Matrices and Determinants

Beifang Chen

Fall 2006

1 Linear Transformations

Definition 1.1. Let X and Y be nonempty sets. A **function** from X to Y is a rule, written $f : X \to Y$, such that each element x in X is assigned a unique element y in Y; the element y is denoted by $f(x)$, written

$$y = f(x),$$

called the **image** of x under f; and the element x is called the **preimage** of $f(x)$. Functions are also called **maps**, or **mappings**, or **transformations**.

Definition 1.2. A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is said to be a **linear transformation** if, for any vectors $\boldsymbol{u}, \boldsymbol{v}$ in $\mathbb{R}^n$ and scalar c,

(a) $T(\boldsymbol{u} + \boldsymbol{v}) = T(\boldsymbol{u}) + T(\boldsymbol{v})$,

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

Example 1.1. (a) The function $T : \mathbb{R}^2 \to \mathbb{R}^2$, defined by $T(x_1, x_2) = (x_1 + 2x_2, x_2)$, is a linear transformation, see Figure 1

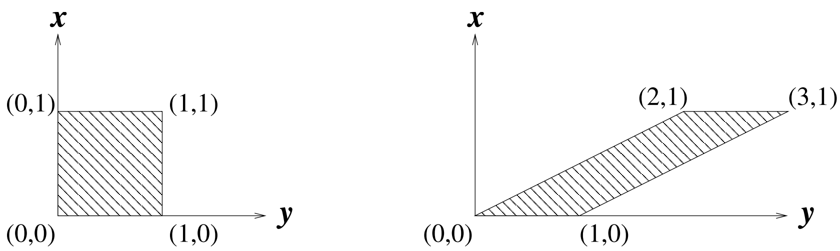

Figure 1: The geometric shape under a linear transformation.

(b) The function $T : \mathbb{R}^2 \to \mathbb{R}^2$, defined by $T(x_1, x_2) = (x_1 + 2x_2, 3x_1 + 4x_2)$, is a linear transformation.

(c) The function $T : \mathbb{R}^3 \to \mathbb{R}^2$, defined by $T(x_1, x_2, x_3) = (x_1 + 2x_2 + 3x_3, 3x_1 + 2x_2 + x_3)$, is a linear transformation.

Example 1.2. The transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ by $T(\boldsymbol{x}) = A\boldsymbol{x}$, where A is an $m \times n$ matrix, is a linear transformation.

Example 1.3. The map $T : \mathbb{R}^n \to \mathbb{R}^n$, defined by $T(\boldsymbol{x}) = \lambda\boldsymbol{x}$, where λ is a constant, is a linear transformation, and is called the **dilation** by λ.

Example 1.4. The **refection** $T : \mathbb{R}^2 \to \mathbb{R}^2$ about a straight line through the origin is a linear transformation.

Example 1.5. The **rotation** $T : \mathbb{R}^2 \to \mathbb{R}^2$ about an angle θ is a linear transformation, see Figure 2.

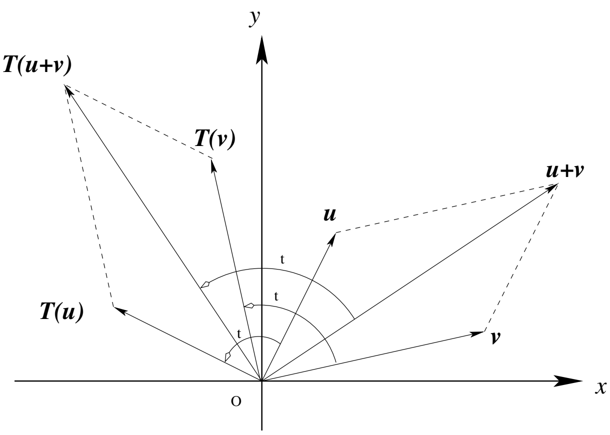

Figure 2: Rotation about an angle θ.

