Linear Algebra II

渕野 昌 (Sakaé Fuchino)

June 11, 2015

The following is (a draft of) the lecture note of “Linear Algebra II”. The most recent version of this text is downloadable as:

`http://kurt.scitec.kobe-u.ac.jp/~fuchino/kobe/lin-alg2-ws13-14LN.pdf`

The present text is

June 11, 2015 (21:10JST) version.

The text will be improved and extended over and again during (and possibly even after) the Fall Semester 2013/2014. Any comments or suggestions including notices of the simplest slips are appreciated.

Other materials connected with the lecture will be posted at:

`http://kurt.scitec.kobe-u.ac.jp/~fuchino/kobe/index.html`

Contents

1 Determinant

determinant

A mapping assigning to $n \times n$ (square) matrix $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$ a scalar $\det(A)$ [1] is said to be the determinant, if this assignment satisfies the following properties for any $n \times n$-matrix $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$:

(1.1) (Multilinearity) For any $1 \leq i \leq n$, det-1

(a) $\det[\mathbf{a}_1 \cdots \mathbf{a}_i + \mathbf{b}_i \cdots \mathbf{a}_n] = \det[\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_n] + \det[\mathbf{a}_1 \cdots \mathbf{b}_i \cdots \mathbf{a}_n]$ for any n-dimensional vector $\mathbf{b}_i$;

(b) $\det[\mathbf{a}_1 \cdots c\mathbf{a}_i \cdots \mathbf{a}_n] = c \det[\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_n]$ for any scalar c;

(1.2) (Alternation) If $\mathbf{a}_i = \mathbf{a}_j$ for some $i \neq j$ then $\det[\mathbf{a}_1 \cdots \mathbf{a}_n] = 0$; det-2

(1.3) (Unit) $\det(E_n) = 1$ where E_n denotes the $n \times n$ unit matrix. det-3

Later we show that the mapping det : $n \times n$-*matrices* $\rightarrow$ *scalars* with the properties (1.1) – (1.3) actually exists and is determined uniquely by these properties.

We also say that $\det(A)$ (or $|A|$) is the determinant of the matrix A.

Form the properties (1.1) – (1.3), we can also drive the following properties.

Recall that vectors n-dimensional vectors $\mathbf{a}_1,\ldots,\mathbf{a}_k$ are said to be linearly dependent there are scalars $c_1,\ldots,c_k$ which are not all equal to zero such that $c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k = \mathbf{0}$. If $\mathbf{a}_1,\ldots,\mathbf{a}_k$ are not linearly dependent they are said to be linearly independent.

Proposition 1.1 For $1 \leq i, j \leq n$, $i \neq j$ and a scalar c we have P-1

(1.4) $\det[\mathbf{a}_1 \cdots \mathbf{a}_j \cdots \mathbf{a}_i \cdots \mathbf{a}_n] = -\det[\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_j \cdots \mathbf{a}_n]$ (here we assume additionally $i < j$); det-4

(1.5) $\det[\mathbf{a}_1 \cdots \underbrace{\mathbf{a}_i + c\mathbf{a}_j}_{i\text{'th entry}} \cdots \mathbf{a}_n] = \det[\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_n]$; det-5

(1.6) $\det[\mathbf{a}_1 \cdots \mathbf{a}_n] = 0$ if $\mathbf{a}_1,\ldots,\mathbf{a}_n$ are linearly dependent. det-6

Proof. (1.4): We have to show:

---

[1] We consider det as a mapping det : $n \times n$-*matrices* $\rightarrow$ *scalars* ($= \mathbb{R}$, $\mathbb{C}$ etc.); $A \mapsto \det(A)$. More precisely, we define det for each $n \in \mathbb{N}$ simultaneously as such mapping that satisfies the following conditions (1.1), (1.2) and (1.3). That these conditions really define a unique mapping will be shown soon.

$\det(A)$ is also often denoted by $|A|$, if $A = [\mathbf{a}_1 \cdots \mathbf{a}_n]$, we write $\det[\mathbf{a}_1 \cdots \mathbf{a}_n]$ or $|\mathbf{a}_1 \cdots \mathbf{a}_n|$ instead of $\det([\mathbf{a}_1 \cdots \mathbf{a}_n])$ or $|[\mathbf{a}_1 \cdots \mathbf{a}_n]|$ respectively.

$$\det[\mathbf{a}_1 \cdots \mathbf{a}_j \cdots \mathbf{a}_i \cdots \mathbf{a}_n] + \det[\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_j \cdots \mathbf{a}_n] = 0.$$

By multiple applications of (1.2) and (1.1), we obtain:

$$\begin{aligned}
&\det[\mathbf{a}_1 \cdots \mathbf{a}_j \cdots \mathbf{a}_i \cdots \mathbf{a}_n] + \det[\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_j \cdots \mathbf{a}_n] \\
&= \det[\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_i \cdots \mathbf{a}_n] + \det[\mathbf{a}_1 \cdots \mathbf{a}_j \cdots \mathbf{a}_i \cdots \mathbf{a}_n] && \text{(by (1.2))} \\
&\quad + \det[\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_j \cdots \mathbf{a}_n] + \det[\mathbf{a}_1 \cdots \mathbf{a}_j \cdots \mathbf{a}_j \cdots \mathbf{a}_n] \\
&= \det[\mathbf{a}_1 \cdots \underbrace{\mathbf{a}_i + \mathbf{a}_j}_{i\text{'th entry}} \cdots \underbrace{\mathbf{a}_i}_{j\text{'th entry}} \cdots \mathbf{a}_n] + \det[\mathbf{a}_1 \cdots \underbrace{\mathbf{a}_i + \mathbf{a}_j}_{i\text{'th entry}} \cdots \mathbf{a}_j \cdots \mathbf{a}_n] \\
& && \text{(by (1.1), (a))} \\
&= \det[\mathbf{a}_1 \cdots \mathbf{a}_i + \mathbf{a}_j \cdots \mathbf{a}_i + \mathbf{a}_j \cdots \mathbf{a}_n] && \text{(by (1.1), (a))} \\
&= 0. && \text{(by (1.2))}
\end{aligned}$$

(1.5):

$$\begin{aligned}
&\det[\mathbf{a}_1 \cdots \mathbf{a}_i + c\mathbf{a}_j \cdots \mathbf{a}_n] \\
&= \det[\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_n] + c \det[\mathbf{a}_1 \cdots \underbrace{\mathbf{a}_j}_{i\text{'th entry}} \cdots \mathbf{a}_n] && \text{(by (1.1), (a),(b))} \\
&= \det[\mathbf{a}_1 \cdots \mathbf{a}_i \cdots \mathbf{a}_n]. && \text{(by (1.2))}
\end{aligned}$$

(1.6): Suppose that $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are linearly dependent, say, $c_1\mathbf{a}_1 + \cdots c_n\mathbf{a}_n = \mathbf{0}$ where at least one of $c_1, \ldots, c_n$ is non-zero. Without loss of generality, we may assume that $c_1 \neq 0$. Then $\mathbf{a}_1 = d_2\mathbf{a}_2 + \cdots + d_n\mathbf{a}_n$ where $d_i = -\dfrac{c_i}{c_1}$ for $i = 2, \ldots, n$. Then we have

$$\begin{aligned}
&\det[\mathbf{a}_1 \cdots \mathbf{a}_n] = \det[(d_2\mathbf{a}_2 + \cdots + d_n\mathbf{a}_n)\, \mathbf{a}_2 \cdots \mathbf{a}_n] \\
&= d_2 \det[\mathbf{a}_2\mathbf{a}_2 \cdots \mathbf{a}_n] + \cdots + d_n \det[\mathbf{a}_n\, \mathbf{a}_2 \cdots \mathbf{a}_n] && \text{(by (1.1), (a),(b))} \\
&= 0. && \text{(by (1.2))}
\end{aligned}$$

❑ (Proposition 1.1)

P-1-0

Theorem 1.2 For each n the mapping det : $n \times n$-*matrices* $\rightarrow$ *scalars* with properties (1.1), (1.2), (1.3) exists uniquely.

Proof. Suppose that det : $n \times n$-*matrices* $\rightarrow$ *scalars* and $\det^*$: $n \times n$-*matrices* $\rightarrow$ *scalars* are two mappings both satisfying (1.1), (1.2), (1.3). For $1 \leq i \leq n$, let $\mathbf{e}_i$ denote the i'th unit vector:

$$\mathbf{e}_i = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i.$$

For $n \times n$-matrix $A = [a_{i,j}]$, let

$$\mathbf{a}_j = \begin{bmatrix} a_{1,j} \\ a_{2,j} \\ \vdots \\ a_{n,j} \end{bmatrix}$$

for $1 \leq j \leq n$. Thus we have $A = [\mathbf{a}_1\mathbf{a}_2 \cdots \mathbf{a}_n]$. Since $\mathbf{a}_j = \sum_{1\leq i\leq n} a_{i,j}\mathbf{e}_i$, we have

$$\begin{aligned} \det(A) &= \textstyle\sum_{1\leq i_1\leq n} a_{i_1,1} \det[\mathbf{e}_{i_1}\mathbf{a}_2 \cdots \mathbf{a}_n] \\ &= \textstyle\sum_{1\leq i_1\leq n} \sum_{1\leq i_2\leq n} a_{i_1,1}a_{i_2,2} \det[\mathbf{e}_{i_1}\mathbf{e}_{i_2}\mathbf{a}_3 \cdots \mathbf{a}_n] = \cdots \\ &= \textstyle\sum_{1\leq i_1\leq n} \sum_{1\leq i_2\leq n} \cdots \sum_{1\leq i_n\leq n} a_{i_1,1}a_{i_2,2} \cdots a_{i_n,n} \det[\mathbf{e}_{i_1}\mathbf{e}_{i_2} \cdots \mathbf{e}_{i_n}] \end{aligned}$$

by (1.1). Likewise we have

$$\det{}^*(A) = \textstyle\sum_{1\leq i_1\leq n} \sum_{1\leq i_2\leq n} \cdots \sum_{1\leq i_n\leq n} a_{i_1,1}a_{i_2,2} \cdots a_{i_n,n} \det{}^*[\mathbf{e}_{i_1}\mathbf{e}_{i_2} \cdots \mathbf{e}_{i_n}].$$

By (1.2) and (1.4), $\det[\mathbf{e}_{i_1}\mathbf{e}_{i_2} \cdots \mathbf{e}_{i_n}]$ is uniquely determined: if any two of i_k, $1 \leq k \leq n$ are the same then $\det[\mathbf{e}_{i_1}\mathbf{e}_{i_2} \cdots \mathbf{e}_{i_n}] = 0$ by (1.2). Otherwise the sequence $\langle i_1, i_2, ..., i_n\rangle$ is a permutation of the sequence $\langle 1, 2, ..., n\rangle$. If this permutation can be obtained by product of odd number of exchanges of two numbers in $\langle 1, ..., n\rangle$, then $\det[\mathbf{e}_{i_1}\mathbf{e}_{i_2} \cdots \mathbf{e}_{i_n}] = -1$ by (1.4). Otherwise $\det[\mathbf{e}_{i_1}\mathbf{e}_{i_2} \cdots \mathbf{e}_{i_n}] = 1$ again by (1.4).

