cdots, \mathbf{x}_k\}$ 가 W의 기저이므로 선형독립이다.)

Since $\{\mathbf{v}_1, \cdots, \mathbf{v}_{i+1}\}$ is an orthogonal set of nonzero vectors, by Theorem 5.1, $\{\mathbf{v}_1, \cdots, \mathbf{v}_{i+1}\}$ is linearly independent set. Note that $\mathbf{v}_1, \cdots, \mathbf{v}_i \in W_i \subset W_{i+1}$, and $\mathbf{v}_{i+1} \in W_{i+1}$. Since $\dim W_{i+1} = i+1$, $\{\mathbf{v}_1, \cdots, \mathbf{v}_{i+1}\}$ is an orthogonal, linearly independent set of $i+1$ vectors, so it forms an orthogonal basis for W_{i+1}.

(i) and (ii) gives the proof for the Gram-Schmidt Process.

Orthogonal projection은 orthogonal basis가 존재할 때 정의할 수 있다는 점에서 (ii)의 귀납 가정이 필요하다는 사실을 숙지하자. □

Theorem 5.16 : The QR Factorization
Let A be an $m \times n$ matrix with linearly independent columns. Then there exists an $m \times n$ matrix Q with orthogonal columns, and an invertible triangular matrix R.

Proof. Let $A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}$, and let $\{\mathbf{q}_1, \cdots, \mathbf{q}_n\}$ be the orthonormal basis of $\mathrm{col}(A)$ obtained by Gram-Schmidt Process and normalization. By Theorem 5.15, $\{\mathbf{q}_1, \cdots, \mathbf{q}_i\}$ forms an orthonormal basis for $\text{span}(\mathbf{a}_1, \cdots, \mathbf{a}_i)$ for $i = 1, \cdots, n$. Thus, there exist scalars $r_{1i}, r_{2i}, \cdots, r_{ii}$ such that

$$\mathbf{a}_i = r_{1i}\mathbf{q}_1 + r_{2i}\mathbf{q}_2 + \cdots + r_{ii}\mathbf{q}_i$$

where $r_{ii} \neq 0$. (If $r_{ii} = 0$, $\mathbf{a}_i \in \text{span}(\mathbf{a}_1, \cdots, \mathbf{a}_{i-1})$ so it is a contradiction.) Then

$$\begin{aligned} A &= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix} \end{aligned}$$

Therefore $Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}$ and $R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix}$ satisfies the condition. □

5.4 Orthogonal Diagonalization of Symmetric Matrices

Definition. A square matrix A is **orthogonally diagonalizable** if there exists an orthogonal matrix Q and a diagonal matrix D such that $Q^T AQ = D$.

Theorem 5.17
If A is orthogonally diagonalizable, then A is symmetric.

Proof. If A is orthogonally diagonalizable, then there exists an orthogonal matrix Q and a diagonal matrix D such that $Q^T AQ = D$. Since $Q^{-1} = Q^T$,

$$A = (Q^T)^{-1} DQ^{-1} = QDQ^T$$

thus

$$A^T = (QDQ^T)^T = (Q^T)^T D^T Q^T = QDQ^T = A$$

Therefore A is symmetric. □

Theorem 5.18
If A is a real symmetric matrix, then the eigenvalues of A are real.

Proof. Suppose that λ is an eigenvalue of A with corresponding eigenvector $\mathbf{v}$. Since A is a real matrix, $A = \overline{A}$, so

$$A\overline{\mathbf{v}} = \overline{A}\overline{\mathbf{v}} = \overline{A\mathbf{v}} = \overline{\lambda\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$$

Thus,

$$\begin{aligned} \lambda(\overline{\mathbf{v}}^T\mathbf{v}) &= \overline{\mathbf{v}}^T(\lambda\mathbf{v}) = \overline{\mathbf{v}}^T(A\mathbf{v}) \\ &= (\overline{\mathbf{v}}^T A)\mathbf{v} = (\overline{\mathbf{v}}^T A^T)\mathbf{v} = (A\overline{\mathbf{v}})^T\mathbf{v} = (\overline{\lambda}\overline{\mathbf{v}})^T\mathbf{v} = \overline{\lambda}(\overline{\mathbf{v}}^T\mathbf{v}) \end{aligned}$$

Since $\mathbf{v} \neq \mathbf{0}$, $\overline{\mathbf{v}}^T\mathbf{v} = \mathbf{v} \cdot \mathbf{v} \neq 0$. (5.4 끝의 Remarks 참조) Therefore, $\lambda = \overline{\lambda}$, which indicates that λ is real. □

Theorem 5.19
If A is a symmetric 'real' matrix, then any two eigenvectors corresponding to distinct eigenvalues of A are orthogonal.

Proof. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors corresponding to the distinct eigenvalues λ_1, λ_2 of A. Then,

$$\begin{aligned}\lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_2) &= (\lambda_1\mathbf{v}_1) \cdot \mathbf{v}_2 = (A\mathbf{v}_1) \cdot \mathbf{v}_2 = (A\mathbf{v}_1)^T\mathbf{v}_2 \\ &= \mathbf{v}_1^T A^T \mathbf{v}_2 = \mathbf{v}_1^T(A\mathbf{v}_2) = \mathbf{v}_1^T(\lambda_2\mathbf{v}_2) = \lambda_2(\mathbf{v}_1 \cdot \mathbf{v}_2)\end{aligned}$$

Thus $(\lambda_1 - \lambda_2)(\mathbf{v}_1 \cdot \mathbf{v}_2) = 0$. Since $\lambda_1 - \lambda_2 \neq 0$, $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. Therefore, $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal. $\square$

Theorem 5.20 : The Spectral Theorem
Let A be an $n \times n$ real matrix. Then A is symmetric if and only if it is orthogonally diagonalizable.

Proof. ($\Leftarrow$) Theorem 5.17.

($\Rightarrow$) We prove the proposition through inductive process.
(i) If A is a 1×1 real matrix, then A is diagonal form, so A is diagonalizable since $A = IAI$.

(ii) Suppose that every $n \times n$ matrices are orthogonally diagonalizable. (귀납 가정)

Let A be an $(n+1)\times(n+1)$ real matrix, and let λ_1 be an eigenvalue of A. By Theorem 5.18, λ_1 is real, so there exists an real eigenvalue $\mathbf{v}_1$ corresponding to λ_1. ($\mathbf{v}_1$이 항상 real vector가 아니라, λ_1에 대응되는 eigenvector들 중 real vector가 존재한다는 서술이 정확하다. 예를 들어, I_2의 eigenvalue는 $\lambda = 1$이지만 eigenvector는 $\mathbf{v} = \begin{bmatrix} i \\ i \end{bmatrix}$처럼 real vector가 아닐 수도 있기 때문. real eigenvector가 존재한다는 것에 대한 증명은: "$\mathbf{v}_1$은 $(A - \lambda_1 I)\mathbf{x} = \mathbf{0}$의 해이고, $A - \lambda_1 I$가 real matrix이므로 실수해가 존재한다.")

We may assume that $\mathbf{v}$ is a unit vector, as if not, we can normalize it. Using the Gram-Schmidt Process, we can construct an orthonormal basis $\{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ of $\mathbb{R}^n$. Let $Q_1 = \begin{bmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{bmatrix}$ then Q_1 is an orthogonal matrix, and let

$$\begin{aligned} B = Q_1^T A Q_1 &= \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} A \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \begin{bmatrix} A\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n \end{bmatrix} = \\ &= \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \begin{bmatrix} \lambda_1\mathbf{v}_1 & A\mathbf{v}_2 & \cdots & A\mathbf{v}_n \end{bmatrix} = \begin{bmatrix} \lambda_1 & ? \\ \mathbf{0} & A_1 \end{bmatrix} \end{aligned}$$

since $\mathbf{v}_1^T(\lambda_1\mathbf{v}_1) = \lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_1) = \lambda_1$ and $\mathbf{v}_i^T(\lambda_1\mathbf{v}_1) = \lambda_1(\mathbf{v}_i \cdot \mathbf{v}_1) = 0$ if $i \neq 1$. Since

$$B^T = (Q_1^T A Q_1)^T = Q_1^T A^T (Q_1^T)^T = Q_1^T A Q_1 = B$$

B is symmetric real matrix as it is a product of real matrices, so $B = \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & A_1 \end{bmatrix}$ and A_1 is also a $n \times n$ symmetric real matrix. By inductive hypothesis, A_1 is orthogonally diagonalizable, so there exists an $n \times n$ orthogonal matrix P and $n \times n$ diagonal matrix D such that $P^T A_1 P = D$.

Let $Q_2 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P \end{bmatrix}$, then Q_2 is orthogonal since columns of P are orthonormal, and also the first column is orthogonal to every other columns. Then,

$$\begin{aligned}(Q_1Q_2)^T A(Q_1Q_2) &= Q_2^T Q_1^T A Q_1 Q_2 = Q_2^T B Q_2 \\ &= \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P^T \end{bmatrix}\begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & A_1 \end{bmatrix}\begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & P^T A_1 P \end{bmatrix} = \begin{bmatrix} \lambda_1 & \mathbf{0} \\ \mathbf{0} & D \end{bmatrix}\end{aligned}$$

where Q_1Q_2 is an orthogonal matrix. Therefore, A is diagonalizable.

(i) and (ii) gives the inductive proof that every symmetric real matrices are orthogonally diagonalizable. □

Remarks. 아까 Theorem 5.18부터 갑자기 왜 real matrix인지, real eigenvalue인지 따지는 것에 대해 궁금했을 것 같다. 우리가 orthogonality, orthonormality 등등을 모두 $\mathbb{R}^n$에서 정의했기 때문에 생기는 일이다.
사실 두 vector $\mathbf{u}, \mathbf{v}$가 $\mathbb{R}^n$의 vector가 아니라 $\mathbb{C}^n$의 vector들이라면 dot product의 정의부터 다르다. $\mathbb{C}^n$에서 두 vector $\mathbf{u}, \mathbf{v}$의 dot product는 다음과 같이 정의된다.

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \cdots \\ u_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \cdots \\ v_n \end{bmatrix}$$

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^{n} u_i \overline{v_i} = \overline{\mathbf{v}}^T \mathbf{u}$$

여기에서는 dot product의 '교환법칙'조차 성립하지 않는다. $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v} \cdot \mathbf{u}}$ 이다. 그래서 $\mathbf{u} \cdot \mathbf{v} = 0$으로 정의되는 orthogonality의 성질도 달라지니까, 결론적으로 처음부터 다시 해야 한다는 말이다. 원한다면 해보고 Spectral Theorem을 복소수 범위까지 확장시켜 보자.
확장시켜 보면, $A = \overline{A}^T$일때 Spectral Theorem이 성립한다 카더라...