Proposition 1.3. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ be vectors in $\mathbb{R}^n$ and let $c_1, c_2, \ldots, c_k$ be scalars. Then*

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

$$T(c_1\boldsymbol{v}_1 + \boldsymbol{v}_2 + \cdots + c_k\boldsymbol{v}_k) = c_1T(\boldsymbol{v}_1) + c_2T(\boldsymbol{v}_2) + \cdots + c_kT(\boldsymbol{v}_k).$$

Theorem 1.4. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Let $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ be vectors in $\mathbb{R}^n$ and let $c_1, c_2, \ldots, c_k$ be real numbers.*

(a) *If $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are linearly dependent, then $T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_k)$ are linearly dependent.*

(b) *If $T(\boldsymbol{v}_1), T(\boldsymbol{v}_2), \ldots, T(\boldsymbol{v}_k)$ are linearly independent, then $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k$ are linearly independent.*

2 Standard Matrix of Linear Transformation

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Consider the following vectors

$$\boldsymbol{e}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \boldsymbol{e}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \boldsymbol{e}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

in the coordinate axis of $\mathbb{R}^n$. It is clear that any vector

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n$$

is a linear combination of $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, …, $\boldsymbol{e}_n$. Thus

$$T(\boldsymbol{x}) = x_1T(\boldsymbol{e}_1) + x_2T(\boldsymbol{e}_2) + \cdots + x_nT(\boldsymbol{e}_n).$$

This means that $T(\boldsymbol{x})$ is completely determined by the images $T(\boldsymbol{e}_1)$, $T(\boldsymbol{e}_2)$, …, $T(\boldsymbol{e}_n)$. The ordered set $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ is called the **standard basis** of $\mathbb{R}^n$.

Definition 2.1. The **standard matrix** of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is the $m \times n$ matrix

$$A = \big[T(\boldsymbol{e}_1), T(\boldsymbol{e}_2), \ldots, T(\boldsymbol{e}_n)\big] = \big[T\boldsymbol{e}_1, T\boldsymbol{e}_2, \ldots, T\boldsymbol{e}_n\big].$$

Proposition 2.2. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation whose standard matrix is A. Then*

$$T(\boldsymbol{x}) = A\boldsymbol{x}.$$

Proof.

$$\begin{aligned} T(\boldsymbol{x}) &= T(x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n) \\ &= x_1T(\boldsymbol{e}_1) + x_2T(\boldsymbol{e}_2) + \cdots + x_nT(\boldsymbol{e}_n) \\ &= [T(\boldsymbol{e}_1), T(\boldsymbol{e}_2), \ldots, T(\boldsymbol{e}_n)]\boldsymbol{x} = A\boldsymbol{x}. \end{aligned}$$

□

Example 2.1. (a) The linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T(x_1, x_2) = (x_1 + 2x_2, 3x_1 + 4x_2)$, can be written as the matrix form

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{bmatrix} = x_1\begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2\begin{bmatrix} 2 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

(b) Let $T : \mathbb{R}^3 \to \mathbb{R}^2$, $T(x_1, x_2, x_3) = (x_1 + 2x_2 + 3x_3, 3x_1 + 2x_2 + 1x_3)$. Then T is a linear transformation and its standard matrix is given by

$$\begin{aligned} T\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) &= \begin{bmatrix} x_1 + 2x_2 + 3x_3 \\ 3x_1 + 2x_2 + x_3 \end{bmatrix} \\ &= x_1\begin{bmatrix} 1 \\ 3 \end{bmatrix} + x_2\begin{bmatrix} 2 \\ 2 \end{bmatrix} + x_3\begin{bmatrix} 3 \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \end{aligned}$$

Example 2.2. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a rotation about an angle θ counterclockwise. Then

$$T(\boldsymbol{e}_1) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}, \quad T(\boldsymbol{e}_2) = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}.$$

Thus

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.$$

Figure

Proposition 2.3. *Let $f, g : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations with the standard matrices A and B, respectively, that is,*

$$f(\boldsymbol{x}) = A\boldsymbol{x}, \quad g(\boldsymbol{x}) = B\boldsymbol{x}, \quad \boldsymbol{x} \in \mathbb{R}^n.$$

Then $f \pm g : \mathbb{R}^n \to \mathbb{R}^m$ are linear transformation with standard matrix $A \pm B$, that is,

$$(f \pm g)(\boldsymbol{x}) = f(\boldsymbol{x}) \pm g(\boldsymbol{x}) = (A \pm B)\boldsymbol{x}, \quad \boldsymbol{x} \in \mathbb{R}^n;$$

and for any scalar c, $cf : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation with the standard matrix cA, that is,

$$(cf)(\boldsymbol{x}) = cf(\boldsymbol{x}) = (cA)\boldsymbol{x}, \quad \boldsymbol{x} \in \mathbb{R}^n.$$

