

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

L^AT_EX Technician : 14041 박승원

Last Compilation Time : Thursday 23rd June, 2016 22:57

Contents

Table of Contents	1
List of Theorems	2
List of Extra Theorems	2
4 Eigenvalues and Eigenvectors	4
4.1 Introduction to Eigenvalues and Eigenspaces	4
4.2 Determinants	4
4.3 Eigenvalues and Eigenvectors of $n \times n$ Matrices	14
4.4 Similarity and Diagonalization	17
4.5 Solutions of Exercises from Chapter 4	20
5 Orthogonality	32
5.1 Orthogonality in \mathbb{R}^n	32
5.2 Orthogonal Complements and Orthogonal Projections	35
5.3 The QR Factorization	39
5.4 Orthogonal Diagonalization of Symmetric Matrices	40
5.5 Solution for Exercises from Chapter 5	42
6 Vector Spaces	54
6.1 Vector Spaces and Subspaces	54
6.2 Linear Independence, Basis and Dimension	57
6.3 Change of Basis	65
6.4 Linear Transformations + 3.6 Part I	67
6.5 Kernel and Range	69
6.6 The Matrix of a Linear Transformation + 3.6 Part II	74
6.7 Solutions for Exercises from Chapter 6	78

List of Theorems

4.1	Theorem (The Laplace Expansion Theorem)	5	5.11	Theorem (The Orthogonal Decomposition Theorem)	37
4.2	Theorem	6	5.12	Corollary	38
4.3	Theorem (Properties of Determinants)	6	5.13	Theorem	38
4.4	Theorem	8	5.14	Corollary	38
4.6	Theorem	9	5.15	Theorem (The Gram-Schmidt Process)	39
4.7	Theorem	10	5.16	Theorem (The QR Factorization) .	39
4.8	Theorem	10	5.17	Theorem	40
4.9	Theorem	10	5.18	Theorem	40
4.10	Theorem	11	5.19	Theorem	41
4.11	Theorem (Cramer's Rule)	11	5.20	Theorem (The Spectral Theorem)	41
4.12	Theorem	12	6.1	Theorem	55
4.15	Theorem	14	6.2	Theorem	56
4.16	Theorem	14	6.3	Theorem	57
4.17	Theorem (The Fundamental Theorem of Invertible Matrices : Version 3)	14	6.4	Theorem	58
4.18	Theorem	15	6.5	Theorem	59
4.19	Theorem	16	6.6	Theorem	60
4.20	Theorem	16	6.7	Theorem	60
4.21	Theorem	17	6.8	Theorem	61
4.22	Theorem	17	6.9	Theorem (The Basis Theorem) . .	62
4.23	Theorem	18	6.10	Theorem	62
4.24	Theorem	19	6.11	Theorem	64
4.25	Theorem	19	6.12	Theorem	66
4.27	Theorem (The Diagonalization Theorem)	20	6.13	Theorem	67
5.1	Theorem	32	6.14	Theorem	68
5.2	Theorem	32	6.15	Theorem	68
5.3	Theorem	33	6.16	Theorem	69
5.4	Theorem	33	6.17	Theorem	69
5.5	Theorem	33	6.18	Theorem	70
5.6	Theorem	34	6.19	Theorem (The Rank Theorem : for Linear Transformation)	70
5.7	Theorem	34	6.20	Theorem	71
5.8	Theorem	34	6.21	Theorem	71
5.9	Theorem	35	6.22	Theorem	72
5.10	Theorem	36	6.23	Corollary	72
			6.24	Theorem	72
			6.25	Theorem	73
			6.26	Theorem	74
			6.27	Theorem	76
			6.28	Theorem	76
			6.29	Theorem	76
			6.30	Theorem (The Fundamental Theorem of Invertible Matrices: Version 4)	77

Extra Theorems

4.13	Lemma	4
4.5	Lemma	9
4.1	Extra Theorem	13
4.26	Lemma	19
6.1	Extra Theorem	57
6.2	Extra Theorem	61
6.3	Extra Theorem	75

L^AT_EX technician's note : Theorems from chapter 3 are omitted here, due to T_EX-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 i th row and j th 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, k th rows and j, l th columns from A .

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

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 i th row**) and also as

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

(which is the **cofactor expansion along the j th column**)

Proof. Let $A = [a_{ij}]$ be an $n \times n$ matrix. Let B be the matrix obtained by moving i th 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 . \square

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 square matrices.

- If A has a zero row (column), then $\det A = 0$.
- (**Lemma 4.14**) If B is obtained by interchanging two rows (columns) of A , then $\det B = -\det A$.
- If A has two identical rows (columns), then $\det A = 0$.
- If B is obtained by multiplying a row (column) of A by k , then $\det B = k \det A$.
- If A , B , and C are identical except that the i th row (column) of C is the sum of the i th rows (columns) of A and B , then $\det C = \det A + \det B$.
- 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 i th 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 r th and $(r+1)$ th row of A , where $r \in \{1, 2, \dots, n-1\}$. 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 r th row and s th row ($r < s$) of A . This operation is equivalent to following sequence of operations:

$$\begin{aligned} R_r &\leftrightarrow R_{r+1}, R_{r+1} \leftrightarrow R_{r+2}, \dots, R_{s-1} \leftrightarrow R_s, \\ R_{s-2} &\leftrightarrow R_{s-1}, \dots, R_r \leftrightarrow R_{r+1} \end{aligned}$$

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 i th 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 i th 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.

- If E results from interchanging two rows of I_n , then $\det E = -1$.
- If E results from multiplying one row of I_n by k , then $\det E = k$.
- 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.

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

- b. $E_i(k)$ is the elementary matrix resulting from multiplying i th row of I by a nonzero scalar k .
- c. $E_{ij}(k)$ is the elementary matrix resulting from adding a multiple of j th row with k to i th row of I .

That is,

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

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, \dots, 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. \square

Theorem 4.7

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

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

.

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

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

\square

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, \dots, 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, $\det(AB) = (\det A)(\det B) = 0$. \square

Theorem 4.9

If A is invertible, then

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

Proof. Let A be an 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 $(E_i(k))^T = E_i(k)$, $\det E_i(k) = \det (E_i(k))^T$.

(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, \dots, 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 i th 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

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

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 $A\mathbf{x} = \mathbf{b}$. For $i = 1, 2, \dots, n$,

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

Then, by Laplace Expansion Theorem,

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

□

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

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

Theorem 4.12

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

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

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

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

□

Extra Theorem

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

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

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

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

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

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

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

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

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

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

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

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

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

Consequently, we have

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

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

A 가 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 $a_{11}, a_{22}, \dots, a_{nn}$ 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. **(Exercise 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 $\mathbf{v}_1, \dots, \mathbf{v}_m$ with $\lambda_1, \lambda_2, \dots, \lambda_m$. 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, \dots, c_m exist such that

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

then, for any integer k ,

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

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

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

For any integer k ,

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

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,

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

□

Theorem 4.20

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

Proof. Suppose that $\mathbf{v}_1, \dots, \mathbf{v}_m$ 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, \dots, c_{k-1} such that

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

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

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

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

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

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

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

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. $A = I_n^{-1}AI_n$, 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, $A \sim C$.

\square

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 $P = [\mathbf{p}_1 \ \mathbf{p}_2 \ \cdots \ \mathbf{p}_n]$ such that $P^{-1}AP = D$. Since $AP = PD$,

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

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

ible. Since $\mathbf{p}_1, \dots, \mathbf{p}_n$ are eigenvectors of A with corresponding eigenvalue $\lambda_1, \dots, \lambda_n$, $AP = PD$. 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 $\lambda_1, \dots, \lambda_k$ be distinct eigenvalues of A . If \mathcal{B}_i is a basis for the eigenspace E_{λ_i} , then $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_k$ (i.e., the total collection of basis vectors for all of the eigenspaces) is linearly independent.

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

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

Let $\mathbf{x}_i = c_{i1}\mathbf{v}_{i1} + \dots + c_{in_i}\mathbf{v}_{in_i}$, then $\mathbf{x}_1 + \dots + \mathbf{x}_k = \mathbf{0}$. Since $\mathbf{x}_i \in \text{span}(\mathcal{B}_i)$, \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, $\mathbf{x}_1 + \dots + \mathbf{x}_k = \mathbf{0}$ gives $\mathbf{x}_1 = \dots = \mathbf{x}_k = \mathbf{0}$. (If at least one of $\mathbf{x}_1, \dots, \mathbf{x}_k$ 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. \square

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. \square

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 $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_p\}$ be a basis for E_{λ_0} .

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

$$Q = [\mathbf{v}_1 \ \dots \ \mathbf{v}_p \ \mathbf{v}_{p+1} \ \dots \ \mathbf{v}_n] = [U \ V]$$

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

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

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

Since $AU = A[\mathbf{v}_1 \ \mathbf{v}_2 \ \dots \ \mathbf{v}_p] = [A\mathbf{v}_1 \ A\mathbf{v}_2 \ \dots \ A\mathbf{v}_p] = [\lambda_0\mathbf{v}_1 \ \lambda_0\mathbf{v}_2 \ \dots \ \lambda_0\mathbf{v}_p] = \lambda_0 U$,

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

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 $\lambda_1, \lambda_2, \dots, \lambda_k$. 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, $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 \cup \dots \cup \mathcal{B}_k$ contains at least $n_1 + n_2 + \dots + n_k = n$ 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 $\lambda_1, \lambda_2, \dots, \lambda_k$ be m_1, m_2, \dots, m_k . By Theorem 4.26, the geometric multiplicity of each eigenvalue is lesser or equal to the algebraic multiplicity. Thus,

$$\begin{aligned}n &= \dim E_{\lambda_1} + \dim E_{\lambda_2} + \dots + \dim E_{\lambda_k} \leq m_1 + m_2 + \dots + m_k = n \\ \therefore (m_1 - \dim E_{\lambda_1}) + \dots + (m_k - \dim E_{\lambda_k}) &= 0\end{aligned}$$

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 $\lambda_1, \lambda_2, \dots, \lambda_k$ be m_1, m_2, \dots, m_k . Then \mathcal{B} has $\dim E_{\lambda_1} + \dots + \dim E_{\lambda_k} = m_1 + m_2 + \dots + m_k = n$ vectors. □

4.5 Solutions of Exercises from Chapter 4

(4.1 35 + 4.1 36 + 4.1 37) : Eigenvalues and Characteristic Equations of 2×2 matrices

35(a) Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $\text{tr}(A) = a + d$, and $\det A = ad - bc$.

The eigenvalues of A are the solutions of the characteristic equation of A , which is

$$\begin{aligned}\det(A - \lambda I) &= \begin{vmatrix} a - \lambda & b \\ c & d - \lambda \end{vmatrix} = (a - \lambda)(d - \lambda) - bc \\ &= \lambda^2 - (a + d)\lambda + (ad - bc) = \lambda^2 - \operatorname{tr}(A)\lambda + \det A = 0\end{aligned}$$

35(b) The solution of the characteristic equation of A , $\lambda^2 - (a + d)\lambda + (ad - bc) = 0$, is given as

$$\begin{aligned}\lambda &= \frac{1}{2}((a + d) \pm \sqrt{(a + d)^2 - 4(ad - bc)}) \\ &= \frac{1}{2}((a + d) \pm \sqrt{(a - d)^2 + 4bc})\end{aligned}$$

35(c) Since λ_1, λ_2 are the solutions of $\lambda^2 - \operatorname{tr}(A)\lambda + \det A$, $\lambda_1 + \lambda_2 = \operatorname{tr}(A)$ and $\lambda_1\lambda_2 = \det A$.

36(a)(b)(c) Since the eigenvalues of A are the solutions of $\lambda^2 - \operatorname{tr}(A)\lambda + \det A$, the condition is given from the determinant of the quadratic equation. Therefore,

$$\begin{aligned}D &= (a + d)^2 - 4(ad - bc) \\ &= (a - d)^2 + 4bc \begin{cases} > 0 \text{ then } A \text{ has two distinct real eigenvalues.} \\ = 0 \text{ then } A \text{ has one real eigenvalue.} \\ < 0 \text{ then } A \text{ has two distinct complex eigenvalues.} \end{cases}\end{aligned}$$

37 Exercise 4.1 35(a) gives that the characteristic polynomial of $A = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}$ is $\lambda^2 - (a + d)\lambda + ad = 0$. Therefore, the eigenvalues of A are $\lambda = a$ and $\lambda = d$.

(4.1 38)

From the equation given from Exercise 4.1 35(b), the eigenvalues of $A = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ are

$$\lambda = \frac{1}{2}(2a \pm \sqrt{-4b^2}) = a \pm bi$$

(i) If $b = 0$, $\lambda = a$ is the only eigenvalue of A . The eigenspace E_0 is equal to the $\operatorname{null}(A - aI) = \operatorname{null}(O)$, which is equivalent to \mathbb{C}^2 .

(ii) If $b \neq 0$, then A has two distinct eigenvalues $\lambda_1 = a + bi$ and $\lambda_2 = a - bi$. The eigenspace E_{a+bi} is equal to the null space of $A - (a + bi)I = \begin{bmatrix} -bi & b \\ -b & -bi \end{bmatrix}$. The RREF of $\begin{bmatrix} -bi & b \\ -b & -bi \end{bmatrix}$ is $\begin{bmatrix} 1 & i \\ 0 & 0 \end{bmatrix}$ so $\left\{ \begin{bmatrix} 1 \\ i \end{bmatrix} \right\}$ forms basis for $\operatorname{null}(A - (a + bi)I) = E_{a+bi}$.

Similarly, the eigenspace E_{a-bi} is equal to the null space of $A - (a - bi)I = \begin{bmatrix} bi & b \\ -b & bi \end{bmatrix}$, whose RREF is $\begin{bmatrix} 1 & -i \\ 0 & 0 \end{bmatrix}$. Thus, $\left\{ \begin{bmatrix} i \\ 1 \end{bmatrix} \right\}$ forms basis for $\operatorname{null}(A - (a - bi)I) = E_{a-bi}$.

(4.2 53)

For $n \times n$ matrices A and B ,

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

(4.2 54)

For $n \times n$ matrices A and B ,

$$\det(B^{-1}AB) = (\det B^{-1})(\det A)(\det B) = \left(\frac{1}{\det B}\right)(\det A)(\det B) = \det A$$

(4.2 65)

For $n \times n$ invertible matrix A , Theorem 4.12 states that

$$A^{-1} = \left(\frac{1}{\det A}\right)\text{adj } A \text{ and } A = (A^{-1})^{-1} = \left(\frac{1}{\det A^{-1}}\right)\text{adj}(A^{-1}) = (\det A)\text{adj}(A^{-1})$$

Therefore,

$$(\text{adj } A)^{-1} = ((\det A)A^{-1})^{-1} = \left(\frac{1}{\det A}\right)(A^{-1})^{-1} = \left(\frac{1}{\det A}\right)A = \text{adj}(A^{-1})$$

(4.2 66)

For $n \times n$ matrix A ,

(i) If A is invertible, by Theorem 4.7 and Theorem 4.12,

$$\det(\text{adj } A) = \det((\det A)A^{-1}) = (\det A)^n \left(\frac{1}{\det A}\right) = (\det A)^{n-1}$$

(ii) If A is not invertible, by the Extra Theorem (see page 12),

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

Since $\det A = 0$, $A(\text{adj } A) = O$.

(i) If $A \neq O$, suppose that $\text{adj } A$ is invertible. Then $A = A(\text{adj } A)(\text{adj } A)^{-1} = O(\text{adj } A)^{-1} = O$, which is a contradiction. Therefore, $\text{adj } A$ is not invertible.

(ii) If $A = O$, $\text{adj } A = O$ so $\text{adj } A$ is not invertible.

Since $\text{adj } A$ is not invertible, by F.T.I.M,

$$\det(\text{adj } A) = 0 = (\det A)^{n-1}$$

(4.2 69)

We prove the proposition by mathematical induction over the size of $P = [p_{ij}]$.

(i) If P is a 1×1 matrix, then by the Laplace Expansion Theorem, the cofactor expansion along the first column gives

$$\det A = p_{11}(\det S) = (\det P)(\det S)$$

So the proposition holds.

(ii) Suppose that every matrices that can be partitioned as $\begin{bmatrix} P & Q \\ O & S \end{bmatrix}$, where P is a $n \times n$ matrix, has its determinant as $(\det P)(\det S)$.

Let $A = \begin{bmatrix} P & Q \\ O & S \end{bmatrix}$, where $P = [p_{ij}]$ is a $(n+1) \times (n+1)$ matrix. By the Laplace Expansion Theorem, the cofactor expansion along the first column of A gives

$$\det A = \sum_{k=1}^{n+1} p_{k1}(-1)^{1+k} \begin{vmatrix} P_{k1} & Q'_k \\ O & S \end{vmatrix}$$

where Q'_k is a matrix obtained by removing the k th row of Q . Since P_{k1} is an $n \times n$ matrix, the inductive hypothesis gives

$$\begin{aligned} \det A &= \sum_{k=1}^{n+1} p_{k1}(-1)^{1+k} \begin{vmatrix} P_{k1} & Q'_k \\ O & S \end{vmatrix} = \sum_{k=1}^{n+1} p_{k1}(-1)^{1+k} (\det P_{k1})(\det S) \\ &= (\det S) \sum_{k=1}^{n+1} p_{k1}(-1)^{1+k} (\det P_{k1}) \end{aligned}$$

Again, $\sum_{k=1}^{n+1} p_{k1}(-1)^{1+k} (\det P_{k1})$ is the cofactor expansion of P along the first column. Therefore,

$$\det A = (\det S)(\det P)$$

(4.3 19)

(a) The Laplace Expansion Theorem gives that for any square matrix A , $\det A = \det A^T$. The characteristic polynomial of A is equal to $\det(A - \lambda I)$, and the characteristic polynomial of A^T is equal to $\det(A^T - \lambda I) = \det(A^T - \lambda I^T) = \det(A - \lambda I)^T = \det(A - \lambda I)$. Therefore, the characteristic polynomial of A and A^T are the same.

(b) Let $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix}$, then the characteristic polynomial of A is given by $\lambda^2 - 3\lambda$, so the eigenvalues of A and A^T are $\lambda = 0, \lambda = 3$.

First, consider the eigenspaces of A . $E_0 = \text{null}(A)$, so $\left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$ forms the basis for E_0 . $E_3 = \text{null}(A - 3I)$, so $\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}$ forms the basis for E_3 .

Now consider the eigenspaces of A^T . $E_0 = \text{null}(A^T)$, so $\left\{ \begin{bmatrix} 1 \\ -2 \end{bmatrix} \right\}$ forms the basis for E_0 . $E_3 = \text{null}(A^T - 3I)$, so $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ forms the basis for E_3 .

This example shows that the eigenvalues of A and A^T are the same, but the eigenspaces could be different.

(4.3 22)

Suppose that \mathbf{v} is an eigenvector of A with corresponding eigenvalue λ , then $A\mathbf{v} = \lambda\mathbf{v}$. For any scalar c ,

$$(A - cI)\mathbf{v} = A\mathbf{v} - cI\mathbf{v} = \lambda\mathbf{v} - c\mathbf{v} = (\lambda - c)\mathbf{v}$$

Therefore, \mathbf{v} is an eigenvector of $A - cI$ with corresponding eigenvalue $\lambda - c$.

(4.3 24)

(a)(b) Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$, then $\lambda = 1$ is an eigenvalue of A with corresponding eigenvector $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, and it is also an eigenvalue of B with corresponding eigenvector $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. However, $1 + 1 = 2$ is not an eigenvalue of $A + B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, since $[A + B - 2I \mid \mathbf{0}] = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}$ only has the trivial solution.

Also, $1 \times 1 = 1$ is not an eigenvalue of $AB = O$, since $[AB - I \mid \mathbf{0}] = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \end{bmatrix}$ only has the trivial solution.

(c) Suppose that λ is an eigenvalue of A with corresponding eigenvector \mathbf{x} , and μ is an eigenvalue of B with corresponding eigenvector \mathbf{x} . Then

$$(A + B)\mathbf{x} = A\mathbf{x} + B\mathbf{x} = \lambda\mathbf{x} + \mu\mathbf{x} = (\lambda + \mu)\mathbf{x}$$

so $\lambda + \mu$ is an eigenvalue of $A + B$. Also,

$$AB\mathbf{v} = A(\mu\mathbf{v}) = \mu(A\mathbf{v}) = \mu(\lambda\mathbf{v}) = (\lambda\mu)\mathbf{v}$$

so $\lambda\mu$ is an eigenvalue of AB .

(4.3 28, 4.3 29 + 4.3 32) : Companion Matrices of Polynomials

For a polynomial $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$, the companion matrix of $p(x)$ is denoted by $C(p)$, and defined as

$$C(p) = \begin{bmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_1 & -a_0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

28(a) The companion matrix of $p(x) = x^2 + ax + b$ is

$$C(p) = \begin{bmatrix} -a & -b \\ 1 & 0 \end{bmatrix}$$

Therefore, the characteristic polynomial of $C(p)$ is given as

$$\begin{aligned} |C(p) - \lambda I| &= \begin{vmatrix} -a - \lambda & -b \\ 1 & -\lambda \end{vmatrix} \\ &= \lambda(a + \lambda) + b = \lambda^2 + a\lambda + b \end{aligned}$$

28(b) Suppose that λ is an eigenvalue of $C(p)$. By 28(a), $\lambda^2 + a\lambda + b = 0$. Then,

$$C(p) \begin{bmatrix} \lambda \\ 1 \end{bmatrix} = \begin{bmatrix} -a & -b \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda \\ 1 \end{bmatrix} = \begin{bmatrix} -a\lambda - b \\ \lambda \end{bmatrix} = \begin{bmatrix} \lambda^2 \\ \lambda \end{bmatrix} = \lambda \begin{bmatrix} \lambda \\ 1 \end{bmatrix}$$

Therefore, $\begin{bmatrix} \lambda \\ 1 \end{bmatrix}$ is an eigenvector of $C(p)$ corresponding to λ .