참고 문헌 : en.wikipedia.org/wiki/Spectral_theorem

Chapter 6

Vector Spaces

6.1 Vector Spaces and Subspaces

Definition. Let V be a set on which two operations, called addition and scalar multiplication, have been defined. If $\mathbf{u}, \mathbf{v} \in V$, the sum of $\mathbf{u}$ and $\mathbf{v}$ is denoted as $\mathbf{u} \oplus \mathbf{v}$, and if c is a scalar, the scalar multiple of $\mathbf{u}$ is denoted by $c \odot \mathbf{u}$. If the following axioms hold for all $\mathbf{u}, \mathbf{v}$, and $\mathbf{w}$ in V and all scalars c and d, then V is a vector space and its elements are vectors.

a. $\mathbf{u} \oplus \mathbf{v} \in V$

b. $\mathbf{u} \oplus \mathbf{v} = \mathbf{v} \oplus \mathbf{u}$

c. $(\mathbf{u} \oplus \mathbf{v}) \oplus \mathbf{w} = \mathbf{u} \oplus (\mathbf{v} \oplus \mathbf{w})$

d. $\exists \mathbf{0} \in V$ such that $\forall \mathbf{u} \in V, \mathbf{u} \oplus \mathbf{0} = \mathbf{u}$ ($\mathbf{0}$ is the zero vector in V)

e. $\exists(-\mathbf{u}) \in V$ such that $\mathbf{u} \oplus (-\mathbf{u}) = \mathbf{0}$

f. $c \odot \mathbf{u} \in V$

g. $c \odot (\mathbf{u} \oplus \mathbf{v}) = (c \odot \mathbf{u}) \oplus (c \odot \mathbf{v})$

h. $(c + d) \odot \mathbf{u} = (c \odot \mathbf{u}) \oplus (d \odot \mathbf{u})$

i. $c \odot (d \odot \mathbf{u}) = (cd) \odot \mathbf{u}$

j. $1 \odot \mathbf{u} = \mathbf{u}$

정의에서 'scalar'는 아무 말도 없으면 실수를 뜻한다. 따라서 아무 말도 없으면 V는 real vector space 이다. 하지만 complex vector space나 vector space over $\mathbb{Z}_p$ 등도 등장할 수 있으니 유의 바람.

다음은 책에서 등장하는 따로 정의 없이 사용 가능한 여러 가지 subspace들이다.

a. For any $n \geq 1$, $\mathbb{R}^n$ is a vector space with the usual vector addition and scalar multiplication.

b. For any $m, n \geq 1$, M_{mn}, a set of all $m \times n$ matrices, is a vector space with the usual operations of matrix addition and scalar multiplication.

c. For any $n \geq 0$, $\mathscr{P}_n$, a set of all polynomials of degree *less than or equal to* n, is a vector space with the usual operations of addition and scalar multiplication. $\mathscr{P}$, a set of all polynomials, is a vector space.

d. $\mathscr{F}$, a set of all real valued functions, is a vector space with the addition and scalar multiplication defined as follows:

$$\text{For } f, g \in \mathscr{F} \text{ and scalar} c, (f+g)(x) = f(x) + g(x) \text{ and } (cf)(x) = cf(x)$$

e. For any $n \geq 1$, $\mathbb{C}^n$ is a *complex* vector space with the usual vector addition and scalar multiplication. (Note that the scalars are complex numbers.)

f. If p is prime, for any $n \geq 1$, $\mathbb{Z}_p^n$ is a vector space with the usual vector addition and scalar multiplication. (Note that the scalars are from $\mathbb{Z}_p$.)

g. $\mathscr{C}$, the set of all continuous real valued functions, and $\mathscr{D}$, the set of all differentiable real valued functions, are vector spaces with the addition and scalar multiplication defined as same as $\mathscr{F}$. Therefore, $\mathscr{C}$ and $\mathscr{D}$ are subspaces of $\mathscr{F}$. Note that $\mathscr{P} \subset \mathscr{D} \subset \mathscr{C} \subset \mathscr{F}$.
Similarly, $\mathscr{C}_{[a,b]}$, the set of all real valued functions which are continuous on $[a, b]$, and $\mathscr{D}_{[a,b]}$, the set of all read valued functions which are differentiable on $[a, b]$, are also vector spaces.

Theorem 6.1
Let V be a vector space, $\mathbf{u}$ a vector in V, and c a scalar.

a. $0 \odot \mathbf{u} = \mathbf{0}$

b. $c \odot \mathbf{0} = \mathbf{0}$

c. $(-1) \odot \mathbf{u} = -\mathbf{u}$

d. $c \odot \mathbf{u} = 0 \Rightarrow c = 0$ or $\mathbf{u} = \mathbf{0}$

Proof. Let V be a vector space, $\mathbf{u}$ a vector in V, and c a scalar.

a. **(Exercise 6.1 22)**

$$\begin{aligned}
\mathbf{u} \oplus (0 \odot \mathbf{u}) &= (1 \odot \mathbf{u}) \oplus (0 \odot \mathbf{u}) \text{ (Axiom 10)} \\
&= (1+0) \odot \mathbf{u} \text{ (Axiom 8)} \\
&= 1 \odot \mathbf{u} = \mathbf{u} \text{ (Axiom 10)}
\end{aligned}$$

Therefore, $0 \odot \mathbf{u} = \mathbf{0}$. (Axiom 4)

b.

$$\begin{aligned}
\mathbf{u} \oplus (c \odot \mathbf{0}) &= \mathbf{u} \oplus (c \odot (0 \odot \mathbf{u})) \text{ (Theorem 6.1(a))} \\
&= (1 \odot \mathbf{u}) \oplus ((c \times 0) \odot \mathbf{u}) \text{ (Axiom 9, 10)} \\
&= (1 + c \times 0) \odot \mathbf{u} \text{ (Axiom 8)} \\
&= 1 \odot \mathbf{u} = \mathbf{u} \text{ (Axiom 10)}
\end{aligned}$$

Therefore, $c \odot \mathbf{0} = \mathbf{0}$. (Axiom 4)

c. **(Exercise 6.1 23)**

$$\begin{aligned}\mathbf{u} \oplus ((-1) \odot \mathbf{u}) &= (1 \odot \mathbf{u}) \oplus ((-1) \odot \mathbf{u}) \text{ (Axiom 10)} \\ &= (1 + (-1)) \odot \mathbf{u} \text{ (Axiom 8)} \\ &= 0 \odot \mathbf{u} = \mathbf{0} \text{ (Theorem 6.1(a))}\end{aligned}$$

Therefore, $(-1) \odot \mathbf{u} = -\mathbf{u}$ (Axiom 5)

d. Suppose that $c \odot \mathbf{u} = \mathbf{0}$. If $c \neq 0$,

$$\begin{aligned}\mathbf{u} &= 1 \odot \mathbf{u} \text{ (Axiom 10)} \\ &= (\frac{1}{c}c) \odot \mathbf{u} = \frac{1}{c} \odot (c \odot \mathbf{u}) \text{ (Axiom 9)} \\ &= \frac{1}{c} \odot \mathbf{0} = \mathbf{0} \text{ (Theorem 6.1(b))}\end{aligned}$$

Therefore, $c = 0$ or $\mathbf{u} = \mathbf{0}$.

□

Definition. A subset W of a vector subspace V is a subspace of V if W is itself a vector space with the same scalars, addition, and scalar multiplication as V.

Note. W가 $\mathbb{R}^n$의 부분집합으로 주어질 때, W의 연산들이 일반적인 vector addition, scalar multiplication과 동일하게 주어졌다면, Axiom 1, 4, 5, 6만 확인하면 된다. 연산들이 새로 정의되었다면, 나머지 Axiom들도 확인하자.

Theorem 6.2
Let V be a vector space and let W be a nonempty subset of V. Then W is a subspace of V if and only if the following propositions hold;

a. $\mathbf{u}, \mathbf{v} \in W \Rightarrow \mathbf{u} \oplus \mathbf{v} \in W$

b. $\mathbf{u} \in W$ and c is a scalar $\Rightarrow c \odot \mathbf{u} \in W$

Proof. ($\Rightarrow$) Proposition (a) is equal to Axiom 1, and proposition (b) is equal to Axiom 6.

($\Leftarrow$) Proposition (a) is equal to Axiom 1, and proposition (b) is equal to Axiom 6. Also, axioms 2, 3, 7, 8, 9, 10 hold since W is a subset of V and the axioms hold for all vectors in V. Since W is a nonempty set, it contains at least one vector $\mathbf{u}$. Proposition (b) gives $0 \odot \mathbf{u} = \mathbf{0} \in W$, so Axiom 4 holds. Also, for any vector $\mathbf{u} \in W$, proposition (b) gives $(-1) \odot \mathbf{u} = -\mathbf{u} \in W$, so Axiom 5 holds. Therefore, W is a subspace of $\mathbb{R}^n$. □

If V is a vector space, then V and $\{\mathbf{0}\}$ are subspaces of V, which are called the **trivial subspaces**.

Definition. If $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is a set of vectors in a vector space V, then the set of all linear combination of $\mathbf{v}_1, \cdots, \mathbf{v}_k$ is called the **span** of $\mathbf{v}_1, \cdots, \mathbf{v}_k$ and denoted by $\text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_k)$ or $\text{span}(S)$. If $V = \text{span}(S)$, then S is called a **spanning set** of V and V is **spanned** by S.

Note. $V = \text{span}(S)$여야 S가 V의 spanning set이고, $V \subset \text{span}(S)$이면 S가 V의 spanning set이 아니라는 주장이 있는데, 사실 S가 V의 부분집합이어야 span과 spanning set이 정의되니까 동치인 표현이 맞다.

여기부터는 vector addition과 scalar multiplication 기호로 쓰던 $\oplus$와 $\odot$ 대신 기존에 쓰던 방식으로 쓰기로 한다.

Extra Theorem
If $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ spans V and $S \subset V$, then S is a spanning set for V.

Proof. (i) Since S is a spans V, $V \subset \text{span}(S)$.

(ii) Let $\mathbf{v}$ be a vector in span(S). Since $\mathbf{v}$ is a linear combination of $\mathbf{v}_1, \cdots, \mathbf{v}_k$ and $S \subset V$, $\mathbf{v} \in V$ as V is a closed under addition and scalar multiplication.

By (i) and (ii), $V = \text{span}(S)$ so S is a spanning set for V. □

Theorem 6.3
Let $\mathbf{v}_1, \cdots, \mathbf{v}_k$ be vectors in a vector space V.

a. span$(\mathbf{v}_1, \cdots, \mathbf{v}_k)$ is a subspace of V.

b. span$(\mathbf{v}_1, \cdots, \mathbf{v}_k)$ is a smallest subspace of V that contains $\mathbf{v}_1, \cdots, \mathbf{v}_k$.

Proof. a. (i) Let $\mathbf{u}, \mathbf{v}$ be vectors in span$(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$. Then there exist scalars $c_1, \cdots, c_k$ and $d_1, \cdots, d_k$ such that $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$, and $\mathbf{v} = d_1\mathbf{v}_1 + \cdots + d_k\mathbf{v}_k$. Then

$$\mathbf{u} + \mathbf{v} = (c_1 + d_1)\mathbf{v}_1 + \cdots + (c_k + d_k)\mathbf{v}_k \in \text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_k)$$

(ii) Let $\mathbf{u}$ be a vector in span$(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$ and c be a scalar. Then there exist scalars $c_1, \cdots, c_k$ such that $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$. Then

$$c\mathbf{u} = cc_1\mathbf{v}_1 + \cdots + cc_k\mathbf{v}_k \in \text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_k)$$

By Theorem 6.2, (i) and (ii) gives that span$(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$ is a subspace of V.

b. Let W be a subspace of V which contains $\mathbf{v}_1, \cdots, \mathbf{v}_k$. Let $\mathbf{x}$ be a vector in span$(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$, then $\mathbf{x}$ is in W since W is closed under addtion and scalar multiplication. Thus, span$(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$ $\subset W$.

Since span$(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$ is a subspace of V, span$(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_k)$ is the smallest subspace of V that contains $\mathbf{v}_1, \cdots, \mathbf{v}_k$.