Example 2.3. Let $f, g : \mathbb{R}^3 \to \mathbb{R}^2$ be linear transformations defined by (writing in coordinates)

$$f(x_1, x_2, x_3) = (x_1 - x_2, x_2 - x_3),$$

$$g(x_1, x_2, x_3) = (x_1 + x_2, x_2 + x_3).$$

Writing in vectors,

$$\begin{aligned} f\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) &= \begin{bmatrix} x_1 - x_2 \\ x_2 - x_3 \end{bmatrix} \\ &= \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \\ g\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) &= \begin{bmatrix} x_1 + x_2 \\ x_2 + x_3 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \end{aligned}$$

We then have

$$\begin{aligned} (f+g)\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) &= \begin{bmatrix} x_1 - x_2 \\ x_2 - x_3 \end{bmatrix} + \begin{bmatrix} x_1 + x_2 \\ x_2 + x_3 \end{bmatrix} \\ &= \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \\ (f-g)\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) &= \begin{bmatrix} x_1 - x_2 \\ x_2 - x_3 \end{bmatrix} - \begin{bmatrix} x_1 + x_2 \\ x_2 + x_3 \end{bmatrix} \\ &= \begin{bmatrix} -2x_2 \\ -2x_3 \end{bmatrix} = \begin{bmatrix} 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \end{aligned}$$

Hence

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix};$$

$$\begin{bmatrix} 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} + \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}.$$

Definition 2.4. The **composition** of a function $f : X \to Y$ and a function $g : Y \to Z$ is a function

$$g \circ f : X \to Z \quad \text{given by}$$

$$(g \circ f)(x) = g(f(x)), \quad x \in X.$$

Proposition 2.5. *Let $f : X \to Y$, $g : Y \to Z$, and $h : Z \to W$. Then*

$$h \circ (g \circ f) = (h \circ g) \circ f.$$

That is, for any $x \in X$,

$$\big(h \circ (g \circ f)\big)(x) = \big((h \circ g) \circ f\big)(x).$$

Proof.

$$\big(h \circ (g \circ f)\big)(x) = h\big((g \circ f)(x)\big) = h\Big(g\big(f(x)\big)\Big),$$

$$\big((h \circ g) \circ f\big)(x) = (h \circ g)\big(f(x)\big) = h\Big(g\big(f(x)\big)\Big).$$

$\square$

Theorem 2.6. *Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A. Let $g : \mathbb{R}^p \to \mathbb{R}^n$ be a linear transformation with standard matrix B. Then the composition $f \circ g : \mathbb{R}^p \to \mathbb{R}^m$ is a linear transformation with the standard matrix AB. Symbolically, we have*

$$\mathbb{R}^p \xrightarrow[B]{g} \mathbb{R}^n \xrightarrow[A]{f} \mathbb{R}^m \quad \Longrightarrow \quad \mathbb{R}^p \xrightarrow[AB]{f \circ g} \mathbb{R}^m.$$

Proof. For vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^p$ and scalars $a, b \in \mathbb{R}$, we have

$$\begin{aligned} (f \circ g)(a\boldsymbol{u} + b\boldsymbol{v}) &= f(g(a\boldsymbol{u} + b\boldsymbol{v})) \\ &= f(ag(\boldsymbol{u}) + bg(\boldsymbol{v})) \\ &= af(g\boldsymbol{u}) + bf(g(\boldsymbol{v})) \\ &= a(f \circ g)(\boldsymbol{u}) + b(f \circ g)(\boldsymbol{v}). \end{aligned}$$

Then the composition map $f \circ g$ is a linear transformation.