29(a) The companion matrix of $p(x) = x^3 + ax^2 + bx + c$ is

$$C(p) = \begin{bmatrix} -a & -b & -c \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

Therefore, the characteristic polynomial of $C(p)$ is given as

$$\begin{aligned} |C(p) - \lambda I| &= \begin{vmatrix} -a - \lambda & -b & -c \\ 1 & -\lambda & 0 \\ 0 & 1 & -\lambda \end{vmatrix} \\ &= -(a + \lambda) \begin{vmatrix} -\lambda & 0 \\ 1 & -\lambda \end{vmatrix} - \begin{vmatrix} -b & -c \\ 1 & -\lambda \end{vmatrix} \\ &= \lambda^2(a + \lambda) - (b\lambda + c) = -(\lambda^3 + a\lambda^2 + b\lambda + c) \end{aligned}$$

29(b) Suppose that λ is an eigenvalue of $C(p)$. By 29(a), $\lambda^3 + a\lambda^2 + b\lambda + c = 0$. Then,

$$C(p) \begin{bmatrix} \lambda^2 \\ \lambda \\ 1 \end{bmatrix} = \begin{bmatrix} -a & -b & -c \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda^2 \\ \lambda \\ 1 \end{bmatrix} = \begin{bmatrix} -(a\lambda^2 + b\lambda + c) \\ \lambda^2 \\ \lambda \end{bmatrix} = \begin{bmatrix} \lambda^3 \\ \lambda^2 \\ \lambda \end{bmatrix} = \lambda \begin{bmatrix} \lambda^2 \\ \lambda \\ 1 \end{bmatrix}$$

Therefore, $\begin{bmatrix} \lambda^2 \\ \lambda \\ 1 \end{bmatrix}$ is an eigenvector of $C(p)$ corresponding to λ .

32(a) We prove the proposition using the mathematical induction on the degree of the polynomial.

(i) If the degree of $p(x)$ is 2, the proposition holds. (28(a) gives the proof.)

(ii) Suppose that if the degree of $p(x)$ is n , the companion matrix $C(p)$ of $p(x)$ has characteristic polynomial $(-1)^n p(x)$.

Let $p(x) = x^{n+1} + a_n x^n + \cdots + a_1 x + a_0$. Then the companion matrix of $p(x)$ is

$$C(p) = \begin{bmatrix} -a_n & -a_{n-1} & \cdots & -a_1 & a_0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}$$

Therefore, the characteristic polynomial of $C(p)$ is given as

$$\begin{aligned}
|C(p) - \lambda I| &= \begin{vmatrix} -a_n - \lambda & -a_{n-1} & \cdots & -a_1 & a_0 \\ 1 & -\lambda & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -\lambda & 0 \\ 0 & 0 & \cdots & 1 & -\lambda \end{vmatrix} \\
&= a_0(-1)^{n+1} \begin{vmatrix} 1 & -\lambda & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\lambda \\ 0 & 0 & \cdots & 0 & 1 \end{vmatrix} + (-\lambda)(-1)^{2n} \begin{vmatrix} -a_n - \lambda & -a_{n-1} & \cdots & -a_2 & -a_1 \\ 1 & -\lambda & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -\lambda & 0 \\ 0 & 0 & \cdots & 1 & -\lambda \end{vmatrix} \\
&= a_0(-1)^{n+1} - \lambda \begin{vmatrix} -a_n - \lambda & -a_{n-1} & \cdots & -a_2 & -a_1 \\ 1 & -\lambda & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -\lambda & 0 \\ 0 & 0 & \cdots & 1 & -\lambda \end{vmatrix} = a_0(-1)^{n+1} - \lambda(*)
\end{aligned}$$

where $(*)$ is the determinant of the matrix above. Let $q(x) = \frac{1}{x}(p(x) - a_0) = x^n + a_n x^{n-1} + \cdots + a_2 x + a_1$, then $(*)$ is the characteristic polynomial of $C(q)$. Since the degree of the $q(x)$ is n , by inductive hypothesis, $(*) = (-1)^n q(\lambda)$. Therefore,

$$\begin{aligned}
|C(p) - \lambda I| &= a_0(-1)^{n+1} - \lambda(-1)^n q(\lambda) \\
&= a_0(-1)^{n+1} - \lambda(-1)^n \frac{1}{\lambda}(p(\lambda) - a_0) \\
&= (-1)^{n+1} p(\lambda)
\end{aligned}$$

By (i) and (ii), the proposition holds for all polynomials.

32(b) Suppose that λ is an eigenvalue of the companion matrix of $p(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0$. By 32(a), $(-1)^{n+1}p(\lambda) = 0$, so $p(\lambda) = 0$. Then,

$$\begin{aligned}
C(p) \begin{bmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \vdots \\ \lambda \\ 1 \end{bmatrix} &= \begin{bmatrix} -a_{n-1} & -a_{n-2} & \cdots & -a_1 & a_0 \\ 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \vdots \\ \lambda \\ 1 \end{bmatrix} \\
&= \begin{bmatrix} -(a_{n-1}\lambda^{n-1} + \cdots + a_0) \\ \lambda^{n-1} \\ \vdots \\ \lambda^2 \\ \lambda \end{bmatrix} = \begin{bmatrix} \lambda^n \\ \lambda^{n-1} \\ \vdots \\ \lambda^2 \\ \lambda \end{bmatrix} = \lambda \begin{bmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \vdots \\ \lambda \\ 1 \end{bmatrix}
\end{aligned}$$

Therefore, the given vector is an eigenvector of $C(p)$ corresponding to λ .

(4.3 39)

The characteristic polynomial of A is given as

$$\det(A - \lambda I) = \begin{vmatrix} P - \lambda I & Q \\ O & S - \lambda I \end{vmatrix}$$

By Exercise 4.2 69,

$$\begin{vmatrix} P - \lambda I & Q \\ O & S - \lambda I \end{vmatrix} = |P - \lambda I| |S - \lambda I|$$

which are characteristic polynomials of P and S . Therefore,

$$c_A(\lambda) = \det(A - \lambda I) = \det(P - \lambda I) \det(S - \lambda I) = c_P(\lambda) c_S(\lambda)$$

(4.3 40)

Since the eigenvalues $\lambda_1, \dots, \lambda_n$ are the solutions of the characteristic equation of A ,

$$\det(A - \lambda I) = (-1)^n (\lambda - \lambda_1) \cdots (\lambda - \lambda_n)$$

So substituting $\lambda = 0$ gives $\det A = \lambda_1 \lambda_2 \cdots \lambda_n$.

Next, consider the coefficient of λ^{n-1} on $\det(A - \lambda I)$.

Claim : If $A = [a_{ij}]$ is $n \times n$ matrix, the coefficient of λ^{n-1} on $\det(A - \lambda I)$ is equal to its coefficient on $(a_{11} - \lambda) \cdots (a_{nn} - \lambda)$, which is $(-1)^{n-1} \text{tr}(A)$.

We prove this claim by mathematical induction.

(i) If A is 1×1 matrix, $\det(A - \lambda I) = a_{11} - \lambda$, so the coefficient of λ^0 should be the same.

(ii) Suppose that the claim holds for any $n \times n$ matrices. Let A be a $(n+1) \times (n+1)$ matrix. Then

$$\begin{aligned} \det(A - \lambda I) &= \begin{vmatrix} a_{11} - \lambda & \cdots & a_{1,(n+1)} \\ \vdots & & \vdots \\ a_{(n+1),1} & \cdots & a_{(n+1),(n+1)} - \lambda \end{vmatrix} \\ &= (a_{11} - \lambda) \det(A - \lambda I)_{11} + \sum_{i=2}^{n+1} a_{1i} (-1)^{1+i} \det(A - \lambda I)_{1i} \end{aligned}$$

Since $(A - \lambda I)_{1i}$ is a matrix obtained by removing the first row and i th column from $A - \lambda I$, its determinant cannot have terms of λ^n or terms with higher degrees, if $i \neq 1$. Thus, the coefficient of λ^n on $\det(A - \lambda I)$ is equal to its coefficient on $(a_{11} - \lambda) \det(A - \lambda I)_{11}$.

Note that $(A - \lambda I)_{11} = A_{11} - \lambda I$ and A_{11} is an $n \times n$ matrix, so by inductive hypothesis, the coefficient of λ^{n-1} on $\det(A - \lambda I)_{11}$ is equal to its coefficient on $(a_{22} - \lambda) \cdots (a_{(n+1),(n+1)} - \lambda)$. Therefore,

$$\begin{aligned} (a_{11} - \lambda) \det(A - \lambda I) &= (a_{11} - \lambda) \{ (-1)^n \lambda^n + (-1)^{n-1} (a_{22} + \cdots + a_{(n+1),(n+1)}) \lambda^{n-1} + \cdots \} \\ &= (-1)^{n+1} \lambda^{n+1} + (-1)^n (a_{11} + a_{22} + \cdots + a_{(n+1),(n+1)}) \lambda^n + \cdots \end{aligned}$$

(i) and (ii) gives the proof for the claim.

Since $\det(A - \lambda I) = (-1)^n(\lambda - \lambda_1)(\cdots)(\lambda - \lambda_n)$, the coefficient of λ^{n-1} on both sides of the equation are

$$(-1)^{n-1}\text{tr}(A) = (-1)^{n+1}(\lambda_1 + \cdots + \lambda_n)$$

Therefore, $\text{tr}(A) = \lambda_1 + \cdots + \lambda_n$.

(4.3 41)

Let A and B be $n \times n$ matrix. By Exercise 4.3 40, $\text{tr}(A + B)$ is the sum of all the eigenvalues of $A + B$. Since $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$, the sum of all the eigenvalues of $A + B$ is equal to the sum of all the eigenvalues of A and B . Also, $\det(AB)$ is the product of all the eigenvalues of AB . Since $\det(AB) = (\det A)(\det B)$, the product of all the eigenvalues of AB is equal to the product of all the eigenvalues of A and B .

(4.4 34)

Let A and B be invertible matrices. Since $BA = A^{-1}(AB)A$, $AB \sim BA$.

(4.4 35)

First we prove that $\text{tr}(AB) = \text{tr}(BA)$. (which is at Exericse 3.2 45) If $A = [a_{ij}]$ and $B = [b_{ij}]$ are $n \times n$ matrices,

$$\begin{aligned}\text{tr}(AB) &= \sum_{i=1}^n (AB)_{ii} \\ &= \sum_{i=1}^n \sum_{j=1}^n a_{ij} b_{ji} \\ &= \sum_{j=1}^n \sum_{i=1}^n b_{ji} a_{ij} \\ &= \sum_{j=1}^n (BA)_{jj} = \text{tr}(BA)\end{aligned}$$

If A and B are similar, there exists an invertible matrix P such that $B = P^{-1}AP$, so

$$\text{tr}(B) = \text{tr}(P^{-1}(AP)) = \text{tr}((AP)P^{-1}) = \text{tr}(A)$$

(4.4 40)

If A is similar to B , there exists an invertible matrix P such that $B = P^{-1}AP$. Then

$$B^T = (P^{-1}AP)^T = P^T A^T (P^{-1})^T = ((P^T)^{-1})^{-1} A^T (P^T)^{-1}$$

where $(P^T)^{-1}$ is invertible. Therefore, B^T is similar to A^T .

(4.4 41)

If A is diagonalizable, then there exists a diagonal matrix D which is similar to A . By Exercise 4.4 40, A^T is also similar to $D^T = D$. Therefore, A^T is diagonalizable.

(4.4 42)

If A is diagonalizable, then there exists a diagonal matrix D and an invertible matrix P such that $P^{-1}AP = D$. Then D is invertible and

$$D^{-1} = (P^{-1}AP)^{-1} = P^{-1}A^{-1}(P^{-1})^{-1} = P^{-1}A^{-1}P$$

where D^{-1} is also a diagonal matrix. (Note that the diagonal entries of D are nonzero, since they are eigenvalues of A and they cannot be zero by Theorem 4.16.) Therefore, $A^{-1} \sim D^{-1}$ and A^{-1} is diagonalizable.

(4.4 43)

Let A be a diagonalizable matrix with single eigenvalue λ . Then A is similar to a diagonal matrix with its diagonal entries as eigenvalues of A , which is λI . Since there exists an invertible matrix P such that $\lambda I = P^{-1}AP$,

$$A = P(\lambda I)P^{-1} = \lambda(PIP^{-1}) = \lambda I$$

Therefore, $A = \lambda I$.

(4.4 44)

(\Rightarrow) Suppose that $\mathbf{v}_1, \dots, \mathbf{v}_n$ are eigenvectors of both A and B . Let λ_{Ai} be an eigenvalue of A corresponding to \mathbf{v}_i , and λ_{Bi} be an eigenvalue of B corresponding to \mathbf{v}_i , where $i = 1, 2, \dots, n$. Then for any \mathbf{v}_i ,

$$AB\mathbf{v}_i = A(\lambda_{Bi}\mathbf{v}_i) = \lambda_{Bi}(A\mathbf{v}_i) = (\lambda_{Ai}\lambda_{Bi})\mathbf{v}_i = \lambda_{Ai}(B\mathbf{v}_i) = B(\lambda_{Ai}\mathbf{v}_i) = BA\mathbf{v}_i$$

Thus $(AB - BA)\mathbf{v}_i = \mathbf{0}$ for $i = 1, 2, \dots, n$. By Theorem 4.24, $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent. So for any $\mathbf{x} \in \mathbb{R}^n = \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_n)$, $(AB - BA)\mathbf{x} = \mathbf{0}$, which indicates that $AB - BA = \mathbf{0}$. Therefore, $AB = BA$.

(\Rightarrow 다른 증명) Suppose that $\mathbf{v}_1, \dots, \mathbf{v}_n$ are eigenvalues of both A and B , and let $P = [\mathbf{v}_1 \ \dots \ \mathbf{v}_n]$. Since A and B have n distinct eigenvalues, they are diagonalizable. Thus, there exist diagonal matrices D_1, D_2 such that $P^{-1}AP = D_1$ and $P^{-1}BP = D_2$. (Note that the diagonal entries of D_1 and D_2 will be the corresponding eigenvalues of $\mathbf{v}_1, \dots, \mathbf{v}_n$) Since $D_1D_2 = D_2D_1$,

$$AB = (PD_1P^{-1})(PD_2P^{-1}) = PD_1D_2P^{-1} = PD_2D_1P^{-1} = (PD_2P^{-1})(PD_1P^{-1}) = BA$$

(\Leftarrow) Suppose that $AB = BA$, and \mathbf{v} is an eigenvector of A corresponding to λ . Then

$$AB\mathbf{v} = BA\mathbf{v} = B(\lambda\mathbf{v}) = \lambda(B\mathbf{v})$$

Also, $B\mathbf{v} \neq \mathbf{0}$ since $\mathbf{v} \neq \mathbf{0}$ and 0 is not an eigenvalue of B as B is invertible. Thus, $B\mathbf{v}$ is also an eigenvector of A corresponding to λ . Since A has n distinct eigenvalues, the geometric multiplicity of each eigenvalues should be 1. Therefore, $B\mathbf{v} = k\mathbf{v}$ for some nonzero scalar k , which indicates that \mathbf{v} is also an eigenvector of B . In other words, all eigenvectors of A is also an eigenvector of B . Since A and B have n eigenvectors, we can conclude that they have same eigenvectors.

(4.4 45)

Since A and B are similar, their eigenvalues and characteristic polynomials are the same. Therefore,

their algebraic multiplicities of each eigenvalues should be the same.

(4.4 46)

Let A and B be similar matrices. Then there exists an invertible matrix P such that $B = P^{-1}AP$, so $PB = AP$ and $BP^{-1} = P^{-1}A$. Let \mathbf{u} be an eigenvector of B corresponding to an eigenvalue λ . Then

$$AP\mathbf{u} = PB\mathbf{u} = P(\lambda\mathbf{u}) = \lambda(P\mathbf{u})$$

So $P\mathbf{u}$ is an eigenvector of A corresponding to λ .

Let the eigenspace of A corresponding to λ be $E_\lambda(A)$, and the eigenspace of B corresponding to λ be $E_\lambda(B)$, then $\dim E_\lambda(A) \geq \dim E_\lambda(B)$.

Similarly, let \mathbf{v} be an eigenvector of A corresponding to an eigenvalue λ . Then

$$BP^{-1}\mathbf{v} = P^{-1}A\mathbf{v} = P^{-1}(\lambda\mathbf{v}) = \lambda(P^{-1}\mathbf{v})$$

So $P^{-1}\mathbf{v}$ is an eigenvector of B corresponding to λ . Thus, $\dim E_\lambda(A) \leq \dim E_\lambda(B)$.

Therefore, $\dim E_\lambda(A) = \dim E_\lambda(B)$.

(4.2 55 + 4.3 21 + 4.4 47) : Idempotent Matrices

4.2 55 Suppose that A is an idempotent matrix. Then

$$\det(A^2) = (\det A)^2 = \det A$$

Therefore, $\det A = 0$ or $\det A = 1$.

4.3 21 Let A be an idempotent matrix, and let λ be an eigenvalue of A with corresponding eigenvector \mathbf{v} . Then

$$A\mathbf{v} = A^2\mathbf{v} = A(A\mathbf{v}) = A(\lambda\mathbf{v}) = \lambda(A\mathbf{v})$$

so $(\lambda - 1)A\mathbf{v} = \mathbf{0}$. If $\lambda \neq 1$, $A\mathbf{v} = \mathbf{0} = 0\mathbf{v}$ so $\lambda = 0$. Therefore, $\lambda = 0$ or $\lambda = 1$.

4.4 47 Let A be a diagonalizable matrix such that every eigenvalue of A is 0 or 1. Then there exists a diagonal matrix D which is similar to A , and its diagonal entries should be eigenvalues of A . Thus, $D^2 = D$. There also exists an invertible matrix P such that $P^{-1}AP = D$. Therefore,

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD^2P^{-1} = PDP^{-1} = A$$

so A is an idempotent matrix.

(4.2 56 + 4.3 20 + 4.4 48) : Nilpotent Matrices

4.2 56 Suppose that A is a nilpotent matrix. Then there exists m such that $A^m = O$. Then

$$\det A^m = (\det A)^m = \det O = 0$$

Therefore, $\det A = 0$.

4.3 20 For nilpotent matrix A , there exists m such that $A^m = O$. Suppose that λ is an eigenvalue of A . By Theorem 4.18(a), λ^m should be an eigenvalue of $A^m = O$, so $\lambda^m = 0$. Therefore, $\lambda = 0$.

4.4 48 Let A be a nilpotent matrix which is diagonalizable. Then the eigenvalues of A are all zero. Since A is diagonalizable, there exists a diagonal matrix D whose diagonal entries are eigenvalues of A , so $D = O$. There also exists an invertible matrix P such that $P^{-1}AP = D = O$. Therefore,

$$A = PDP^{-1} = POP^{-1} = O$$

Chapter 5

Orthogonality

5.1 Orthogonality in \mathbb{R}^n

Definition. A set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ 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 $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is an orthogonal set of nonzero vectors in \mathbb{R}^n , then these vectors are linearly independent.

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

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

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

Theorem 5.2

Let $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ 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, \dots, c_k such that

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

are given by

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

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

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

□

Theorem 5.3

Let $\{\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_k\}$ 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

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

and the representation is unique.

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

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

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

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

□

Theorem 5.4

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

Proof. Let $Q = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \dots \quad \mathbf{q}_n]$, then

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

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

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

which holds if and only if $Q^T Q = I_n$.

□

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 $Q^T Q = I_n$, 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 $Q = [\mathbf{q}_1 \quad \mathbf{q}_2 \quad \cdots \quad \mathbf{q}_n]$, 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_1Q_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 $Q^TQ = I_n$,

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

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 $Q_2^{-1} = Q_2^T$.

$$(Q_1Q_2)^{-1} = Q_2^{-1}Q_1^{-1} = Q_2^TQ_1^T = (Q_1Q_2)^T$$

Therefore Q_1Q_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 \mathbf{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 $W = \text{span}(\mathbf{w}_1, \dots, \mathbf{w}_k)$, 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 $\mathbf{w}_1, \dots, \mathbf{w}_k$.

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

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

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

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

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

□

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 $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ 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

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

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

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 $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ 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 $\mathbf{u}_1, \dots, \mathbf{u}_k$, $\mathbf{w} \in W$. Also,

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

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 $\mathbf{w}^\perp = \mathbf{w}_1^\perp$. \square

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 $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ 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). \square

Theorem 5.13

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

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

Proof. Let $\mathcal{B}_W = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ be an orthogonal basis for W and let $\mathcal{B}_{W^\perp} = \{\mathbf{v}_1, \dots, \mathbf{v}_l\}$ 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,

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

\square

Corollary 5.14

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

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

Proof. By Theorem 5.13,

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

\square

5.3 The QR Factorization

Theorem 5.15 : The Gram-Schmidt Process

Let $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ be a basis for a subspace W of \mathbb{R}^n . Define $\mathbf{v}_1, \dots, \mathbf{v}_k$ and W_1, \dots, W_k as following:

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{x}_1 & W_1 &= \text{span}(\mathbf{x}_1) \\ \mathbf{v}_2 &= \mathbf{x}_2 - \left(\frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \right) \mathbf{v}_1 = \mathbf{x}_2 - \text{proj}_{W_1}(\mathbf{x}_2) & W_2 &= \text{span}(\mathbf{x}_1, \mathbf{x}_2) \\ &\vdots & &\vdots \\ \mathbf{v}_k &= \mathbf{x}_k - \left(\frac{\mathbf{x}_k \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \right) \mathbf{v}_1 - \dots - \left(\frac{\mathbf{x}_k \cdot \mathbf{v}_{k-1}}{\mathbf{v}_{k-1} \cdot \mathbf{v}_{k-1}} \right) \mathbf{v}_{k-1} = \mathbf{x}_k - \text{proj}_{W_{k-1}}(\mathbf{x}_k) & W_k &= \text{span}(\mathbf{x}_1, \dots, \mathbf{x}_k) \end{aligned}$$

Then $\{\mathbf{v}_1, \dots, \mathbf{v}_i\}$ 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 $\{\mathbf{v}_1, \dots, \mathbf{v}_i\}$ is an orthogonal basis for $W_i = \text{span}(\mathbf{x}_1, \dots, \mathbf{x}_i)$. Let

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

which is orthogonal to W_i . (Orthogonal Decomposition Theorem 증명 과정에서 보였다.)