□

6.2 Linear Independence, Basis and Dimension

Definition. A *finite* set of vectors $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ in a vector space V is **linearly dependent** if there exist scalars $c_1, \cdots, c_k$, at least one of which is nonzero, such that

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

An *infinite* set of vectors is linearly dependent if it contains a linearly dependent subset. A set of vectors that is not linearly dependent is said to be **linearly independent.**

Note. 무한집합이 일차종속임을 보일 때에는 무한집합의 부분집합들 중 일차종속인 유한부분집합을 찾는다. 반대로 무한집합이 일차독립임을 보일 때에는 일차종속인 유한부분집합이 존재한다고 가정한 후, 모순을 이끌어내어 귀류법으로 증명한다.

(Example 6.28) In $\mathscr{P}$, show that $S = \{1, x, x^2, \cdots\}$ is linearly independent.

Suppose that S contains a finite linearly dependent subset T. Let x^m be the highest power of x in T, and let x^n be the lowest power of x in T. Since T is linearly dependent, $\{x^n, x^{n+1}, \cdots, x^m\}$ is also linearly dependent so there exist scalars $c_n, \cdots, c_m$, at least one of which is nonzero, such that

$$c_n x^n + c_{n+1}x^{n+1} + \cdots + c_m x^m = 0$$

However, this implies that $c_n = \cdots = c_m = 0$, since the equation is an identity with respect to x. Therefore, such T cannot exist, hence S is linearly independent.

Theorem 6.4
A set of vectors $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ in a vector space V is linearly dependent if and only if at least one of the vectors can be expressed as a linear combination of the other vectors.

Proof. ($\Rightarrow$) If vectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly dependent, there exist scalars $c_1, c_2, \cdots, c_n$, at least one of which is nonzero, such that $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$. Without loss of generality, we may assume $c_1 \neq 0$, then

$$\mathbf{v}_1 = -\frac{c_2}{c_1}\mathbf{v}_2 - \cdots - \frac{c_n}{c_1}\mathbf{v}_n$$

so $\mathbf{v}_1$ is represented with a linear combination of other vectors.

($\Leftarrow$) Without loss of generality, suppose that $\mathbf{v}_1$ can be expressed as linear combination of $\mathbf{v}_2, \mathbf{v}_3, \cdots, \mathbf{v}_n$. Then there exist scalars $c_2, c_3, \cdots, c_n$ such that $\mathbf{v}_1 = c_2\mathbf{v}_2 + c_3\mathbf{v}_3 + \cdots + c_n\mathbf{v}_n$. Then

$$\mathbf{v}_1 - c_2\mathbf{v}_2 - \cdots - c_n\mathbf{v}_n = \mathbf{0}$$

Therefore, $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_n$ are linearly dependent. □

Definition. A subset $\mathcal{B}$ of a vector space V is a **basis** for V if

a. $\mathcal{B}$ spans V. (= $\mathcal{B}$ is a spanning set for V.)

b. $\mathcal{B}$ is linearly independent.

다음은 책에서 정의된 standard basis들이다.

a. $\{\mathbf{e}_1, \cdots, \mathbf{e}_n\}$ is a standard basis for $\mathbb{R}^n$.

b. The set $\mathcal{E} = \{E_{11}, \cdots, E_{1n}, \cdots, E_{m1}, \cdots, E_{mn}\}$ is a standard basis for M_{mn}, where E_{ab} is defined as

$$[E_{ab}]_{ij} = \begin{cases} 1 \text{ , if } i = a \text{ and } j = b \\ 0 \text{ , otherwise} \end{cases}$$

c. The set $\{1, x, x^2, x^3, \cdots, x^n\}$ is a standard basis for $\mathscr{P}_n$.

Note. Coordinate vector는 원소들의 순서가 바뀌면 달라지기 때문에 기저는 그냥 set이 아니라 ordered set이다.

Theorem 6.5
Let V be a vector space. Then $\mathcal{B}$ is a basis for V if and only if $\mathcal{B}$ is a subset of V, and there is a unique way to represent $\mathbf{v}$ as a linear combination of the vectors in $\mathcal{B}$ for every vector $\mathbf{v}$ in V.

Proof. ($\Rightarrow$) *Existence.* Since $\mathcal{B}$ is a basis for V, $\mathcal{B}$ spans V so every vector $\mathbf{v}$ in V can be expressed as a linear combination of the vectors in $\mathcal{B}$.

Uniqueness. Suppose that there are more than two ways to represent $\mathbf{v} \in V$ as a linear combination of $\mathcal{B}$. Since $\mathbf{v}$ is a finite linear combination of the basis vectors, let $\mathbf{v}_1, \cdots .\mathbf{v}_k \in \mathcal{B}$ such that there are more than two ways to represent $\mathbf{v}$ as a linear combination of $\mathbf{v}_1, \cdots, \mathbf{v}_k$. ($V$가 무한 차원일때를 고려) Then there exist scalars $c_1, \cdots, c_k$ and $d_1, \cdots, d_k$, at least one of $c_i - d_i \neq 0$, such that

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = d_1\mathbf{v}_1 + \cdots + d_k\mathbf{v}_k$$

Then

$$(c_1 - d_1)\mathbf{v}_1 + \cdots + (c_k - d_k)\mathbf{v}_k = \mathbf{0}$$

so the vectors $\mathbf{v}_1, \cdots, \mathbf{v}_k$ are linearly dependent. It is a contradiction since $\mathcal{B}$ is a basis. Therefore, there is only one way to represent $\mathbf{v}$ as a linear combination of the basis vectors.

($\Leftarrow$) **(Exercise 6.2 30)** (i) Let $\mathbf{v}$ be a vector in V, then $\mathbf{v}$ can be expressed as a linear combination of vectors in $\mathcal{B}$, so $\mathcal{B}$ spans V.

(ii) Suppose that there exists a finite linearly dependent subset $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ of $\mathcal{B}$. Then there exist scalars $c_1, \cdots, c_k$, at least one of which is nonzero, such that

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0} = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_k$$

This implies that $\mathbf{0} \in V$ has two different representation of linear combination, is a contradiction. Therefore, $\mathcal{B}$ is linearly independent.

(i) and (ii) gives the proof that $\mathcal{B}$ is a basis for V. $\square$

Note. 한 쪽 방향의 증명에서 동치인 명제로 확장되었다. 그리고 $\mathcal{B}$가 V의 부분집합이라는 조건이 있어야 정확하다. 또 3장의 정리와 다른 점은, $\mathcal{B}$가 무한집합일때도 성립한다는 점이다.

Definition. Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a basis for a vector space V. Let $\mathbf{v}$ be a vector in V, and let scalars $c_1, \cdots, c_n$ such that $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Then $c_1, \cdots, c_n$ are called the **coordinates of v with respect to** $\mathcal{B}$, and the vector

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

is called the **coordinate vector of v with respect to** $\mathcal{B}$.

Theorem 6.6
Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a basis for a vector space V. Let $\mathbf{u}, \mathbf{v}$ be vectors in V, and let c be a scalar.

a. $[\mathbf{u}+\mathbf{v}]_{\mathcal{B}} = [\mathbf{u}]_{\mathcal{B}} + [\mathbf{v}]_{\mathcal{B}}$

b. $[c\mathbf{u}]_{\mathcal{B}} = c[\mathbf{u}]_{\mathcal{B}}$

Proof. Let $\mathbf{u}, \mathbf{v}$ be vectors in V, then there exist scalars $c_1, \cdots, c_n$ and $d_1, \cdots, d_n$ such that

$$\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \text{ and } \mathbf{v} = d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n$$

a.
$$\mathbf{u}+\mathbf{v} = (c_1+d_1)\mathbf{v}_1 + \cdots + (c_n+d_n)\mathbf{v}_n$$

So
$$[\mathbf{u}+\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1+d_1 \\ \vdots \\ c_n+d_n \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} + \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix} = [\mathbf{u}]_{\mathcal{B}} + [\mathbf{v}]_{\mathcal{B}}$$

b. For any scalar c,
$$c\mathbf{u} = cc_1\mathbf{v}_1 + \cdots + cc_k\mathbf{v}_n$$

So
$$[c\mathbf{u}]_{\mathcal{B}} = \begin{bmatrix} cc_1 \\ \vdots \\ cc_n \end{bmatrix} = c\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = c[\mathbf{u}]_{\mathcal{B}}$$

□

Theorem 6.7
Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a basis for a vector space V, and let $\mathbf{u}_1, \cdots, \mathbf{u}_k$ be vectors in V. Then $\mathbf{u}_1, \cdots, \mathbf{u}_k$ are linearly independent in V if and only if $[\mathbf{u}_1]_{\mathcal{B}}, \cdots, [\mathbf{u}_k]_{\mathcal{B}}$ is linearly independent in $\mathbb{R}^n$.

Proof. ($\Rightarrow$) Suppose that $\mathbf{u}_1, \cdots, \mathbf{u}_n$ are linearly independent. Consider the linear system

$$c_1[\mathbf{u}_1]_{\mathcal{B}} + \cdots + c_k[\mathbf{u}_k]_{\mathcal{B}} = \mathbf{0}$$

in $\mathbb{R}^n$. By Theorem 6.6,

$$c_1[\mathbf{u}_1]_{\mathcal{B}} + \cdots + c_k[\mathbf{u}_k]_{\mathcal{B}} = [c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k]_{\mathcal{B}} = \mathbf{0}$$

Thus
$$c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k = 0\mathbf{v}_1 + \cdots + 0\mathbf{v}_k = \mathbf{0}$$

(위 식의 $\mathbf{0}$은 $\mathbb{R}^n$의 zero vector이고, 아래 식의 $\mathbf{0}$은 V의 zero vector임에 유의하자) Since $\mathbf{u}_1, \cdots, \mathbf{u}_k$ are linearly independent, $c_1 = \cdots = c_k = 0$. Therefore, $c_1[\mathbf{u}_1]_{\mathcal{B}} + \cdots + c_k[\mathbf{u}_k]_{\mathcal{B}} = \mathbf{0}$ has only the trivial solution, hence $[\mathbf{u}_1]_{\mathcal{B}}, \cdots, [\mathbf{u}_k]_{\mathcal{B}}$ are also linearly independent.

($\Leftarrow$) **(Exercise 6.2 32)** Conversely, suppose that $[\mathbf{u}_1]_{\mathcal{B}}, \cdots, [\mathbf{u}_k]_{\mathcal{B}}$ are linearly independent. Consider the linear system
$$c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k = \mathbf{0}$$

in V. By Theorem 6.6,

$$[c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k]_{\mathcal{B}} = c_1[\mathbf{u}_1]_{\mathcal{B}} + \cdots + c_k[\mathbf{u}_k]_{\mathcal{B}} = \mathbf{0}$$

Since $[\mathbf{u}_1]_{\mathcal{B}}, \cdots, [\mathbf{u}_k]_{\mathcal{B}}$ are linearly independent, $c_1 = \cdots = c_k = 0$. Therefore, $c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k = \mathbf{0}$ only has the trivial solution, hence $\mathbf{u}_1, \cdots, \mathbf{u}_k$ are linearly independent. $\square$

Extra Theorem
Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a basis for a vector space V and let $\mathbf{u}_1, \cdots, \mathbf{u}_k$ be vectors in V. Then $\text{span}(\mathbf{u}_1, \cdots, \mathbf{u}_k) = V$ if and only if $\text{span}([\mathbf{u}_1]_{\mathcal{B}}, \cdots, [\mathbf{u}_k]_{\mathcal{B}}) = \mathbb{R}^n$.