Write $A = [a_{ij}]_{m \times n}$, $B = [b_{jk}]_{n \times p}$. Let $C = [c_{ik}]_{m \times p}$ be the standard matrix of $f \circ g$. Let us write

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{bmatrix}, \quad \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \boldsymbol{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_m \end{bmatrix}.$$

Then the linear transformations f, g, $f \circ g$ can be written as

$$\boldsymbol{y} = g(\boldsymbol{x}) = B\boldsymbol{x}, \quad \boldsymbol{z} = f(\boldsymbol{y}) = A\boldsymbol{y}, \quad \boldsymbol{z} = (f \circ g)(\boldsymbol{x}) = C\boldsymbol{x}.$$

Writing in coordinates, we have

$$\begin{aligned} y_j &= \sum_{k=1}^{p} b_{jk} x_k, \quad 1 \le j \le n, \\ z_i &= \sum_{j=1}^{n} a_{ij} y_j = \sum_{k=1}^{p} c_{ik} x_k, \quad 1 \le i \le m. \end{aligned}$$

Substitute $y_j = \sum_{k=1}^{p} b_{jk} x_k$ into $z_i = \sum_{j=1}^{n} a_{ij} y_j$. We obtain

$$z_i = \sum_{j=1}^{n} a_{ij} \sum_{k=1}^{p} b_{jk} x_k = \sum_{k=1}^{p} \left(\sum_{j=1}^{n} a_{ij} b_{jk} \right) x_k.$$

Comparing the coefficients of the variables x_k, we conclude that

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk}, \quad 1 \le i \le m, \ 1 \le k \le p.$$

□

Example 2.4. Consider the linear transformations

$$S : \mathbb{R}^2 \to \mathbb{R}^3, \quad S(\boldsymbol{x}) = \begin{bmatrix} 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 - x_2 \\ -x_1 + x_2 \\ x_1 + x_2 \end{bmatrix},$$

$$T : \mathbb{R}^3 \to \mathbb{R}^2, \quad T(\boldsymbol{x}) = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ x_2 + x_3 \end{bmatrix}.$$

Then $S \circ T : \mathbb{R}^3 \to \mathbb{R}^3$ and $T \circ S : \mathbb{R}^2 \to \mathbb{R}^2$ are both linear transformations, and

$$\begin{aligned}
(S \circ T)(\boldsymbol{x}) &= S(T(\boldsymbol{x})) = S\left(\begin{bmatrix} x_1 + x_2 \\ x_2 + x_3 \end{bmatrix}\right) \\
&= \begin{bmatrix} x_1 - x_3 \\ -x_1 + x_3 \\ x_1 + 2x_2 + x_3 \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & -1 \\ -1 & 0 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},
\end{aligned}$$

$$\begin{aligned}
(T \circ S)(\boldsymbol{x}) &= T(S(\boldsymbol{x})) = T\left(\begin{bmatrix} x_1 - x_2 \\ -x_1 + x_2 \\ x_1 + x_2 \end{bmatrix}\right) \\
&= \begin{bmatrix} 0 \\ 2x_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}.
\end{aligned}$$

Hence

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

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

Proposition 2.7. *Let A be an $m \times n$ matrix and let B be an $n \times p$ matrix. Write $B = [\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_p]$. Then*

$$AB = A[\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_p] = [A\boldsymbol{b}_1, A\boldsymbol{b}_2, \ldots, A\boldsymbol{b}_p].$$

Example 2.5. Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}$. Write $B = [\boldsymbol{b}_1, \boldsymbol{b}_2, \boldsymbol{b}_3]$. Then

$$A\boldsymbol{b}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix},$$

$$A\boldsymbol{b}_2 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \end{bmatrix},$$

$$A\boldsymbol{b}_3 = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \end{bmatrix},$$

and

$$AB = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 5 & 4 \\ 5 & 8 & 5 \end{bmatrix}.$$

Example 2.6. For any real number r, the mapping $D_r : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$D_r(\boldsymbol{x}) = r\boldsymbol{x},$$

is a linear transformation, called the **dilation** by r. Its standard matrix is D_r is the matrix.

$$\begin{bmatrix} r & 0 & \cdots & 0 \\ 0 & r & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r \end{bmatrix}$$

Example 2.7. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A. Then

$$T \circ D_r = D_r \circ T.$$

The standard matrices of $T \circ D_r$ and $D_r \circ T$ are equal to rA.