Since det^* also satisfies (1.1), (1.2) and (1.3) (and hence also (1.4)), $\det^*[\mathbf{e}_{i_1}\mathbf{e}_{i_2} \cdots \mathbf{e}_{i_n}]$ is calculated exactly in the same way. Thus

$$\begin{aligned} \det(A) &= \textstyle\sum_{1\leq i_1\leq n} \sum_{1\leq i_2\leq n} \cdots \sum_{1\leq i_n\leq n} a_{i_1,1}a_{i_2,2} \cdots a_{i_n,n} \det[\mathbf{e}_{i_1}\mathbf{e}_{i_2} \cdots \mathbf{e}_{i_n}] \\ &= \textstyle\sum_{1\leq i_1\leq n} \sum_{1\leq i_2\leq n} \cdots \sum_{1\leq i_n\leq n} a_{i_1,1}a_{i_2,2} \cdots a_{i_n,n} \det{}^*[\mathbf{e}_{i_1}\mathbf{e}_{i_2} \cdots \mathbf{e}_{i_n}] \\ &= \det{}^*(A). \end{aligned}$$

This shows the uniqueness of det.

The existence is also shown by the same idea as above. For $1 \leq i_1, ..., i_n \leq n$, let

$$d_{i_1,i_2,\ldots,i_n} = \begin{cases} 0, & \text{if } i_k = i_{k'} \text{ for some } 1 \leq k < k' \leq n; \\ -1, & \text{if } \langle i_1, ..., i_n\rangle \text{ is a permutation of } \langle 1, ..., n\rangle \\ & \text{obtained as a product of odd number of} \\ & \text{exchanges of numbers in } \langle 1, ..., n\rangle; \\ 1, & \text{otherwise.} \end{cases}$$

Then we can check that the mapping $\det(A)$ defined by

(1.7) $\det(A) = \sum_{1\le i_1\le n}\sum_{1\le i_2\le n}\cdots\sum_{1\le i_n\le n} d_{i_1,i_2,\ldots,i_n} a_{i_1,1}a_{i_2,2}\cdots a_{i_n,n}$

satisfies (1.1), (1.2) and (1.3). □ (Theorem 1.2)

Example 1.1 For 2×2-matrices A, $\det(A)$ coincides with the determinant as is learned in high school math. This can be seen as follows:

$$\begin{aligned}
&\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = a\det\begin{bmatrix} 1 & b \\ 0 & d \end{bmatrix} + c\det\begin{bmatrix} 0 & b \\ 1 & d \end{bmatrix}\\
&= ab\det\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + ad\det\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + bc\det\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + dc\det\begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}\\
&= ad - bc.
\end{aligned}$$

Similarly we can also check that $\det(A)$ for 3×3-matrices A coincides with the determinant as it is learned in high school math.

Determinant of elementary matrices.

An $n\times n$-matrix T is said to be an elementary matrix if A is in one of the forms $T_n(i,j)$, $M_n(i,c)$ or $A_n(i,j,c)$ in the following **(a)**, **(b)**, **(c)**.

(a) Elementary matrices for column (or row) exchange:

For $1\le i<j\le n$, the $n\times n$ elementary matrix $T_n(i,j)$ is obtained from the unit matrix E_n by exchanging the i'th and j'th columns of E_n.

For an $m\times n$-matrix A, $AT_n(i,j)$ is a matrix obtained from A by exchanging i'th and j'th columns while $T_n(i,j)A$ is the matrix obtained from A by exchanging i'th and j'th rows.

By (1.3) and (1.4), we have

(1.8) $\det(T_{i,j}) = -1\cdot\det(E_n) = -1$ and $\det(AT^n_{i,y}) = -\det(A)$ for any $n\times n$-matrix A. In particular $\det(AT_n(i,j)) = \det(A)\det(T_n(i,j))$ always holds. det-7

(b) Elementary matrices for column (or row) multiplication:

For a scalar c, $c\neq 0$, let $M_n(i,c)$ be the $n\times n$ elementary matrix obtained from E_n by changing the (i,i)-entry of E_n (from 1) to c.

For an $m\times n$-matrix A, $AM_n(i,c)$ is a matrix obtained from A by multiplying i'th column by c while $M_n(i,c)A$ is the matrix obtained from A by multiplying i'th row by c.

By (1.1), (b), we have

(1.9) $\det(M_n(i,c)) = c$ and $AM_n(i,c) = c\det(A)$ for any $n\times n$-matrix A. det-8

Also in this case, we have $\det(AM_n(i,c)) = \det(A)\det(M_n(i,c))$.

(c) Elementary matrices for addition of a multiple of a column (or row) to another column (or row):

For $1 \leq i, j \leq n$ $(i \neq j)$ and a scalar c, $c \neq 0$, the $n \times n$ matrix $A_n(i,j,c)$ is obtained from the unit matrix E_n by changing the (i,j)-entry of E_n (from 0) to c.

For an $m \times n$-matrix A, $AA_n(i,j,c)$ is a matrix obtained from A adding the c times of i'th column of A to the j'th column of A, while $A_n(i,j,c)A$ is the matrix obtained from A by adding the c times of j'th row of A to the i'th row of A.

By (1.5), we have

(1.10) $\det(A_n(i,j,c)) = 1$ and $\det(AA_n(i,j,c)) = \det(A)$ for any $n \times n$-matrix A. det-9
In particular, we have $\det(AA_n(i,j,c)) = \det(A)\det(A_n(i,j,c))$.

L-0

Lemma 1.3 (1) For an $n \times n$-matrix A and an $n \times n$ elementary matrix T, we have $\det(AT) = \det(A)\det(T)$.

(2) $\det(T) \neq 0$ for any elementary matrix T. ❑

Proof. By (1.8), (1.9) and (1.10). ❑ (Lemma 1.3)

An $n \times n$-matrix A is invertible if there is an $n \times n$-matrix B such that $AB = BA = E_n$. Such B is called the inverse of A.[2)] If A is invertible, its (unique) inverse is denoted by A^{-1}: we have $AA^{-1} = A^{-1}A = E_n$.

L-1

Lemma 1.4 For an $n \times n$-matrix $A = [\mathbf{a}_1 ... \mathbf{a}_n]$, the following are equivalent:

(a) A is invertible.

(b) $\mathbf{a}_1, ..., \mathbf{a}_n$ are linearly independent.

(c) there are elementary matrices $T_1, T_2, ..., T_k$ such that $A = T_1 T_2 \cdots T_k$. ❑

P-2

Proposition 1.5 Suppose that A is an invertible $n \times n$-matrix and $A = T_1 T_2 \cdots T_k$ where $T_1, T_2, ..., T_k$ are some elementary matrices. Then we have

(1) $\det(A) = \det(T_1)\det(T_2) \cdots \det(T_k)$.

(2) $\det(A) \neq 0$.

Proof. (1): By k-times application of Lemma 1.3 (1) yields

[2)] Actually the inverse of an $n \times n$-matrix A is unique if it exists[3)]: if B and B' satisfy $AB = BA = E_n$ and $AB' = B'A = E_n$ then we have $B = E_n B = (B'A)B = B'(AB) = B'E_n = B'$. Note that this is not any more true if we consider matrices of infinite size: for infinite matrix (say of size $\mathbb{N} \times \mathbb{N}$ only one of $BA = E$ oar $AB = E$ is not sufficient to deduce that B is the inverse of A).

$$\begin{aligned}\det(A) &= \det(E_nT_1\cdots T_k) = \det(E_nT_1\cdots T_{k-1})\det(T_k) = \cdots \\ &= \det(E_n)\det(T_1)\cdots\det(T_{k-1})\det(T_k) \\ &= \det(T_1)\cdots\det(T_{k-1})\det(T_k).\end{aligned}$$

(2): By Lemma 1.3, (2) $\det(T_i) \neq 0$ for all $1 \leq i \leq k$. By (1), it follows that $\det(A) \neq 0$. □ (Proposition 1.5)

P-3

Corollary 1.6 An $n \times n$-matrix A is invertible if and only if $\det(A) \neq 0$.

Proof. If A is invertible then $\det(A) \neq 0$ by Proposition 1.5, (2). If A is not invertible then, by Lemma 1.4, the columns of A are linearly dependent. By Proposition 1.1, (1.6), it follows that $\det(A) = 0$. □ (Corollary 1.6)

P-4

Proposition 1.7 For any $n\times n$-matrices A, B, we have $\det(AB) = \det(A)\det(B)$.

Proof. If B is not invertible then AB is neither invertible[4]. Thus, in this case, we have $\det(AB) = 0 = \det(A)\det(B)$ by Corollary 1.6.

Otherwise, there are elementary matrices $T_1,\dots,T_k$ such that $B = T_1\cdots T_k$. By Proposition 1.5, (1) (and by (1.3)), it follows that

$$\begin{aligned}\det(AB) &= \det(AT_1\cdots T_k) = \det(A)\det(T_1)\cdots\det(T_k) \\ &= \det(A)(\det(E_n)\det(T_1)\cdots\det(T_k)) \\ &= \det(A)\det(E_nT_1\cdots T_k) \\ &= \det(A)\det(B).\end{aligned}$$

□ (Proposition 1.7)

P-5

Corollary 1.8 For an invertible $n \times n$-matrix A, we have $\det(A^{-1}) = \dfrac{1}{\det(A)}$.