Proof. **(Exercise 6.2 33)**

($\Rightarrow$) Suppose that $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ is a spanning set for V. Let $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ be a vector in $\mathbb{R}^n$, and let $\mathbf{v} = x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n$ be a vector in V. Then $[\mathbf{v}]_{\mathcal{B}} = \mathbf{x}$. Since $\mathbf{u}_1, \cdots, \mathbf{u}_n$ spans V, there exist scalars $c_1, \cdots, c_k$ such that

$$\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k$$

By Theorem 6.6,

$$\mathbf{x} = [\mathbf{v}]_{\mathcal{B}} = [c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k]_{\mathcal{B}} = c_1[\mathbf{u}_1]_{\mathcal{B}} + \cdots + c_k[\mathbf{u}_k]_{\mathcal{B}}$$

which is a linear combination of $[\mathbf{u}_1]_{\mathcal{B}}, \cdots, [\mathbf{u}_k]_{\mathcal{B}}$. Therefore, $\{[\mathbf{u}_1]_{\mathcal{B}}, \cdots, [\mathbf{u}_k]_{\mathcal{B}}\}$ is a spanning set for $\mathbb{R}^n$.

($\Leftarrow$) Suppose that $\{[\mathbf{u}_1]_{\mathcal{B}}, \cdots, [\mathbf{u}_k]_{\mathcal{B}}\}$ is a spanning set for $\mathbb{R}^n$. Let $\mathbf{v}$ be a vector in V. Since $[\mathbf{v}]_{\mathcal{B}} \in \mathbb{R}^n$, there exist scalars $c_1, \cdots, c_k$ such that

$$[\mathbf{v}]_{\mathcal{B}} = c_1[\mathbf{u}_1]_{\mathcal{B}} + \cdots + c_k[\mathbf{u}_k]_{\mathcal{B}} = [c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k]_{\mathcal{B}}$$

This implies that $\mathbf{v} = c_1\mathbf{u}_1 + \cdots + c_k\mathbf{u}_k$, thus $\mathbf{u}_1, \cdots, \mathbf{u}_k$ span V, hence $\{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ is a spanning set for V. $\square$

Theorem 6.8
Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a basis for a vector space V.

a. Any set of more than n vectors in V is linearly dependent in V.

b. Any set of fewer than n vectors in V cannot span V.

Proof. Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a basis for a vector space V.

a. Let $\mathbf{u}_1, \cdots, \mathbf{u}_m$ be vectors in V, where $m > n$. Since $[\mathbf{u}_1]_{\mathcal{B}}, \cdots, [\mathbf{u}_m]_{\mathcal{B}}$ are m vectors in $\mathbb{R}^n$, they are linearly dependent in $\mathbb{R}^n$. (Remind Theorem 2.8.) Therefore, by Theorem 6.7, $\mathbf{u}_1, \cdots, \mathbf{u}_m$ are also linearly dependent in V.

b. Let $\mathbf{u}_1, \cdots, \mathbf{u}_m$ be vectors in V, where $m < n$. Since $[\mathbf{u}_1]_{\mathcal{B}}, \cdots, [\mathbf{u}_m]_{\mathcal{B}}$ are m vectors in $\mathbb{R}^n$, they cannot span $\mathbb{R}^n$. Therefore, by the Extra Theorem, $\mathbf{u}_1, \cdots, \mathbf{u}_n$ also cannot span V.

$\square$

Theorem 6.9 : The Basis Theorem
If a vector space V has a basis with n vectors, then every basis for V contains exactly n vectors.

Proof. Let $\mathcal{B}$ be a basis for V with n vectors, and let $\mathcal{B}'$ be another basis for V with m vectors. Since $\mathcal{B}'$ is linearly independent in V and is a spanning set of V, $m \geq n$ by Theorem 6.8(a), and $m \leq n$ by Theorem 6.8(b). Therefore, $m = n$. □

Definition. A vector space V is **finite-dimensional** if it has a basis consisting of finitely many vectors. For such vector spaces, the **dimension** of V, denoted as $\dim V$, is the number of vectors in a basis for V. The dimension of the zero vector space is defined to be zero. A vector space that has no finite basis is called **infinite-dimensional**.

Theorem 6.10
Let V a vector space with $\dim V = n$.

a. Any linearly independent set in V contains at most n vectors.

b. Any spanning set of V contains at least n vectors.

c. Any linearly independent set of exactly n vectors in V is a basis for V.

d. Any spanning set for V consisting of exactly n vectors is a basis for V.

e. **(Plus Theorem)** Any linearly independent set in V can be extended to a basis for V.

f. **(Minus Theorem)** Any spanning set for V can be reduced to a basis for V.

Proof. Let V be a vector space with $\dim V = n$.

a. Theorem 6.8(a)

b. Theorem 6.8(b)

c. Let $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a linearly independent set of n vectors in V. If S does not span V, then there exists $\mathbf{v} \in V$ which is not in span(S).

   (Exercise 6.2 54) Suppose that $S' = S \cup \{\mathbf{v}\}$ is linearly dependent. Then there exist scalars $c_1, \cdots, c_n, c$, at least one of which is nonzero, such that

$$c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n + c\mathbf{v} = \mathbf{0}$$

   If $c = 0$, then $c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n = \mathbf{0}$ where at least one of $c_1, \cdots, c_n$ is nonzero. This implies that $\mathbf{v}_1, \cdots, \mathbf{v}_n$ is linearly dependent, which is a contradiction. If $c \neq 0$, then

$$\mathbf{v} = -\frac{c_1}{c}\mathbf{v}_1 - \cdots - \frac{c_n}{c}\mathbf{v}_n$$

   so $\mathbf{v}$ is in span(S), which is also a contradiction. Therefore, S' is still a linearly independent set in V.

   However, this is impossible since $\dim V = n$. Therefore, S spans V, hence S is a basis for V.

d. Let $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a spanning set for V. If S is linearly dependent, there exists $\mathbf{v} \in S$ which is a linear combination of the other vectors in S. (Theorem 6.4) Without loss of generality, we may assume that $\mathbf{v} = \mathbf{v}_n$ and $\mathbf{v}_n \in \text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_{n-1})$.

(Exercise 6.2 55) Since $\mathbf{v}_n \in \text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_{n-1})$, there exist scalars $d_1, \cdots, d_{n-1}$ such that

$$\mathbf{v}_n = d_1\mathbf{v}_1 + \cdots + d_{n-1}\mathbf{v}_{n-1}$$

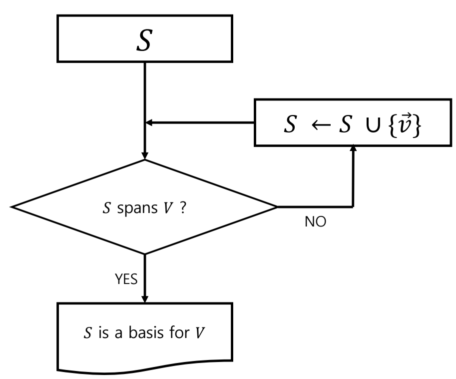
Let $\mathbf{x}$ be a vector in V. Since S is a spanning set for V, there exist scalars $c_1, \cdots, c_n$ such that

$$\begin{aligned}\mathbf{x} &= c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n \\ &= c_1\mathbf{v}_1 + \cdots + c_{n-1}\mathbf{v}_{n-1} + c_n(d_1\mathbf{v}_1 + \cdots + d_{n-1}\mathbf{v}_{n-1}) \\ &= (c_1 + c_n d_1)\mathbf{v}_1 + \cdots + (c_{n-1} + c_n d_{n-1})\mathbf{v}_{n-1}\end{aligned}$$

so $\mathbf{x} \in \text{span}(\mathbf{v}_1, \cdots, \mathbf{v}_{n-1})$. Therefore, $S' = \{\mathbf{v}_1, \cdots, \mathbf{v}_{n-1}\}$ is still a spanning set for V.

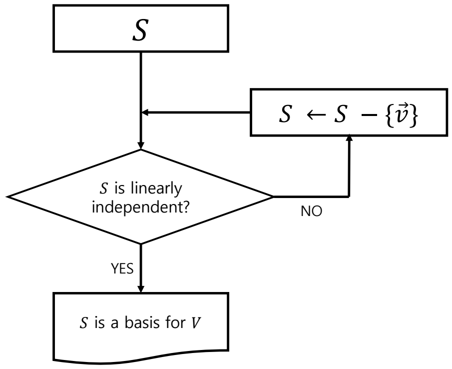
However, this is impossible since $\dim V = n$. Therefore, S is linearly independent, hence S is a basis for V.

e. Let S be a linearly independent set of vectors in V. If S does not span V, there exist $\mathbf{v} \in V$ which is not in span(S). Redefine $S = S \cup \mathbf{v}$, then S is still a linearly independent set in V. Repeat the process until S becomes a spanning set for V.



This process terminates in finite operations since a linearly independent set in V has at most n vectors. The final S forms a basis for V.

f. **(Exercise 6.2 56)** Let S be a spanning set for V. If S is not linearly independent, there exist $\mathbf{v} \in S$ which can be expressed as the linear combination of the other vectors. Redefine $S = S - \{\mathbf{v}\}$, then S is still a spanning set for V. (Exercise 6.2 55) Repeat the process until S becomes a linearly independent set in V.



This process terminates in finite operations since a spanning set for V has at least n vectors. The final S forms a basis for V.

$\square$

Theorem 6.11
Let W be a subspace of a finite-dimensional vector space V.

a. W is a finite-dimensional and $\dim W \leq \dim V$.

b. $\dim W = \dim V$ if and only if $W = V$.

Proof. If W is infinite-dimensional, it contains a infinite set which is linearly independent. However, by Theorem 6.10(a), a linearly independent set in V has at most n vectors. Thus W is finite-dimensional, so let $\mathcal{B}$ be a basis for W.

a. (i) If $W = \{\mathbf{0}\}$, then $\dim W = 0 \leq \dim V$.

   (ii) Suppose that W is a nonzero subspace. Since $\mathcal{B}$ is a linearly independent set in V, by Theorem 6.10(e), $\mathcal{B}$ can be extended to a basis for V, which has n vectors. Therefore, a basis for W has lesser or equal than n vectors, hence $\dim W \leq n = \dim V$.

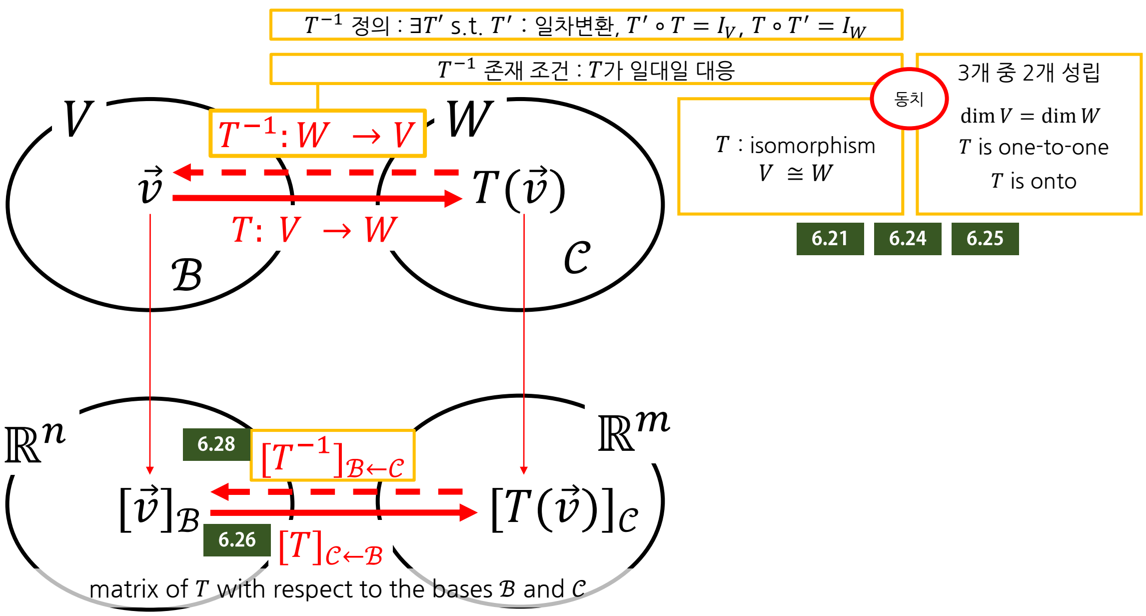
b. ($\Rightarrow$) If $\dim W = \dim V = n$, then any basis $\mathcal{B}$ consists of n vectors. Since $\mathcal{B}$ is a linearly independent set of n vectors in V, it also forms a basis for V by Theorem 6.10(c). Therefore, $V = W$.