Also, $\mathbf{v}_{i+1} \neq \mathbf{0}$, because if $\mathbf{v}_{i+1} = \mathbf{0}$, $\mathbf{x}_{i+1} = \text{proj}_{W_i}(\mathbf{x}_{i+1}) \in W_i = \text{span}(\mathbf{x}_1, \dots, \mathbf{x}_i)$, which is impossible since $\{\mathbf{x}_1, \dots, \mathbf{x}_{i+1}\}$ is a linearly independent set. ($\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$ 가 W 의 기저이므로 선형독립이다.)

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

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 $A = [\mathbf{a}_1 \ \dots \ \mathbf{a}_n]$, and let $\{\mathbf{q}_1, \dots, \mathbf{q}_n\}$ be the orthonormal basis of $\text{col}(A)$ obtained by Gram-Schmidt Process and normalization. By Theorem 5.15, $\{\mathbf{q}_1, \dots, \mathbf{q}_i\}$ forms an orthonormal

basis for $\text{span}(\mathbf{a}_1, \dots, \mathbf{a}_i)$ for $i = 1, \dots, n$. Thus, there exist scalars $r_{1i}, r_{2i}, \dots, r_{ii}$ such that

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

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

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

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

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 $Q^T A Q = D$.

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 $Q^T A Q = D$. Since $Q^{-1} = Q^T$,

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

thus

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

Therefore A is symmetric. \square

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

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

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. \square

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 a real eigenvector \mathbf{v}_1 corresponding to λ_1 . (\mathbf{v}_1 이 항상 real vector가 아니라, λ_1 에 대응되는 eigenvector들 중 real vector가 존재한다는 서술이 정확하다. 예를 들어, I_2 의 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 $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of \mathbb{R}^n . Let $Q_1 = [\mathbf{v}_1 \ \dots \ \mathbf{v}_n]$ then Q_1 is an orthogonal matrix, and let

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

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,

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

where $Q_1 Q_2$ is an orthogonal matrix. Therefore, A is diagonalizable.

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

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.

$$\begin{aligned} \mathbf{u} &= \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\ \mathbf{u} \cdot \mathbf{v} &= \sum_{i=1}^n u_i \overline{v_i} = \overline{\mathbf{v}}^T \mathbf{u} \end{aligned}$$

여기에서는 dot product의 ‘교환법칙’조차 성립하지 않는다. $\mathbf{u} \cdot \mathbf{v} = \overline{\mathbf{v}}^T \mathbf{u}$ 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

5.5 Solution for Exercises from Chapter 5

(5.1 25)

참고) Theorem 3.17에 $P^{-1} = P^T$ 라는 명제이고, 이 풀이의 일부는 이 정리의 증명에 해당한다.

If P is a permutation matrix, then P is a product of elementary matrices which are obtained by interchanging two rows of I . So

$$P = E_1 \cdots E_k$$

where E_1, \dots, E_k are such elementary matrices. Note that for each E_i , $E_i^{-1} = E_i^T$. (Remind Theorem 3.11) Thus,

$$P^{-1} = (E_1 \cdots E_k)^{-1} = E_k^{-1} \cdots E_1^{-1} = E_k^T \cdots E_1^T = (E_1 \cdots E_k)^T = P^T$$

Therefore, P is orthogonal by Theorem 5.5.

Note. P 가 $n \times n$ 행렬일 때 P 의 열들이 $\mathbf{e}_1, \dots, \mathbf{e}_n$ 의 재배열임을 이용하여 증명하는 방법도 있다.

(5.1 26)

Let Q be an orthogonal matrix and let Q' be the matrix obtained by rearranging the rows of Q . Then there exists a permutation matrix P such that $Q' = PQ$. (Remind Theorem 3.10) Since P is orthogonal by Exercise 5.1 25, Q' is also orthogonal by Theorem 5.8(d).

(5.1 27)

Let \mathbf{x} and \mathbf{y} be the vectors in \mathbb{R}^2 . If θ is the angle between \mathbf{x} and \mathbf{y} ,

$$\cos \theta = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|}$$

Let Q be an orthogonal 2×2 matrix. Similarly, if θ' is the angle between $Q\mathbf{x}$ and $Q\mathbf{y}$,

$$\cos \theta' = \frac{(Q\mathbf{x}) \cdot (Q\mathbf{y})}{\|Q\mathbf{x}\| \|Q\mathbf{y}\|} = \frac{\mathbf{x} \cdot \mathbf{y}}{\|\mathbf{x}\| \|\mathbf{y}\|} = \cos \theta$$

by Theorem 5.6(b) and Theorem 5.6(c). Therefore, $\theta = \theta'$.

(5.1 28)

(a) Suppose that $Q = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$ is an orthogonal matrix. Since the columns of Q are orthonormal set of vectors,

$$\left\| \begin{bmatrix} a \\ b \end{bmatrix} \right\| = \sqrt{a^2 + b^2} = 1, \left\| \begin{bmatrix} c \\ d \end{bmatrix} \right\| = \sqrt{c^2 + d^2} = 1, \begin{bmatrix} a \\ b \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} = ac + bd = 0$$

From $ac + bd = 0$,

$$a^2 c^2 = b^2 d^2 = (1 - a^2)(1 - c^2) = a^2 c^2 - a^2 - c^2 + 1 \\ \therefore a^2 + c^2 = 1$$

Since $a^2 + b^2 = 1$ and $c^2 + d^2 = 1$, $b^2 = c^2$ and $a^2 = d^2$. From $ab + cd = 0$, if $c = b$, then $d = -a$ and if $c = -b$, then $d = a$. Therefore,

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} a & b \\ b & -a \end{bmatrix} \text{ or } \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

(b) By (a), every orthogonal 2×2 matrix has the form

$$\begin{bmatrix} a & b \\ b & -a \end{bmatrix} \text{ or } \begin{bmatrix} a & -b \\ b & a \end{bmatrix}$$

where $a^2 + b^2 = 1$. So there exists $\theta \in [0, 2\pi)$ such that $a = \cos \theta$ and $b = \sin \theta$, then

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

(c) Since Q is an orthogonal matrix, for any vector $\mathbf{x} \in \mathbb{R}^2$, $\|Q\mathbf{x}\| = \|\mathbf{x}\|$.

Now, suppose that the 2×2 orthogonal matrix Q has the form $Q = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix}$. Let $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ be a vector in \mathbb{R}^2 , then the new vector $Q\mathbf{x}$ is given as

$$Q\mathbf{x} = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \cos \theta + y \sin \theta \\ x \sin \theta - y \cos \theta \end{bmatrix}$$

Then, if the angle between \mathbf{x} and $\begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix}$ is ϕ ($0 \leq \phi \leq \pi$),

$$\cos \phi = \frac{x \cos(\theta/2) + y \sin(\theta/2)}{\sqrt{x^2 + y^2}}$$

And if the angle between $Q\mathbf{x}$ and $\begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix}$ is ϕ' ($0 \leq \phi' \leq \pi$),

$$\begin{aligned} \cos \phi' &= \frac{(x \cos \theta + y \sin \theta) \cos(\theta/2) + (x \sin \theta - y \cos \theta) \sin(\theta/2)}{\sqrt{(x \cos \theta + y \sin \theta)^2 + (x \sin \theta - y \cos \theta)^2}} \\ &= \frac{x[\cos \theta \cos(\theta/2) + \sin \theta \sin(\theta/2)] + y[\sin \theta \cos(\theta/2) - \cos \theta \sin(\theta/2)]}{\sqrt{x^2 + y^2}} \\ &= \frac{x \cos(\theta/2) + y \sin(\theta/2)}{\sqrt{x^2 + y^2}} = \cos \phi \end{aligned}$$

Therefore, Q corresponds to a reflection to the line $\mathbf{x} = t \begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix}$, $t \in \mathbb{R}$.

(Another proof for reflection) We already have $\|Q\mathbf{x}\| = \|\mathbf{x}\|$. So our proof can be done by proving $Q\mathbf{x} + \mathbf{x}$ is in specific line.

$$\begin{aligned} Q\mathbf{x} + \mathbf{x} &= \begin{bmatrix} x(1 + \cos \theta) + y \sin \theta \\ x \sin \theta + y(1 - \cos \theta) \end{bmatrix}. \text{ Since } \frac{\sin \theta}{1 + \cos \theta} = \frac{1 - \cos \theta}{\sin \theta} = \tan \frac{\theta}{2}, \\ \frac{x \sin \theta}{x(1 + \cos \theta)} &= \frac{y(1 - \cos \theta)}{y \sin \theta} = \frac{x \sin \theta + y(1 - \cos \theta)}{x(1 + \cos \theta) + y \sin \theta} = \tan \frac{\theta}{2}. (\because \text{가비의 리}) \\ \therefore Q\mathbf{x} + \mathbf{x} &\text{ is in the line } \mathbf{x} = t \begin{bmatrix} \cos(\theta/2) \\ \sin(\theta/2) \end{bmatrix}, t \in \mathbb{R}. \end{aligned}$$

Suppose that the 2×2 orthogonal matrix Q has the form $Q = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$. Let $\mathbf{x} = \begin{bmatrix} x \\ y \end{bmatrix}$ be a vector in \mathbb{R}^2 , then the new vector $Q\mathbf{x}$ is given as

$$Q\mathbf{x} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \cos \theta + y \sin \theta \\ -x \sin \theta + y \cos \theta \end{bmatrix}$$

If the angle between \mathbf{x} and $Q\mathbf{x}$ is ϕ ,

$$\begin{aligned}\cos \phi &= \frac{x(x \cos \theta + y \sin \theta) + y(-x \sin \theta + y \cos \theta)}{\sqrt{x^2 + y^2} \sqrt{(x \cos \theta + y \cos \theta)^2 + (-x \sin \theta + y \cos \theta)^2}} \\ &= \frac{\cos \theta (x^2 + y^2)}{x^2 + y^2} = \cos \theta\end{aligned}$$

Therefore, $\theta = \phi$ and Q corresponds to a rotation by θ .

(d) We have shown that if Q is the form $Q = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$, then Q corresponds to a rotation, and if Q is the form $Q = \begin{bmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{bmatrix}$, then Q corresponds to a reflection. Since

$$\begin{vmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{vmatrix} = \cos^2 \theta + \sin^2 \theta = 1 \text{ and } \begin{vmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{vmatrix} = -(\cos^2 \theta + \sin^2 \theta) = -1$$

if $\det Q = 1$ then Q corresponds to a rotation, and if $\det Q = -1$ then Q corresponds to a reflection.

(5.1 33)

Let A and B be orthogonal matrices.

(a) Since $A^{-1} = A^T$ and $B^{-1} = B^T$,

$$A(A^T + B^T)B = AA^TB + AB^TB = IB + AI = A + B$$

(b) By (a), $A(A^T + B^T)B = A + B$ so

$$\begin{aligned}\det(A(A^T + B^T)B) &= \det A \det(A^T + B^T) \det B \\ &= \det A \det(A + B)^T \det B = \det A \det(A + B) \det B = \det(A + B) \\ \therefore \det(A + B)(\det A \det B - 1) &= 0\end{aligned}$$

Note that $\det A = \pm 1$ and $\det B = \pm 1$ since they are orthogonal matrices. Since $\det A + \det B = 0$,

$$\det A \det B = (-\det B) \det B = -(\det B)^2 = -1$$

Therefore, $\det(A + B) = 0$, so $A + B$ is not invertible.

(5.1 34)

$$Q^T = \begin{bmatrix} x_1 & \mathbf{y}^T \\ \mathbf{y} & I^T - (\frac{1}{1-x_1})(\mathbf{y}\mathbf{y}^T)^T \end{bmatrix} = \begin{bmatrix} x_1 & \mathbf{y}^T \\ \mathbf{y} & I - (\frac{1}{1-x_1})\mathbf{y}\mathbf{y}^T \end{bmatrix} = Q$$

Now consider $Q^T Q = Q^2$.

$$\begin{aligned}Q^T Q = Q^2 &= \begin{bmatrix} x_1^2 + \mathbf{y}^T \mathbf{y} & x_1 \mathbf{y}^T + \mathbf{y}^T (I - (\frac{1}{1-x_1})\mathbf{y}\mathbf{y}^T) \\ x_1 \mathbf{y} + (I - (\frac{1}{1-x_1})\mathbf{y}\mathbf{y}^T) \mathbf{y} & \mathbf{y}\mathbf{y}^T + (I - (\frac{1}{1-x_1})\mathbf{y}\mathbf{y}^T)^2 \end{bmatrix} \\ &= \begin{bmatrix} x_1^2 + \mathbf{y}^T \mathbf{y} & (1+x_1)\mathbf{y}^T - (\frac{1}{1-x_1})\mathbf{y}^T \mathbf{y}\mathbf{y}^T \\ (1+x_1)\mathbf{y} - (\frac{1}{1-x_1})\mathbf{y}\mathbf{y}^T \mathbf{y} & I + (1 - \frac{2}{1-x_1})\mathbf{y}\mathbf{y}^T + (\frac{1}{1-x_1})^2 \mathbf{y}\mathbf{y}^T \mathbf{y}\mathbf{y}^T \end{bmatrix}\end{aligned}$$

Since $\mathbf{y}^T \mathbf{y} = x_2^2 + \cdots + x_n^2 = 1 - x_1^2$,

$$\begin{aligned} Q^T Q &= Q^2 = \begin{bmatrix} x_1^2 + 1 - x_1^2 & (1 + x_1)\mathbf{y}^T - (1 + x_1)\mathbf{y}^T \\ (1 + x_1)\mathbf{y} - (1 + x_1)\mathbf{y} & I - (\frac{1+x_1}{1-x_1})\mathbf{y}\mathbf{y}^T + (\frac{1+x_1}{1-x_1})\mathbf{y}\mathbf{y}^T \end{bmatrix} \\ &= \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix} = I \end{aligned}$$

Therefore, Q is orthogonal.

(5.1 35)

Solution 1.

We prove this with mathematical induction on each columns of $A = [a_{ij}]$.

Let $A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_n]$ be an $n \times n$ matrix.

(i) In the first column of A , $a_{21} = \cdots = a_{n1} = 0$ since A is upper triangular.

(ii) Suppose that for $i = 1, 2, \dots, k$ ($k < n$), $a_{ji} = 0$ if $i \neq j$, and $a_{ii} \neq 0$.

Since A is orthogonal, for $i = 1, 2, \dots, k$,

$$\mathbf{a}_i \cdot \mathbf{a}_{k+1} = \sum_{j=1}^n a_{ji} a_{j(k+1)} = a_{ii} a_{i(k+1)} = 0$$

so $a_{1(k+1)} = \cdots = a_{k(k+1)} = 0$. Since A is upper triangular, $a_{(k+2)(k+1)} = \cdots = a_{n(k+1)} = 0$. Since \mathbf{a}_{k+1} is a unit vector, $a_{(k+1)(k+1)} \neq 0$ so the induction proposition holds for $(k+1)$ th column.

(i) and (ii) give that A is a diagonal matrix.

Solution 2.

If λ is an eigenvalue of A with corresponding eigenvector \mathbf{v} , A is orthogonal so

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

thus $|\lambda| = 1$. Since the diagonal entries of A are the eigenvalues of A , the diagonal entries should be either 1 or -1. Since the norm of each columns of A are 1, the non-diagonal entries of A are zero. Therefore, A is diagonal matrix.

Solution 3. Since $A^{-1} = A^T$ and A^{-1} is also an upper triangular matrix (이 부분은 추가 증명이 필요.), A is diagonal matrix.

(5.1 36)

Let A be an $m \times n$ matrix. Since $n > m$, $\text{rank}(A) \leq m < n$ so there exists a nonzero vector $\mathbf{x} \in \mathbb{R}^n$ which is in $\text{null}(A)$. That is, $A\mathbf{x} = \mathbf{0}$ so

$$\|A\mathbf{x}\| = \|\mathbf{0}\| = 0 \neq \|\mathbf{x}\|$$

Therefore, an $m \times n$ matrix which satisfies $\|A\mathbf{x}\| = \|\mathbf{x}\|$ for all $\mathbf{x} \in \mathbb{R}^n$ cannot exist if $m < n$.

(5.1 37)

(a) Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an orthonormal basis for \mathbb{R}^n . Let $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ then there exist scalars c_1, \dots, c_n and d_1, \dots, d_n such that

$$\mathbf{x} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n \text{ and } \mathbf{y} = d_1\mathbf{v}_1 + \dots + d_n\mathbf{v}_n$$

So the dot product of \mathbf{x} and \mathbf{y} is given as

$$\begin{aligned} \mathbf{x} \cdot \mathbf{y} &= (c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n) \cdot (d_1\mathbf{v}_1 + \dots + d_n\mathbf{v}_n) \\ &= \sum_{i=1}^n \sum_{j=1}^n c_i d_j (\mathbf{v}_i \cdot \mathbf{v}_j) \\ &= \sum_{i=1}^n c_i d_i (\mathbf{v}_i \cdot \mathbf{v}_i) = \sum_{i=1}^n ((c_i\mathbf{v}_i) \cdot (d_i\mathbf{v}_i)) = \sum_{i=1}^n c_i d_i \\ &= \sum_{i=1}^n (\mathbf{x} \cdot \mathbf{v}_i)(\mathbf{y} \cdot \mathbf{v}_i) \end{aligned}$$

since $\mathbf{v}_1, \dots, \mathbf{v}_n$ are unit vectors and are orthogonal.

(b) Since $\mathbf{x} \cdot \mathbf{v}_i$ and $\mathbf{y} \cdot \mathbf{v}_i$ are the i th components of the coordinate vector $[\mathbf{x}]_{\mathcal{B}}$ and $[\mathbf{y}]_{\mathcal{B}}$,

$$\mathbf{x} \cdot \mathbf{y} = [\mathbf{x}]_{\mathcal{B}} \cdot [\mathbf{y}]_{\mathcal{B}}$$

(5.2 25)

Counterexample : Let $W = \mathbb{R}^2$ and $\mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Let $\mathbf{w} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{w}' = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ then \mathbf{w} is a vector of \mathbb{R}^2 which is in W and orthogonal to \mathbf{w}' , and $\mathbf{v} = \mathbf{w} + \mathbf{w}'$. However, since $W' = \{\mathbf{0}\}$, $\mathbf{w}' \notin W^\perp$.

(5.2 26)

Let $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be an orthogonal basis for \mathbb{R}^n and let $W = \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$.

(i) For any vector $\mathbf{u} \in W^\perp$, there exist scalars c_1, \dots, c_n such that

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

Since \mathbf{u} is orthogonal to all vectors in W , for $i = 1, 2, \dots, k$,

$$\mathbf{u} \cdot \mathbf{v}_i = (c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n) \cdot \mathbf{v}_i = c_i(\mathbf{v}_i \cdot \mathbf{v}_i) = 0$$

Since $\mathbf{v}_i \neq \mathbf{0}$, $c_1 = c_2 = \dots = c_k = 0$. Therefore, $\mathbf{u} = c_{k+1}\mathbf{v}_{k+1} + \dots + c_n\mathbf{v}_n$ so $\mathbf{u} \in \text{span}(\mathbf{v}_{k+1}, \dots, \mathbf{v}_n)$. Thus, $W^\perp \subset \text{span}(\mathbf{v}_{k+1}, \dots, \mathbf{v}_n)$.

(ii) Since $\mathbf{v}_1, \dots, \mathbf{v}_k$ are orthogonal, $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ forms a basis for W . For any vector $\mathbf{u} \in \text{span}(\mathbf{v}_{k+1}, \dots, \mathbf{v}_n)$, there exist scalars c_{k+1}, \dots, c_n such that

$$\mathbf{u} = c_{k+1}\mathbf{v}_{k+1} + \dots + c_n\mathbf{v}_n$$

so

$$\mathbf{u} \cdot \mathbf{v}_i = (c_{k+1}\mathbf{v}_{k+1} + \cdots + c_n\mathbf{v}_n) \cdot \mathbf{v}_i = 0$$

for $i = 1, 2, \dots, k$. Therefore, by Theorem 5.9(d), $\mathbf{u} \in W^\perp$. Thus, $\text{span}(\mathbf{v}_{k+1}, \dots, \mathbf{v}_n) \subset W^\perp$.

By (i) and (ii), $W^\perp = \text{span}(\mathbf{v}_{k+1}, \dots, \mathbf{v}_n)$.

Note. 6장 개념 이용시, vector space V, W 에 대해 $V \subset W, W \subset V, \dim V = \dim W$ 중 2개가 성립하면 나머지 하나도 성립하므로 (i), (ii) 중 하나만 증명해도 무방하다.

(5.2 27 + 5.2 28) : Projection to a Subspace

Note that by the Orthogonal Decomposition Theorem, $\mathbf{w} = \text{proj}_W(\mathbf{x})$ and $\mathbf{w}^\perp = \text{perp}_W(\mathbf{x})$ are the unique vectors satisfying $\mathbf{w} \in W, \mathbf{w}^\perp \in W^\perp$, and $\mathbf{w} + \mathbf{w}^\perp = \mathbf{x}$.

27 (\Rightarrow) Suppose that $\mathbf{x} \in W$. Then $\mathbf{x} = \mathbf{x} + \mathbf{0}$ is the unique orthogonal decomposition of \mathbf{x} with respect to W , since $\mathbf{x} \in W$ and $\mathbf{0} \in W^\perp$. Therefore, $\mathbf{x} = \text{proj}_W(\mathbf{x})$.

(\Leftarrow) Since $\text{proj}_W(\mathbf{x}) \in W$, \mathbf{x} is in W .

28 (\Rightarrow) Suppose that \mathbf{x} is orthogonal to W , then $\mathbf{x} \in W^\perp$. Then $\mathbf{x} = \mathbf{0} + \mathbf{x}$ is the unique orthogonal decomposition of \mathbf{x} with respect to W , since $\mathbf{0} \in W$ and $\mathbf{x} \in W^\perp$. Therefore, $\text{proj}_W(\mathbf{x}) = \mathbf{0}$.