Proof. By Proposition 1.5, (2), $\det(A) \neq 0$. By (1.3) and Proposition 1.7, $1 = \det(E_n) = \det(AA^{-1}) = \det(A)\det(A^{-1})$. By dividing the both sides of this equation by $\det(A)$, we obtain $\det(A^{-1}) = \dfrac{1}{\det(A)}$. □ (Corollary 1.8)

For a matrix A, ${}^tA$ denotes the transposition of A. That is, if A is an $m \times n$-matrix with $A = [a_{i,j}]$ then ${}^tA$ is the $n \times m$-matrix whose (i, j)-entry is $a_{j,i}$.

P-6

Proposition 1.9 For any $n \times n$-matrix A, we have $\det({}^tA) = \det(A)$.

---

[4] A simple proof of this fact will be given in the next section.

Proof. If A is not invertible then ${}^tA$ is neither invertible[5]. Hence, in this case, we have $\det(A) = 0 = \det({}^tA)$.

For all of the elementary matrices T, we can check easily that the equation $\det(T) = \det({}^tT)$ holds. Hence, for an arbitrary invertible matrix A, letting $A = T_1 \cdots T_k$ for some elementary matrices $T_1, ..., T_k$, we have

$$\begin{aligned}\det(A) &= \det(T_1 \cdots T_{k-1}T_k) = \det(T_1) \cdots \det(T_{k-1}) \det(T_k) \\ &= \det({}^tT_k) \det({}^tT_{k-1}) \cdots \det({}^tT_1) = \det({}^tT_k{}^tT_{k-1} \cdots {}^tT_1) = \det({}^tA).\end{aligned}$$

❑ (Proposition 1.9)

By the Proposition above, we can translate assertions about determinant on columns to the corresponding assertions about determinant on rows. The following Proposition is such an example:

Proposition 1.10 P-7

Suppose that $A = \begin{bmatrix} \mathbf{a}^1 \\ \vdots \\ \mathbf{a}^i \\ \vdots \\ \mathbf{a}^j \\ \vdots \\ \mathbf{a}^n \end{bmatrix}$ where $\mathbf{a}^1$,..., $\mathbf{a}^n$ are n-dimensional row vectors.

Then we have

$$\det \begin{bmatrix} \mathbf{a}^1 \\ \vdots \\ \mathbf{a}^j \\ \vdots \\ \mathbf{a}^i \\ \vdots \\ \mathbf{a}^n \end{bmatrix} = -\det \begin{bmatrix} \mathbf{a}^1 \\ \vdots \\ \mathbf{a}^i \\ \vdots \\ \mathbf{a}^j \\ \vdots \\ \mathbf{a}^n \end{bmatrix}.$$

Proof. By Proposition 1.9 and (1.4), we have

$$\det \begin{bmatrix} \mathbf{a}^1 \\ \vdots \\ \mathbf{a}^j \\ \vdots \\ \mathbf{a}^i \\ \vdots \\ \mathbf{a}^n \end{bmatrix} = \det[{}^t\mathbf{a}^1 \cdots {}^t\mathbf{a}^j \cdots {}^t\mathbf{a}^i \cdots {}^t\mathbf{a}^n]$$

$$= -\det[{}^t\mathbf{a}^1 \cdots {}^t\mathbf{a}^i \cdots {}^t\mathbf{a}^j \cdots {}^t\mathbf{a}^n] = -\det \begin{bmatrix} \mathbf{a}^1 \\ \vdots \\ \mathbf{a}^i \\ \vdots \\ \mathbf{a}^j \\ \vdots \\ \mathbf{a}^n \end{bmatrix}.$$

❑ (Proposition 1.10)

L-2

Lemma 1.11 Suppose that A is an $n \times n$-matrix, $\mathbf{b}$ a row vector of dimension n and $\mathbf{0}$ the column zero vector of dimension n. Then for any scalar c, we have

[5] If ${}^tA$ is invertible and ${}^tAB = E_n$ then we have ${}^tBA = {}^tE_n = E_n$. Thus A is also invertible.

$$\det \begin{bmatrix} c & \mathbf{b} \\ \mathbf{0} & A \end{bmatrix} = c \det(A).$$

Proof. By (1.1), (b), it is enough to show that $\det \begin{bmatrix} 1 & \mathbf{b} \\ \mathbf{0} & A \end{bmatrix} = \det(A)$.

Now, let $d : n \times n$-*matrices* $\rightarrow$ *scalars* be defined by

$$d(B) = \det \begin{bmatrix} 1 & \mathbf{b} \\ \mathbf{0} & B \end{bmatrix}.$$

Then it is easy to check that d satisfies properties corresponding to (1.1), (1.2), (1.3). By the uniqueness of determinant (Theorem 1.2), it follows that $d = \det$. In other words, we have $\det \begin{bmatrix} 1 & \mathbf{b} \\ \mathbf{0} & A \end{bmatrix} = \det(A)$. ❑ (Lemma 1.11)

For an $m \times n$-matrix A for some m,$n > 1$, $1 \leq i \leq m$ and $1 \leq j \leq n$, $A_{i,j}$ denotes the $(m-1) \times (n-1)$-matrix which is obtained from A by deleting i'th row and j'th column of A.

Lemma 1.12 Suppose that A is an $n \times n$-matrix A for $n > 1$ and A' is the matrix obtained from A by deleting the first column of A. Let $A' = \begin{bmatrix} \mathbf{a}^1 \\ \vdots \\ \mathbf{a}^n \end{bmatrix}$ where $\mathbf{a}^1$,..., $\mathbf{a}^n$ are $(n-1)$-dimensional row vectors. Then, for any scalar c and $1 \leq i \leq n$, we have

$$\det[c\mathbf{e}_i^n A'] = (-1)^{i+1} c \det(A_{i,1}).$$

Proof. The matrix $\begin{bmatrix} c & \mathbf{a}^i \\ \mathbf{0} & A_{i,1} \end{bmatrix}$ can be obtained by $i-1$ times exchanges of rows starting from $[c\mathbf{e}_i^n A']$. By Proposition 1.10, we have

$$\det[c\mathbf{e}_i^n A'] = (-1)^{i-1} \det \begin{bmatrix} c & \mathbf{a}^i \\ \mathbf{0} & A_{i,1} \end{bmatrix}.$$

Since

$$\det \begin{bmatrix} c & \mathbf{a}^i \\ \mathbf{0} & A_{i,1} \end{bmatrix} = c \det(A_{i,1})$$

by Lemma 1.11, and since $i-1$ and $i+1$ have the same parity, it follows that $\det[c\mathbf{e}_i^n A'] = (-1)^{i-1} c \det(A_{i,1}) = (-1)^{i+1} c \det(A_{i,1})$. ❑ (Lemma 1.12)

Theorem 1.13 Suppose that $A = [a_{i,j}]$ is an $n \times n$-matrix. Then we have

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

More generally, if $1 \leq j \leq n$, then we have

$$\text{(1.11)}\quad \det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{i,j} \det(A_{i,j}).$$

Proof. $\square$ (Theorem 1.13)

By exchanging rows and columns of the matrix A as in the proof of Proposition 1.10, we obtain the following:

Corollary 1.14 Suppose that $A = [a_{i,j}]$ is an $n \times n$-matrix and $1 \leq i \leq n$. Then we have

$$\text{(1.12)}\quad \det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{i,j} \det A_{i,j}.$$

2 Linear mappings

lin

For $m, n \in \mathbb{N}$, a mapping $\varphi : \mathbb{R}^m \to \mathbb{R}^n$ is said to be a linear mapping[6)] if for any $\mathbf{a}$, $\mathbf{b} \in \mathbb{R}^m$ c, $d \in \mathbb{R}$, we have

$$\text{(2.1)}\quad \varphi(c\mathbf{a} + d\mathbf{b}) = c\varphi(\mathbf{a}) + d\varphi(\mathbf{b}).$$ lin-1

The following is easy to show:

L-2-a-0

Lemma 2.1 $\varphi : \mathbb{R}^m \to \mathbb{R}^n$ is a linear mapping if and only if the following two conditions hold:

(2.2) $\varphi(\mathbf{a} + \mathbf{b}) = \varphi(\mathbf{a}) + \varphi(\mathbf{b})$ for all $\mathbf{a}$, $\mathbf{b} \in \mathbb{R}^m$; lin-2

(2.3) $\varphi(c\mathbf{a}) = c\varphi(\mathbf{a})$ for all $\mathbf{a} \in \mathbb{R}^m$ and $c \in \mathbb{R}$. $\square$ lin-3

L-2-a

Lemma 2.2 If $\varphi : \mathbb{R}^m \to \mathbb{R}^n$ is a linear mapping, then $\varphi(\mathbf{0}) = \mathbf{0}$.

Proof. $\varphi(\mathbf{0}) = \varphi(\mathbf{0} + \mathbf{0}) = \varphi(\mathbf{0}) + \varphi(\mathbf{0})$. Subtracting $\varphi(\mathbf{0})$ from the both sides of the equation, we obtain $\mathbf{0} = \varphi(\mathbf{0})$. $\square$ (Lemma 2.2)

Ex-1

Example 2.1 (1) If $m = n = 1$, then $\varphi : \mathbb{R} \to \mathbb{R}$ is a linear mapping if and only if there is some $c \in \mathbb{R}$ such that $\varphi(b) = cb$ for all $b \in \mathbb{R}$.

(2) Let A be an $n \times m$-matrix (with real numbers as entries) then the mapping $\varphi_A : \mathbb{R}^m \to \mathbb{R}^n$ defined by $\varphi_A(\mathbf{b}) = A\mathbf{b}$ for $\mathbf{b} \in \mathbb{R}^m$ is a linear mapping. $\square$

---

6) We can treat also mappings $\varphi : \mathbb{C}^m \to \mathbb{C}^n$, $\varphi : \mathbb{Q}^m \to \mathbb{Q}^n$ etc. in exactly the same manner but to be concrete we shall consider here only the mappings on real vector spaces, i.e. the vector spaces of the form $\mathbb{R}^m$, $\mathbb{R}^n$.

Proof. (1): Suppose that $\varphi : \mathbb{R} \to \mathbb{R}$ is a linear mapping. Let $c = \varphi(1)$. For any $b \in \mathbb{R}$ since $b = b \cdot 1$ we have $\varphi(b) = b\varphi(1) = bc = cb$.