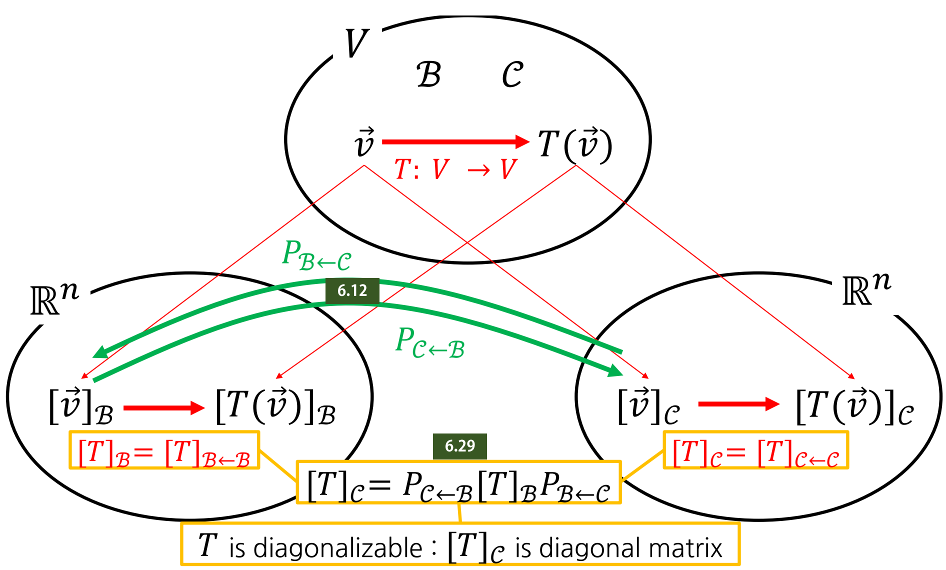
   ($\Leftarrow$) If $W = V$, clearly $\dim W = \dim V$.

$\square$

6.3 Change of Basis

6.3, 6.4, 6.5, 6.6 보기 전에 그림을 한 번 보고 가자.





Definition. Let $\mathcal{B} = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be bases for a vector space V. The $n \times n$ matrix whose columns are the coordinate vectors $[\mathbf{u}_1]_{\mathcal{C}}, \cdots, [\mathbf{u}_n]_{\mathcal{C}}$ is denoted by $P_{\mathcal{C}\leftarrow\mathcal{B}}$ and is called the **change-of-basis matrix** from $\mathcal{B}$ to $\mathcal{C}$.

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{bmatrix}[\mathbf{u}_1]_{\mathcal{C}} & \cdots & [\mathbf{u}_n]_{\mathcal{C}}\end{bmatrix}$$

Theorem 6.12
Let $\mathcal{B} = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be bases for a vector space V.

a. $P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = [\mathbf{x}]_{\mathcal{C}}$ for all $\mathbf{x} \in V$.

b. $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is the unique matrix P with the property that $P[\mathbf{x}]_{\mathcal{B}} = [\mathbf{x}]_{\mathcal{C}}$ for all $\mathbf{x} \in V$.

c. $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible and $(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1} = P_{\mathcal{B}\leftarrow\mathcal{C}}$.

Proof. Let $\mathcal{B} = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be bases for a vector space V.

a. Let $\mathbf{x}$ be a vector in V. Then there exist scalars $c_1, \cdots, c_n$ such that

$$\mathbf{x} = c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n$$

which implies that

$$[\mathbf{x}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}$$

Then

$$\begin{aligned}
[\mathbf{x}]_{\mathcal{C}} &= [c_1\mathbf{u}_1 + \cdots + c_n\mathbf{u}_n]_{\mathcal{C}} \\
&= c_1[\mathbf{u}_1]_{\mathcal{C}} + \cdots + c_n[\mathbf{u}_n]_{\mathcal{C}} \\
&= \begin{bmatrix}[\mathbf{u}_1]_{\mathcal{C}} & \cdots & [\mathbf{u}_n]_{\mathcal{C}}\end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = P_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{x}]_{\mathcal{B}}
\end{aligned}$$

b. Suppose that $P = \begin{bmatrix}\mathbf{p}_1 & \cdots & \mathbf{p}_n\end{bmatrix}$ is an $n \times n$ matrix such that $P[\mathbf{x}]_{\mathcal{B}} = [\mathbf{x}]_{\mathcal{C}}$ for all $\mathbf{x} \in V$. Since $[\mathbf{u}_i]_{\mathcal{B}} = \mathbf{e}_i$, for $i = 1, 2, \cdots, n$,

$$\mathbf{p}_i = P\mathbf{e}_i = P[\mathbf{u}_i]_{\mathcal{B}} = [\mathbf{u}_i]_{\mathcal{C}}$$

Therefore,

$$P = \begin{bmatrix}\mathbf{p}_1 & \cdots & \mathbf{p}_n\end{bmatrix} = \begin{bmatrix}[\mathbf{u}_1]_{\mathcal{C}} & \cdots & [\mathbf{u}_n]_{\mathcal{C}}\end{bmatrix} = P_{\mathcal{C}\leftarrow\mathcal{B}}$$

c. Since $\mathcal{B}$ is a linearly independent set in V, the columns of $P_{\mathcal{C}\leftarrow\mathcal{B}}$ are also linearly independent in $\mathbb{R}^n$ by Theorem 6.7, so $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible by F.T.I.M. Since $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible, for every $\mathbf{x}$ in V, $[\mathbf{x}]_{\mathcal{B}} = (P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1}[\mathbf{x}]_{\mathcal{C}}$. Therefore, by Theorem 6.12(b), $(P_{\mathcal{C}\leftarrow\mathcal{B}})^{-1} = P_{\mathcal{B}\leftarrow\mathcal{C}}$.

$\square$

Theorem 6.13
Let $\mathcal{B} = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ are basis for a vector space V. Let $B = \begin{bmatrix} [\mathbf{u}_1]_\mathcal{E} & \cdots & [\mathbf{u}_n]_\mathcal{E} \end{bmatrix}$ and $C = \begin{bmatrix} [\mathbf{v}_1]_\mathcal{E} & \cdots & [\mathbf{v}_n]_\mathcal{E} \end{bmatrix}$, where $\mathcal{E}$ is any basis for V. Then the row reduction applied to $n \times 2n$ augmented matrix $\begin{bmatrix} C \mid B \end{bmatrix}$ produces

$$\begin{bmatrix} C \mid B \end{bmatrix} \to \begin{bmatrix} I \mid P_{\mathcal{C}\leftarrow\mathcal{B}} \end{bmatrix}$$

Proof. Let scalars p_{ij} such that

$$[\mathbf{u}_i]_\mathcal{C} = \begin{bmatrix} p_{1i} \\ \vdots \\ p_{ni} \end{bmatrix} \text{ for } i = 1, 2, \cdots, n$$

Then

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{bmatrix} [\mathbf{u}_1]_\mathcal{C} & \cdots & [\mathbf{u}_n]_\mathcal{C} \end{bmatrix} = \begin{bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & & \vdots \\ p_{n1} & \cdots & p_{nn} \end{bmatrix}$$

Let $\mathcal{E}$ be a basis for V. Since $\mathbf{u}_i = p_{1i}\mathbf{v}_1 + \cdots + p_{ni}\mathbf{v}_n$, for $i = 1, 2, \cdots, n$,

$$\begin{aligned} [\mathbf{u}_i]_\mathcal{E} &= [p_{1i}\mathbf{v}_1 + \cdots + p_{ni}\mathbf{v}_n]_\mathcal{E} = p_{1i}[\mathbf{v}_1]_\mathcal{E} + \cdots + p_{ni}[\mathbf{v}_n]_\mathcal{E} \\ &= \begin{bmatrix} [\mathbf{v}_1]_\mathcal{E} & \cdots & [\mathbf{v}_n]_\mathcal{E} \end{bmatrix} \begin{bmatrix} p_{1i} \\ \vdots \\ p_{ni} \end{bmatrix} \end{aligned}$$

Thus, $\begin{bmatrix} p_{1i} \\ \vdots \\ p_{ni} \end{bmatrix}$ is the solution of the linear system with the augmented matrix given as

$$\begin{bmatrix} [\mathbf{v}_1]_\mathcal{E} & \cdots & [\mathbf{v}_n]_\mathcal{E} \mid [\mathbf{u}_i]_\mathcal{E} \end{bmatrix}$$

Since the coefficient matrix is equal for each $i = 1, 2, \cdots, n$, P can be obtained by row reducing the $n \times 2n$ augmented matrix

$$\begin{bmatrix} [\mathbf{v}_1]_\mathcal{E} & \cdots & [\mathbf{v}_n]_\mathcal{E} \mid [\mathbf{u}_1]_\mathcal{E} & \cdots & [\mathbf{u}_n]_\mathcal{E} \end{bmatrix} = \begin{bmatrix} C \mid B \end{bmatrix}$$

Since $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are linearly independent, $[\mathbf{v}_1]_\mathcal{E}, \cdots, [\mathbf{v}_n]_\mathcal{E}$ are also linearly independent by Theorem 6.7. Thus, by F.T.I.M, RREF of C is I. Therefore, row reducing the augmented matrix will result in

$$\begin{bmatrix} C \mid B \end{bmatrix} \to \begin{bmatrix} I \mid P_{\mathcal{C}\leftarrow\mathcal{B}} \end{bmatrix}$$

□

6.4 Linear Transformations + 3.6 Part I

Definition. A **linear transformation** from a vector space V to a vector space W is a mapping $T : V \to W$ such that for all $\mathbf{u}$ and $\mathbf{v}$ in V and for all scalars c,

a. $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$

b. $T(c\mathbf{u}) = cT(\mathbf{u})$

Theorem 6.14
Let $T : V \to W$ be a linear transformation.

a. $T(\mathbf{0}) = \mathbf{0}$

b. $T(-\mathbf{v}) = -T(\mathbf{v})$ for all $\mathbf{v} \in V$

c. $T(\mathbf{u}-\mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in V$

Proof. Let $T : V \to W$ be a linear transformation.

a. Let $\mathbf{v}$ be a vector in V. Then

$$T(\mathbf{0}) = T(0\mathbf{v}) = 0T(\mathbf{v}) = \mathbf{0}$$

b. **(Exercise 6.4 21)** For any $\mathbf{v} \in V$,

$$T(-\mathbf{v}) = T((-1)\mathbf{v}) = (-1)T(\mathbf{v}) = -T(\mathbf{v})$$

c. For any $\mathbf{u}, \mathbf{v} \in V$,

$$T(\mathbf{u}-\mathbf{v}) = T(\mathbf{u}+(-1)\mathbf{v}) = T(\mathbf{u}) + (-1)T(\mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v})$$

□

Note. Theorem 6.14(a)에서 V의 zero vector와 W의 zero vector를 잘 구분하자.