(\Leftarrow) Since $\text{perp}_W \mathbf{x} \in W^\perp$, $\mathbf{x} = \mathbf{x} - \text{proj}_W(\mathbf{x}) = \text{perp}_W(\mathbf{x})$ is in W^\perp . Therefore, \mathbf{x} is orthogonal to W .

(5.2 29)

For all $\mathbf{x} \in \mathbb{R}^n$, $\text{proj}_W(\mathbf{x}) \in W$ by Exercise 5.2 27. Therefore, $\text{proj}_W(\text{proj}_W(\mathbf{x})) = \text{proj}_W(\mathbf{x})$.

(5.2 30)

Solution 1.

Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be an orthonormal set in \mathbb{R}^n , and let \mathbf{x} be a vector in \mathbb{R}^n . Let $W = \text{span}(S)$, where W is a subspace of \mathbb{R}^n , then S is a basis for W .

(a) By the Orthogonal Decomposition Theorem,

$$\begin{aligned} \|\mathbf{x}\|^2 &= \mathbf{x} \cdot \mathbf{x} \\ &= (\text{proj}_W(\mathbf{x}) + \text{perp}_W(\mathbf{x})) \cdot (\text{proj}_W(\mathbf{x}) + \text{perp}_W(\mathbf{x})) \\ &= \|\text{proj}_W(\mathbf{x})\|^2 + \|\text{perp}_W(\mathbf{x})\|^2 \geq \|\text{proj}_W(\mathbf{x})\|^2 \end{aligned}$$

The definition of $\text{proj}_W(\mathbf{x})$ gives that

$$\begin{aligned}
\|\text{proj}_W(\mathbf{x})\|^2 &= \text{proj}_W(\mathbf{x}) \cdot \text{proj}_W(\mathbf{x}) \\
&= \left(\left(\frac{\mathbf{x} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \right) \mathbf{v}_1 + \cdots + \left(\frac{\mathbf{x} \cdot \mathbf{v}_k}{\mathbf{v}_k \cdot \mathbf{v}_k} \right) \mathbf{v}_k \right) \cdot \left(\left(\frac{\mathbf{x} \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \right) \mathbf{v}_1 + \cdots + \left(\frac{\mathbf{x} \cdot \mathbf{v}_k}{\mathbf{v}_k \cdot \mathbf{v}_k} \right) \mathbf{v}_k \right) \\
&= ((\mathbf{x} \cdot \mathbf{v}_1)\mathbf{v}_1 + \cdots + (\mathbf{x} \cdot \mathbf{v}_k)\mathbf{v}_k) \cdot ((\mathbf{x} \cdot \mathbf{v}_1)\mathbf{v}_1 + \cdots + (\mathbf{x} \cdot \mathbf{v}_k)\mathbf{v}_k) \\
&= |\mathbf{x} \cdot \mathbf{v}_1|^2(\mathbf{v}_1 \cdot \mathbf{v}_1) + \cdots + |\mathbf{x} \cdot \mathbf{v}_k|^2(\mathbf{v}_k \cdot \mathbf{v}_k) \\
&= |\mathbf{x} \cdot \mathbf{v}_1|^2 + \cdots + |\mathbf{x} \cdot \mathbf{v}_k|^2
\end{aligned}$$

Since S is an orthonormal set of vectors. Therefore,

$$\|\mathbf{x}\|^2 \geq |\mathbf{x} \cdot \mathbf{v}_1|^2 + \cdots + |\mathbf{x} \cdot \mathbf{v}_k|^2$$

(b) The proof of (a) gives

$$\|\mathbf{x}\|^2 \geq \|\text{proj}_W(\mathbf{x})\|^2 = |\mathbf{x} \cdot \mathbf{v}_1|^2 + \cdots + |\mathbf{x} \cdot \mathbf{v}_k|^2$$

so the equality holds if and only if $\mathbf{x} = \text{proj}_W(\mathbf{x})$. Therefore, by Exercise 5.2 27, the equality holds if and only if $\mathbf{x} \in W = \text{span}(S)$.

Solution 2.

(a) Using the Gram-Schmidt Process, we can extend S to an orthonormal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of \mathbb{R}^n . Then for any vector $\mathbf{x} \in \mathbb{R}^n$,

$$\begin{aligned}
\|\mathbf{x}\|^2 &= \mathbf{x} \cdot \mathbf{x} \\
&= |\mathbf{x} \cdot \mathbf{v}_1|^2 + \cdots + |\mathbf{x} \cdot \mathbf{v}_n|^2 \\
&\geq |\mathbf{x} \cdot \mathbf{v}_1|^2 + \cdots + |\mathbf{x} \cdot \mathbf{v}_k|^2
\end{aligned}$$

(b) The equality holds when $\mathbf{x} \cdot \mathbf{v}_{k+1} = \cdots = \mathbf{x} \cdot \mathbf{v}_n = 0$. If then, $\mathbf{x} \in \text{span}(S)$.

(5.3 19)

A QR factorization of A is $A = AI$, since A is an orthogonal matrix, and I is an invertible upper triangular matrix.

(5.3 20)

A is invertible.

\Updownarrow (F.T.I.M)

Columns of A are linearly independent.

\Updownarrow (Theorem 5.16)

A can be factorized as $A = QR$, where Q is an orthogonal matrix and R is an invertible upper triangular matrix.

\Updownarrow (F.T.I.M)

R has nonzero diagonal entries, since the eigenvalues of R are not zero.

(5.3 23)

Let A be an $m \times n$ matrix with linearly independent columns. Let Q be an $m \times n$ matrix with orthonormal columns, and R an $n \times n$ upper triangular matrix such that $A = QR$.

Suppose that R is not invertible, then by F.T.I.M, there exists a nonzero vector $\mathbf{x} \in \mathbb{R}^n$ such that $R\mathbf{x} = \mathbf{0}$. Since $A\mathbf{x} = QR\mathbf{x} = \mathbf{0}$, there exists a nontrivial solution of $A\mathbf{x} = \mathbf{0}$, thus the columns of A cannot be linearly independent. Therefore, R is invertible.

(5.3 24)

Let A be an $m \times n$ matrix with linearly independent columns, which can be factorized as $A = QR$ where Q is an $m \times n$ matrix with orthonormal columns of R is an $n \times n$ invertible upper triangular matrix.

Since $A = QR$, the columns of A can be expressed as a linear combination of the columns of Q . (Remind Exercise 3.1 26.) Also, since $Q = AR^{-1}$, the columns of Q can be expressed as a linear combination of the columns of A . Therefore, $\text{col}(A) = \text{col}(Q)$.

(5.4 13)

The Spectral Theorem states that a matrix is orthogonally diagonalizable if and only if it is symmetric. Since A and B are orthogonally diagonalizable, A and B are symmetric.

(a) Since $A + B$ is symmetric, $A + B$ is orthogonally diagonalizable.

(b) For any scalar c , cA is symmetric, so cA is orthogonally diagonalizable.

(c) Since A^2 is symmetric, A^2 is orthogonally diagonalizable.

(5.4 14)

Let A be an orthogonally diagonalizable invertible matrix, then A is symmetric by the Spectral Theorem. Since $(A^{-1})^T = (A^T)^{-1} = A^{-1}$, A^{-1} is also symmetric and A^{-1} is orthogonally diagonaliz-

able.

(5.4 15)

Let A and B be orthogonally diagonalizable matrices, then A and B are symmetric by the Spectral Theorem. Since $AB = BA$,

$$(AB)^T = (BA)^T = A^T B^T = AB$$

AB is also symmetric, so AB is orthogonally diagonalizable.

(5.4 16)

Since A is symmetric, by the Spectral Theorem, A is orthogonally diagonalizable so there exists an orthogonal matrix Q and a diagonal matrix D such that $A = QDQ^T$.

(\Rightarrow) If every eigenvalues of A are nonnegative, the diagonal entries of D are nonnegative so there exists a diagonal matrix D_1 such that $D_1^2 = D$. (We take the square roots of the entries of D for each entries of D_1 .) Let $B = QD_1Q^T$, then

$$A = QDQ^T = QD_1^2Q^T = QD_1Q^T QD_1Q^T = B^2$$

Since B is orthogonally diagonalizable, B is symmetric.

(\Leftarrow) Suppose that there exists a symmetric matrix B such that $A = B^2$. By the Spectral Theorem, B is orthogonally diagonalizable so there exists an orthogonal matrix Q and a diagonal matrix D such that $B = QDQ^T$. Then

$$A = B^2 = (QDQ^T)(QDQ^T) = QD^2Q^T$$

and D^2 is also a diagonal matrix with nonnegative diagonal entries. Since $A \sim D^2$, the eigenvalues of A are the eigenvalues of D^2 , so they are nonnegative.

(5.4 25)

For any vector $\mathbf{v} \in \mathbb{R}^n$,

$$(\mathbf{q}\mathbf{q}^T)\mathbf{v} = \mathbf{q}(\mathbf{q}^T\mathbf{v}) = (\mathbf{q} \cdot \mathbf{v})\mathbf{q} = \left(\frac{\mathbf{q} \cdot \mathbf{v}}{\mathbf{q} \cdot \mathbf{q}} \right) \mathbf{q} = \text{proj}_W(\mathbf{v})$$

(5.4 26)

Let $\{\mathbf{q}_1, \dots, \mathbf{q}_k\}$ be an orthonormal basis of W .

(a) For any vector $\mathbf{v} \in \mathbb{R}^n$,

$$\begin{aligned} P\mathbf{v} &= (\mathbf{q}_1\mathbf{q}_1^T + \dots + \mathbf{q}_k\mathbf{q}_k^T)\mathbf{v} \\ &= \mathbf{q}_1\mathbf{q}_1^T\mathbf{v} + \dots + \mathbf{q}_k\mathbf{q}_k^T\mathbf{v} \\ &= (\mathbf{q}_1 \cdot \mathbf{v})\mathbf{q}_1 + \dots + (\mathbf{q}_k \cdot \mathbf{v})\mathbf{q}_k \\ &= \left(\frac{\mathbf{q}_1 \cdot \mathbf{v}}{\mathbf{q}_1 \cdot \mathbf{q}_1} \right) \mathbf{q}_1 + \dots + \left(\frac{\mathbf{q}_k \cdot \mathbf{v}}{\mathbf{q}_k \cdot \mathbf{q}_k} \right) \mathbf{q}_k = \text{proj}_W(\mathbf{v}) \end{aligned}$$

(b) Since $(\mathbf{q}_i \mathbf{q}_i^T)^T = \mathbf{q}_i \mathbf{q}_i^T$ for $\mathbf{q}_1, \dots, \mathbf{q}_k$, $P = \mathbf{q}_1 \mathbf{q}_1^T + \dots + \mathbf{q}_k \mathbf{q}_k^T$ is symmetric. Also, for any vector $\mathbf{v} \in \mathbb{R}^n$,

$$P^2 \mathbf{v} = \text{proj}_W(\text{proj}_W(\mathbf{v})) = \text{proj}_W(\mathbf{v}) = P \mathbf{v}$$

Since $(P^2 - P) \mathbf{v} = \mathbf{0}$ for any vector $\mathbf{v} \in \mathbb{R}^n$, $P^2 - P = O$.

(c) Let $Q = [\mathbf{q}_1 \ \dots \ \mathbf{q}_k]$, then

$$QQ^T = [\mathbf{q}_1 \ \dots \ \mathbf{q}_k] \begin{bmatrix} \mathbf{q}_1^T \\ \vdots \\ \mathbf{q}_k^T \end{bmatrix} = \mathbf{q}_1 \mathbf{q}_1^T + \dots + \mathbf{q}_k \mathbf{q}_k^T = P$$

Since $P = QQ^T$, $\text{rank}(P) = \text{rank}(QQ^T) \leq \text{rank}(Q) = k$. (Remind Exercise 3.5 60a) Also, for $i = 1, 2, \dots, k$,

$$\begin{aligned} P \mathbf{q}_i &= (\mathbf{q}_1 \mathbf{q}_1^T + \dots + \mathbf{q}_k \mathbf{q}_k^T) \mathbf{q}_i \\ &= (\mathbf{q}_1 \cdot \mathbf{q}_i) \mathbf{q}_1 + \dots + (\mathbf{q}_k \cdot \mathbf{q}_i) \mathbf{q}_k = (\mathbf{q}_i \cdot \mathbf{q}_i) \mathbf{q}_i = \mathbf{q}_i \end{aligned}$$

Thus, $\mathbf{q}_i = P \mathbf{q}_i \in \text{col}(P)$. Since $\mathbf{q}_1, \dots, \mathbf{q}_k$ are linearly independent, $\text{rank}(P) \geq k$.

Therefore, $\text{rank}(P) = k$.

(5.4 27)

We prove the proposition by mathematical induction.

(i) If A is a 1×1 real matrix, then $A = I^T A I$, where A is an upper triangular matrix and I is an orthogonal matrix. Therefore, the proposition holds.

(ii) Suppose that for any real $(n-1) \times (n-1)$ matrix A with real eigenvalues, there exist an $(n-1) \times (n-1)$ orthogonal matrix Q and an $(n-1) \times (n-1)$ upper triangular matrix T such that $Q^T A Q = T$.

Let A be an $n \times n$ real matrix with real eigenvalues. Let λ_1 be an eigenvalue of A , then there exists a real unit eigenvector \mathbf{v}_1 corresponding to λ_1 . Since we can extend \mathbf{v}_1 to an orthonormal basis $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ of \mathbb{R}^n using the Gram-Schmidt Process, let Q_1 be an orthogonal matrix such that

$$Q_1 = [\mathbf{v}_1 \ \dots \ \mathbf{v}_n]$$

Since $\mathbf{v}_1, \dots, \mathbf{v}_n$ are orthonormal,

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

as $\mathbf{v}_1^T(\lambda_1 \mathbf{v}_1) = \lambda_1(\mathbf{v}_1 \cdot \mathbf{v}_1) = \lambda_1$ 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 $A \sim B$, the characteristic polynomial of B is equal to that of A . Since $c_A(\lambda) = c_B(\lambda) = (\lambda - \lambda_1)c_{A_1}(\lambda)$ (Exercise 4.3 39), the eigenvalues of A_1 are also eigenvalues of A , so they are real. Therefore, A_1 is an $(n-1) \times (n-1)$ matrix with real eigenvalues.

By inductive hypothesis, there exist an $(n-1) \times (n-1)$ orthogonal matrix P_2 and an $(n-1) \times (n-1)$ upper triangular matrix T_1 such that $P_2^T A_1 P_2 = T_1$. Let Q_2 be

$$Q_2 = \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P_2 \end{bmatrix}$$

Then

$$\begin{aligned}
(Q_1 Q_2)^T A (Q_1 Q_2) &= Q_2^T Q_1^T A Q_1 Q_2 = Q_2^T B Q_2 \\
&= \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P_2^T \end{bmatrix} \begin{bmatrix} \lambda_1 & ? \\ \mathbf{0} & A_1 \end{bmatrix} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & P_2 \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & ? \\ \mathbf{0} & P_2^T A_1 P_2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & ? \\ \mathbf{0} & T_1 \end{bmatrix}
\end{aligned}$$

which is an upper triangular matrix.

By (i) and (ii), for any size of real matrix A , there exist an orthogonal matrix Q and an upper triangular matrix T such that $Q^T A Q = T$.

(5.4 28)

By Exercise 5.4 27, there exist an orthogonal matrix Q and an upper triangular matrix T such that $Q^T A Q = T$. Note that $\lambda = 0$ is the only eigenvalue of A since A is nilpotent. (Exercise 4.3 20) Since $A \sim T$, $\lambda = 0$ is also the only eigenvalue of T , so the diagonal entries of T should be zero.

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 V 는 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 $n \geq 0$, \mathcal{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. \mathcal{P} , a set of all polynomials, is a vector space.

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

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

- 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. \mathcal{C} , the set of all continuous real valued functions, and \mathcal{D} , the set of all differentiable real valued functions, are vector spaces with the addition and scalar multiplication defined as same as \mathcal{F} . Therefore, \mathcal{C} and \mathcal{D} are subspaces of \mathcal{F} . Note that $\mathcal{P} \subset \mathcal{D} \subset \mathcal{C} \subset \mathcal{F}$. Similarly, $\mathcal{C}_{[a,b]}$, the set of all real valued functions which are continuous on $[a, b]$, and $\mathcal{D}_{[a,b]}$, the set of all real 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. $c \odot \mathbf{u} = \mathbf{0} \Rightarrow c = 0 \text{ or } \mathbf{u} = \mathbf{0}$

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 $c \neq 0$,

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

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. W 가 \mathbb{R}^n 의 부분집합으로 주어질 때, W 의 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 $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ is a set of vectors in a vector space V , then the set of all linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$ is called the **span** of $\mathbf{v}_1, \dots, \mathbf{v}_k$ and denoted by $\text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ 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 S 가 V 의 spanning set이고, $V \subset \text{span}(S)$ 이면 S 가 V 의 spanning set이 아니라는 주장이 있는데, 사실 S 가 V 의 부분집합이어야 span 과 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 $S = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ spans V and $S \subset V$, then S is a spanning set for V .

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

(ii) Let \mathbf{v} be a vector in $\text{span}(S)$. Since \mathbf{v} is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$ and $S \subset V$, $\mathbf{v} \in V$ as V is 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 $\mathbf{v}_1, \dots, \mathbf{v}_k$ be vectors in a vector space V .

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

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

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

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

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

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

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

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

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

6.2 Linear Independence, Basis and Dimension

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

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

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 \mathcal{P} , show that $S = \{1, x, x^2, \dots\}$ 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, $\{x^n, x^{n+1}, \dots, x^m\}$ is also linearly dependent so there exist scalars c_n, \dots, c_m , at least one of which is nonzero, such that

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

However, this implies that $c_n = \dots = c_m = 0$, 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 $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ 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 $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are linearly dependent, there exist scalars c_1, c_2, \dots, c_n , at least one of which is nonzero, such that $c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 + \dots + c_n \mathbf{v}_n = \mathbf{0}$. Without loss of generality, we may assume $c_1 \neq 0$, then

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

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 $\mathbf{v}_2, \mathbf{v}_3, \dots, \mathbf{v}_n$. Then there exist scalars c_2, c_3, \dots, c_n such that $\mathbf{v}_1 = c_2 \mathbf{v}_2 + c_3 \mathbf{v}_3 + \dots + c_n \mathbf{v}_n$. Then

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

Therefore, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ 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. $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ is a standard basis for \mathbb{R}^n .
- b. The set $\mathcal{E} = \{E_{11}, \dots, E_{1n}, \dots, E_{m1}, \dots, E_{mn}\}$ 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 $\{1, x, x^2, x^3, \dots, x^n\}$ is a standard basis for \mathcal{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 $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathcal{B}$ such that there are more than two ways to represent \mathbf{v} as a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$. (V 가 무한 차원일때를 고려) Then there exist scalars c_1, \dots, c_k and d_1, \dots, d_k , at least one of $c_i - d_i \neq 0$, such that

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

Then

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

so the vectors $\mathbf{v}_1, \dots, \mathbf{v}_k$ 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 $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ of \mathcal{B} . Then there exist scalars c_1, \dots, c_k , at least one of which is nonzero, such that

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

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 . □

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

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

Theorem 6.6

Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ 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, \dots, c_n and d_1, \dots, d_n such that

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

a.

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

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 ,

$$c\mathbf{u} = cc_1\mathbf{v}_1 + \dots + cc_n\mathbf{v}_n$$

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

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

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

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

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

Thus

$$c_1\mathbf{u}_1 + \dots + c_k\mathbf{u}_k = 0\mathbf{v}_1 + \dots + 0\mathbf{v}_n = \mathbf{0}$$

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

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

$$c_1\mathbf{u}_1 + \dots + c_k\mathbf{u}_k = \mathbf{0}$$

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 $[\mathbf{u}_1]_{\mathcal{B}}, \dots, [\mathbf{u}_k]_{\mathcal{B}}$ 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 $\mathbf{u}_1, \dots, \mathbf{u}_k$ are linearly independent. \square

Extra Theorem

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

Proof. (Exercise 6.2 33)

(\Rightarrow) Suppose that $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ 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 $\mathbf{u}_1, \dots, \mathbf{u}_k$ spans V , there exist scalars c_1, \dots, 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 $[\mathbf{u}_1]_{\mathcal{B}}, \dots, [\mathbf{u}_k]_{\mathcal{B}}$. Therefore, $\{[\mathbf{u}_1]_{\mathcal{B}}, \dots, [\mathbf{u}_k]_{\mathcal{B}}\}$ is a spanning set for \mathbb{R}^n .

(\Leftarrow) Suppose that $\{[\mathbf{u}_1]_{\mathcal{B}}, \dots, [\mathbf{u}_k]_{\mathcal{B}}\}$ 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, \dots, 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 $\mathbf{u}_1, \dots, \mathbf{u}_k$ span V , hence $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ is a spanning set for V . \square

Theorem 6.8

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

- Any set of more than n vectors in V is linearly dependent in V .
- Any set of fewer than n vectors in V cannot span V .

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

- Let $\mathbf{u}_1, \dots, \mathbf{u}_m$ be vectors in V , where $m > n$. Since $[\mathbf{u}_1]_{\mathcal{B}}, \dots, [\mathbf{u}_m]_{\mathcal{B}}$ are m vectors in \mathbb{R}^n , they are linearly dependent in \mathbb{R}^n . (Remind Theorem 2.8.) Therefore, by Theorem 6.7, $\mathbf{u}_1, \dots, \mathbf{u}_m$ are also linearly dependent in V .
- Let $\mathbf{u}_1, \dots, \mathbf{u}_m$ be vectors in V , where $m < n$. Since $[\mathbf{u}_1]_{\mathcal{B}}, \dots, [\mathbf{u}_m]_{\mathcal{B}}$ are m vectors in \mathbb{R}^n , they cannot span \mathbb{R}^n . Therefore, by the Extra Theorem, $\mathbf{u}_1, \dots, \mathbf{u}_m$ 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, $m = n$. \square

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 $\dim V = n$.