Theorem 2.8. (a) *If A is an $m \times n$ matrix, B and C are $n \times p$ matrices, then*

$$A(B + C) = AB + AC.$$

(b) *If A and B are $m \times n$ matrices, C is an $n \times p$ matrix, then*

$$(A + B)C = AC + BC.$$

(c) *If A is an $m \times n$ matrix and B is an $n \times p$ matrix, then for any scalar a,*

$$a(AB) = (aA)B = A(aB).$$

(d) *If A is an $m \times n$ matrix, then*

$$I_m A = A = AI_n.$$

Definition 2.9. The **transpose** of an $m \times n$ matrix A is the $n \times m$ matrix A^T whose (i, j)-entry is the (j, i)-entry of A, that is,

$$A^T = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nm} \end{bmatrix}.$$

Example 2.8.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 3 & 4 & 5 & 6 \\ 5 & 6 & 7 & 8 \end{bmatrix}^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \\ 3 & 5 & 7 \\ 4 & 6 & 8 \end{bmatrix}.$$

Proposition 2.10. *Let A be an $m \times n$ matrix and B an $n \times p$ matrix. Then*

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

Proof. Let c_{ij} be the (i, j)-entry of AB. Let b'_{ik} be the (i, k)-entry of B^T be and a'_{kj} the (k, j)-entry of A^T. Then $b'_{ik} = b_{ki}$ and $a'_{kj} = a_{jk}$. Thus the (i, j)-entry of $(AB)^T$ is

$$c_{ji} = \sum_{k=1}^{n} a_{jk} b_{ki} = \sum_{k=1}^{n} b'_{ik} a'_{kj},$$

which is the (i, j)-entry of $B^T A^T$ by the matrix multiplication. □

Theorem 2.11. *Let $h : X \to Y$, $g : Y \to Z$, $f : Z \to W$ be functions. Then*

$$(f \circ g) \circ h = f \circ (g \circ h).$$

Proof. For any $x \in X$,

$$\big((f \circ g) \circ h\big)(x) = (f \circ g)\big(h(x)\big) = f\big(g(h(x))\big) = f\big((g \circ h)(x)\big) = \big(f \circ (g \circ h)\big)(x).$$

□

Corollary 2.12. *Let A be an $m \times n$ matrix, B an $n \times p$ matrix, and C an $p \times q$ matrix. Then $(AB)C$ and $A(BC)$ are $m \times q$ matrices, and*

$$(AB)C = A(BC).$$

Definition 2.13. Let $f : X \to Y$ be a function.

(a) The function f is called **one-to-one** if distinct elements in X are mapped to distinct elements in Y.

(b) The function f is called **onto** if every element y in Y is the image of some element x in X under f.

Theorem 2.14 (Characterization of One-to-One and Onto). *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation with standard matrix A. Then*

(a) *T is one-to-one $\iff$ $T(\boldsymbol{x}) = \mathbf{0}$ has the only trivial solution.*
$\iff$ *The column vectors of A are linearly independent.*
$\iff$ *Every column of A has a pivot position.*

(b) *T is onto $\iff$ The column vectors of A span $\mathbb{R}^m$.*
$\iff$ *Every row of A has a pivot position.*

Example 2.9. (a) The linear transformation $T_1 : \mathbb{R}^2 \to \mathbb{R}^3$, defined by

$$T_1\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix},$$

is one-to-one but not onto. The column vectors are linearly independent, but can not span $\mathbb{R}^3$.

(b) The linear transformation $T_2 : \mathbb{R}^3 \to \mathbb{R}^2$, defined by

$$T_2\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

is not one-to-one but onto. The column vectors are linearly dependent but span $\mathbb{R}^2$.

(c) The linear transformation $T_3 : \mathbb{R}^3 \to \mathbb{R}^3$, defined by

$$T_3\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & 4 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

is neither one-to-one nor onto. The column vectors are linearly dependent and can not span $\mathbb{R}^3$.

(d) The linear transformation $T_4 : \mathbb{R}^3 \to \mathbb{R}^3$, defined by

$$T_4\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 4 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

is neither one-to-one nor onto. The column vectors are linearly dependent and can not span $\mathbb{R}^3$.

(e) The linear transformation $T_5 : \mathbb{R}^3 \to \mathbb{R}^3$, defined by

$$T_4\left(\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix},$$

is both one-to-one and onto. The column vectors are linearly independent and span $\mathbb{R}^3$.