Theorem 6.15
Let $T : V \to W$ be a linear transformation and let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ is a spanning set for V. Then $T(\mathcal{B}) = \{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_n)\}$ is the spanning set for range of T.

Proof. Let $\mathbf{w}$ be a vector in range(T), then there exists a vector $\mathbf{v} \in V$ such that $\mathbf{w} = T(\mathbf{v})$. Since $\mathcal{B}$ spans V, there exist scalars $c_1, \cdots, c_n$ such that

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

Thus,

$$\mathbf{w} = T(\mathbf{v}) = T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) = c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n)$$

Therefore, $\mathbf{w}$ is in span($T(\mathcal{B})$), hence $T(\mathcal{B})$ is a spanning set for range(T). □

Note. Spanning set은 이렇게 되는데, linear independence는 T가 one-to-one이어야 성립한다. (Theorem 6.22)

Definition. If $T : U \to V$ and $S : V \to W$ are linear transformations, then the **composition of S with T** is the mapping $S \circ T$, defined by

$$(S \circ T)(\mathbf{u}) = S(T(\mathbf{u}))$$

Theorem 6.16
If $T : U \to V$ and $S : V \to W$ are linear transformations, then $S \circ T : U \to W$ is a linear transformation.

Theorem 3.32 : Part 1
Let $T : \mathbb{R}^m \to \mathbb{R}^n$ and $S : \mathbb{R}^n \to \mathbb{R}^p$ be linear transformations. Then $S \circ T : \mathbb{R}^m \to \mathbb{R}^p$ is a linear transformation.

Proof. Let $\mathbf{u}, \mathbf{v}$ be vectors in U, and let c be a scalar. Then

$$\begin{aligned}(S \circ T)(\mathbf{u}+\mathbf{v}) &= S(T(\mathbf{u}+\mathbf{v}))\\ &= S(T(\mathbf{u})+T(\mathbf{v}))\\ &= S(T(\mathbf{u}))+S(T(\mathbf{v})) = (S\circ T)(\mathbf{u})+(S\circ T)(\mathbf{v})\end{aligned}$$

Also,

$$\begin{aligned}(S \circ T)(c\mathbf{u}) &= S(T(c\mathbf{u}))\\ &= S(cT(\mathbf{u}))\\ &= cS(T(\mathbf{u})) = c(S\circ T)(\mathbf{u})\end{aligned}$$

Therefore, $S \cdot T$ is a linear transformation. $\square$

Definition. A linear transformation $T : V \to W$ is **invertible** if there exists a linear transformation $T' : W \to V$ such that

$$T' \circ T = I_V \text{ and } T \circ T' = I_W$$

and such linear transformation T' is called an **inverse** for T.

Theorem 6.17
If T is an invertible linear transformation, then its inverse is unique.

Proof. **(Exercise 6.4 31)** Let $T : V \to W$ be an invertible linear transformation. Suppose that $T' : W \to V$ and $T'' : W \to V$ are both inverses of T, then

$$T \circ T' = T \circ T'' = I_W \text{ and } T' \circ T = T'' \circ T = I_V$$

Thus,

$$T' = I_V \circ T' = (T'' \circ T) \circ T' = T'' \circ (T \circ T') = T'' \circ I_W = T''$$

Therefore, the inverse of T is unique. $\square$

6.5 Kernel and Range

Definition. Let $T : V \to W$ be a linear transformation. The **kernel** of T, denoted by $\ker(T)$, is defined by

$$\ker(T) = \{\mathbf{v} \in V | T(\mathbf{v}) = \mathbf{0}\}$$

The **range** of T, denoted by $\text{range}(T)$, is defined by

$$\text{range}(T) = \{\mathbf{w} \in W | \mathbf{w} = T(\mathbf{v}) \text{ for some } \mathbf{v} \in V\}$$

Note. $\ker(T)$는 V의 vector들의 집합이고, $\text{range}(T)$는 W의 vector들의 집합임을 상기하자.

Theorem 6.18
Let $T : V \to W$ be a linear transformation.

a. $\ker(T)$ is a subspace of V.

b. $\text{range}(T)$ is a subspace of W.

Proof. Let $T : V \to W$ be a linear transformation.

a. Since $T(\mathbf{0}) = \mathbf{0}$ by Theorem 6.14(a), $\mathbf{0} \in \ker(T)$ so $\ker(T)$ is a nonempty set. Let $\mathbf{u}, \mathbf{v}$ be vectors in $\ker(T)$, and let c be a scalar, then

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) &= T(\mathbf{u}) + T(\mathbf{v}) = \mathbf{0} + \mathbf{0} = \mathbf{0} \\ T(c\mathbf{u}) &= cT(\mathbf{u}) = c\mathbf{0} = \mathbf{0} \end{aligned}$$

Therefore $\mathbf{u} + \mathbf{v} \in \ker(T)$ and $c\mathbf{u} \in \ker(T)$, hence $\ker(T)$ is a subspace of V by Theorem 6.2.

b. Since $T(\mathbf{0}) = \mathbf{0}$ by Theorem 6.14(a), $\mathbf{0} \in \text{range}(T)$ so $\text{range}(T)$ is a nonempty set. Let $\mathbf{w}_1, \mathbf{w}_2$ be vectors in $\text{range}(T)$ and let c be a scalar. Then there exist $\mathbf{v}_1, \mathbf{v}_2 \in V$ such that $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$. Then

$$\begin{aligned} \mathbf{w}_1 + \mathbf{w}_2 &= T(\mathbf{v}_1) + T(\mathbf{v}_2) = T(\mathbf{v}_1 + \mathbf{v}_2) \\ c\mathbf{w}_1 &= cT(\mathbf{v}_1) = T(c\mathbf{v}_1) \end{aligned}$$

Therefore $\mathbf{w}_1 + \mathbf{w}_2 \in \text{range}(T)$ and $c\mathbf{w}_1 \in \text{range}(T)$, hence $\text{range}(T)$ is a subspace of V by Theorem 6.2.

□

이제 $\ker(T)$와 $\text{range}(T)$가 vector space임을 증명했으니 그 dimension으로 rank와 nullity를 정의할 수 있다.

Definition. Let $T : V \to W$ be a linear transformation. Then the rank and nullity of T, denoted by $\text{rank}(T)$ and $\text{nullity}(T)$, is defined by

$$\text{rank}(T) = \dim(\text{range}(T)) \text{ and } \text{nullity}(T) = \dim(\ker(T))$$

Theorem 6.19 : The Rank Theorem : for Linear Transformation
Let $T : V \to W$ be a linear transformation from a finite-dimensional vector space V to a vector space W. Then

$$\text{rank}(T) + \text{nullity}(T) = \dim V$$

Proof. Let $n = \dim V$, and let $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ be a basis for $\ker(T)$. By Theorem 6.10(e), $\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ can be extended to a basis $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_k, \mathbf{v}_{k+1}, \cdots, \mathbf{v}_n\}$ for V.

Let $\mathcal{C} = \{T(\mathbf{v}_{k+1}), \cdots, T(\mathbf{v}_n)\}$, then $\mathcal{C}$ is a subset of $\text{range}(T)$. Let $\mathbf{w}$ be a vector in $\text{range}(T)$, then there exists $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$. Then there exist scalars $c_1, \cdots, c_n$ such that

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k + c_{k+1}\mathbf{v}_{k+1} + \cdots + c_n\mathbf{v}_n$$

Then

$$\begin{aligned}\mathbf{w} = T(\mathbf{v}) &= T(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k + c_{k+1}\mathbf{v}_{k+1} + \cdots + c_n\mathbf{v}_n) \\ &= c_1T(\mathbf{v}_1) + \cdots + c_kT(\mathbf{v}_k) + c_{k+1}T(\mathbf{v}_{k+1}) + \cdots + c_nT(\mathbf{v}_n) \\ &= c_{k+1}T(\mathbf{v}_{k+1}) + \cdots + c_nT(\mathbf{v}_n)\end{aligned}$$

Thus $\mathbf{w} \in \text{span}(\mathcal{C})$, which implies that $\mathcal{C}$ is a spanning set for range(T).

Now, consider the equation

$$c_{k+1}T(\mathbf{v}_{k+1}) + \cdots + c_nT(\mathbf{v}_n) = \mathbf{0}$$

Then $T(c_{k+1}\mathbf{v}_{k+1} + \cdots + c_n\mathbf{v}_n) = \mathbf{0}$, so $c_{k+1}\mathbf{v}_{k+1} + \cdots + c_n\mathbf{v}_n \in \ker(T)$. Since $\mathcal{B}$ is a basis for $\ker(T)$, there exist scalars $c_1, \cdots, c_k$ such that

$$c_{k+1}\mathbf{v}_{k+1} + \cdots + c_n\mathbf{v}_n = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$$

Since $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are linearly independent, this implies that $c_1 = \cdots = c_k = c_{k+1} = \cdots = c_n = 0$. Therefore, $T(\mathbf{v}_{k+1}), \cdots, T(\mathbf{v}_n)$ are linearly independent.

Since $\mathcal{C}$ is linearly independent and is a spanning set for range(T), it forms a basis for range(T). Therefore,

$$\text{rank}(T) + \text{nullity}(T) = (n - k) + k = n = \dim V$$

□

Definition. A linear transformation $T : V \to W$ is called **one-to-one** if T maps distinct vectors in V to distinct vectors in W. That is, for vectors $\mathbf{v}, \mathbf{w}$ in V,

$$\mathbf{v} \neq \mathbf{w} \Rightarrow T(\mathbf{v}) \neq T(\mathbf{w})$$

Definition. A linear transformation $T : V \to W$ is called **onto** if range$(T) = W$. That is,

$$^{\forall}\mathbf{w} \in W, ^{\exists}\mathbf{v} \in V \text{ such that } T(\mathbf{v}) = \mathbf{w}$$

Theorem 6.20
A linear transformation $T : V \to W$ is one-to-one if and only if $\ker(T) = \{\mathbf{0}\}$.

Proof. ($\Rightarrow$) Suppose that there exists a nonzero vector $\mathbf{v}$ in $\ker(T)$. Then $T(\mathbf{v}) = T(\mathbf{0}) = \mathbf{0}$, which is a contradiction since T is one-to-one. Therefore, $\ker(T) = \{\mathbf{0}\}$.

($\Leftarrow$) Let $\mathbf{u}, \mathbf{v} \in V$ such that $T(\mathbf{u}) = T(\mathbf{v})$. Since $\ker(T) = \{\mathbf{0}\}$ and $T(\mathbf{u} - \mathbf{v}) = T(\mathbf{u}) - T(\mathbf{v}) = \mathbf{0}$, $\mathbf{u} - \mathbf{v} = \mathbf{0}$. Therefore, $\mathbf{u} = \mathbf{v}$ and hence, T is one-to-one. □

Theorem 6.21
Let $\dim V = \dim W = n$. Then a linear transformation $T : V \to W$ is one-to-one if and only if it is onto.