(2): For $\mathbf{a}$, $\mathbf{b} \in \mathbb{R}^m$ and c, $d \in \mathbb{R}$, we have $\varphi_A(c\mathbf{a} + d\mathbf{b}) = A(c\mathbf{a} + d\mathbf{b}) = cA\mathbf{a} + cA\mathbf{b} = c\varphi_A(\mathbf{a}) + d\varphi_A(\mathbf{b})$. ❑

For a linear mapping $\varphi : \mathbb{R}^m \to \mathbb{R}^n$, the kernel of φ (denoted by $\mathrm{Ker}(\varphi)$) is the set $\mathrm{Ker}(\varphi) = \varphi^{-1}{}''\{\mathbf{0}\} = \{\mathbf{a} \in \mathbb{R}^m : \varphi(\mathbf{a}) = \mathbf{0}\}$. By Lemma 2.2, we always have $\{\mathbf{0}\} \subseteq \mathrm{Ker}(\varphi)$.

A mapping $f : X \to Y$ is said to be 1-1 (one to one) if, for any x, $x' \in X$ with $x \neq x'$, we have $f(x) \neq f(x')$. f is 1-1 if and only if $f^{-1}{}''\{y\} = \{x \in X : f(x) = y\}$ has at most one element for all $y \in Y$.

L-2-0

Lemma 2.3 Suppose that $\varphi : \mathbb{R}^m \to \mathbb{R}^n$ is a linear mapping. Then the following are equivalent:

(a) f is 1-1;

(b) $\mathrm{Ker}(f) = \{\mathbf{0}\}$;

(c) for any linearly independent $\mathbf{a}_1,\ldots,\mathbf{a}_k \in \mathbb{R}^m$, we have that $\varphi(\mathbf{a}_1),\ldots,\varphi(\mathbf{a}_k)$ are also linearly independent.

Proof. (a) $\Rightarrow$ (b): This is clear by the remark before the theorem.

(b) $\Rightarrow$ (c): Assume that (b) holds and suppose that $\mathbf{a}_1,\ldots,\mathbf{a}_k \in \mathbb{R}^m$ are linearly independent (in $\mathbb{R}^m$) and

$$(2.4)\qquad c_1\varphi(\mathbf{a}_1) + \cdots c_k\varphi(\mathbf{a}_k) = \mathbf{0}$$ lin-3-a

for some $c_1,\ldots,c_k \in \mathbb{R}$. We have to show that $c_1 = \cdots = c_k = 0$.

By (2.4) and since φ is linear, it follows that

$$(2.5)\qquad \varphi(c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k) = \mathbf{0}.$$ lin-3-a-0

By (b), it follows that $c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k = \mathbf{0}$. But since $\mathbf{a}_1,\ldots,\mathbf{a}_k$ are linearly independent, it follows that $c_1 = \cdots = c_k = 0$ as desired.

(c) $\Rightarrow$ (a): We shall prove the contraposition of the implication. Suppose that φ is not 1-1. Then there are $\mathbf{a}$, $\mathbf{a}' \in \mathbb{R}^m$, $\mathbf{a} \neq \mathbf{a}'$ such that $\varphi(\mathbf{a}) = \varphi(\mathbf{a}')$. Since φ is linear, it follows that $\varphi(\mathbf{a} - \mathbf{a}') = \varphi(\mathbf{a}) - \varphi(\mathbf{a}') = \mathbf{0}$ but $\mathbf{a} - \mathbf{a}' \neq \mathbf{0}$. Thus $\mathbf{a} - \mathbf{a}'$ is linearly independent (as a sequence of elements of $\mathbb{R}^m$ of length 1) but $\varphi(\mathbf{a} - \mathbf{a}')$ is not. Thus (c) does not hold. ❑ (Lemma 2.3)

T-0

Theorem 2.4 (1) For any linear mapping $\psi : \mathbb{R}^m \to \mathbb{R}^n$, there is a unique $n \times m$-matrix[7] M_ψ such that, using the notation of Example 2.1, (2), we have $\psi = \varphi_{M_\psi}$.

[7] Note that this is really $n \times m$ and not $m \times n$!

(2) For any $n \times m$-matrix A we have $A = M_{\varphi_A}$.

Proof. (1): Let $\mathbf{e}_i^m$, $1 \leq i \leq m$ be the m-dimensional unit vectors. Let $\mathbf{a}_i = \psi(\mathbf{e}_i^m)$ for $1 \leq i \leq m$. Note that $\mathbf{a}_i \in \mathbb{R}^n$.

Let $M_\psi = [\mathbf{a}_1\mathbf{a}_2 \cdots \mathbf{a}_m]$. We show that this M_ψ is as desired. Let $\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$ be an arbitrary vector in $\mathbb{R}^m$. We have to show that $\psi(\mathbf{b}) = \varphi_{M_\psi}(\mathbf{b})$ $(= M_\psi \mathbf{b})$.

Since ψ is linear, we have

$$\begin{aligned} \psi(\mathbf{b}) &= \psi(b_1\mathbf{e}_1^m + b_2\mathbf{e}_2^m + \cdots + b_m\mathbf{e}_m^m) \\ &= b_1\psi(\mathbf{e}_1^m) + b_2\psi(\mathbf{e}_2^m) + \cdots + b_m\psi(\mathbf{e}_m^m) \\ &= b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + \cdots + b_m\mathbf{a}_m \\ &= M_\psi\mathbf{b}. \end{aligned}$$

The uniqueness of M_ψ follows from the fact that, if $A \neq B$, then $\varphi_A \neq \varphi_B$. This can be seen as follows: suppose $A = [\mathbf{a}_1\mathbf{a}_2 \cdots \mathbf{a}_m]$ and $B = [\mathbf{b}_1\mathbf{b}_2 \cdots \mathbf{b}_m]$. If $A \neq B$ then there is $1 \leq j^* \leq m$ such that $\mathbf{a}_{j^*} \neq \mathbf{b}_{j^*}$. Then we have

$$\varphi_A(\mathbf{e}_{j^*}^m) = A\mathbf{e}_{j^*}^m = \mathbf{a}_{j^*} \neq \mathbf{b}_{j^*} = B\mathbf{e}_{j^*}^m = \varphi_B(\mathbf{e}_{j^*}^m).$$

Thus $\varphi_A \neq \varphi_B$. □ (Theorem 2.4)

Recall that, for any mappings $f : X \to Y$ and $g : Y \to Z$, the composition of g and f is the mapping $g \circ f : X \to Z$ defined by $(g \circ f)(x) = g(f(x))$ for all $x \in X$.

A mapping $f : X \to Y$ is said to be onto if, for any $y \in Y$ there is at least one $x \in X$ such that $f(x) = y$.

A mapping $f : X \to Y$ is said to be bijective (or 1-1 onto) if f is 1-1 and onto. For a bijective mapping $f : X \to Y$ the inverse of f is the mapping $g : Y \to X$ defined by

$$g(y) = \text{the unique } x \in X \text{ such that } f(x) = y$$

for all $y \in Y$. The inverse of f is denoted by f^{-1}. It is characterized by $f \circ f^{-1} = id_Y$ and $f^{-1} \circ f = id_X$ where id_X and id_Y denote identity mapping on X and Y respectively[8].

L-2-1

---

[8] The identity mapping on X is the mapping $id_X : X \to X$ defined by $id_X(x) = x$ for all $x \in X$.

Lemma 2.5 (1) For linear mappings $\varphi : \mathbb{R}^m \to \mathbb{R}^n$ and $\psi : \mathbb{R}^n \to \mathbb{R}^r$, $\psi \circ \varphi$ is linear mapping from $\mathbb{R}^m$ to $\mathbb{R}^r$ and $M_{\psi\circ\varphi} = M_\psi M_\varphi$.

(2) If A and B are $n \times m$ and $r \times n$-matrices then, we have $\varphi_{BA} = \varphi_B \circ \varphi_A$.

(3) $\varphi_{E_n} = id_{\mathbb{R}^n}$ and $M_{id_{\mathbb{R}^n}} = E_n$.

(4) A linear mapping $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is bijective then the inverse φ^{-1} of φ is also a linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$. We shall say such a linear mapping φ is invertible. An invertible linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^n$ is also said be an automorphism on $\mathbb{R}^n$.

(5) A linear mapping $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ is invertible if and only if M_φ is invertible. If φ is invertible then we have $\varphi^{-1} = \varphi_{(M_\varphi)^{-1}}$.

(6) An $n \times n$ matrix A is invertible if and only if φ_A is an invertible linear mapping. If A is invertible then we have $A^{-1} = M_{(\varphi_A)^{-1}}$.

Proof.

□ (Lemma 2.5)

$X \subseteq \mathbb{R}^n$ is a *linear subspace* (also called *linear subalgebra*) of $\mathbb{R}^n$ if X is closed with respect to vector addition and scalar multiplication (i.e. if, for any $\mathbf{a}$, $\mathbf{b} \in X$ we have $\mathbf{a} + \mathbf{b} \in V$, and, for any $\mathbf{a} \in X$ and $c \in \mathbb{R}$, we have $c\mathbf{a} \in V$).

Ex-2

Example 2.2 (a) $\{\mathbf{0}\} \subseteq \mathbb{R}^n$ is a linear subspace of $\mathbb{R}^n$.

(b) $\mathbb{R}^n \subseteq \mathbb{R}^n$ is a linear subspace of $\mathbb{R}^n$.

(c) For $\mathbf{a}_1,\ldots, \mathbf{a}_k \in \mathbb{R}^n$, let

$$[\mathbf{a}_1, \ldots, \mathbf{a}_k]_{\mathbb{R}^n} = \{c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k \,:\, c_1, \ldots, c_k \in \mathbb{R}\}.$$

Then $[\mathbf{a}_1, \ldots, \mathbf{a}_k]_{\mathbb{R}^n}$ is a linear subspace of $\mathbb{R}^n$.

$\{\mathbf{0}\} \subseteq \mathbb{R}^n$ is called the zero dimensional subspace of $\mathbb{R}^n$. $[\mathbf{a}_1, \ldots, \mathbf{a}_k]_{\mathbb{R}^n}$ as above is called the subspace of $\mathbb{R}^n$ spanned by $\mathbf{a}_1,\ldots, \mathbf{a}_k$.