- Any linearly independent set in V contains at most n vectors.
- Any spanning set of V contains at least n vectors.
- Any linearly independent set of exactly n vectors in V is a basis for V .
- Any spanning set for V consisting of exactly n vectors is a basis for V .
- (Plus Theorem)** Any linearly independent set in V can be extended to a basis for V .
- (Minus Theorem)** Any spanning set for V can be reduced to a basis for V .

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

- Theorem 6.8(a)
- Theorem 6.8(b)
- Let $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ 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 $\text{span}(S)$.

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

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

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

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

so \mathbf{v} is in $\text{span}(S)$, 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 $S = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ 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 $\mathbf{v}_n \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_{n-1})$.

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

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

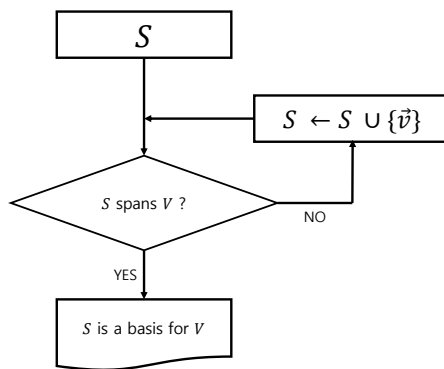
Let \mathbf{x} be a vector in V . Since S is a spanning set for V , there exist scalars c_1, \dots, c_n such that

$$\begin{aligned} \mathbf{x} &= c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n \\ &= c_1\mathbf{v}_1 + \dots + c_{n-1}\mathbf{v}_{n-1} + c_n(d_1\mathbf{v}_1 + \dots + d_{n-1}\mathbf{v}_{n-1}) \\ &= (c_1 + c_nd_1)\mathbf{v}_1 + \dots + (c_{n-1} + c_nd_{n-1})\mathbf{v}_{n-1} \end{aligned}$$

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

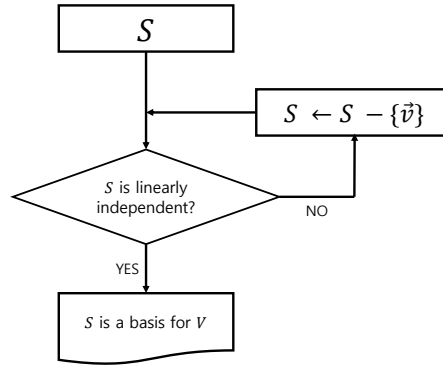
However, this is impossible since $\dim V = n$. Therefore, S is linearly independent, hence S is a basis for V .

- 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 $\text{span}(S)$. 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 .

□

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 $W = V$.

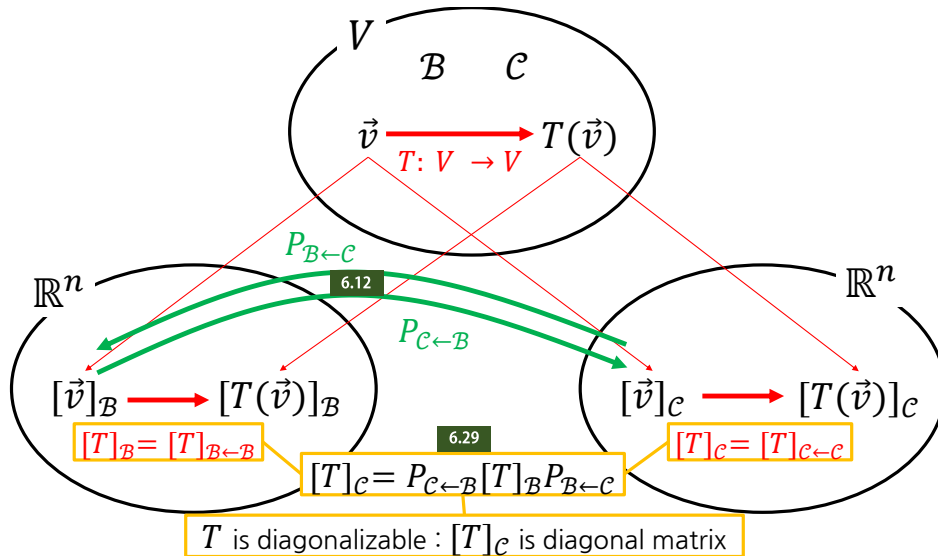
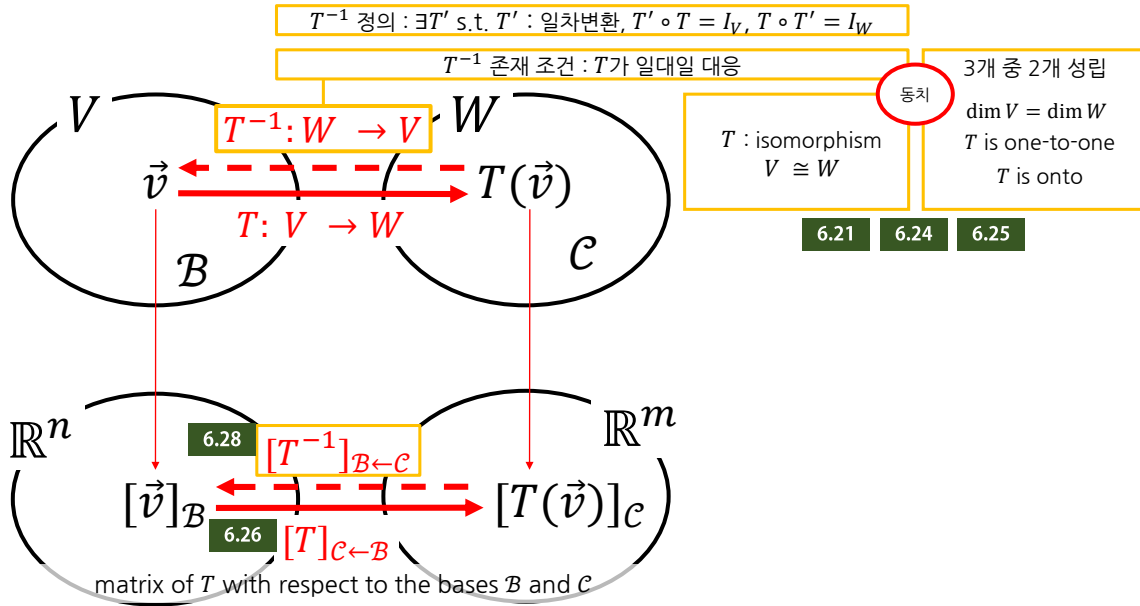
Proof. If W is infinite-dimensional, it contains a infinite set which is linearly independent. However, by Theorem 6.10(a), a linearly independent set in V has at most n vectors. Thus W is finite-dimensional, so let \mathcal{B} be a basis for W .

- 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 $\dim W = \dim V$.

□

6.3 Change of Basis

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



Definition. Let $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be bases for a vector space V . The $n \times n$ matrix whose columns are the coordinate vectors $[\mathbf{u}_1]_{\mathcal{C}}, \dots, [\mathbf{u}_n]_{\mathcal{C}}$ 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 $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ 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 $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$ and $\mathcal{C} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be bases for a vector space V .

- a. Let \mathbf{x} be a vector in V . Then there exist scalars c_1, \dots, 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 $i = 1, 2, \dots, n$,

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

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), $(P_{\mathcal{C} \leftarrow \mathcal{B}})^{-1} = P_{\mathcal{B} \leftarrow \mathcal{C}}$.

□

Theorem 6.13

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

$$[C \mid B] \rightarrow [I \mid P_{\mathcal{C} \leftarrow \mathcal{B}}]$$

Proof. Let scalars p_{ij} such that

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

Then

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

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

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

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

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

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

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

Since $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent, $[\mathbf{v}_1]_{\mathcal{E}}, \dots, [\mathbf{v}_n]_{\mathcal{E}}$ 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

$$[C \mid B] \rightarrow [I \mid P_{\mathcal{C} \leftarrow \mathcal{B}}]$$

□

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 $T : V \rightarrow W$ 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 $T : V \rightarrow W$ 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 $T : V \rightarrow W$ 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 V 의 zero vector와 W 의 zero vector를 잘 구분하자.

Theorem 6.15

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

Proof. Let \mathbf{w} be a vector in $\text{range}(T)$, 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, \dots, c_n such that

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

Thus,

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

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

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

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

Theorem 3.32 : Part 1

Let $T : \mathbb{R}^m \rightarrow \mathbb{R}^n$ and $S : \mathbb{R}^n \rightarrow \mathbb{R}^p$ be linear transformations. Then $S \circ T : \mathbb{R}^m \rightarrow \mathbb{R}^p$ 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, $S \circ T$ is a linear transformation. □

Definition. A linear transformation $T : V \rightarrow W$ is **invertible** if there exists a linear transformation $T' : W \rightarrow V$ 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 $T : V \rightarrow W$ be an invertible linear transformation. Suppose that $T' : W \rightarrow V$ and $T'' : W \rightarrow V$ 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. □

6.5 Kernel and Range

Definition. Let $T : V \rightarrow W$ 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 V 의 vector들의 집합이고, $\text{range}(T)$ 는 W 의 vector들의 집합임을 상기하자.

Theorem 6.18

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

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

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

Let $\mathcal{C} = \{T(\mathbf{v}_{k+1}), \dots, T(\mathbf{v}_n)\}$, 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, \dots, c_n such that

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

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 $\text{range}(T)$.

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

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

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

□

Definition. A linear transformation $T : V \rightarrow W$ 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 $T : V \rightarrow W$ is called **onto** if $\text{range}(T) = W$. 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 $T : V \rightarrow W$ 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 $T : V \rightarrow W$ 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 $\text{nullity}(T) = 0$. By the Rank Theorem, $\text{rank}(T) = \dim(\text{range}(T)) = n$. Since $\text{range}(T)$ is a subspace of W , $\text{range}(T) = W$ by Theorem 6.11(b). Therefore, T is onto.

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

Theorem 6.22

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

Proof. Consider the equation

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

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

Corollary 6.23

Let $\dim V = \dim W = n$. Then a one-to-one linear transformation $T : V \rightarrow W$ 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 . \square

Theorem 6.24

A linear transformation $T : V \rightarrow W$ 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 $T^{-1} : W \rightarrow V$ 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 $T : V \rightarrow W$ 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 $T' : W \rightarrow V$ 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

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

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. \square

Definition. A linear transformation $T : V \rightarrow W$ 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 $T : V \rightarrow W$ which is an isomorphism. Since T is one-to-one, $\ker(T) = \{\mathbf{0}\}$ so nullity(T) = 0, and by Rank Theorem, 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 $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for V and let $\mathcal{C} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ be a basis for W . For every vector \mathbf{v} in V , there exist scalars c_1, \dots, c_n such that

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

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

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

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

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

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

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

Thus T is a linear transformation.

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

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

Since $\mathbf{w}_1, \dots, \mathbf{w}_n$ are linearly independent, $c_1 = \dots = c_n = 0$ 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.

$T : V \rightarrow W$	one-to-one	onto
$\dim V > \dim W$	항상 False	직접 판별
$\dim V = \dim W$	$\ker(T) \stackrel{?}{=} \{\vec{0}\} \begin{cases} \text{둘다 True} \\ \text{둘다 False} \end{cases} (\because \text{Thm 6.21})$	
$\dim V < \dim W$	직접 판별($\ker(T) \stackrel{?}{=} \{\vec{0}\}$)	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 $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$. 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

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

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 $T : V \rightarrow W$ 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 $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$ 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, \dots, 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 $i = 1, 2, \dots, n$,

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

where $A\mathbf{e}_i$ is the i th 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. \square

Extra Theorem

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

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

Proof. Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$, then the i th 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 i th 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 $T : V \rightarrow W$ 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 $T : V \rightarrow V$ 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 $T : \mathbb{R}^n \rightarrow \mathbb{R}^n$ 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 $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_n\}$,

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

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 $T : V \rightarrow V$ 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 $T : V \rightarrow W$ 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

6.7 Solutions for Exercises from Chapter 6

(6.1 1 to 11)

True : **7**

Let $x, y, z \in \mathbb{R}^+$ and c, d be scalars.

- a. (Axiom 1)

$$x \oplus y = xy \in \mathbb{R}^+$$

- b. (Axiom 2)

$$x \oplus y = xy = yx = y \oplus x$$

- c. (Axiom 3)

$$(x \oplus y) \oplus z = xy \oplus z = xyz = x \oplus yz = x \oplus (y \oplus z)$$

- d. (Axiom 4)

$$\forall x \in \mathbb{R}^+, x \oplus 1 = x$$

which implies that $\mathbf{0} = 1$.

- e. (Axiom 5)

$$\forall x \in \mathbb{R}^+, x \oplus \frac{1}{x} = 1 = \mathbf{0}$$

which implies that $-x = \frac{1}{x}$.

- f. (Axiom 6)

$$c \odot x = x^c \in \mathbb{R}^+$$

- g. (Axiom 7)

$$c \odot (x \oplus y) = c \odot xy = (xy)^c = x^c y^c = x^c \oplus y^c = (c \odot x) \oplus (c \odot y)$$

h. (Axiom 8)

$$(c + d) \odot x = x^{c+d} = x^c x^d = x^c \oplus x^d = (c \odot x) \oplus (d \odot x)$$

i. (Axiom 9)

$$c \odot (d \odot x) = c \odot x^d = (x^d)^c = x^{cd} = (cd) \odot x$$

j. (Axiom 10)

$$1 \odot x = x^1 = x$$

Therefore, the given set is a vector space.

False : **6**

Let $\mathbf{u} = \mathbf{v} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \in \mathbb{R}^2$. Then

$$2(\mathbf{u} \oplus \mathbf{v}) = 2\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \oplus \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) = 2\begin{bmatrix} 3 \\ 3 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \end{bmatrix} \neq \begin{bmatrix} 5 \\ 5 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \oplus \begin{bmatrix} 2 \\ 2 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} \oplus 2\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 2\mathbf{u} \oplus 2\mathbf{v}$$

Therefore Axiom 7 fails to hold, hence the given set is not a vector space.

(6.1 24 to 45)

True : **26**

Since $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \in W$, W is a nonempty subset of V . Let $\mathbf{u} = \begin{bmatrix} u \\ 0 \\ u \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} v \\ 0 \\ v \end{bmatrix}$ be vectors in W where $u, v \in \mathbb{R}$, and c a scalar. Then

$$\begin{aligned} \mathbf{u} + \mathbf{v} &= \begin{bmatrix} u \\ 0 \\ u \end{bmatrix} + \begin{bmatrix} v \\ 0 \\ v \end{bmatrix} = \begin{bmatrix} u+v \\ 0 \\ u+v \end{bmatrix} \in W \\ c\mathbf{u} &= c \begin{bmatrix} u \\ 0 \\ u \end{bmatrix} = \begin{bmatrix} cu \\ 0 \\ cu \end{bmatrix} \in W \end{aligned}$$

Therefore, by Theorem 6.2, W is a subspace of V .

False : **27**

Let $\mathbf{w} = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \in W$. However,

$$(-1)\mathbf{w} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \notin W$$

so W is not a subspace of V .

(6.1 46)

Since U and W are subspaces of V , $\mathbf{0} \in U$ and $\mathbf{0} \in W$, so $\mathbf{0} \in U \cap W$. Thus, $U \cap W$ is a nonempty

subset of V .

Let $\mathbf{v}_1, \mathbf{v}_2$ be vectors in $U \cap W$ and c be a scalar. Since U, W are subspaces of V and $\mathbf{v}_1, \mathbf{v}_2 \in U$ and $\mathbf{v}_1, \mathbf{v}_2 \in W$,

$$\begin{aligned}\mathbf{v}_1 + \mathbf{v}_2 &\in U \text{ and } \mathbf{v}_1 + \mathbf{v}_2 \in W \Rightarrow \mathbf{v}_1 + \mathbf{v}_2 \in U \cap W \\ c\mathbf{v}_1 &\in U \text{ and } c\mathbf{v}_1 \in W \Rightarrow c\mathbf{v}_1 \in U \cap W\end{aligned}$$

Therefore, by Theorem 6.2, $U \cap W$ is a subspace of V .

(6.1 47)

Let $V = \mathbb{R}^2$ and let $U = \text{span}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right)$, $W = \text{span}\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right)$ then U and W are subspaces of V . Then $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \in U \cup W$, but $\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \notin U \cup W$. Therefore, $U \cup W$ is not a subspace of V .

(6.1 63) : Uniqueness of the Identity Element

Let V be a vector space. Suppose that there are more than two zero vectors in V , and let $\mathbf{0}$ and $\mathbf{0}'$ be any two of them. By Axiom 2 and Axiom 4,

$$\mathbf{0} = \mathbf{0} + \mathbf{0}' = \mathbf{0}' + \mathbf{0} = \mathbf{0}'$$

Therefore, V has a unique zero vector.

(6.1 64) : Uniqueness of the Inverse Element

Let V be a vector space, and let $\mathbf{0}$ be a zero vector of V . Suppose that for $\mathbf{v} \in V$, there exist two vectors \mathbf{v}_1 and \mathbf{v}_2 such that $\mathbf{v} + \mathbf{v}_1 = \mathbf{v} + \mathbf{v}_2 = \mathbf{0}$. Then

$$\begin{aligned}\mathbf{v}_1 &= \mathbf{v}_1 + \mathbf{0} \text{ (Axiom 4)} \\ &= \mathbf{v}_1 + (\mathbf{v} + \mathbf{v}_2) \\ &= (\mathbf{v} + \mathbf{v}_1) + \mathbf{v}_2 \text{ (Axiom 2 and Axiom 3)} \\ &= \mathbf{0} + \mathbf{v}_2 = \mathbf{v}_2 \text{ (Axiom 4)}\end{aligned}$$

Therefore, there is a unique \mathbf{v}' such that $\mathbf{v} + \mathbf{v}' = \mathbf{0}$.

(6.2 15)

Suppose that f and g are linearly dependent, then one of f and g is a scalar multiple of the other, by Theorem 6.4. Without loss of generality, we may assume that

$$\forall x, f(x) = cg(x) \text{ for some } c \in \mathbb{R}$$

Then

$$\forall x, f'(x) = cg'(x)$$

So the Wronskian of f and g are given by

$$W(x) = \begin{vmatrix} f(x) & g(x) \\ f'(x) & g'(x) \end{vmatrix} = f(x)g'(x) - g(x)f'(x) = cg(x)g'(x) - cg(x)g'(x) = 0$$

Therefore, the contraposition - if $\exists x$ such that $W(x) \neq 0$, then f and g are linearly independent - is also true.

(6.2 17)

(a) Suppose that $\mathbf{u} + \mathbf{v}, \mathbf{v} + \mathbf{w}, \mathbf{w} + \mathbf{u}$ are linearly dependent, then there exist scalars c_1, c_2, c_3 , at least one of which is nonzero, such that

$$c_1(\mathbf{u} + \mathbf{v}) + c_2(\mathbf{v} + \mathbf{w}) + c_3(\mathbf{w} + \mathbf{u}) = \mathbf{0}$$

Then

$$(c_1 + c_3)\mathbf{u} + (c_1 + c_2)\mathbf{v} + (c_2 + c_3)\mathbf{w} = \mathbf{0}$$

Since $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are linearly independent, $c_1 + c_3 = c_1 + c_2 = c_2 + c_3 = 0$. This implies that $c_1 = c_2 = c_3 = 0$, which is a contradiction. Therefore, $\{\mathbf{u} + \mathbf{v}, \mathbf{v} + \mathbf{w}, \mathbf{w} + \mathbf{u}\}$ is linearly independent.

(b) The given set $\{\mathbf{u} - \mathbf{v}, \mathbf{v} - \mathbf{w}, \mathbf{u} - \mathbf{w}\}$ is linearly dependent since

$$(\mathbf{u} - \mathbf{v}) + (\mathbf{v} - \mathbf{w}) - (\mathbf{u} - \mathbf{w}) = \mathbf{0}$$

(6.1 48 + 6.2 42) : Subspace $U + W$

Let V be a vector space, and let U, W be subspaces of V . Then the **sum of U and W** , denoted by $U + W$, is defined by

$$U + W = \{\mathbf{u} + \mathbf{w} | \mathbf{u} \in U \text{ and } \mathbf{w} \in W\}$$

6.1 48a

(i) Let $\begin{bmatrix} x \\ y \\ 0 \end{bmatrix}$ be a vector in $\left\{ \begin{bmatrix} t_1 \\ t_2 \\ 0 \end{bmatrix} \middle| t_1, t_2 \in \mathbb{R} \right\}$. Since $\begin{bmatrix} x \\ y \\ 0 \end{bmatrix} = \begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ y \\ 0 \end{bmatrix}$ and $\begin{bmatrix} x \\ 0 \\ 0 \end{bmatrix} \in U, \begin{bmatrix} 0 \\ y \\ 0 \end{bmatrix} \in W$,

$$\begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \in U + W.$$

(ii) Let $\mathbf{x} \in U + W$, then there exist $\mathbf{u} = \begin{bmatrix} u \\ 0 \\ 0 \end{bmatrix} \in U$ and $\mathbf{w} = \begin{bmatrix} 0 \\ w \\ 0 \end{bmatrix} \in W$ such that $\mathbf{x} = \mathbf{u} + \mathbf{w}$, where

$$u, w \in \mathbb{R}. \text{ Then } \mathbf{x} = \begin{bmatrix} u \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ w \\ 0 \end{bmatrix} = \begin{bmatrix} u \\ w \\ 0 \end{bmatrix} \in \left\{ \begin{bmatrix} t_1 \\ t_2 \\ 0 \end{bmatrix} \middle| t_1, t_2 \in \mathbb{R} \right\}.$$

$$\text{By (i) and (ii), } U + W = \left\{ \begin{bmatrix} t_1 \\ t_2 \\ 0 \end{bmatrix} \middle| t_1, t_2 \in \mathbb{R} \right\}.$$

6.1 48b

Since $\mathbf{0} \in U$ and $\mathbf{0} \in W$, $\mathbf{0} = \mathbf{0} + \mathbf{0} \in U + W$ so $U + W$ is a nonempty subset of V .