Proof. ($\Rightarrow$) If T is one-to-one, $\ker(T) = \{\mathbf{0}\}$ by Theorem 6.20, so nullity$(T) = 0$. By the Rank Theorem, rank$(T) = \dim(\text{range}(T)) = 0$. Since range$(T)$ is a subspace of W, range$(T) = W$ by Theorem 6.11(b). Therefore, T is onto.

($\Leftarrow$) If T is onto, then rank$(T) = \dim(\text{range}(T)) = \dim(W) = n$. By the Rank Theorem, nullity(T) $= \dim(\ker(T)) = 0$, so $\ker(T) = \{\mathbf{0}\}$. Therefore, by Theorem 6.20, T is one-to-one. □

Theorem 6.22
Let $T : V \to W$ be a one-to-one linear transformation. If $S = \{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$ is a linearly independent set in V, Then $T(S) = \{T(\mathbf{v}_1), \cdots, T(\mathbf{v}_k)\}$ is a linearly independent set in W.

Proof. Consider the equation

$$c_1 T(\mathbf{v}_1) + \cdots + c_k T(\mathbf{v}_k) = \mathbf{0}$$

Then $T(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) = \mathbf{0}$, so $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \in \ker(T)$. Since T is one-to-one, $\ker(T) = \{\mathbf{0}\}$ by Theorem 6.20, so $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$. This implies that $c_1 = \cdots = c_k = 0$, since $\mathbf{v}_1, \cdots, \mathbf{v}_k$ are linearly independent. Therefore, $T(\mathbf{v}_1), \cdots, T(\mathbf{v}_k)$ are also linearly independent. □

Corollary 6.23
Let $\dim V = \dim W = n$. Then a one-to-one linear transformation $T : V \to W$ maps a basis for V to a basis for W.

Proof. If $\mathcal{B}$ is a basis of V, Theorem 6.15 and Theorem 6.22 together gives that $T(\mathcal{B})$ is a linearly independent set in W and is a spanning set for W. Therefore, $T(\mathcal{B})$ is a basis for W. □

Theorem 6.24
A linear transformation $T : V \to W$ is invertible if and only if it is one-to-one and onto.

Proof. ($\Rightarrow$) Suppose that T is invertible. Then there exists a linear transformation $T^{-1} : W \to V$ such that

$$T^{-1} \circ T = I_V \text{ and } T \circ T^{-1} = I_W$$

(i) Let $\mathbf{v}$ be a vector in $\ker(T)$, then $T(\mathbf{v}) = \mathbf{0}$. Then,

$$I_V(\mathbf{v}) = (T^{-1} \circ T)(\mathbf{v}) = T^{-1}(T(\mathbf{v})) = T^{-1}(\mathbf{0}) = \mathbf{0}$$

Thus, $\mathbf{v} = \mathbf{0}$, which implies that $\ker(T) = \{\mathbf{0}\}$. Therefore, T is one-to-one by Theorem 6.20.

(ii) Let $\mathbf{w}$ be a vector in W, and let $\mathbf{v} = T^{-1}(\mathbf{w})$. Then

$$T(\mathbf{v}) = T(T^{-1}(\mathbf{w})) = (T \circ T^{-1})(\mathbf{w}) = I_W(\mathbf{w}) = \mathbf{w}$$

Therefore, $\mathbf{w} \in \text{range}(T)$. Therefore, $\text{range}(T) = W$ and T is onto.

($\Leftarrow$) Suppose that $T : V \to W$ is one-to-one and onto. For every vector $\mathbf{w}$ in W, there exists a unique vector $\mathbf{v} \in V$ such that $T(\mathbf{v}) = \mathbf{w}$, since T is one-to-one and onto. Let $T' : W \to V$ be a transformation which maps $\mathbf{w}$ into such $\mathbf{v}$.

Let $\mathbf{w}_1, \mathbf{w}_2$ be vectors in W and c a scalar. Let $\mathbf{v}_1, \mathbf{v}_2$ be vectors such that $T(\mathbf{v}_1) = \mathbf{w}_1$ and $T(\mathbf{v}_2) = \mathbf{w}_2$, then $T'(\mathbf{w}_1) = \mathbf{v}_1$ and $T'(\mathbf{w}_2) = \mathbf{v}_2$. Then

$$T(\mathbf{v}_1 + \mathbf{v}_2) = T(\mathbf{v}_1) + T(\mathbf{v}_2) = \mathbf{w}_1 + \mathbf{w}_2$$
$$T(c\mathbf{v}_1) = cT(\mathbf{v}_1) = c\mathbf{w}_1$$

so $T'(\mathbf{w}_1 + \mathbf{w}_2) = \mathbf{v}_1 + \mathbf{v}_2 = T'(\mathbf{w}_1) + T'(\mathbf{w}_2)$ and $T'(c\mathbf{w}_1) = c\mathbf{v}_1 = cT(\mathbf{w}_1)$. Therefore, T' is a linear transformation.

Also, for any vector $\mathbf{v} \in V$, let $\mathbf{w} = T(\mathbf{v})$ then $T'(\mathbf{w}) = \mathbf{v}$, so

$$\begin{aligned}(T' \circ T)(\mathbf{v}) &= T'(T(\mathbf{v})) = T'(\mathbf{w}) = \mathbf{v} \\ (T \circ T')(\mathbf{w}) &= T(T'(\mathbf{w})) = T(\mathbf{v}) = \mathbf{w}\end{aligned}$$

This implies that $T' \circ T = I_V$ and $T \circ T' = I_W$.

Therefore, T' is an inverse of T, hence T is invertible. □

Definition. A linear transformation $T : V \to W$ is called **isomorphism** if it is one-to-one and onto. If V and W are two vector spaces such that there exists an isomorphism from V to W, then V is called to be **isomorphic** to W and denoted by $V \cong W$.

Theorem 6.25
Let V and W be two finite-dimensional vector spaces, over the same field of scalars. Then $V \cong W$ if and only if $\dim V = \dim W$.

Proof. Let $n = \dim V$.

($\Rightarrow$) If $V \cong W$, then there exists a linear transformation $T : V \to W$ which is an isomorphism. Since T is one-to-one, $\ker(T) = \{\mathbf{0}\}$ so $\text{nullity}(T) = 0$, and by Rank Theorem, $\text{rank}(T) = n$. Since T is onto, $W = \text{range}(T)$ so $\dim W = \dim(\text{range}(T)) = \text{rank}(T) = n$.

($\Leftarrow$) Let $n = \dim V = \dim W$. Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$ be a basis for V and let $\mathcal{C} = \{\mathbf{w}_1, \cdots, \mathbf{w}_n\}$ be a basis for W. For every vector $\mathbf{v}$ in V, there exist scalars $c_1, \cdots, c_n$ such that

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

then let $T : V \to W$ be defined by

$$T(\mathbf{v}) = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n$$

(i) First we prove that T is a linear transformation. Let $\mathbf{u}, \mathbf{v} \in V$, then there exist scalars $c_1, \cdots, c_n$ and $d_1, \cdots, d_n$ such that $\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$ and $\mathbf{v} = d_1\mathbf{v}_1 + \cdots + d_n\mathbf{v}_n$. Since $\mathbf{u} + \mathbf{v} = (c_1 + d_1)\mathbf{v}_1 + \cdots + (c_n + d_n)\mathbf{v}_n$,

$$\begin{aligned}T(\mathbf{u} + \mathbf{v}) &= (c_1 + d_1)\mathbf{w}_1 + \cdots + (c_n + d_n)\mathbf{w}_n \\ &= (c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n) + (d_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n) = T(\mathbf{u}) + T(\mathbf{v})\end{aligned}$$

And for any scalar c, $c\mathbf{u} = cc_1\mathbf{v}_1 + \cdots + cc_n\mathbf{v}_n$, so

$$\begin{aligned}T(c\mathbf{u}) &= cc_1\mathbf{w}_1 + \cdots + cc_n\mathbf{w}_n \\ &= c(c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n) = cT(\mathbf{u})\end{aligned}$$

Thus T is a linear transformation.

(ii) Suppose that $\mathbf{v} \in \ker(T)$, then there exist scalars $c_1, \cdots, c_n$ such that $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Then

$$T(\mathbf{v}) = c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n = \mathbf{0}$$

Since $\mathbf{w}_1, \cdots, \mathbf{w}_n$ are linearly independent, $c_1 = \cdots = c_n = 0$ so $\mathbf{v} = \mathbf{0}$. Therefore, $\ker(T) = \{\mathbf{0}\}$, hence T is one-to-one by Theorem 6.20.

(iii) Since $\dim V = \dim W$ and T is one-to-one, T is onto by Theorem 6.21.

By (i), (ii), and (iii), T is an isomorphism, therefore V and W are isomorphic. $\square$

Note. 6.5 요약 : '항상 False'의 증명은 Exercise 6.5 35에 있음.

$T : V \to W$	one-to-one	onto
$\dim V > \dim W$	항상 False	직접 판별
$\dim V = \dim W$	$\ker(T) \overset{?}{=} \{\vec{0}\} \begin{cases} \text{둘다 True} \\ \text{둘다 False} \end{cases}$	$(\because$ Thm 6.21)
$\dim V < \dim W$	직접 판별($\ker(T) \overset{?}{=} \{\vec{0}\}$)	항상 False

6.6 The Matrix of a Linear Transformation + 3.6 Part II

Definition. Let V and W be vector spaces with $\dim V = n$ and $\dim W = m$. Let $\mathcal{B}$ and $\mathcal{C}$ be bases for V and W, respectively, where $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$. The **matrix of T with respect to the bases $\mathcal{B}$ and $\mathcal{C}$**, denoted by $[T]_{\mathcal{C}\leftarrow\mathcal{B}}$, is the $m \times n$ matrix defined by

$$[T]_{\mathcal{C}\leftarrow\mathcal{B}} = \big[[T(\mathbf{v}_1)]_{\mathcal{C}} \quad \cdots \quad [T(\mathbf{v}_n)]_{\mathcal{C}}\big]$$

If $V = W$ and $\mathcal{B} = \mathcal{C}$, we write $[T]_{\mathcal{B}\leftarrow\mathcal{B}}$ as $[T]_{\mathcal{B}}$ for short.

Theorem 6.26
Let V and W be two finite-dimensional vector spaces with bases $\mathcal{B}$ and $\mathcal{C}$, respectively. If $T : V \to W$ is a linear transformation,

$$[T]_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{v}]_{\mathcal{B}} = [T(\mathbf{v})]_{\mathcal{C}}$$

for every vector $\mathbf{v}$ in V. Moreover, $[T]_{\mathcal{C}\leftarrow\mathcal{B}}$ is the unique matrix with such property.

Theorem 3.30 & Theorem 3.31
The transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if T is a matrix transformation: that is, T is defined by

$$T(\mathbf{x}) = A\mathbf{x}$$

where A is an $m \times n$ matrix.