Corollary 2.15. *Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be defined by $T(\boldsymbol{x}) = A\boldsymbol{x}$.*

(a) *If T is one-to-one, then $n \le m$.*

(b) *If T is onto, then $n \ge m$.*

(c) *If T is one-to-one and onto, then $m = n$.*

Proof. (a) Since T is one-to-one, the system $A\boldsymbol{x} = \mathbf{0}$ has only the trivial solution. Then the reduced row echelon form of A must be of the form

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}.$$

This shows that every column of A has a pivot position. Thus $n \le m$.

(b) Since T is onto, the reduced row echelon form B for A can not have zero rows. This means that every row of B has a pivot position. Note that every pivot positions must be in different columns. Hence $n \ge m$. □

Corollary 2.16. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation. The following statements are equivalent.*

(a) *T is one-to-one.*

(b) *T is onto.*

(c) *T is one-to-one and onto.*

3 Invertible Matrices

Definition 3.1. A function $f : X \to Y$ is called **invertible** if there is a function $g : Y \to X$ such that

$$g \circ f(x) = x \quad \text{for all} \quad x \in X,$$

$$f \circ g(y) = y \quad \text{for all} \quad y \in Y.$$

If so, the function g is called the **inverse** of f, and is denoted by $g = f^{-1}$. The function $Id : X \to X$, defined by $Id(x) = x$, is called the **identity function**.

Theorem 3.2. *A function $f : X \to Y$ is invertible if and only if f is one-to-one and onto.*

Proof. Assume that f is invertible. For two distinct elements u and v of X, if $f(u)$ and $f(v)$ are not distinct, that is, $f(u) = f(v)$, then $u = g(f(u)) = g(f(v)) = v$, a contradiction. This means that f is 1-1. For any element w of Y, consider the element $u = g(w)$ of X. We have $f(u) = f(g(w)) = w$. This shows that f is onto.

Conversely, assume that f is 1-1 and onto. For any element y of Y, there exists a unique element x in X such that $f(x) = y$. Define $g : Y \to X$ by $g(y) = x$, where $f(x) = y$. Then g is the inverse of f. □

Note. If a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is invertible, then T must be $1-1$ and onto. Hence $m = n$.

Definition 3.3. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is called **invertible** if there is a linear transformation $S : \mathbb{R}^n \to \mathbb{R}^n$ such that

$$S \circ T = \mathrm{Id}_n \quad \text{and} \quad T \circ S = \mathrm{Id}_n,$$

where $\mathrm{Id}_n : \mathbb{R}^n \to \mathbb{R}^n$ is the **identity transformation** defined by $\mathrm{Id}_n(\boldsymbol{x}) = \boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{R}^n$. The standard matrix of Id_n is the **identity matrix**

$$I_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

Definition 3.4. A square $n \times n$ matrix A is called **invertible** if there is a square $n \times n$ matrix B such that

$$AB = I_n = BA.$$

If A is invertible, the matrix B is called the **inverse** of A, and is denoted by $B = A^{-1}$.

Theorem 3.5. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation.*

(a) *If T is 1-1, then T is invertible.*

(b) *If T is onto, then T is invertible.*

In order to find out whether an $n \times n$ matrix A is invertible or not, it is to decide whether the matrix equation

$$AX = I_n$$

has a solution, and if it has a solution, the solution matrix is the inverse of A. Write $X = [\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n]$. Then $AX = I_n$ is equivalent to solving the following linear systems

$$A\boldsymbol{x}_1 = \boldsymbol{e}_1, \quad A\boldsymbol{x}_2 = \boldsymbol{e}_2, \quad \ldots, \quad A\boldsymbol{x}_n = \boldsymbol{e}_n.$$

Perform row operation to the corresponding augmented matrices; we have

$$[A \,|\, \boldsymbol{e}_1] \overset{\ell}{\sim} [I_n \,|\, \boldsymbol{b}_1], \quad [A \,|\, \boldsymbol{e}_2] \overset{\ell}{\sim} [I_n \,|\, \boldsymbol{b}_2], \quad \ldots, \quad [A \,|\, \boldsymbol{e}_n] \overset{\ell}{\sim} [I_n \,|\, \boldsymbol{b}_n].$$

The corresponding solutions $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n$ can be obtained simultaneously by applying the same row operations to $\big[A \,|\, I\big]$, i.e.,

$$\big[A \,|\, I\big] = \big[A \,|\, \boldsymbol{e}_1, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\big] \overset{\ell}{\sim} \big[A \,|\, \boldsymbol{b}_1, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_n\big] = \big[I \,|\, B\big].$$

Then B is the inverse of A.