Let $\mathbf{x}_1, \mathbf{x}_2$ be vectors in $U + W$, then there exist vectors $\mathbf{u}_1, \mathbf{u}_2 \in U$ and $\mathbf{w}_1, \mathbf{w}_2 \in W$ such that $\mathbf{x}_1 = \mathbf{u}_1 + \mathbf{w}_1$ and $\mathbf{x}_2 = \mathbf{u}_2 + \mathbf{w}_2$. Let c be a scalar, then

$$\begin{aligned} \mathbf{x}_1 + \mathbf{x}_2 &= (\mathbf{u}_1 + \mathbf{w}_1) + (\mathbf{u}_2 + \mathbf{w}_2) = (\mathbf{u}_1 + \mathbf{u}_2) + (\mathbf{w}_1 + \mathbf{w}_2) \\ c\mathbf{x}_1 &= c(\mathbf{u}_1 + \mathbf{w}_1) = c\mathbf{u}_1 + c\mathbf{w}_1 \end{aligned}$$

Since U and W are subspaces of V , $\mathbf{u}_1 + \mathbf{u}_2, c\mathbf{u}_1 \in U$ and $\mathbf{w}_1 + \mathbf{w}_2, c\mathbf{w}_1 \in W$. Therefore, $\mathbf{x}_1 + \mathbf{x}_2 \in U + W$ and $c\mathbf{x}_1 \in U + W$, hence $U + W$ is a subspace of V .

6.2 42

Since $U \cap W$ is a vector space (Exercise 6.1 46), let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a basis for $U \cap W$. Since \mathcal{B} is a linearly independent set in U and W , by Theorem 6.10(e), let $\mathcal{C} = \{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{u}_1, \dots, \mathbf{u}_m\}$ be a basis for U extended from \mathcal{B} , and let $\mathcal{D} = \{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_1, \dots, \mathbf{w}_n\}$ be a basis for W extended from \mathcal{B} .

Let \mathbf{x} be a vector in $U + W$, then there exist vectors $\mathbf{u} \in U$ and $\mathbf{w} \in W$ such that $\mathbf{x} = \mathbf{u} + \mathbf{w}$. Then \mathbf{u} is expressed as a linear combination of the vectors in \mathcal{C} and \mathbf{w} is expressed as a linear combination of the vectors in \mathcal{D} . Thus, \mathbf{x} is a linear combination of the vectors in $\mathcal{C} \cup \mathcal{D} = \{\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{u}_1, \dots, \mathbf{u}_m, \mathbf{w}_1, \dots, \mathbf{w}_n\}$. This implies that $\mathcal{C} \cup \mathcal{D}$ spans $U + W$.

Now, suppose that $\mathcal{C} \cup \mathcal{D}$ is linearly dependent. Then there exist scalars $c_1, \dots, c_k, d_1, \dots, d_m, e_1, \dots, e_n$, at least one of which is nonzero, such that

$$c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k + d_1\mathbf{u}_1 + \dots + d_m\mathbf{u}_m + e_1\mathbf{w}_1 + \dots + e_n\mathbf{w}_n = \mathbf{0}$$

Let $\mathbf{v} = c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k$, $\mathbf{u} = d_1\mathbf{u}_1 + \dots + d_m\mathbf{u}_m$, $\mathbf{w} = e_1\mathbf{w}_1 + \dots + e_n\mathbf{w}_n$, then $\mathbf{v} + \mathbf{u} + \mathbf{w} = \mathbf{0}$. If $\mathbf{u} = \mathbf{0}$, then $\mathbf{v} + \mathbf{w} = \mathbf{0}$, which is a contradiction since \mathcal{D} is a linearly independent set in V . Thus, $\mathbf{u} \neq \mathbf{0}$, and by similar procedure, $\mathbf{w} \neq \mathbf{0}$.

Since $\mathbf{w} = -\mathbf{u} - \mathbf{v}$, \mathbf{w} is in both $\text{span}(\mathcal{D}) = W$ and $\text{span}(\mathcal{C}) = U$, so \mathbf{w} is in $U \cap W$. However, \mathbf{w} is not a linear combination of vectors in \mathcal{B} , since it is already a linear combination of $\mathbf{w}_1, \dots, \mathbf{w}_n$. Therefore, $\mathcal{C} \cup \mathcal{D}$ should be linearly independent.

Therefore, $\mathcal{C} \cup \mathcal{D}$ is a basis for $U + W$. So

$$\dim(U + W) = k + m + n = (k + m) + (k + n) - k = \dim U + \dim W - \dim(U \cap V)$$

(6.1 49 + 6.1 50 + 6.2 43) : Vector Space $U \times W$

문제의 조건이 불완전하다. $U \times W$ 가 vector space임을 보이기 위해서는 상위의 vector space가 주어졌거나, vector addition과 scalar multiplication이 정의되어야 하는데 둘 다 없으므로 원래 증명을 할 수 없네요 ππ

If U and V are vector spaces, the **Cartesian Product** of U and V , denoted by $U \times V$, is defined by

$$U \times V = \{(\mathbf{u}, \mathbf{v}) \mid \mathbf{u} \in U \text{ and } \mathbf{v} \in V\}$$

The addition between two vectors $(\mathbf{u}_1, \mathbf{v}_1)$ and $(\mathbf{u}_2, \mathbf{v}_2)$ in $U \times V$ is defined by

$$(\mathbf{u}_1, \mathbf{v}_1) + (\mathbf{u}_2, \mathbf{v}_2) = (\mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}_1 + \mathbf{v}_2)$$

The scalar multiplication between a vector (\mathbf{u}, \mathbf{v}) in $U \times V$ and a scalar c is defined by

$$c(\mathbf{u}, \mathbf{v}) = (c\mathbf{u}, c\mathbf{v})$$

6.1 49

Let $(\mathbf{u}_1, \mathbf{v}_1), (\mathbf{u}_2, \mathbf{v}_2), (\mathbf{u}_3, \mathbf{v}_3)$ be vectors of $U \times V$ and let c, d be scalars.

a. (Axiom 1) Since U and V are vector spaces, $\mathbf{u}_1 + \mathbf{u}_2 \in U$ and $\mathbf{v}_1 + \mathbf{v}_2 \in V$. So

$$(\mathbf{u}_1, \mathbf{v}_1) + (\mathbf{u}_2, \mathbf{v}_2) = (\mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}_1 + \mathbf{v}_2) \in U \times V$$

b. (Axiom 2)

$$(\mathbf{u}_1, \mathbf{v}_1) + (\mathbf{u}_2, \mathbf{v}_2) = (\mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}_1 + \mathbf{v}_2) = (\mathbf{u}_2 + \mathbf{u}_1, \mathbf{v}_2 + \mathbf{v}_1) = (\mathbf{u}_2, \mathbf{v}_2) + (\mathbf{u}_1, \mathbf{v}_1)$$

c. (Axiom 3)

$$\begin{aligned} ((\mathbf{u}_1, \mathbf{v}_1) + (\mathbf{u}_2, \mathbf{v}_2)) + (\mathbf{u}_3, \mathbf{v}_3) &= (\mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}_1 + \mathbf{v}_2) + (\mathbf{u}_3, \mathbf{v}_3) \\ &= (\mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3, \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3) \\ &= (\mathbf{u}_1, \mathbf{v}_1) + (\mathbf{u}_2 + \mathbf{u}_3, \mathbf{v}_2 + \mathbf{v}_3) \\ &= (\mathbf{u}_1, \mathbf{v}_1) + ((\mathbf{u}_2, \mathbf{v}_2) + (\mathbf{u}_3, \mathbf{v}_3)) \end{aligned}$$

d. (Axiom 4) For any vector $(\mathbf{u}, \mathbf{v}) \in U \times V$,

$$(\mathbf{u}, \mathbf{v}) + (\mathbf{0}, \mathbf{0}) = (\mathbf{u} + \mathbf{0}, \mathbf{v} + \mathbf{0}) = (\mathbf{u}, \mathbf{v})$$

So the zero vector of $U \times V$ is given by $\mathbf{0} = (\mathbf{0}, \mathbf{0})$.

e. (Axiom 5) For any vector $(\mathbf{u}, \mathbf{v}) \in U \times V$,

$$(\mathbf{u}, \mathbf{v}) + (-\mathbf{u}, -\mathbf{v}) = (\mathbf{u} - \mathbf{u}, \mathbf{v} - \mathbf{v}) = (\mathbf{0}, \mathbf{0}) = \mathbf{0}$$

So the inverse element of (\mathbf{u}, \mathbf{v}) is given by $-(\mathbf{u}, \mathbf{v}) = (-\mathbf{u}, -\mathbf{v})$.

f. (Axiom 6) Since U and V are vector spaces, $c\mathbf{u}_1 \in U$ and $c\mathbf{v}_1 \in V$. So

$$c(\mathbf{u}_1, \mathbf{v}_1) = (c\mathbf{u}_1, c\mathbf{v}_1) \in U \times V$$

g. (Axiom 7)

$$\begin{aligned} c((\mathbf{u}_1, \mathbf{v}_1) + (\mathbf{u}_2, \mathbf{v}_2)) &= c(\mathbf{u}_1 + \mathbf{u}_2, \mathbf{v}_1 + \mathbf{v}_2) \\ &= (c\mathbf{u}_1 + c\mathbf{u}_2, c\mathbf{v}_1 + c\mathbf{v}_2) \\ &= (c\mathbf{u}_1, c\mathbf{v}_1) + (c\mathbf{u}_2, c\mathbf{v}_2) = c(\mathbf{u}_1, \mathbf{v}_1) + c(\mathbf{u}_2, \mathbf{v}_2) \end{aligned}$$

h. (Axiom 8)

$$\begin{aligned} (c + d)(\mathbf{u}_1, \mathbf{v}_1) &= ((c + d)\mathbf{u}_1, (c + d)\mathbf{v}_1) \\ &= (c\mathbf{u}_1 + d\mathbf{u}_1, c\mathbf{v}_1 + d\mathbf{v}_1) = (c\mathbf{u}_1, c\mathbf{v}_1) + (d\mathbf{u}_1, d\mathbf{v}_1) = c(\mathbf{u}_1, \mathbf{v}_1) + d(\mathbf{u}_1, \mathbf{v}_1) \end{aligned}$$

i. (Axiom 9)

$$c(d(\mathbf{u}, \mathbf{v})) = c(d\mathbf{u}, d\mathbf{v}) = (cd\mathbf{u}, cd\mathbf{v}) = (cd)(\mathbf{u}, \mathbf{v})$$

j. (Axiom 10)

$$1(\mathbf{u}, \mathbf{v}) = (1\mathbf{u}, 1\mathbf{v}) = (\mathbf{u}, \mathbf{v})$$

Therefore, $U \times V$ is a vector space.

6.1 50

Since $\mathbf{0} \in W$, $(\mathbf{0}, \mathbf{0}) \in \Delta$ so Δ is a nonempty subset of $V \times V$.

Let $(\mathbf{w}_1, \mathbf{w}_1), (\mathbf{w}_2, \mathbf{w}_2)$ be vectors in Δ and c be a scalar. Since W is a subspace of V , $\mathbf{w}_1 + \mathbf{w}_2 \in W$ and $c\mathbf{w}_1 \in W$. Then

$$\begin{aligned}(\mathbf{w}_1, \mathbf{w}_1) + (\mathbf{w}_2, \mathbf{w}_2) &= (\mathbf{w}_1 + \mathbf{w}_2, \mathbf{w}_1 + \mathbf{w}_2) \in \Delta \\ c(\mathbf{w}_1, \mathbf{w}_1) &= (c\mathbf{w}_1, c\mathbf{w}_1) \in \Delta\end{aligned}$$

Therefore, Δ is a subspace of $V \times V$ by Theorem 6.2.

6.2 43a

Let $\mathcal{B} = \{\mathbf{u}_1, \dots, \mathbf{u}_m\}$ be a basis for U and let $\mathcal{C} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for V . Let $\mathcal{D} = \{(\mathbf{u}_1, \mathbf{0}), \dots, (\mathbf{u}_m, \mathbf{0}), (\mathbf{0}, \mathbf{v}_1), \dots, (\mathbf{0}, \mathbf{v}_n)\}$, then \mathcal{D} is a subset of $U \times V$.

Let (\mathbf{u}, \mathbf{v}) be a vector in $U \times V$. Then there exist scalars $c_1, \dots, c_m, d_1, \dots, d_n$ such that

$$\mathbf{u} = c_1\mathbf{u}_1 + \dots + c_m\mathbf{u}_m \text{ and } \mathbf{v} = d_1\mathbf{v}_1 + \dots + d_n\mathbf{v}_n$$

Then

$$\begin{aligned}(\mathbf{u}, \mathbf{v}) &= (c_1\mathbf{u}_1 + \dots + c_m\mathbf{u}_m, d_1\mathbf{v}_1 + \dots + d_n\mathbf{v}_n) \\ &= c_1(\mathbf{u}_1, \mathbf{0}) + \dots + c_m(\mathbf{u}_m, \mathbf{0}) + d_1(\mathbf{0}, \mathbf{v}_1) + \dots + d_n(\mathbf{0}, \mathbf{v}_n)\end{aligned}$$

so $\mathbf{u} \in \text{span}(\mathcal{D})$, hence \mathcal{D} spans $U \times V$.

Consider the scalars $c_1, \dots, c_m, d_1, \dots, d_n$ such that

$$c_1(\mathbf{u}_1, \mathbf{0}) + \dots + c_m(\mathbf{u}_m, \mathbf{0}) + d_1(\mathbf{0}, \mathbf{v}_1) + \dots + d_n(\mathbf{0}, \mathbf{v}_n) = \mathbf{0} = (\mathbf{0}, \mathbf{0})$$

Since

$$c_1(\mathbf{u}_1, \mathbf{0}) + \dots + c_m(\mathbf{u}_m, \mathbf{0}) + d_1(\mathbf{0}, \mathbf{v}_1) + \dots + d_n(\mathbf{0}, \mathbf{v}_n) = (c_1\mathbf{u}_1 + \dots + c_m\mathbf{u}_m, d_1\mathbf{v}_1 + \dots + d_n\mathbf{v}_n)$$

$c_1\mathbf{u}_1 + \dots + c_m\mathbf{u}_m = \mathbf{0}$ and $d_1\mathbf{v}_1 + \dots + d_n\mathbf{v}_n = \mathbf{0}$. Since \mathcal{B} and \mathcal{C} are linearly independent, this implies that $c_1 = \dots = c_m = d_1 = \dots = d_n = 0$. Thus, \mathcal{D} is linearly independent.

Therefore, \mathcal{D} is a basis for $U \times V$, hence $\dim(U \times V) = m + n = \dim U + \dim V$.

6.2 43b

Let $\mathcal{B} = \mathbf{w}_1, \dots, \mathbf{w}_k$ be a basis for W , and let $\mathcal{C} = \{(\mathbf{w}_1, \mathbf{w}_1), \dots, (\mathbf{w}_k, \mathbf{w}_k)\}$, then \mathcal{C} is a subset for Δ .

Let (\mathbf{w}, \mathbf{w}) be vector in Δ , then since $\mathbf{w} \in W$, there exist scalars c_1, \dots, c_k such that

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

Then

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

so $(\mathbf{w}, \mathbf{w}) \in \text{span}(\mathcal{C})$, hence \mathcal{C} spans Δ .

Consider the scalars c_1, \dots, c_k such that

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

Then

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

so $c_1 \mathbf{w}_1 + \cdots + c_k \mathbf{w}_k = \mathbf{0}$. Since \mathcal{B} is linearly independent, this implies that $c_1 = \cdots = c_k = 0$. Thus, \mathcal{C} is linearly independent.

Therefore, \mathcal{C} is a basis for Δ , hence $\dim \Delta = k = \dim W$.

(6.2 44)

Let \mathcal{B} be a finite set of polynomials. Suppose that \mathcal{B} is a basis for \mathcal{P} , and let n be the highest degree of the polynomials in \mathcal{B} . Then x^{n+1} cannot be represented as linear combination of polynomials in \mathcal{B} , so \mathcal{B} cannot span \mathcal{P} , which is a contradiction. Therefore, \mathcal{P} cannot have finite basis, hence \mathcal{P} is infinite-dimensional.

(6.2 57)

Suppose that $c_1 \mathbf{v}_1, \dots, c_n \mathbf{v}_n$ are linearly dependent. Then there exist scalars d_1, \dots, d_n , at least one of which is nonzero, such that

$$d_1(c_1 \mathbf{v}_1) + \cdots + d_n(c_n \mathbf{v}_n) = c_1 d_1 \mathbf{v}_1 + \cdots + c_n d_n \mathbf{v}_n = \mathbf{0}$$

Since $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent, this implies that $c_1 d_1 = c_2 d_2 = \cdots = c_n d_n = 0$. Since c_1, \dots, c_n are nonzero scalars, $d_1 = \cdots = d_n = 0$, which is a contradiction. Thus, $\{c_1 \mathbf{v}_1, \dots, c_n \mathbf{v}_n\}$ is linearly independent. Since $\dim V = n$ and $\{c_1 \mathbf{v}_1, \dots, c_n \mathbf{v}_n\}$ is a linearly independent set in V with n vectors, $\{c_1 \mathbf{v}_1, \dots, c_n \mathbf{v}_n\}$ is a basis for V by Theorem 6.10(c).

(6.2 58)

Consider the scalars c_1, \dots, c_n such that

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

Then

$$\begin{aligned} & c_1 \mathbf{v}_1 + c_2(\mathbf{v}_1 + \mathbf{v}_2) + \cdots + c_n(\mathbf{v}_1 + \cdots + \mathbf{v}_n) \\ &= (c_1 + \cdots + c_n) \mathbf{v}_1 + (c_2 + \cdots + c_n) \mathbf{v}_2 + \cdots + c_n \mathbf{v}_n = \mathbf{0} \end{aligned}$$

Since $\mathbf{v}_1, \dots, \mathbf{v}_n$ are linearly independent, this implies that

$$c_1 + \cdots + c_n = c_2 + \cdots + c_n = \cdots = c_n = 0$$

Thus $c_1 = c_2 = \cdots = c_n = 0$, hence $\{\mathbf{v}_1, \mathbf{v}_1 + \mathbf{v}_2, \cdots, \mathbf{v}_1 + \cdots + \mathbf{v}_n\}$ is linearly independent. Since $\dim V = n$ and $\{\mathbf{v}_1, \mathbf{v}_1 + \mathbf{v}_2, \cdots, \mathbf{v}_1 + \cdots + \mathbf{v}_n\}$ is a linearly independent set in V with n vectors, $\{\mathbf{v}_1, \mathbf{v}_1 + \mathbf{v}_2, \cdots, \mathbf{v}_1 + \cdots + \mathbf{v}_n\}$ is a basis for V by Theorem 6.10(c).

Note. Theorem 6.10을 이용해서 주어진 집합이 일차독립 / spanning set 중 하나만 보여도 basis임을 증명할 수 있다. 57번에서는 귀류법을 사용해서 일차독립임을 보였고, 58번에서는 직접 증명했는데 서술 방식이 다르니 참고.

(6.2 59 + 6.2 60 + 6.2 61 + 6.2 62) : Lagrange Polynomials

Let a_0, \cdots, a_n be distinct real numbers. The **Lagrange Polynomials** associated with a_i , denoted by $p_i(x)$, is defined by

$$p_i(x) = \frac{(x - a_0) \cdots (x - a_{i-1})(x - a_{i+1}) \cdots (x - a_n)}{(a_i - a_0) \cdots (a_i - a_{i-1})(a_i - a_{i+1}) \cdots (a_i - a_n)}$$

59a

$$p_0(x) = \frac{(x - 2)(x - 3)}{(1 - 2)(1 - 3)} = \frac{1}{2}x^2 - \frac{5}{2}x + 3$$

$$p_1(x) = \frac{(x - 1)(x - 3)}{(2 - 1)(2 - 3)} = -x^2 + 4x - 3$$

$$p_2(x) = \frac{(x - 1)(x - 2)}{(3 - 1)(3 - 2)} = \frac{1}{2}x^2 - \frac{3}{2}x + 1$$

59b

If $i \neq j$, $(x - a_j)$ is a factor of $p_i(x)$, so $p_i(a_j) = 0$. If $i = j$,

$$p_i(a_i) = \frac{(a_i - a_0) \cdots (a_i - a_{i-1})(a_i - a_{i+1}) \cdots (a_i - a_n)}{(a_i - a_0) \cdots (a_i - a_{i-1})(a_i - a_{i+1}) \cdots (a_i - a_n)} = 1$$

60a

Let $\mathcal{B} = \{p_0(x), \cdots, p_n(x)\}$. Consider the scalars c_0, \cdots, c_n such that

$$c_0 p_0(x) + \cdots + c_n p_n(x) = \mathbf{0} = 0$$

Then for $i = 0, 1, \cdots, n$,

$$c_0 p_0(a_i) + \cdots + c_n p_n(a_i) = c_i = 0$$

since $p_j(a_i) = 0$ if $i \neq j$, and $p_i(a_i) = 1$. Thus, $c_0 = \cdots = c_n = 0$, which implies that \mathcal{B} is linearly independent.

60b

Since \mathcal{P}_n is an $(n + 1)$ -dimensional vector space and \mathcal{B} is a linearly independent set in \mathcal{P}_n with $n + 1$ vectors, \mathcal{B} is a basis for \mathcal{P}_n by Theorem 6.10(c).

61a

Let $q(x) = c_0 p_0(x) + c_1 p_1(x) + \cdots + c_n p_n(x)$. For $i = 0, 1, \cdots, n$,

$$q(a_i) = c_0 p_0(a_i) + c_1 p_1(a_i) + \cdots + c_n p_n(a_i) = c_i$$

since $p_j(a_i) = 0$ if $i \neq j$, and $p_i(a_i) = 1$. Also, \mathcal{B} is a basis for \mathcal{P}_n , so $q(x) = c_0p_0(x) + c_1p_1(x) + \cdots + c_np_n(x)$ is the only way to represent $q(x)$ as the linear combination of the polynomials in \mathcal{B} by Theorem 6.5.