Proof. Let $\mathbf{v}$ be a vector in V, then there exist scalars $c_1, \cdots, c_n$ such that

$$\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$$

since $\mathcal{B}$ is a basis for V. Thus,

$$\begin{aligned}
[T(\mathbf{v})]_{\mathcal{C}} &= [T(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n)]_{\mathcal{C}} \\
&= [c_1T(\mathbf{v}_1) + \cdots + c_nT(\mathbf{v}_n)]_{\mathcal{C}} \\
&= c_1[T(\mathbf{v}_1)]_{\mathcal{C}} + \cdots + c_n[T(\mathbf{v}_n)]_{\mathcal{C}} \\
&= \begin{bmatrix} [T(\mathbf{v}_1)]_{\mathcal{C}} & \cdots & [T(\mathbf{v}_n)]_{\mathcal{C}} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = [T]_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{v}]_{\mathcal{B}}
\end{aligned}$$

(Exercise 6.6 39) Let A be an $m \times n$ matrix such that

$$A[\mathbf{v}]_{\mathcal{B}} = [T(\mathbf{v})]_{\mathcal{C}}$$

for all $\mathbf{v} \in V$. For $i = 1, 2, \cdots, n$,

$$A[\mathbf{v}_i]_{\mathcal{B}} = A\mathbf{e}_i = [T(\mathbf{v}_i)]_{\mathcal{C}}$$

where $A\mathbf{e}_i$ is the ith column of A. Therefore, A is given as

$$A = \begin{bmatrix} [T(\mathbf{v}_1)]_{\mathcal{C}} & \cdots & [T(\mathbf{v}_n)]_{\mathcal{C}} \end{bmatrix} = P_{\mathcal{C}\leftarrow\mathcal{B}}$$

hence $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is the unique matrix with the given property. □

Extra Theorem
Let $T : V \to W$ be a linear transformation between finite-dimensional vector spaces V and W. Let $\mathcal{B}$ and $\mathcal{C}$ be bases for V and W, respectively.

a. $\text{nullity}(T) = \text{nullity}([T]_{\mathcal{C}\leftarrow\mathcal{B}})$

b. $\text{rank}(T) = \text{rank}([T]_{\mathcal{C}\leftarrow\mathcal{B}})$

Proof. Let $T : V \to W$ be a linear transformation between finite-dimensional vector spaces V and W. Let $\mathcal{B}$ and $\mathcal{C}$ be bases for V and W, respectively.

a. **(Exercise 6.6 40)**

$$\mathbf{v} \in V \text{ is in } \ker(T).$$

$$\Updownarrow$$

$$[T(\mathbf{v})]_{\mathcal{C}} = [\mathbf{0}]_{\mathcal{C}} = \mathbf{0} = [T]_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{v}]_{\mathcal{B}} \text{ (Theorem 6.26)}$$

$$\Updownarrow$$

$$[\mathbf{v}]_{\mathcal{B}} \text{ is in } \text{null}([T]_{\mathcal{C}\leftarrow\mathcal{B}}).$$

Since the coordinate vectors of distinct vectors in V are distinct, $\text{nullity}(T) = \dim(\ker(T)) = \dim(\text{null}([T]_{\mathcal{C}\leftarrow\mathcal{B}})) = \text{nullity}([T]_{\mathcal{C}\leftarrow\mathcal{B}})$.

b. **(Exercise 6.6 41)** Let $n = \dim V$. By the Rank Theorem for matrices and linear transformations,

$$\text{rank}(T) = n - \text{nullity}(T) = n - \text{nullity}([T]_{\mathcal{C}\leftarrow\mathcal{B}}) = \text{rank}([T]_{\mathcal{C}\leftarrow\mathcal{B}})$$

□

Theorem 6.27
Let U, V, and W be finite-dimensional vector spaces with bases $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{D}$, respectively. Let $T : U \to V$ and $S : V \to W$ be linear transformations. Then

$$[S \circ T]_{\mathcal{D}\leftarrow\mathcal{B}} = [S]_{\mathcal{D}\leftarrow\mathcal{C}}[T]_{\mathcal{C}\leftarrow\mathcal{B}}$$

Theorem 3.32 : Part 2

Let $T : \mathbb{R}^m \to \mathbb{R}^n$ and $S : \mathbb{R}^n \to \mathbb{R}^p$ be linear transformations. Then

$$[S \circ T] = [S][T]$$

Proof. Let $\mathcal{B} = \{\mathbf{v}_1, \cdots, \mathbf{v}_n\}$, then the ith column of $[S \circ T]_{\mathcal{D}\leftarrow\mathcal{B}}$ is $[(S \circ T)(\mathbf{v}_i)]_{\mathcal{D}}$, which is given as

$$\begin{aligned}
[(S \circ T)(\mathbf{v}_i)]_{\mathcal{D}} &= [S(T(\mathbf{v}_i))]_{\mathcal{D}} \\
&= [S]_{\mathcal{D}\leftarrow\mathcal{C}}[T(\mathbf{v}_i)]_{\mathcal{C}} \\
&= [S]_{\mathcal{D}\leftarrow\mathcal{C}}[T]_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{v}_i]_{\mathcal{B}} = [S]_{\mathcal{D}\leftarrow\mathcal{C}}[T]_{\mathcal{C}\leftarrow\mathcal{B}}\mathbf{e}_i
\end{aligned}$$

which is the ith column of $[S]_{\mathcal{D}\leftarrow\mathcal{C}}[T]_{\mathcal{C}\leftarrow\mathcal{B}}$. Therefore, $[S]_{\mathcal{D}\leftarrow\mathcal{C}}[T]_{\mathcal{C}\leftarrow\mathcal{B}} = [S \circ T]_{\mathcal{D}\leftarrow\mathcal{B}}$. □

Theorem 6.28
Let $T : V \to W$ be a linear transformation between n-dimensional vector spaces V and W, and let $\mathcal{B}$ and $\mathcal{C}$ be bases for V and W, respectively. Then T is invertible if and only if the matrix $[T]_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible. Also,

$$([T]_{\mathcal{C}\leftarrow\mathcal{B}})^{-1} = [T^{-1}]_{\mathcal{B}\leftarrow\mathcal{C}}$$

Proof. ($\Rightarrow$) If T is invertible, then there exists a linear transformation T^{-1} such that $T^{-1} \circ T = I_V$. Then

$$I = [I_V]_{\mathcal{B}\leftarrow\mathcal{B}} = [T^{-1} \circ T]_{\mathcal{B}\leftarrow\mathcal{B}} = [T^{-1}]_{\mathcal{B}\leftarrow\mathcal{C}}[T]_{\mathcal{C}\leftarrow\mathcal{B}}$$

Thus $([T]_{\mathcal{C}\leftarrow\mathcal{B}})^{-1} = [T^{-1}]_{\mathcal{B}\leftarrow\mathcal{C}}$.

($\Leftarrow$) Suppose that $[T]_{\mathcal{C}\leftarrow\mathcal{B}}$ is invertible. Let $\mathbf{v}$ be a vector in $\ker(T)$, then

$$[T]_{\mathcal{C}\leftarrow\mathcal{B}}[\mathbf{v}]_{\mathcal{B}} = [T(\mathbf{v})]_{\mathcal{C}} = [\mathbf{0}]_{\mathcal{C}} = \mathbf{0}$$

Thus, $[\mathbf{v}]_{\mathcal{B}} = \mathbf{0}$ by F.T.I.M, $\mathbf{v} = \mathbf{0}$ and $\ker(T) = \{\mathbf{0}\}$. Therefore, T is one-to-one (Theorem 6.20) and since $\dim V = \dim W = n$, T is onto (Theorem 6.21), hence T is invertible (Theorem 6.24). □

Theorem 6.29
Let V be a finite-dimensional vector space with bases $\mathcal{B}$ and $\mathcal{C}$ and let $T : V \to V$ be a linear transformation. Then

$$[T]_{\mathcal{C}} = (P_{\mathcal{B}\leftarrow\mathcal{C}})^{-1}[T]_{\mathcal{B}}P_{\mathcal{B}\leftarrow\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[T]_{\mathcal{B}}P_{\mathcal{B}\leftarrow\mathcal{C}}$$

Theorem 3.33
The linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is invertible if and only if $[T]$ is an invertible matrix. If T is invertible linear transformation, then

$$[T^{-1}] = [T]^{-1}$$

Proof. Note that $P_{\mathcal{C}\leftarrow\mathcal{B}}$ is the matrix of I_V with respect to the bases $\mathcal{B}$ and $\mathcal{C}$, since if $\mathcal{B} = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$,

$$P_{\mathcal{C}\leftarrow\mathcal{B}} = \begin{bmatrix}[\mathbf{u}_1]_{\mathcal{C}} & \cdots & [\mathbf{u}_n]_{\mathcal{C}}\end{bmatrix} = \begin{bmatrix}[I_V(\mathbf{u}_1)]_{\mathcal{C}} & \cdots & [I_V(\mathbf{u}_n)]_{\mathcal{C}}\end{bmatrix} = [I_V]_{\mathcal{C}\leftarrow\mathcal{B}}$$

Similarly, $P_{\mathcal{B}\leftarrow\mathcal{C}}$ is the matrix of I_V with respect to the bases $\mathcal{C}$ and $\mathcal{B}$. Therefore, by Theorem 6.27,

$$\begin{aligned}[T]_{\mathcal{C}\leftarrow\mathcal{C}} &= [I_V \circ T \circ I_V]_{\mathcal{C}\leftarrow\mathcal{C}} \\ &= [I_V]_{\mathcal{C}\leftarrow\mathcal{B}}[T]_{\mathcal{B}\leftarrow\mathcal{B}}[I_V]_{\mathcal{B}\leftarrow\mathcal{C}} = P_{\mathcal{C}\leftarrow\mathcal{B}}[T]_{\mathcal{B}\leftarrow\mathcal{B}}P_{\mathcal{B}\leftarrow\mathcal{C}}\end{aligned}$$

□

Note. 교과서의 Theorem 3.33을 동치인 명제로 확장하였다. 그리고 $P_{\mathcal{C}\leftarrow\mathcal{B}}$를 identity transformation의 행렬로 나타내는 것은 유용하니 알아 두자.

Definition. Let V be a finite-dimensional vector space and let $T : V \to V$ be a linear transformation. Then T is called **diagonalizable** if there is a basis $\mathcal{C}$ for V such that $[T]_{\mathcal{C}}$ is a diagonal matrix.

Theorem 6.30 : The Fundamental Theorem of Invertible Matrices: Version 4
Let A be an $n \times n$ matrix, and let $T : V \to W$ be a linear transformation such that its matrix $[T]_{\mathcal{C}\leftarrow\mathcal{B}}$ with respect to the bases $\mathcal{B}$ and $\mathcal{C}$ of V and W, respectively, is equal to A. The following propositions are equivalent:

a. A is invertible.

b. $A\mathbf{x} = \mathbf{b}$ has a unique solution for every $\mathbf{b} \in \mathbb{R}^n$.

c. $A\mathbf{x} = \mathbf{0}$ has only the trivial solution.

d. The RREF of A is I_n.

e. A is a product of elementary matrices.

f. $\text{rank}(A) = n$

g. $\text{nullity}(A) = 0$

h. The columns of A are linearly independent.

i. The columns of A span $\mathbb{R}^n$.

j. The columns of A form a basis for $\mathbb{R}^n$.

k. The rows of A are linearly independent.

l. The rows of A span $\mathbb{R}^n$.

m. The rows of A form a basis for $\mathbb{R}^n$.

n. $\det A \neq 0$

o. 0 is not an eigenvalue of A.

p. T is invertible.

q. T is one-to-one.

r. T is onto.

s. $\ker T = \{\mathbf{0}\}$

t. $\text{range}(T) = W$

Proof. Since A is $n \times n$ matrix, $\dim V = \dim W = n$, so Theorem 6.21 gives (q) $\Leftrightarrow$ (r). Theorem 6.24 gives (q) and (r) $\Leftrightarrow$ (p), and Theorem 6.20 gives (q) $\Leftrightarrow$ (s). (r) $\Leftrightarrow$ (t) holds since it is the definition of onto. Finally, Theorem 6.28 gives (a) $\Leftrightarrow$ (p). $\square$