Example 3.1. The matrix $A = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}$ is invertible.

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 & 1 & 0 \\ 1 & 2 & 4 & 0 & 0 & 1 \end{array}\right] \quad \begin{array}{c} R_2 - 2R_1 \\ \sim \\ R_3 - R_1 \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & -1 & -3 & -2 & 1 & 0 \\ 0 & 1 & 2 & -1 & 0 & 1 \end{array}\right] \quad \begin{array}{c} R_3 + R_2 \\ \sim \\ (-1)R_2 \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 3 & 2 & -1 & 0 \\ 0 & 0 & -1 & -3 & 1 & 1 \end{array}\right] \quad \begin{array}{c} R_2 + 3R_3 \\ \sim \\ R_1 + 2R_3 \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 0 & -5 & 2 & 2 \\ 0 & 1 & 0 & -7 & 2 & 3 \\ 0 & 0 & -1 & -3 & 1 & 1 \end{array}\right] \quad \begin{array}{c} R_1 - R_2 \\ \sim \\ (-1)R_3 \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & 2 & 0 & -1 \\ 0 & 1 & 0 & -7 & 2 & 3 \\ 0 & 0 & 1 & 3 & -1 & -1 \end{array}\right],$$

$$\begin{bmatrix} 1 & 1 & 2 \\ 2 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & 0 & -1 \\ -7 & 2 & 3 \\ 3 & -1 & -1 \end{bmatrix}.$$

Example 3.2. The matrix $\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 2 \\ 3 & 4 & 1 \end{bmatrix}$ is not invertible.

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 2 & 3 & 2 & 0 & 1 & 0 \\ 3 & 4 & 1 & 0 & 0 & 1 \end{array}\right] \quad \begin{array}{c} R_2 - 2R_1 \\ \to \\ R_3 - 3R_1 \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & -4 & -2 & 1 & 0 \\ 0 & -2 & -8 & -3 & 0 & 1 \end{array}\right] \quad \begin{array}{c} (-1)R_2 \\ \to \\ R_3 - 2R_2 \end{array}$$

$$\left[\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & 1 & 4 & 2 & -1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{array}\right].$$

Proposition 3.6. *If A and B are $n \times n$ invertible matrices, then*

$$(A^T)^{-1} = (A^{-1})^T,$$

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Proof. Since taking transposition reverses the order of multiplication, we have

$$A^T(A^{-1})^T = (A^{-1}A)^T = I^T = I,$$

$$(A^{-1})^T A^T = (AA^{-1})^T = I^T = I.$$

By definition of inverse matrix, we conclude $(A^T)^{-1} = (A^{-1})^T$.

Since

$$B^{-1}A^{-1}AB = B^{-1}B = I = AA^{-1} = ABB^{-1}A^{-1}.$$

By definition of invertible matrix, we conclude $(AB)^{-1} = B^{-1}A^{-1}$. □

Proposition 3.7. *For any 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, if $\det A := ad - bc \neq 0$, then*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Proof. Assume $ad - bc = 0$, that is, $ad = bc$. If $ad = 0$, then $a = 0$ or $d = 0$; $b = 0$ or $c = 0$. The matrix contains either a zero row or zero column; so it is not invertible. If $ad \neq 0$, then a, b, c, and d are all nonzero. Thus $a/b = c/d$. This means that the two columns are linearly dependent. So the matrix is not invertible.