61b

Let $q_1(x)$ be the polynomial with degree less or equal than n which passes through points $(a_0, c_0), \dots, (a_n, c_n)$. Let $r(x) = q(x) - q_1(x)$, then for $i = 0, 1, \dots, n$, $r(a_i) = q(a_i) - q_1(a_i) = c_i - c_i = 0$. Thus, the equation $r(x) = 0$ has at least $n + 1$ zeros. Since the degree of $r(x)$ is at most n , this implies that $r(x)$ is a zero polynomial, so $q(x) = q_1(x)$. Therefore, $q(x)$ is the unique polynomial with degree less or equal than n which passes through given points.

61c

(i) The Lagrange Polynomials are given by

$$\begin{aligned} p_0(x) &= \frac{(x-2)(x-3)}{(1-2)(1-3)} = \frac{1}{2}x^2 - \frac{5}{2}x + 3 \\ p_1(x) &= \frac{(x-1)(x-3)}{(2-1)(2-3)} = -x^2 + 4x - 3 \\ p_2(x) &= \frac{(x-1)(x-2)}{(3-1)(3-2)} = \frac{1}{2}x^2 - \frac{3}{2}x + 1 \end{aligned}$$

so

$$\begin{aligned} q(x) &= c_0p_0(x) + c_1p_1(x) + c_2p_2(x) \\ &= 6\left(\frac{1}{2}x^2 - \frac{5}{2}x + 3\right) - (-x^2 + 4x - 3) - 2\left(\frac{1}{2}x^2 - \frac{3}{2}x + 1\right) = 3x^2 - 16x + 19 \end{aligned}$$

(ii) The Lagrange Polynomials are given by

$$\begin{aligned} p_0(x) &= \frac{(x-0)(x-3)}{(-1-0)(-1-3)} = \frac{1}{4}x^2 - \frac{3}{4}x \\ p_1(x) &= \frac{(x+1)(x-3)}{(0+1)(0-3)} = -\frac{1}{3}x^2 + \frac{2}{3}x + 1 \\ p_2(x) &= \frac{(x+1)(x-0)}{(3+1)(3-0)} = \frac{1}{12}x^2 + \frac{1}{12}x \end{aligned}$$

so

$$\begin{aligned} q(x) &= c_0p_0(x) + c_1p_1(x) + c_2p_2(x) \\ &= 10\left(\frac{1}{4}x^2 - \frac{3}{4}x\right) + 5\left(-\frac{1}{3}x^2 + \frac{2}{3}x + 1\right) + 2\left(\frac{1}{12}x^2 + \frac{1}{12}x\right) = x^2 - 4x + 5 \end{aligned}$$

62

Let $q(x) \in \mathcal{P}_n$. If a_0, \dots, a_n are zeros of $q(x) = 0$, $q(x)$ passes through $(a_0, 0), \dots, (a_n, 0)$. So the Lagrange interpolation formula is given as

$$q(x) = 0p_0(x) + \cdots + 0p_n(x) = 0$$

Therefore, $q(x)$ is a zero polynomial.

(6.2 63)

F.T.I.M에 의해 역행렬이 존재하는 행렬인 것과 열들이 일차독립인 것이 동치이므로, 일차독립인 열을 n 개 선택하는 가짓수를 구하는 것으로 답을 구할 수 있다. 첫 번째 열부터 $i-1$ 번째 열까지 이미 선택했다고 가정하고, i 번째 열을 선택하는 가짓수를 생각해 보자. i 번째 열은 나머지 열들과 일차독립이어야 하므로, 가능한 전체 가짓수에서 1열 ($i-1$)열의 spanning set의 원소의 갯수를 빼면 된다. 이 spanning set의 원소는 각 열에 곱해주는 scalar가 정해지면 유일하게 결정되므로, 원소의 갯수는 p^{i-1} . 전체 가짓수는 p^n 이므로 i 번째 열을 선택하는 가짓수는 $p^n - p^{i-1}$ 이다. 이 과정을 모든 열에 대해 적용하면, 답은

$$(p^n - p^0)(p^n - p^1)(p^n - p^2) \cdots (p^n - p^{n-1})$$

(6.3 21)

For every vector \mathbf{x} in V ,

$$P_{\mathcal{D} \leftarrow \mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = [\mathbf{x}]_{\mathcal{D}}$$

and

$$P_{\mathcal{D} \leftarrow \mathcal{C}} P_{\mathcal{C} \leftarrow \mathcal{B}}[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{D} \leftarrow \mathcal{C}}[\mathbf{x}]_{\mathcal{C}} = [\mathbf{x}]_{\mathcal{D}}$$

by Theorem 6.12(a). Note that $P_{\mathcal{D} \leftarrow \mathcal{B}}$ is the unique matrix satisfying such property by Theorem 6.12(b). Therefore, $P_{\mathcal{D} \leftarrow \mathcal{C}} P_{\mathcal{C} \leftarrow \mathcal{B}}$ should be equal to $P_{\mathcal{D} \leftarrow \mathcal{B}}$.

(6.3 22)

For $i = 1, 2, \dots, n$, $\mathbf{u}_i = p_{1i}\mathbf{v}_1 + \cdots + p_{ni}\mathbf{v}_n$, so

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

Since P is invertible, the columns of P are linearly independent by F.T.I.M, that is $[\mathbf{u}_1]_{\mathcal{B}}, \dots, [\mathbf{u}_n]_{\mathcal{B}}$ are linearly independent. Thus, by Theorem 6.7, $\mathbf{u}_1, \dots, \mathbf{u}_n$ are also linearly independent. Since \mathcal{C} is a linearly independent set in V with n vectors, \mathcal{C} forms a basis for V by Theorem 6.10(c). Also, since \mathcal{C} is a basis for V , $P = P_{\mathcal{C} \leftarrow \mathcal{B}}$ by its definition.

(6.4 1 to 12)

True : 1

Let $\mathbf{u} = \begin{bmatrix} u_1 & u_2 \\ u_3 & u_4 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix} \in M_{22}$, and c a scalar. Then

$$\begin{aligned} T(\mathbf{u}_1 + \mathbf{u}_2) &= T \begin{bmatrix} u_1 + v_1 & u_2 + v_2 \\ u_3 + v_3 & u_4 + v_4 \end{bmatrix} = \begin{bmatrix} u_1 + u_2 + v_1 + v_2 & 0 \\ 0 & u_3 + u_4 + v_3 + v_4 \end{bmatrix} \\ &= \begin{bmatrix} u_1 + u_2 & 0 \\ 0 & u_3 + u_4 \end{bmatrix} + \begin{bmatrix} v_1 + v_2 & 0 \\ 0 & v_3 + v_4 \end{bmatrix} = T(\mathbf{u}) + T(\mathbf{v}) \\ T(c\mathbf{u}) &= T \begin{bmatrix} cu_1 & cu_2 \\ cu_3 & cu_4 \end{bmatrix} = \begin{bmatrix} cu_1 + cu_2 & 0 \\ 0 & cu_3 + cu_4 \end{bmatrix} \\ &= c \begin{bmatrix} u_1 + u_2 & 0 \\ 0 & u_3 + u_4 \end{bmatrix} = cT(\mathbf{u}) \end{aligned}$$

Therefore, $T : M_{22} \rightarrow M_{22}$ is a linear transformation.

False : **2**

$$T(\mathbf{0}) = T \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \neq \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{0}$$

Therefore, T is not a linear transformation by Theorem 6.14(a).

(6.4 19)

Let $a = T(E_{11})$, $b = T(E_{12})$, $c = T(E_{21})$, and $d = T(E_{22})$, then $a, b, c, d \in \mathbb{R}$. For any vector $\mathbf{v} = \begin{bmatrix} w & x \\ y & z \end{bmatrix} \in M_{22}$,

$$\mathbf{v} = wE_{11} + xE_{12} + yE_{21} + zE_{22}$$

so

$$\begin{aligned} T(\mathbf{v}) &= T(wE_{11} + xE_{12} + yE_{21} + zE_{22}) \\ &= wT(E_{11}) + xT(E_{12}) + yT(E_{21}) + zT(E_{22}) = aw + bx + cy + dz \end{aligned}$$

(6.4 22)

For any vector $\mathbf{v} \in V$, there exist scalars c_1, \dots, c_n such that

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

then

$$\begin{aligned} T(\mathbf{v}) &= T(c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n) \\ &= c_1T(\mathbf{v}_1) + \dots + c_nT(\mathbf{v}_n) \\ &= c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n = \mathbf{v} \end{aligned}$$

so T is the identity transformation on V .

(6.4 24)

(a) Consider the scalars c_1, \dots, c_n such that

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

Then

$$\begin{aligned} \mathbf{0} &= T(\mathbf{0}) = T(c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n) \\ &= c_1T(\mathbf{v}_1) + \dots + c_nT(\mathbf{v}_n) \end{aligned}$$

Since $T(\mathbf{v}_1), \dots, T(\mathbf{v}_n)$ are linearly independent, $c_1 = \dots = c_n = 0$. Therefore, $\mathbf{v}_1, \dots, \mathbf{v}_n$ are also linearly independent.

(b) Let $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ be a transformation defined by

$$T \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix}$$

Let $\mathbf{u} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$, $\mathbf{v} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$ be vectors in \mathbb{R}^2 , and c a scalar. Then

$$\begin{aligned} T(\mathbf{u} + \mathbf{v}) &= T \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ 0 \end{bmatrix} = \begin{bmatrix} x_1 \\ 0 \end{bmatrix} + \begin{bmatrix} x_2 \\ 0 \end{bmatrix} = T(\mathbf{u}) + T(\mathbf{v}) \\ T(c\mathbf{u}) &= T \begin{bmatrix} cx_1 \\ cy_1 \end{bmatrix} = \begin{bmatrix} cx_1 \\ 0 \end{bmatrix} = c \begin{bmatrix} x_1 \\ 0 \end{bmatrix} = cT(\mathbf{u}) \end{aligned}$$

So T is a linear transformation.

Now, $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\}$ is a linearly independent set in \mathbb{R}^2 . However, $\left\{ T \begin{bmatrix} 1 \\ 0 \end{bmatrix}, T \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\} = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$ is linearly dependent in \mathbb{R}^2 . Therefore, the converse of (a) is false.

(6.4 32a)

(\Rightarrow) Let \mathbf{v} be a vector in V . Since \mathbf{v} and $T(\mathbf{v})$ are linearly dependent, there exist scalars c_1, c_2 , at least one of which is nonzero, such that

$$c_1\mathbf{v} + c_2T(\mathbf{v}) = \mathbf{0}$$

Then

$$T(c_1\mathbf{v} + c_2T(\mathbf{v})) = c_1T(\mathbf{v}) + c_2T(T(\mathbf{v})) = c_1T(\mathbf{v}) + c_2\mathbf{v} = T(\mathbf{0}) = \mathbf{0}$$

So we have

$$(c_1\mathbf{v} + c_2T(\mathbf{v})) - (c_1T(\mathbf{v}) + c_2\mathbf{v}) = (c_1 - c_2)(\mathbf{v} - T(\mathbf{v})) = \mathbf{0}$$

thus $c_1 = c_2$ or $\mathbf{v} = T(\mathbf{v})$. If $c_1 = c_2$, $c_1 = c_2 \neq 0$ so $\mathbf{v} + T(\mathbf{v}) = \mathbf{0}$, which gives that $T(\mathbf{v}) = -\mathbf{v}$. Therefore, $T(\mathbf{v}) = \mathbf{v}$ or $T(\mathbf{v}) = -\mathbf{v}$.

(\Leftarrow) If $T(\mathbf{v}) = \mathbf{v}$, $\mathbf{v} - T(\mathbf{v}) = \mathbf{0}$ so \mathbf{v} and $T(\mathbf{v})$ are linearly dependent.

If $T(\mathbf{v}) = -\mathbf{v}$, $\mathbf{v} + T(\mathbf{v}) = \mathbf{0}$ so \mathbf{v} and $T(\mathbf{v})$ are linearly dependent.

(6.4 33)

(\Rightarrow) Let \mathbf{v} be a vector in V . Since \mathbf{v} and $T(\mathbf{v})$ are linearly dependent, there exist scalars c_1, c_2 , at least one of which is nonzero, such that

$$c_1\mathbf{v} + c_2T(\mathbf{v}) = \mathbf{0}$$

Then

$$T(c_1\mathbf{v} + c_2T(\mathbf{v})) = c_1T(\mathbf{v}) + c_2T(T(\mathbf{v})) = c_1T(\mathbf{v}) + c_2T(\mathbf{v}) = T(\mathbf{0}) = \mathbf{0}$$

So we have

$$(c_1\mathbf{v} + c_2T(\mathbf{v})) - (c_1T(\mathbf{v}) + c_2T(\mathbf{v})) = c_1(\mathbf{v} - T(\mathbf{v})) = \mathbf{0}$$

thus $c_1 = 0$ or $\mathbf{v} = T(\mathbf{v})$. If $c_1 = 0$, $c_2T(\mathbf{v}) = \mathbf{0}$ and $c_2 \neq 0$, so $T(\mathbf{v}) = \mathbf{0}$. Therefore, $T(\mathbf{v}) = \mathbf{0}$ or $T(\mathbf{v}) = \mathbf{v}$.

(\Leftarrow) If $T(\mathbf{v}) = \mathbf{v}$, $\mathbf{v} - T(\mathbf{v}) = \mathbf{0}$ so \mathbf{v} and $T(\mathbf{v})$ are linearly dependent.

If $T(\mathbf{v}) = \mathbf{0}$, $0\mathbf{v} + 1T(\mathbf{v}) = \mathbf{0} + \mathbf{0} = \mathbf{0}$ so \mathbf{v} and $T(\mathbf{v})$ are linearly dependent.

(6.4 34 + 6.4 35 + 6.4 36) : Sum and Scalar Multiples of Linear Transformation

For vector spaces V and W , $\mathcal{L}(V, W)$ is the set of all linear transformations from V to W . If $S, T \in \mathcal{L}(V, W)$, the **sum** of S and T , denoted by $S + T$, is defined by

$$(S + T)(\mathbf{v}) = S(\mathbf{v}) + T(\mathbf{v})$$

for any $\mathbf{v} \in V$. If c is a scalar, then the **scalar multiple**, denoted by cT , is defined by

$$(cT)(\mathbf{v}) = cT(\mathbf{v})$$

for any $\mathbf{v} \in V$.

34

Let $\mathbf{u}, \mathbf{v} \in V$, and d a scalar. Then

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

$$\begin{aligned} (S + T)(d\mathbf{u}) &= S(d\mathbf{u}) + T(d\mathbf{u}) = dS(\mathbf{u}) + dT(\mathbf{u}) = d(S(\mathbf{u}) + T(\mathbf{u})) = d(S + T)(\mathbf{u}) \\ (cT)(d\mathbf{u}) &= cT(d\mathbf{u}) = cdT(\mathbf{u}) = d(cT)(\mathbf{u}) \end{aligned}$$

So $S + T$ and cT are linear transformations from V to W , hence $S + T, cT \in \mathcal{L}(V, W)$.

35

Let R, S, T be linear transformations from V to W .

a. (Axiom 1) Since $S + T$ is also a linear transformation from V to W , $S + T \in \mathcal{L}(V, W)$.

b. (Axiom 2) For every vector $\mathbf{v} \in V$,

$$(R + S)(\mathbf{v}) = R(\mathbf{v}) + S(\mathbf{v}) = S(\mathbf{v}) + R(\mathbf{v}) = (S + R)(\mathbf{v})$$

thus $R + S = S + R$.

c. (Axiom 3) For every vector $\mathbf{v} \in V$,

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

thus $(R + S) + T = R + (S + T)$.

d. (Axiom 4) Let $O : V \rightarrow W$ be a transformation defined by

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

for every vector $\mathbf{v} \in V$. Let \mathbf{u}, \mathbf{v} be vectors in V , and c a scalar. Then

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

so O is a linear transformation, hence $O \in \mathcal{L}(V, W)$. Then for any $\mathbf{v} \in V$,

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

thus $T + O = T$.

e. (Axiom 5) For a linear transformation $T : V \rightarrow W$, let $T' : V \rightarrow W$ be a transformation defined by

$$T'(\mathbf{v}) = -T(\mathbf{v})$$

for every vector $\mathbf{v} \in V$. Let \mathbf{u}, \mathbf{v} be vectors in V , and c a scalar. Then

$$\begin{aligned} T'(\mathbf{u} + \mathbf{v}) &= -T(\mathbf{u} + \mathbf{v}) = -(T(\mathbf{u}) + T(\mathbf{v})) = -T(\mathbf{u}) - T(\mathbf{v}) = T'(\mathbf{u}) + T'(\mathbf{v}) \\ T'(c\mathbf{u}) &= -T(c\mathbf{u}) = -cT(\mathbf{u}) = c(-T(\mathbf{u})) = cT'(\mathbf{u}) \end{aligned}$$

so T' is a linear transformation, hence $T' \in \mathcal{L}(V, W)$. Then for any $\mathbf{v} \in V$,

$$(T + T')(\mathbf{v}) = T(\mathbf{v}) + T'(\mathbf{v}) = T(\mathbf{v}) + (-T(\mathbf{v})) = \mathbf{0} = O(\mathbf{v})$$

thus $T + T' = O$.

f. (Axiom 6) Since cT is also a linear transformation from V to W , $cT \in \mathcal{L}(V, W)$.

g. (Axiom 7) For every vector $\mathbf{v} \in V$ and any scalar c ,

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

thus $c(S + T) = cS + cT$.

h. (Axiom 8) For every vector $\mathbf{v} \in V$ and any scalars c, d ,

$$((c + d)T)(\mathbf{v}) = (c + d)T(\mathbf{v}) = cT(\mathbf{v}) + dT(\mathbf{v}) = (cT)(\mathbf{v}) + (dT)(\mathbf{v}) = (cT + dT)(\mathbf{v})$$

thus $(c + d)T = cT + dT$.

i. (Axiom 9) For every vector $\mathbf{v} \in V$ and any scalars c, d ,

$$(c(dT))(\mathbf{v}) = c(dT)(\mathbf{v}) = cdT(\mathbf{v}) = ((cd)T)(\mathbf{v})$$

thus $c(dT) = (cd)T$.

j. (Axiom 10) For every vector $\mathbf{v} \in V$,

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

thus $1T = T$.

Therefore, $\mathcal{L}(V, W)$ is a vector space with given addition and scalar multiplication.

36

Let U , V , and W be vector spaces, let S, T be linear transformations in $\mathcal{L}(U, V)$, and R be a linear transformation in $\mathcal{L}(V, W)$.

(a) For every vector $\mathbf{v} \in U$,

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

thus $R \circ (S + T) = R \circ S + R \circ T$.

(b) For every vector $\mathbf{v} \in U$ and any scalar c ,

$$\begin{aligned} (c(R \circ S))(\mathbf{v}) &= c(R \circ S)(\mathbf{v}) = cR(S(\mathbf{v})) \\ &= (cR)(S(\mathbf{v})) = ((cR) \circ S)(\mathbf{v}) \end{aligned}$$

and since R is a linear transformation, $cR(S(\mathbf{v})) = R(cS(\mathbf{v}))$. Then

$$(c(R \circ S))(\mathbf{v}) = cR(S(\mathbf{v})) = R(cS(\mathbf{v})) = R((cS)(\mathbf{v})) = (R \circ (cS))(\mathbf{v})$$

thus $c(R \circ S) = (cR) \circ S = R \circ (cS)$.

(6.5 5)

Let $\mathbf{v} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_{22}$ be a vector in $\ker(T)$, where $a, b, c, d \in \mathbb{R}$. Then

$$T \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

so $a = d = 0$, hence \mathbf{v} is in $\left\{ \begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix} \mid b, c \in \mathbb{R} \right\}$.

Let $\mathbf{v} = \begin{bmatrix} 0 & x \\ y & 0 \end{bmatrix} \in \left\{ \begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix} \mid b, c \in \mathbb{R} \right\}$ where $x, y \in \mathbb{R}$, then

$$T \begin{bmatrix} 0 & x \\ y & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{0}$$

so \mathbf{v} is in $\ker(T)$. Therefore, $\ker(T) = \left\{ \begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix} \mid b, c \in \mathbb{R} \right\}$.

Let $\mathbf{w} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_{22}$ be a vector in $\text{range}(T)$, where $a, b, c, d \in \mathbb{R}$. Then there exists $\mathbf{v} = \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix} \in M_{22}$, where $v_1, v_2, v_3, v_4 \in \mathbb{R}$, such that $T(\mathbf{v}) = \mathbf{w}$. Then

$$T(\mathbf{v}) = T \begin{bmatrix} v_1 & v_2 \\ v_3 & v_4 \end{bmatrix} = \begin{bmatrix} v_1 & 0 \\ 0 & v_4 \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

so $b = c = 0$, hence \mathbf{w} is in $\left\{ \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \mid a, d \in \mathbb{R} \right\}$.

Let $\mathbf{w} = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \in \left\{ \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \mid a, d \in \mathbb{R} \right\}$, where $x, y \in \mathbb{R}$. then

$$T \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix}$$

so \mathbf{w} is in $\text{range}(T)$. Therefore, $\text{range}(T) = \left\{ \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \mid a, d \in \mathbb{R} \right\}$.

Since $\{E_{12}, E_{21}\}$ forms basis for $\ker(T) = \left\{ \begin{bmatrix} 0 & b \\ c & 0 \end{bmatrix} \mid b, c \in \mathbb{R} \right\}$ and $\{E_{11}, E_{22}\}$ forms basis for $\text{range}(T) = \left\{ \begin{bmatrix} a & 0 \\ 0 & d \end{bmatrix} \mid a, d \in \mathbb{R} \right\}$, $\text{rank}(T) = \text{nullity}(T) = 2$. Therefore, $\text{rank}(T) + \text{nullity}(T) = 4$, which is equal to the dimension of M_{22} .

(6.5 28)

Let $p_1(x)$ and $p_2(x)$ be polynomials in \mathcal{P}_n , and c a scalar. Then

$$\begin{aligned} T(p_1(x) + p_2(x)) &= p_1(x - 2) + p_2(x - 2) = T(p_1(x)) + T(p_2(x)) \\ T(cp_1(x)) &= cp_1(x - 2) = cT(p_1(x)) \end{aligned}$$

thus T is a linear transformation.