Proof of Example 2.2. (a): If $\mathbf{a}$, $\mathbf{b} \in \{\mathbf{0}\}$, then $\mathbf{a} = \mathbf{b} = \mathbf{0}$. But then, for any c, $d \in \mathbb{R}$, $c\mathbf{a} + d\,\mathbf{b} = c\mathbf{0} + d\,\mathbf{0} = \mathbf{0} \in \{\mathbf{0}\}$.

(b): This is clear since $\mathbb{R}^n$ is closed with respect to vector addition and scalar multiplication.

(c): For $\mathbf{a}$, $\mathbf{b} \in [\mathbf{a}_1, \ldots, \mathbf{a}_k]_{\mathbb{R}^n}$ and c, $d \in \mathbb{R}$, if $\mathbf{a} = c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k$ and $\mathbf{b} = d_1\mathbf{b}_1 + \cdots + d_k\mathbf{a}_k$ for some $c_1, \ldots, c_k$, $d_1,\ldots, d_k \in \mathbb{R}$, then we have

$$c\mathbf{a} + d\mathbf{b} = (cc_1 + dd_1)\mathbf{a}_1 + \cdots + (cc_k + dd_k)\mathbf{a}_k.$$

Thus $c\mathbf{a} + d\mathbf{b} \in [\mathbf{a}_1, ..., \mathbf{a}_k]_{\mathbb{R}^n}$. $\square$ (Example 2.5)

The next lemma shows that Example 2.2 actually exhausts all possible linear subspaces of $\mathbb{R}^n$.

L-3

Lemma 2.6 (1) If $X \subseteq \mathbb{R}^n$ is a linear subspace of $\mathbb{R}^n$ and $\mathbf{a}_1,..., \mathbf{a}_k \in X$ then $[\mathbf{a}_1, ..., \mathbf{a}_k]_{\mathbb{R}^n} \subseteq X$.

(2) If $X \subseteq \mathbb{R}^n$ is a linear subspace of $\mathbb{R}^n$, then there are linearly independent $\mathbf{u}_1, ..., \mathbf{u}_\ell \in X$ such that $[\mathbf{u}_1, ..., \mathbf{u}_\ell]_{\mathbb{R}^n} = X$.

We need the following facts from Linear Algebra I for the proof of Lemma 2.6.

Recall that a sequence $\langle \mathbf{b}_1, ..., \mathbf{b}_m \rangle$ of elements of $\mathbb{R}^n$ is called a linear basis of $\mathbb{R}^n$, if $\mathbf{b}_1,..., \mathbf{b}_m$ are independent and $[\mathbf{b}_1, ..., \mathbf{b}_m]_{\mathbb{R}^n} = \mathbb{R}^n$.

T-1

Theorem 2.7 (1) For any linear basis $\langle \mathbf{b}_1, ..., \mathbf{b}_m \rangle$ of $\mathbb{R}^n$, we have $m = n$. This number $m = n$ is called the dimension of $\mathbb{R}^n$.

(2) If $\mathbf{b}_1, ..., \mathbf{b}_\ell \in \mathbb{R}^n$ are linearly independent, then we can find $\mathbf{b}_{\ell+1}, ..., \mathbf{b}_m \in \mathbb{R}^n$ such that $\langle \mathbf{b}_1, ..., \mathbf{b}_\ell, \mathbf{b}_{\ell+1}, ..., \mathbf{b}_m \rangle$ is a linear basis of $\mathbb{R}^n$.

(3) If $\mathbf{b}_1, ..., \mathbf{b}_\ell \in \mathbb{R}^n$ are linearly independent, then we have $\ell \leq n$. $\square$

Note that Theorem 2.7, (3) is an immediate consequence of Theorem 2.7, (1), (2).

L-3-0

Lemma 2.8 $\mathbf{b}_1,..., \mathbf{b}_k \in \mathbb{R}^n$ are linearly independent if and only if,

(2.6) $\mathbf{b}_1 \neq \mathbf{0}$; and lin-3-0

(2.7) for any $1 \leq i < k$, $\mathbf{b}_{i+1} \notin [\mathbf{b}_1, ..., \mathbf{b}_i]_{\mathbb{R}^n}$. $\square$ lin-3-1

Proof of Lemma 2.6. (1): If $\mathbf{a} \in [\mathbf{a}_1, ..., \mathbf{a}_k]_{\mathbb{R}^n}$, then there are $c_1,..., c_k \in \mathbb{R}$ such that $\mathbf{a} = c_1\mathbf{a}_1 + \cdots + c_k\mathbf{a}_k$. Since X is closed with respect to vector addition and scalar multiplication, it follows that $\mathbf{a} \in X$.

(2): If $X = \{\mathbf{0}\}$, $\ell = 0$ will do.

Otherwise, we take inductively the elements $\mathbf{u}_1, \mathbf{u}_2, ... \in X$, such that

(2.8) $\mathbf{u}_1 \neq \mathbf{0}$; lin-4

(2.9) $\mathbf{u}_{i+1} \in X \setminus [\mathbf{u}_1, ..., \mathbf{u}_i]_{\mathbb{R}^n}$, if $[\mathbf{u}_1, ..., \mathbf{u}_i]_{\mathbb{R}^n} \subsetneqq X$; otherwise we let $\ell = i$ and finish the construction. lin-6

By (2.8), (2.9) and by Lemma 2.8, $\mathbf{u}_1$, $\mathbf{u}_2$,... are linearly independent. By the second half of (2.9) and Lemma 2.6, (1), if the construction stops at $\ell + 1$st step, then $[\mathbf{u}_1, ..., \mathbf{u}_\ell]_{\mathbb{R}^n} = X$ and $\mathbf{u}_1, ..., \mathbf{u}_\ell$ are as desired.

Thus we are done by showing that the construction stops at latest after $n+1$ steps: Assume that this were not the case. Then we have linearly independent $\mathbf{u}_1, ..., \mathbf{u}_{n+1} \in X \subseteq \mathbb{R}^n$. This is a contradiction to Theorem 2.7, (3). □ (Lemma 2.6)

A sequence $\langle \mathbf{u}_1, ..., \mathbf{u}_\ell \rangle$ of elements of X as in Lemma 2.6, (2) is said to be a linear basis of X.

For linear subspaces X of $\mathbb{R}^n$, we can also consider the dimension of X:

L-4

Lemma 2.9 For any linear subspace X of $\mathbb{R}^n$, if $\langle \mathbf{u}_1, ..., \mathbf{u}_\ell \rangle$ and $\langle \mathbf{v}_1, ..., \mathbf{v}_m \rangle$ are linear bases of X, then $\ell = m$.

Proof. Let $\varphi : \mathbb{R}^\ell \to X$, $\psi : X \to \mathbb{R}^m$ be defined by

$$\varphi(\begin{bmatrix} a_1 \\ \vdots \\ a_\ell \end{bmatrix}) = a_1\mathbf{u}_1 + \cdots + a_\ell \mathbf{u}_\ell \quad \text{and} \quad \psi(c_1\mathbf{v}_1 + \cdots + c_m\mathbf{v}_m) = \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix}.$$

Claim 2.9.1 $\psi \circ \varphi$ is a 1-1- linear mapping[9)] from $\mathbb{R}^\ell$ to $\mathbb{R}^m$.

⊢ To see that $\psi \circ \varphi$ is a linear mapping, let $c_1, c_2, a_1,..., a_\ell, b_1,..., b_\ell \in \mathbb{R}$ and

$$a_1\mathbf{u}_1 + \cdots + a_\ell\mathbf{u}_\ell = a'_1\mathbf{v}_1 + \cdots + a'_m\mathbf{v}_m \text{ and}$$
$$b_1\mathbf{u}_1 + \cdots + b_\ell\mathbf{u}_\ell = b'_1\mathbf{v}_1 + \cdots + b'_m\mathbf{v}_m$$

for some $a'_1,..., a'_\ell, b'_1,..., b'_\ell \in \mathbb{R}$.

Then we have

$$\begin{aligned}
&\psi \circ \varphi \left(c_1 \begin{bmatrix} a_1 \\ \vdots \\ a_\ell \end{bmatrix} + c_2 \begin{bmatrix} b_1 \\ \vdots \\ b_\ell \end{bmatrix} \right) = \psi \circ \varphi \left(\begin{bmatrix} c_1a_1 + c_2b_1 \\ \vdots \\ c_1a_\ell + c_2b_\ell \end{bmatrix} \right) \\
&= \psi((c_1a_1 + c_2b_1)\mathbf{u}_1 + \cdots + (c_1a_\ell + c_2b_\ell)\mathbf{u}_\ell) \\
&= \psi(c_1(a_1\mathbf{u}_1 + \cdots + a_\ell\mathbf{u}_\ell) + c_2(b_1\mathbf{u}_1 + \cdots + b_\ell\mathbf{u}_\ell)) \\
&= \psi(c_1(a'_1\mathbf{v}_1 + \cdots + a'_m\mathbf{v}_m) + c_2(b'_1\mathbf{v}_1 + \cdots + b'_m\mathbf{v}_m)) \\
&= \psi((c_1a'_1 + c_2b'_1)\mathbf{v}_1 + \cdots + (c_1a'_m + c_2b'_m)\mathbf{v}_m)) \\
&= \begin{bmatrix} c_1a'_1 + c_2b'_1 \\ \vdots \\ c_1a'_m + c_2b'_m \end{bmatrix} \\
&= c_1 \begin{bmatrix} a'_1 \\ \vdots \\ a'_m \end{bmatrix} + c_2 \begin{bmatrix} b'_1 \\ \vdots \\ b'_m \end{bmatrix} = c_1\psi \circ \varphi \left(\begin{bmatrix} a_1 \\ \vdots \\ a_\ell \end{bmatrix} \right) + c_2\psi \circ \varphi \left(\begin{bmatrix} b_1 \\ \vdots \\ b_\ell \end{bmatrix} \right).
\end{aligned}$$

9) Actually we can show that $\psi \circ \varphi$ is a 1-1 onto linear mapping which also implies $\ell = m$.

$\psi \circ \varphi$ is 1-1 since both φ and ψ are 1-1 by the definition. ⊣ (Claim 2.9.1)

By Lemma 2.3, (c) and Theorem 2.7, (3), it follows that $\ell \leq m$. By the same argument, we also obtain $m \leq \ell$. Hence $\ell = m$. ❑ (Lemma 2.9)

The uniquely determined number ℓ ($= m$) as above is said to be the dimension of X and denoted by $\dim(X)$.

For a linear mapping $\varphi : \mathbb{R}^m \to \mathbb{R}^n$, we already defined the kernel of of φ as $\mathrm{Ker}(\varphi) = \varphi^{-1}{}''\{\mathbf{0}\} = \{\mathbf{a} \in \mathbb{R}^m : \varphi(\mathbf{a}) = \mathbf{0}\}$.

The image of φ (notation: $\mathrm{Im}(\varphi)$) is the subset of $\mathbb{R}^n$ defined by

$$\mathrm{Im}(\varphi) = \{\mathbf{b} \in \mathbb{R}^n : \text{there is } \mathbf{a} \in \mathbb{R}^m \text{ such that } \varphi(\mathbf{a}) = \mathbf{b}\}.$$

L-5

Lemma 2.10 Suppose that $\varphi : \mathbb{R}^m \to \mathbb{R}^n$ is a linear mapping. Then:

(1) $\mathrm{Ker}(\varphi)$ is a linear subspace of $\mathbb{R}^m$.

(2) $\mathrm{Im}(\varphi)$ is a linear subspace of $\mathbb{R}^n$.

Proof. (1): Suppose that $\mathbf{a}_1, \mathbf{a}_2 \in \mathrm{Ker}(\varphi)$ and $c_1, c_2 \in \mathbb{R}$. Then

$$\varphi(c_1\mathbf{a}_1 + c_2\mathbf{a}_2) = c_1\varphi(\mathbf{a}_1) + c_2\varphi(\mathbf{a}_2) = c_1\mathbf{0} + c_2\mathbf{0} = \mathbf{0}.$$

This shows that $c_1\mathbf{a}_1 + c_2\mathbf{a}_2 \in \mathrm{Ker}(\varphi)$.

(2): Suppose that $\mathbf{b}_1, \mathbf{b}_2 \in \mathrm{Im}(\varphi)$ and $d_1, d_2 \in \mathbb{R}$. Let $\mathbf{a}_1, \mathbf{a}_2 \in \mathbb{R}^m$ be such that $\varphi(\mathbf{a}_1) = \mathbf{b}_1$ and $\varphi(\mathbf{a}_2) = \mathbf{b}_2$. Then we have

$$\varphi(d_1\mathbf{a}_1 + d_2\mathbf{a}_2) = d_1\varphi(\mathbf{a}_1) + d_2\varphi(\mathbf{a}_2) = d_1\mathbf{b}_1 + d_2\mathbf{b}_2.$$

This shows that $d_1\mathbf{b}_1 + d_2\mathbf{b}_2 \in \mathrm{Im}(\varphi)$. ❑ (Lemma 2.10)

dimension-thm

Theorem 2.11 (Dimension Theorem) For any linear mapping $\varphi : \mathbb{R}^m \to \mathbb{R}^n$, we always have $\dim(\mathrm{Ker}(\varphi)) + \dim(\mathrm{Im}(\varphi)) = m$.

Proof. Let $k = \dim(\mathrm{Ker}(\varphi))$ and $\ell = \dim(\mathrm{Im}(\varphi))$. We have to show that $k + \ell = m$.

Let $\mathbf{a}_1, \ldots, \mathbf{a}_k \in \mathrm{Ker}(\varphi)$ be such that $\langle \mathbf{a}_1, \ldots, \mathbf{a}_k \rangle$ is a linear basis of $\mathrm{Ker}(\varphi)$, let $\mathbf{b}_1, \ldots, \mathbf{b}_\ell \in \mathrm{Im}(\varphi)$ be such that $\langle \mathbf{b}_1, \ldots, \mathbf{b}_\ell \rangle$ is a linear basis of $\mathrm{Im}(\varphi)$ and let $\mathbf{c}_1, \ldots, \mathbf{c}_\ell \in \mathbb{R}^m$ be such that $\mathbf{b}_1 = \varphi(\mathbf{c}_1), \ldots, \mathbf{b}_\ell = \varphi(\mathbf{c}_\ell)$.

By Theorem 2.7, (1), the following claim proves the theorem:

cl-lin-basis

Claim 2.11.1 $\langle \mathbf{a}_1, \ldots, \mathbf{a}_k, \mathbf{c}_1, \ldots, \mathbf{c}_\ell \rangle$ is a linear basis of $\mathbb{R}^m$.

⊢ We have to prove

(2.10) $[\mathbf{a}_1, \ldots, \mathbf{a}_k, \mathbf{c}_1, \ldots, \mathbf{c}_\ell]_{\mathbb{R}^m} = \mathbb{R}^m$ and lin-7

(2.11) $\mathbf{a}_1, ..., \mathbf{a}_k, \mathbf{c}_1, ..., \mathbf{c}_\ell$ are linearly independent. lin-8

Proof of (2.10): For an arbitrary $\mathbf{a} \in \mathbb{R}^m$, let $c_1, ..., c_\ell \in \mathbb{R}$ be such that $\varphi(\mathbf{a}) = c_1\mathbf{b}_1 + \cdots + c_\ell\mathbf{b}_\ell$ (such $c_1, ..., c_\ell$ exist since $\langle \mathbf{b}_1, ..., \mathbf{b}_\ell \rangle$ is a linear basis of $\mathrm{Im}(\varphi)$).

Let

(2.12) $$\mathbf{c} = \mathbf{a} - (c_1\mathbf{c}_1 + \cdots + c_\ell\mathbf{c}_\ell).$$ lin-9

Then we have

(2.13) $$\varphi(\mathbf{c}) = \varphi(\mathbf{a}) - (c_1\varphi(\mathbf{c}_1) + \cdots + c_\ell\varphi(\mathbf{c}_\ell)) = \varphi(\mathbf{a}) - (c_1\mathbf{b}_1 + \cdots + c_\ell\mathbf{b}_\ell) = \mathbf{0}.$$ lin-10

Thus $\mathbf{c} \in \mathrm{Ker}(\varphi)$. Since $\langle \mathbf{a}_1, ..., \mathbf{a}_k \rangle$ is a linear basis of $\mathrm{Ker}(\varphi)$, there are $d_1, ..., d_k \in \mathbb{R}$ such that

(2.14) $$\mathbf{c} = d_1\mathbf{a}_1 + \cdots + d_k\mathbf{a}_k.$$ lin-11

By (2.12) and (2.14), it follows that

(2.15) $$\mathbf{a} = d_1\mathbf{a}_1 + \cdots + d_k\mathbf{a}_k + c_1\mathbf{c}_1 + \cdots + c_\ell\mathbf{c}_\ell.$$

This shows $\mathbf{a} \in [\mathbf{a}_1, ..., \mathbf{a}_k, \mathbf{c}_1, ..., \mathbf{c}_\ell]_{\mathbb{R}^m}$. Since $\mathbf{a} \in \mathbb{R}^m$ was arbitrary, it follows that $[\mathbf{a}_1, ..., \mathbf{a}_k, \mathbf{c}_1, ..., \mathbf{c}_\ell]_{\mathbb{R}^m} \supseteq \mathbb{R}^m$. Since $[\mathbf{a}_1, ..., \mathbf{a}_k, \mathbf{c}_1, ..., \mathbf{c}_\ell]_{\mathbb{R}^m} \subseteq \mathbb{R}^m$ clearly holds, we can conclude that $[\mathbf{a}_1, ..., \mathbf{a}_k, \mathbf{c}_1, ..., \mathbf{c}_\ell]_{\mathbb{R}^m} = \mathbb{R}^m$.

Proof of (2.11): Suppose that

(2.16) $$d_1\mathbf{a}_1 + \cdots + d_k\mathbf{a}_k + c_1\mathbf{c}_1 + \cdots + c_\ell\mathbf{c}_\ell = \mathbf{0}.$$ lin-12

We have to show that $d_1 = \cdots = d_k = c_1 = \cdots = d_\ell = 0$.

By (2.16), it follows from (2.16)

(2.17) $$\begin{aligned} \mathbf{0} = \varphi(\mathbf{0}) &= \varphi(d_1\mathbf{a}_1 + \cdots + d_k\mathbf{a}_k + c_1\mathbf{c}_1 + \cdots + c_\ell\mathbf{c}_\ell) \\ &= d_1\varphi(\mathbf{a}_1) + \cdots + d_k\varphi(\mathbf{a}_k) + c_1\varphi(\mathbf{c}_1) + \cdots + c_\ell\varphi(\mathbf{c}_\ell) \\ &= c_1\mathbf{b}_1 + \cdots + c_\ell\mathbf{b}_\ell. \end{aligned}$$ lin-13

Since $\langle \mathbf{b}_1, ..., \mathbf{b}_\ell \rangle$ is a linear basis of $\mathrm{Im}(\varphi)$, it follows that $c_1 = \cdots = c_\ell = 0$. Thus we have $d_1\mathbf{a}_1 + \cdots + d_k\mathbf{a}_k = \mathbf{0}$. Since $\langle \mathbf{a}_1, ..., \mathbf{a}_k \rangle$ is a linear basis of $\mathrm{Ker}(\varphi)$, it follows that $d_1 = \cdots = d_k = 0$. $\dashv$ (Claim 2.11.1)

□ (Theorem 2.11)

For the next application of the Dimension Theorem, let us review some basic facts about mappings.

L-6

Lemma 2.12 Suppose that X and Y are any sets, $f : X \to Y$ and $g : Y \to X$.

(1) If $g \circ f = id_X$, then f is 1-1 and g is onto[10].

(2) If $g \circ f = id_X$ and f is onto then g is an inverse function of f.

(3) If $g \circ f = id_X$ and g is 1-1 then g is an inverse function of f.

Proof. (1): We prove both of assertions inductively. If f were not 1-1 then there would be two different x, $x' \in X$ with $f(x) = f(x')$. But then $(g \circ f)(x) = g(f(x)) = g(f(x')) = (g \circ f)(x)$. This is a contradiction to $g \circ g = id_X$. If g were not onto then there would be some $x^* \in X$ such that $g(y) \neq x^*$ for all $y \in Y$. It follows that $g(f(x)) \neq x^*$ for all $x \in X$ or $(g \circ f)(x) \neq x^*$ for all x. This is a contradiction to $g \circ g = id_X$.