(i) Suppose that $T(p_1(x)) = T(p_2(x)) = q(x)$. Then $T(p_1(x)) = p_1(x - 2) = q(x)$ and $T(p_2(x)) = p_2(x - 2) = q(x)$, so $p_1(x - 2) = p_2(x - 2)$ for all x . Thus, substituting $x + 2$ to x gives $p_1(x) = p_2(x)$, which implies that T is one-to-one.

(ii) Let $p(x)$ be a polynomial in \mathcal{P}_n , then since $T(p(x + 2)) = p(x)$, $p(x)$ is in $\text{range}(T)$. Therefore, $\text{range}(T) = \mathcal{P}_n$, hence T is onto.

By (i) and (ii), T is an isomorphism.

(6.5 32)

Let $T : \mathcal{C}[a, b] \rightarrow \mathcal{C}[c, d]$ be a transformation defined by

$$(T(f))(x) = f\left(\frac{d - c}{b - a}(x - a) + c\right)$$

Let $f, g \in \mathcal{C}[a, b]$ and c a scalar. Then

$$\begin{aligned}
(T(f+g))(x) &= (f+g) \left(\frac{d-c}{b-a}(x-a) + c \right) \\
&= f \left(\frac{d-c}{b-a}(x-a) + c \right) + g \left(\frac{d-c}{b-a}(x-a) + c \right) \\
&= (T(f))(x) + (T(g))(x) = (T(f) + T(g))(x) \\
(T(cf))(x) &= (cf) \left(\frac{d-c}{b-a}(x-a) + c \right) \\
&= cf \left(\frac{d-c}{b-a}(x-a) + c \right) = c(T(f))(x) = (cT(f))(x)
\end{aligned}$$

thus $T(f+g) = T(f) + T(g)$ and $T(cf) = cT(f)$. Therefore, T is a linear transformation.

Let $T' : \mathcal{C}[c, d] \rightarrow \mathcal{C}[a, b]$ be a transformation defined by

$$(T'(f))(x) = f \left(\frac{b-a}{d-c}(x-c) + a \right)$$

With similar procedure, T' is shown to be a linear transformation.

Let $f \in \mathcal{C}[a, b]$ and $g \in \mathcal{C}[c, d]$, then

$$\begin{aligned}
((T' \circ T)(f))(x) &= (T'(T(f)))(x) = (T(f)) \left(\frac{b-a}{d-c}(x-c) + a \right) \\
&= f \left(\frac{b-a}{d-c} \left(\frac{d-c}{b-a}(x-a) + c - c \right) + a \right) = f(x) \\
((T \circ T')(g))(x) &= (T(T'(g)))(x) = (T'(g)) \left(\frac{d-c}{b-a}(x-a) + c \right) \\
&= g \left(\frac{d-c}{b-a} \left(\frac{b-a}{d-c}(x-c) + a - a \right) + c \right) = g(x)
\end{aligned}$$

Thus $T' \circ T = I_{\mathcal{C}[a, b]}$ and $T \circ T' = I_{\mathcal{C}[c, d]}$, so T' is an inverse of T . Therefore, T is both one-to-one and onto, hence $\mathcal{C}[a, b] \cong \mathcal{C}[c, d]$.

(6.5 33)

(a) Suppose that $(S \circ T)(\mathbf{u}_1) = (S \circ T)(\mathbf{u}_2)$ for some $\mathbf{u}_1, \mathbf{u}_2 \in U$, then $S(T(\mathbf{u}_1)) = S(T(\mathbf{u}_2))$. Since S is one-to-one, this implies that $T(\mathbf{u}_1) = T(\mathbf{u}_2)$. Since T is also one-to-one, $\mathbf{u}_1 = \mathbf{u}_2$. Therefore, $S \circ T$ is also one-to-one.

(b) Let \mathbf{w} be a vector in W . Since S is onto, there exists $\mathbf{v} \in V$ such that $S(\mathbf{v}) = \mathbf{w}$. Also, since T is onto, there exists $\mathbf{u} \in U$ such that $T(\mathbf{u}) = \mathbf{v}$. Then

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

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

(6.5 34)

(a) Let $\mathbf{u}_1, \mathbf{u}_2 \in U$ such that $T(\mathbf{u}_1) = T(\mathbf{u}_2)$, then $(S \circ T)(\mathbf{u}_1) = S(T(\mathbf{u}_1)) = S(T(\mathbf{u}_2)) = (S \circ T)(\mathbf{u}_2)$. Since $S \circ T$ is one-to-one, this implies that $\mathbf{u}_1 = \mathbf{u}_2$. Therefore, T is one-to-one.

(b) Let \mathbf{w} be a vector in W . Since $S \circ T$ is onto, there exists $\mathbf{u} \in U$ such that $\mathbf{w} = (S \circ T)(\mathbf{u})$. Then $\mathbf{w} = S(T(\mathbf{u}))$, so $\mathbf{w} \in \text{range}(S)$. Therefore, $\text{range}(S) = W$ and S is onto.

(6.5 35)

(a) Let $\mathcal{C} = \{\mathbf{w}_1, \dots, \mathbf{w}_n\}$ be a basis for W . Suppose that T is onto, then there exist vectors $\mathbf{v}_1, \dots, \mathbf{v}_n$ such that $T(\mathbf{v}_i) = \mathbf{w}_i$ for each $i = 1, 2, \dots, n$. Since \mathcal{C} is a linearly independent set in W , $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is linearly independent in V . (Exercise 6.4 24(a)) However, since \mathcal{B} contains $n = \dim W > \dim V$ vectors, it should be linearly dependent by Theorem 6.10(a). It is a contradiction, so T cannot be onto.

(b) Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ be a basis for V . Suppose that T is one-to-one. Since \mathcal{B} is a linearly independent set in V , $\mathcal{C} = \{T(\mathbf{v}_1), \dots, T(\mathbf{v}_m)\}$ is linearly independent in W by Theorem 6.22. However, since \mathcal{C} contains $m = \dim V > \dim W$ vectors, it should be linearly dependent by Theorem 6.10(a). It is a contradiction, so T cannot be one-to-one.

(6.5 36)

Let $p(x), q(x) \in \mathcal{P}_n$ and c a scalar. Then

$$\begin{aligned} T((p+q)(x)) &= \begin{bmatrix} (p+q)(a_0) \\ \vdots \\ (p+q)(a_n) \end{bmatrix} = \begin{bmatrix} p(a_0) \\ \vdots \\ p(a_n) \end{bmatrix} + \begin{bmatrix} q(a_0) \\ \vdots \\ q(a_n) \end{bmatrix} = T(p(x)) + T(q(x)) \\ T((cp)(x)) &= \begin{bmatrix} (cp)(a_0) \\ \vdots \\ (cp)(a_n) \end{bmatrix} = \begin{bmatrix} cp(a_0) \\ \vdots \\ cp(a_n) \end{bmatrix} = c \begin{bmatrix} p(a_0) \\ \vdots \\ p(a_n) \end{bmatrix} = cT(p(x)) \end{aligned}$$

thus T is a linear transformation.

Let $\mathbf{v} = \begin{bmatrix} a_0 \\ \vdots \\ a_n \end{bmatrix}$ be a vector in \mathbb{R}^{n+1} , and let $p(x) = a_0 + a_1x + \dots + a_nx^n$, then $p(x) \in \mathcal{P}_n$. Since $T(p(x)) = \mathbf{v}$, $\mathbf{v} \in \text{range}(T)$. Therefore, $\text{range}(T) = \mathbb{R}^{n+1}$, hence T is onto.

Since $\dim \mathcal{P}_n = \dim \mathbb{R}^{n+1} = n+1$ and T is onto, T is one-to-one by Theorem 6.24. Therefore, T is an isomorphism.

(6.5 37)

Let \mathbf{v} be a vector in $\ker(T)$. Then

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

so \mathbf{v} is also in $\ker(T \circ T)$. Thus $\ker(T) \subset \ker(T \circ T)$. Since $\text{nullity}(T) = \text{nullity}(T \circ T)$ by the Rank Theorem, $\ker(T) = \ker(T \circ T)$ by Theorem 6.11(b).

Let \mathbf{v} be a vector in $\ker(T) \cap \text{range}(T)$. Then there exists $\mathbf{u} \in V$ such that $T(\mathbf{u}) = \mathbf{v}$. Then

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

so \mathbf{u} is in $\ker(T \circ T) = \ker(T)$, which implies that $T(\mathbf{u}) = \mathbf{v} = \mathbf{0}$. Therefore, $\ker(T) \cap \text{range}(T) = \{\mathbf{0}\}$.

(6.5 38)

(a) Let $(\mathbf{u}_1, \mathbf{w}_1), (\mathbf{u}_2, \mathbf{w}_2)$ be vectors in vector space $U \times W$, and c be a scalar. Note that $U \times W$ is a vector space. (Exercise 6.1 49) Then

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

Therefore, T is a linear transformation.

(b) Let $\mathbf{v} \in V$ be a vector in $\text{range}(T)$, then there exist $\mathbf{u} \in U$ and $\mathbf{w} \in W$ such that

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

Thus, $\mathbf{v} \in U + W$, hence $\text{range}(T) \subset U + W$.

Let $\mathbf{v} \in V$ be a vector in $U + W$, then there exist $\mathbf{u} \in U$ and $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{u} + \mathbf{w}$. Then

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

Thus, $\mathbf{v} \in \text{range}(T)$, hence $U + W \subset \text{range}(T)$.

Therefore, $\text{range}(T) = U + W$.

(c) Let $\Delta = \{(\mathbf{w}, \mathbf{w}) | \mathbf{w} \in U \cap W\}$ be a subspace of $U \times W$. Let $(\mathbf{w}, \mathbf{w}) \in \Delta$, then

$$T(\mathbf{w}, \mathbf{w}) = \mathbf{w} - \mathbf{w} = \mathbf{0}$$

so $(\mathbf{w}, \mathbf{w}) \in \ker(T)$, hence $\Delta \subset \ker(T)$. Also, for any $(\mathbf{u}, \mathbf{w}) \in \ker(T)$ where $\mathbf{u} \in U$ and $\mathbf{w} \in W$,

$$T(\mathbf{u}, \mathbf{w}) = \mathbf{u} - \mathbf{w} = \mathbf{0}$$

so $\mathbf{u} = \mathbf{w}$, hence \mathbf{u} and \mathbf{w} are both in $U \cap W$. Thus, $(\mathbf{u}, \mathbf{w}) \in \Delta$ and $\ker(T) \subset \Delta$. Therefore, $\Delta = \ker(T)$. Since $\dim(U \cap W) = \dim \Delta = \dim(\ker(T))$ (Exercise 6.2 43(b)), $\ker(T) \cong U \cap W$.

(d) Note that for vector spaces U and V , $\dim(U \times V) = \dim U + \dim V$. (Exercise 6.2 43(a)) By the Rank Theorem,

$$\text{rank}(T) + \text{nullity}(T) = \dim(U + W) + \dim(U \cap W) = \dim(U \times W) = \dim U + \dim W$$

Therefore, $\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W)$.

(6.6 37)

Let $\mathbf{d}' = \begin{bmatrix} -d_2 \\ d_1 \end{bmatrix}$ be a vector in \mathbb{R}^2 . Since \mathbf{d} and \mathbf{d}' are orthogonal and nonzero, they are linearly independent and hence, $\mathcal{D} = \{\mathbf{d}, \mathbf{d}'\}$ forms a basis for \mathbb{R}^2 .

Let T be a linear transformation of reflection in ℓ . Since $T(\mathbf{d}) = \mathbf{d}$ and $T(\mathbf{d}') = -\mathbf{d}'$,

$$[T(\mathbf{d})]_{\mathcal{D}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, [T(\mathbf{d}')]_{\mathcal{D}} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}, [T]_{\mathcal{D}} = \begin{bmatrix} [T(\mathbf{d})]_{\mathcal{D}} & [T(\mathbf{d}')]_{\mathcal{D}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

And the change-of-basis matrix from \mathcal{D} to the standard basis \mathcal{E} is given by

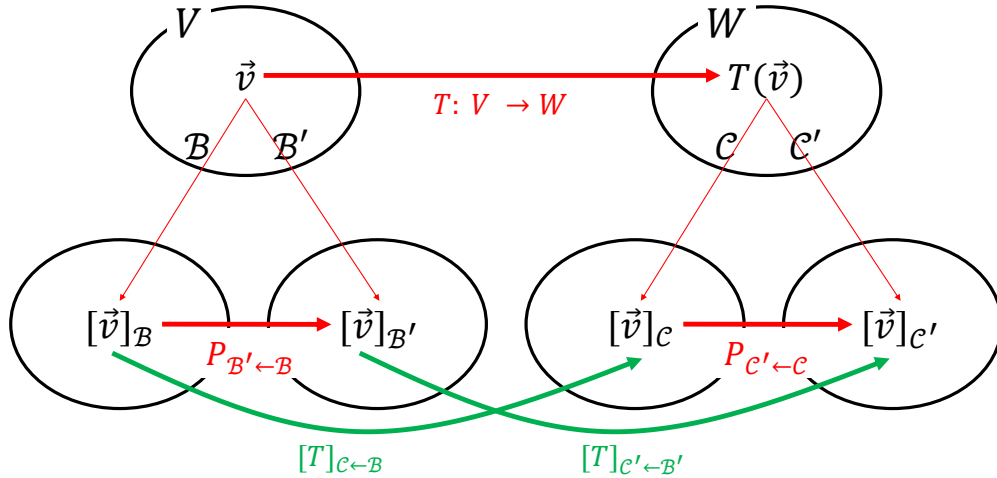
$$P_{\mathcal{E} \leftarrow \mathcal{D}} = \begin{bmatrix} d_1 & -d_2 \\ d_2 & d_1 \end{bmatrix}$$

Therefore, by Theorem 6.29, the standard matrix $[T]_{\mathcal{E}}$ of T is given by

$$\begin{aligned} [T]_{\mathcal{E}} &= P_{\mathcal{E} \leftarrow \mathcal{D}} [T]_{\mathcal{D}} (P_{\mathcal{E} \leftarrow \mathcal{D}})^{-1} \\ &= \begin{bmatrix} d_1 & -d_2 \\ d_2 & d_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \left(\frac{1}{d_1^2 + d_2^2} \begin{bmatrix} d_1 & d_2 \\ -d_2 & d_1 \end{bmatrix} \right) \\ &= \frac{1}{d_1^2 + d_2^2} \begin{bmatrix} d_1^2 - d_2^2 & 2d_1d_2 \\ 2d_1d_2 & d_2^2 - d_1^2 \end{bmatrix} \end{aligned}$$

(6.6 44)

$$\text{직관적 이해} : [T]_{\mathcal{C}' \leftarrow \mathcal{B}'} = P_{\mathcal{C}' \leftarrow \mathcal{C}} [T]_{\mathcal{C} \leftarrow \mathcal{B}} P_{\mathcal{B} \leftarrow \mathcal{B}'} = [T]_{\mathcal{C}' \leftarrow \mathcal{B}'} = P_{\mathcal{C}' \leftarrow \mathcal{C}} [T]_{\mathcal{C} \leftarrow \mathcal{B}} (P_{\mathcal{B}' \leftarrow \mathcal{B}})^{-1}$$



By Theorem 6.12(a) and Theorem 6.12(c), for any vector $\mathbf{v} \in V$ and $\mathbf{w} \in W$,

$$[\mathbf{v}]_{\mathcal{B}} = P_{\mathcal{B} \leftarrow \mathcal{B}'} [\mathbf{v}]_{\mathcal{B}'} = (P_{\mathcal{B}' \leftarrow \mathcal{B}})^{-1} [\mathbf{v}]_{\mathcal{B}'} \text{ and } [\mathbf{w}]_{\mathcal{C}'} = P_{\mathcal{C}' \leftarrow \mathcal{C}} [\mathbf{w}]_{\mathcal{C}}$$

By Theorem 6.26, for every vector $\mathbf{v} \in V$,

$$[T]_{\mathcal{C} \leftarrow \mathcal{B}}[\mathbf{v}]_{\mathcal{B}} = [T(\mathbf{v})]_{\mathcal{C}} \text{ and } [T]_{\mathcal{C}' \leftarrow \mathcal{B}'}[\mathbf{v}]_{\mathcal{B}'} = [T(\mathbf{v})]_{\mathcal{C}'}$$

Therefore, for every vector $\mathbf{v} \in V$,

$$P_{\mathcal{C}' \leftarrow \mathcal{C}}[T]_{\mathcal{C} \leftarrow \mathcal{B}}(P_{\mathcal{B}' \leftarrow \mathcal{B}})^{-1}[\mathbf{v}]_{\mathcal{B}'} = P_{\mathcal{C}' \leftarrow \mathcal{C}}[T]_{\mathcal{C} \leftarrow \mathcal{B}}[\mathbf{v}]_{\mathcal{B}} = P_{\mathcal{C}' \leftarrow \mathcal{C}}[T(\mathbf{v})]_{\mathcal{C}} = [T(\mathbf{v})]_{\mathcal{C}'}$$

Since $[T]_{\mathcal{C}' \leftarrow \mathcal{B}'}$ is the unique matrix satisfying such property (Exercise 6.6 39), this implies that $[T]_{\mathcal{C}' \leftarrow \mathcal{B}'} = P_{\mathcal{C}' \leftarrow \mathcal{C}}[T]_{\mathcal{C} \leftarrow \mathcal{B}}(P_{\mathcal{B}' \leftarrow \mathcal{B}})^{-1}$.

(6.6 45)

Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ be a basis for V , and $\mathcal{C} = \{\mathbf{w}_1, \dots, \mathbf{w}_m\}$ be a basis for W . Let $\Phi : \mathcal{L}(V, W) \rightarrow M_{mn}$ be a transformation defined by

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

for any linear transformation $T : V \rightarrow W$.

(i) For any $S, T \in \mathcal{L}(V, W)$ and any scalar c ,

$$\begin{aligned} \Phi(S + T) &= [S + T]_{\mathcal{C} \leftarrow \mathcal{B}} \\ &= [[(S + T)(\mathbf{v}_1)]_{\mathcal{C}} \cdots [(S + T)(\mathbf{v}_n)]_{\mathcal{C}}] \\ &= [[S(\mathbf{v}_1) + T(\mathbf{v}_1)]_{\mathcal{C}} \cdots [S(\mathbf{v}_n) + T(\mathbf{v}_n)]_{\mathcal{C}}] \\ &= [[S(\mathbf{v}_1)]_{\mathcal{C}} \cdots [S(\mathbf{v}_n)]_{\mathcal{C}}] + [[T(\mathbf{v}_1)]_{\mathcal{C}} \cdots [T(\mathbf{v}_n)]_{\mathcal{C}}] \\ &= [S]_{\mathcal{C} \leftarrow \mathcal{B}} + [T]_{\mathcal{C} \leftarrow \mathcal{B}} = \Phi(S) + \Phi(T) \\ \Phi(cT) &= [cT]_{\mathcal{C} \leftarrow \mathcal{B}} \\ &= [[(cT)(\mathbf{v}_1)]_{\mathcal{C}} \cdots [(cT)(\mathbf{v}_n)]_{\mathcal{C}}] \\ &= [[cT(\mathbf{v}_1)]_{\mathcal{C}} \cdots [cT(\mathbf{v}_n)]_{\mathcal{C}}] \\ &= c [[T(\mathbf{v}_1)]_{\mathcal{C}} \cdots [T(\mathbf{v}_n)]_{\mathcal{C}}] = c[T]_{\mathcal{C} \leftarrow \mathcal{B}} = c\Phi(T) \end{aligned}$$

thus Φ is a linear transformation.

(ii) Let $T_1, T_2 \in \mathcal{L}(V, W)$ such that $\Phi(T_1) = \Phi(T_2)$, then $[T_1]_{\mathcal{C} \leftarrow \mathcal{B}} = [T_2]_{\mathcal{C} \leftarrow \mathcal{B}}$. Then for each $i = 1, 2, \dots, n$, $[T_1(\mathbf{v}_i)]_{\mathcal{C}}$ and $[T_2(\mathbf{v}_i)]_{\mathcal{C}}$ are the same, so $T_1(\mathbf{v}_i) = T_2(\mathbf{v}_i)$. Let $\mathbf{v} \in V$, then there exist scalars c_1, \dots, c_n such that $\mathbf{v} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n$. Then

$$\begin{aligned} T_1(\mathbf{v}) &= T_1(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) \\ &= c_1T_1(\mathbf{v}_1) + \cdots + c_nT_1(\mathbf{v}_n) = c_1T_2(\mathbf{v}_1) + \cdots + c_nT_2(\mathbf{v}_n) \\ &= T_2(c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n) = T_2(\mathbf{v}) \end{aligned}$$

Therefore, $T_1 = T_2$, hence Φ is one-to-one.

(iii) Let A be a $m \times n$ matrix. Let $T : V \rightarrow W$ be a transformation, which is defined by mapping

$$T(\mathbf{v}_i) = A_{1i}\mathbf{w}_1 + \cdots + A_{mi}\mathbf{w}_m$$

for each $i = 1, 2, \dots, n$, and for any vector $\mathbf{v} \in V$, there exist scalars c_1, \dots, c_n such that $\mathbf{v} = c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n$, then $T(\mathbf{v})$ is defined as

$$T(\mathbf{v}) = c_1T(\mathbf{v}_1) + \dots + c_nT(\mathbf{v}_n)$$

It is trivial from the definition that T is a linear transformation. Also, the definition of T also gives that $[T]_{\mathcal{C} \leftarrow \mathcal{B}} = A$, so $\Phi(T) = A$. Therefore, Φ is onto.

By (i), (ii), and (iii), Φ is an isomorphism, thus $\mathcal{L}(V, W) \cong M_{mn}$.

(6.6 46) Let $n = \dim V$, then $V^* = \mathcal{L}(V, W) \cong M_{n1}$ by Exercise 6.6 45. Therefore, $\dim V^* = \dim M_{n1} = n = \dim V$, hence $V^* \cong V$.