(2): f is 1-1 by (1). Hence if f is onto then f is a bijection and the inverse f^{-1} of f exists. Since $g \circ f = id_X$, multiplying both sides of this equation by f^-1 from right side, we obtain $g = g \circ f \circ f^{-1} = id_X \circ f^{-1} = f^{-1}$.

(3): can be proved similarly to (2). ❑ (Lemma 2.12)

T-2

Theorem 2.13 For any $n \times n$-matrices A, B, if $BA = E_n$ then A and B are inverse to each other[11]. In particular, from $BA = E_n$ it follows that $AB = E_n$.

Proof. By Lemma 2.5, (2) and (3), we have $\varphi_B \circ \varphi_A = id_{\mathbb{R}^n}$. By Lemma 2.12, (1), φ_A is 1-1 and hence $\dim(\mathrm{Ker}(\varphi_A)) = 0$.

By Theorem 2.11, it follows that $\dim(\mathrm{Im}(\varphi_A)) = n$. By Theorem 2.7, this means that $\mathrm{Im}(\varphi_A) = \mathbb{R}^n$ or φ_A is onto. Thus, by Lemma 2.12, (2), φ_A and φ_B are inverse functions to each other. By Lemma 2.5, (5), this means that A and B are inverse to each other. ❑ (Theorem 2.13)

3 Change of bases

exch

In this and next sections we often use the following trivial Lemma:

matrix=

Lemma 3.1 For any $m \times n$-matrix A and B, $A\mathbf{c} = B\mathbf{c}$ for all $\mathbf{c} \in \mathbb{R}^n$ if and only if $A = B$.

---

10) id_X denotes the identity function on X, that is, id_X is defined by $id_X : X \to X$ and $id_X(x) = x$ for all $x \in X$.

11) It is essential that matrices A and B here are of finite size. For matrices of infinite size, there are counter-examples to the corresponding assertion.

Proof. $A\mathbf{c} = B\mathbf{c}$ for all $\mathbf{c} \in \mathbb{R}^n$ $\Leftrightarrow$ $\varphi_A(\mathbf{c}) = \varphi_B(\mathbf{c})$ for all $\mathbf{c} \in \mathbb{R}^n$
$\Leftrightarrow$ $\varphi_A = \varphi_B$ $\Leftrightarrow$ $A = B$. ❑ (Lemma 3.1)

Suppose that $\langle \mathbf{b}_1, ..., \mathbf{b}_n \rangle$ is a linear basis of $\mathbb{R}^n$ (remember that all linear bases of $\mathbb{R}^n$ have size n).

For $\mathbf{a} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, suppose that $\mathbf{a}$ is represented as $\mathbf{a} = y_1\mathbf{b}_1 + \cdots + y_n\mathbf{b}_n$.

Then, letting $\mathbf{e}^n_i = \left.\begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} \begin{matrix} \\ \\ \\ \leftarrow i. \\ \\ \\ \end{matrix}\right\} n$

for $1 \leq i \leq n$, we have

exch-0

$$\text{(3.1)} \quad \mathbf{a} = x_1\mathbf{e}^n_1 + \cdots x_n\mathbf{e}^n_n = y_1\mathbf{b}_1 + \cdots + y_n\mathbf{b}_n$$

$$\Leftrightarrow \underbrace{[\mathbf{e}^n_1 \cdots \mathbf{e}^n_n]}_{\substack{\parallel \\ E_n}} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [\mathbf{b}_1 \cdots \mathbf{b}_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \Leftrightarrow \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = [\mathbf{b}_1 \cdots \mathbf{b}_n] \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

$$\Leftrightarrow [\mathbf{b}_1 \cdots \mathbf{b}_n]^{-1} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

Here we can actually take $[\mathbf{b}_1 \cdots \mathbf{b}_n]^{-1}$ by the following observation:

invertible

Lemma 3.2 For $\mathbf{b}_1, ..., \mathbf{b}_n \in \mathbb{R}^n$, $[\mathbf{b}_1 \cdots \mathbf{b}_n]$ is invertible if and only if $\langle \mathbf{b}_1, ..., \mathbf{b}_n \rangle$ is a linear basis of $\mathbb{R}^n$.

Proof. We have:

$B = [\mathbf{b}_1 \cdots \mathbf{b}_n]$ is invertible $\Leftrightarrow$ φ_B is invertible (as a linear mapping)
$\Leftrightarrow$ φ_B is 1-1 and onto
$\Leftrightarrow$ $vecb_1, ..., \mathbf{b}_n$ are linearly independent and $[\{\mathbf{b}_1, ..., \mathbf{b}_n\}]_{\mathbb{R}^n} = \mathbb{R}^n$
$\Leftrightarrow$ $\langle \mathbf{b}_1, ..., \mathbf{b}_n \rangle$ is a linear basis of $\mathbb{R}^n$.

❑ (Lemma 3.2)

Suppose now that $\langle \mathbf{b}_1, ..., \mathbf{b}_n \rangle$ and $\langle \mathbf{c}_1, ..., \mathbf{c}_n \rangle$ are two linear bases of $\mathbb{R}^n$

For $\mathbf{a} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$, suppose that $\mathbf{a}$ is represented as $\mathbf{a} = y_1\mathbf{b}_1 + \cdots + y_n\mathbf{b}_n$ and $\mathbf{a} = z_1\mathbf{c}_1 + \cdots + z_n\mathbf{c}_n$. Then, by (3.1), we have

$$(3.2)\quad [\mathbf{b}_1 \cdots \mathbf{b}_n]^{-1}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad\text{and}\quad [\mathbf{c}_1 \cdots \mathbf{c}_n]^{-1}\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$ exch-1

Hence we have

$$(3.3)\quad \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = [\mathbf{c}_1 \cdots \mathbf{c}_n]^{-1}[\mathbf{b}_1 \cdots \mathbf{b}_n]\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$ exch-2

Thus, we have obtained:

Lemma 3.3 For any linear bases $\langle \mathbf{b}_1, ..., \mathbf{b}_n\rangle$ and $\langle \mathbf{c}_1, ..., \mathbf{c}_n\rangle$ of $\mathbb{R}^n$, if a vector $\mathbf{a} \in \mathbb{R}^n$ is represented as $\mathbf{a} = y_1\mathbf{b}_1 + \cdots y_n\mathbf{b}_n$ as well as $\mathbf{a} = z_1\mathbf{c}_1 + \cdots z_n\mathbf{c}_n$, then we have

$$\begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} = [\mathbf{c}_1 \cdots \mathbf{c}_n]^{-1}[\mathbf{b}_1 \cdots \mathbf{b}_n]\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$ ❏

Now let $\varphi : \mathbb{R}^m \to \mathbb{R}^n$ be an arbitrary linear mapping with $k = \dim(\mathrm{Ker}(\varphi))$ and $\ell = \dim(\mathrm{Im}(\varphi))$. By Dimension Theorem (Theorem 2.11) we have $k + \ell = m$.

Let $\langle \mathbf{a}_1, ..., \mathbf{a}_k\rangle$ be a linear basis of $\mathrm{Ker}(\varphi)$ and $\langle \mathbf{b}_1, ..., \mathbf{b}_\ell\rangle$ a linear basis of $\mathrm{Im}(\varphi)$. Let $\mathbf{c}_1, ..., \mathbf{c}_\ell \in \mathbb{R}^m$ be such that $\mathbf{b}_1 = \varphi(\mathbf{c}_1)$,..., $\mathbf{b}_\ell = \varphi(\mathbf{c}_\ell)$. By Claim 2.11.1, $\langle \mathbf{c}_1, ..., \mathbf{c}_\ell, \mathbf{a}_1, ..., \mathbf{a}_k\rangle$ is a linear basis of $\mathbb{R}^m$.

Let $\langle \mathbf{b}_1, ..., \mathbf{b}_\ell, \mathbf{b}_{\ell+1}, ..., \mathbf{b}_n\rangle$ be any extension of $\langle \mathbf{b}_1, ..., \mathbf{b}_\ell\rangle$ to a linear basis of $\mathbb{R}^n$ (or just $\langle \mathbf{b}_1, ..., \mathbf{b}_\ell\rangle$ itself if $\ell = n$).

For any $\mathbf{a} \in \mathbb{R}^m$ with the representation $\mathbf{a} = [\mathbf{c}_1 \cdots \mathbf{c}_\ell\mathbf{a}_1 \cdots \mathbf{a}_k]\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$, we have

$$(3.4)\quad \varphi(\mathbf{a}) = \varphi\left([\mathbf{c}_1 \cdots \mathbf{c}_\ell\mathbf{a}_1 \cdots \mathbf{a}_k]\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}\right) = M_\varphi[\mathbf{c}_1 \cdots \mathbf{c}_\ell\mathbf{a}_1 \cdots \mathbf{a}_k]\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$ exch-3

$$= [\mathbf{b}_1 \cdots \mathbf{b}_\ell\mathbf{0} \cdots \mathbf{0}]\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = [\mathbf{b}_1 \cdots \mathbf{b}_\ell\mathbf{b}_{\ell+1} \cdots \mathbf{b}_n]D\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix},$$

where D is the $n \times m$-matrix of the form

$$(3.5)\quad \begin{bmatrix} E_\ell & O_1 \\ O_2 & O_3 \end{bmatrix}$$ exch-4

with $\ell \times \ell$ unit matrix E_ℓ, $\ell \times (m-\ell)$-zero matrix O_1, $k \times \ell$-zero matrix O_2 and $k \times (m-\ell)$-zero matrix O_2.

Since this holds for all $\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{R}^m$, we have

$$M_\varphi[\mathbf{c}_1 \cdots \mathbf{c}_\ell \mathbf{a}_1 \cdots \mathbf{a}_k] = [\mathbf{b}_1 \cdots \mathbf{b}_n]D$$

by Lemma 3.1. It follows that

$$(3.6) \quad [\mathbf{b}_1 \cdots \mathbf{b}_n]^{-1} M_\varphi[\mathbf{c}_1 \cdots \mathbf{c}_\ell \mathbf{a}_1 \cdots \mathbf{a}_k] = D.$$ exch-5

Thus we have obtained:

Theorem 3.4 For any linear mapping $\varphi : \mathbb{R}^m \to \mathbb{R}^n$ there are invertible $n \times n$ matrix U and invertible $m \times m$ matrix V such that $UM_\varphi V = D$ where D is of the form (3.5) for $k = \dim(\mathrm{Ker}(\varphi))$ and $\ell = \dim(\mathrm{Im}(\varphi))$. ❑

4 Diagonalization

diagon

An $n \times n$-matrix $D = [d_{i,j}]$ is said to be a diagonal matrix if $d_{i,j} = 0$ holds for all $1 \le i, j \le n$ with $i \ne j$.

An $n \times n$-matrix A is said to be diagonalizable if there is an invertible $n \times n$-matrix U such that $D = U^{-1}AU$ is diagonal. D is said to be a diagonalization of A.

A non zero vector $\mathbf{u} \in \mathbb{R}^n$ is said to be an eigenvector of the $n \times n$-matrix A if there is some $\varepsilon \in \mathbb{R}$ such that $A\mathbf{u} = \varepsilon\mathbf{u}$. $\varepsilon \in \mathbb{R}$ with $A\mathbf{u} = \varepsilon\mathbf{u}$ for some nonzero $\mathbf{u} \in \mathbb{R}^n$ is called an eigenvalue of A. In this case, we shall also say $\mathbf{u}$ is an eigenvector of A with the eigenvalue ε.

Note that the equation $A\mathbf{u} = \varepsilon\mathbf{u}$ is equivalent to $\varphi_A(\mathbf{u}) = \varepsilon\mathbf{u}$. We shall also talk about eigenvector and eigenvalue of a linear mapping $\varphi : \mathbb{R}^n \to \mathbb{R}^n$: $\mathbf{u}$ is an eigenvector of a linear mapping $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ with the eigenvalue ε if $\mathbf{u}$ is an eigenvector of M_φ with the eigenvalue ε.

For an $n \times n$-matrix A and $\varepsilon \in \mathbb{R}$,

$$\begin{aligned} (4.1) \quad & \varepsilon \text{ is an eigenvalue of } A \\ \Leftrightarrow\ & \text{the equation } A\mathbf{x} = \varepsilon\mathbf{x}\ (\Leftrightarrow (A - \varepsilon E_n)\mathbf{x} = \mathbf{0}) \text{ has a non-trivial solution} \\ \Leftrightarrow\ & \text{the } n \times n\text{-matrix } A - \varepsilon E_n \text{ is not invertible} \\ \Leftrightarrow\ & \det(A - \varepsilon E_n) = 0. \end{aligned}$$ diag-0

The polynomial $\det(A - \varepsilon E_n)$ with the unknown ε is called a characteristic polynomial of A. By (4.1), we can obtain all eigenvalues of the matrix A by finding roots of the characteristic polynomial of the $n \times n$-matrix A.

Suppose now that $\mathbf{u}_1,\ldots,\mathbf{u}_n$ are eigenvectors of A for eigenvalues $\varepsilon_1,\ldots,\varepsilon_n$ respecpively such that $\mathbf{u}_1,\ldots,\mathbf{u}_n$ build a linear basis of $\mathbb{R}^n$. By Lemma 3.2, the matrix $U = [\mathbf{u}_1 \cdots \mathbf{u}_n]$ is invertible.

For an arbitrary $\mathbf{c} \in \mathbb{R}^n$, we have

$$(4.2) \qquad A\mathbf{c} = A(UU^{-1})\mathbf{c} = (AU)U^{-1}\mathbf{c} = U \begin{bmatrix} \varepsilon_1 & & 0 \\ & \ddots & \\ 0 & & \varepsilon_n \end{bmatrix} U^{-1}\mathbf{c}$$ diag-1

for arbitrary $\mathbf{c} \in \mathbb{R}^n$. By Lemma 3.1, it follows that $A = U \begin{bmatrix} \varepsilon_1 & & 0 \\ & \ddots & \\ 0 & & \varepsilon_n \end{bmatrix} U^{-1}$.

Conversely, if $D = U^{-}AU$ is a diagonalizaion of the $n \times n$-matrix A with its diagonal entries being $\varepsilon_1,\ldots,\varepsilon_n$, letting $U = [\mathbf{u}_1 \cdots \mathbf{u}_n]$, we have

$$(4.3) \qquad [A\mathbf{u}_1 \cdots A\mathbf{u}_n] = A[\mathbf{u}_1 \cdots \mathbf{u}_n] = AU = UD = [\varepsilon_1\mathbf{u}_1 \cdots \varepsilon_n\mathbf{u}_n].$$ diag-2

Thus we have $A\mathbf{u}_i = \varepsilon_i\mathbf{u}_i$ and $\mathbf{u}_1,\ldots,\mathbf{u}_n$ are eigenvectors of A. $\mathbf{u}_1,\ldots,\mathbf{u}_n$ build a linear basis of $\mathbb{R}^n$ since U is invertible.

Putting the arguments above together, we obtain:

diagonalization-thm

Theorem 4.1 An $n \times n$-matrix A is diagonalizable if and only if there is a linear basis of $\mathbb{R}^n$ consisting of eigenvectors $\mathbf{u}_1,\ldots,\mathbf{u}_n$ of A. If the eigenvalues associated with $\mathbf{u}_1,\ldots,\mathbf{u}_n$ are $\varepsilon_1,\ldots,\varepsilon_n$ respectively, then, letting $U = [\mathbf{u}_1 \cdots \mathbf{u}_n]$, we have

$$A = U \begin{bmatrix} \varepsilon_1 & & 0 \\ & \ddots & \\ 0 & & \varepsilon_n \end{bmatrix} U^{-1} \quad \text{or} \quad U^{-1}AU = \begin{bmatrix} \varepsilon_1 & & 0 \\ & \ddots & \\ 0 & & \varepsilon_n \end{bmatrix}. \qquad \square$$

diagonalization

Corollary 4.2 If an $n \times n$-matrix A is diagonalizable, its diagonalization is unique up to parmutation of diagonal entries.

The condition in Theorem 4.1 can be further reformulated as in the following Corollary 4.5.

diag-3

Theorem 4.3 If $\mathbf{v}_1,\ldots,\mathbf{v}_m \in \mathbb{R}^n$ are eigenvectors of an $n \times n$-matrix A of pairwise different eigenvalues $\varepsilon_1,\ldots,\varepsilon_m$ respectively, then $\mathbf{v}_1,\ldots,\mathbf{v}_m$ are linearly independent.

Proof. Note that $\mathbf{v}_1,\ldots,\mathbf{v}_m$ are non-zero vectors by definition of eigenvectors.

Suppose that $\mathbf{v}_1,\ldots,\mathbf{v}_m$ were linearly dependent. Then we can find an ℓ with $1\ell < m$ such that $\mathbf{v}_1,\ldots,\mathbf{v}_\ell$ are linearly independent but $\mathbf{v}_1,\ldots,\mathbf{v}_{\ell+1}$ are not. Let $c_1\mathbf{v}_1 + \cdots + c_\ell\mathbf{v}_\ell + c_{\ell+1}\mathbf{v}_{\ell+1}$ be a nontrivial linear combination with $c_1\mathbf{v}_1 + \cdots + c_\ell\mathbf{v}_\ell + c_{\ell+1}\mathbf{v}_{\ell+1} = \mathbf{0}$. Then we have $c_{\ell+1} \neq 0$ by the choice of ℓ. Let $d_i = -\frac{c_i}{c_{\ell+1}}$ for $1 \leq i \leq \ell$. Then we have

(4.4) $\mathbf{v}_{\ell+1} = d_1\mathbf{v}_1 + \cdots + d_\ell\mathbf{v}_\ell.$ diag-3-0

Note that $d_1,\ldots, d_\ell$ cannot be all equal to zero since $\mathbf{v}_{\ell+1} \neq \mathbf{0}$. By multiplying the both sides of (4.4) by A from the left side, we obtain

(4.5) $\varepsilon_{\ell+1}\mathbf{v}_{\ell+1} = \varepsilon_1 d_1\mathbf{v}_1 + \cdots + \varepsilon_\ell d_\ell\mathbf{v}_\ell.$ diag-3-1

On the other hand, by multiplying the both sides of (4.4) by $\varepsilon_{\ell+1}$, we obtain

(4.6) $\varepsilon_{\ell+1}\mathbf{v}_{\ell+1} = \varepsilon_{\ell+1} d_1\mathbf{v}_1 + \cdots + \varepsilon_{\ell+1} d_\ell\mathbf{v}_\ell.$ diag-3-2

Thus by subtracting (4.5) from (4.6) we obtain

(4.7) $\mathbf{0} = (\varepsilon_{\ell+1} - \varepsilon_1)d_1\mathbf{v}_1 + \cdots + (\varepsilon_{\ell+1} - \varepsilon_\ell)d_\ell\mathbf{v}_\ell.$ diag-3-3

Since $\varepsilon_{\ell+1} - \varepsilon_1 \neq 0,\ldots, \varepsilon_{\ell+1} - \varepsilon_\ell \neq 0$, the linear combination in the right side of (4.7) is non-trivial. But then this is a contradiction to the linearly independence of $\mathbf{v}_1,\ldots, \mathbf{v}_\ell$. ❑ (Theorem 4.3)

diag-4

Corollary 4.4 If $\varepsilon_1,\ldots, \varepsilon_m$ are pairwise different eigenvalues of an $n \times n$-matrix A and, for each $1 \leq i \leq m$, $\mathbf{u}_{i,1},\ldots, \mathbf{u}_{i,k_i}$ are linearly independent eigenvectors of A for the eigenvalue ε_i, then the vectors $\mathbf{u}_{1,1}, \ldots, \mathbf{u}_{1,k_1}, \mathbf{u}_{2,1}, \ldots, \mathbf{u}_{2,k_2},\ldots, \mathbf{u}_{m,1}, \ldots, \mathbf{u}_{m,k_m}$ are linearly independent. ❑

diag-thm-x

Corollary 4.5 An $n\times n$-matrix A with its all pairwise distinct eigenvalues $\varepsilon_1,\ldots, \varepsilon_m$ is diagonalizable if and only if there are linearly independent eigen vectors $\mathbf{u}_{i,1}, \ldots, \mathbf{u}_{i,k_i}$ of A for each eigenvalue ε_i, $1 \leq i \leq m$ such that $k_1 + \cdots + k_m = n$.

Proof. By Corollary 4.4 and Theorem 4.1. ❑ (Corollary 4.5)