Assume $ad - bc \neq 0$; we have

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} ad - bc & 0 \\ 0 & ad - bc \end{bmatrix} = (ad - bc)\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

□

Theorem 3.8 (Characterization of Invertible Matrices). *Let A be an $n \times n$ square matrix. Then the following statements are equivalent.*

(a) *A is invertible.*

(b) *The reduced row echelon form of A is the identity matrix I_n.*

(c) *A has n pivot positions.*

(d) *$A\boldsymbol{x} = \mathbf{0}$ has the only trivial solution.*

(e) *The column vectors of A are linearly independent.*

(f) *The linear transformation $T(\boldsymbol{x}) = A\boldsymbol{x}$ is one-to-one.*

(g) *The equation $A\boldsymbol{x} = \boldsymbol{b}$ has at least one solution for any $b \in \mathbb{R}^n$.*

(h) *The column vectors of A span $\mathbb{R}^n$.*

(i) *The linear transformation $T(\boldsymbol{x}) = A\boldsymbol{x}$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ is onto.*

(j) *There is an $n \times n$ square matrix B such that $BA = I_n$.*

(k) *There is an $n \times n$ square matrix B such that $AB = I_n$.*

(l) *A^{τ} is invertible.*

Proof. It is straightforward that

$$(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (d) \Rightarrow (e) \Rightarrow (f) \Rightarrow (g) \Rightarrow (h) \Rightarrow (i) \Rightarrow (l) \Rightarrow (T \text{ is invertible}) \Rightarrow (a).$$

Let $S(x) = Bx$. Then

$$\text{(j)} \Rightarrow (T \text{ is one-to-one}) \Rightarrow (S \text{ is the inverse of } T) \Rightarrow (B \text{ is the inverse of } A) \Rightarrow \text{(k)} \Rightarrow \text{(j)}.$$

□

4 Determinants

Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be a linear transformation with the standard matrix

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Recall that the determinant of a 2×2 matrix $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is defined by

$$\det \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Let $A = [a_{ij}]$ be an $n \times n$ square matrix. For a fixed (i, j), where $1 \le i \le m$ and $1 \le j \le n$, let

$$\begin{aligned} A_{ij} \quad = \quad & (n-1) \times (n-1) \text{ submatrix of } A \text{ obtained} \\ & \text{by deleting the } i\text{th row and } j\text{th column of } A. \end{aligned}$$

Definition 4.1. Let $A = [a_{ij}]$ be an $n \times n$ square matrix. The **determinant** of A is a number $\det A$, inductively defined by

$$\begin{aligned} \det A &= a_{11} \det A_{11} - a_{12} \det A_{12} + \cdots + (-1)^{n+1} a_{1n} \det A_{1n} \\ &= a_{11} \det A_{11} - a_{21} \det A_{21} + \cdots + (-1)^{n+1} a_{n1} \det A_{n1}. \end{aligned}$$

Theorem 4.2. *Let $A = [a_{ij}]$ be an $n \times n$ matrix. The (i, j)-**cofactor** of A $(1 \le i, j \le n)$ is the number*

$$C_{ij} = (-1)^{i+j} \det A_{ij}.$$

Then

$$\begin{aligned} \det A &= a_{11}C_{11} + a_{12}C_{12} + \cdots + a_{1n}C_{1n} \\ &= a_{11}C_{11} + a_{21}C_{21} + \cdots + a_{n1}C_{n1}. \end{aligned}$$

Example 4.1. For any 3×3 matrix,

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}.$$

Theorem 4.3 (Cofactor Expansion Formula). *For any $n \times n$ square matrix $A = [a_{ij}]$,*

$$\begin{aligned} \det A &= a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} \\ &= a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}. \end{aligned}$$

In other words,

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

Proposition 4.4.

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} = a_{11}a_{22}\cdots a_{nn}.$$

Theorem 4.5. *Determinant satisfies the following properties.*

(a) *Adding a multiple of one row (column) to another row (column) does not change the determinant.*

(b) *Interchanging two rows (columns) changes the sign of the determinant.*

(c) *If two rows (columns) are the same, then the determinant is zero.*

(d) *If one row (column) of A is multiplied by a scalar γ to produce a matrix B, then*

$$\det B = \gamma \det A.$$

Proof. Consider $n \times n$ matrices. We proceed induction on n.
(a) For $n = 2$, we have

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} + ca_{11} & a_{22} + ca_{12} \end{vmatrix} &= a_{11}(a_{22} + ca_{12}) - a_{12}(a_{21} + ca_{11}) \\ &= a_{11}a_{22} - a_{12}a_{21} \\ &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}; \\ \begin{vmatrix} a_{11} + ca_{21} & a_{12} + ca_{22} \\ a_{21} & a_{22} \end{vmatrix} &= (a_{11} + ca_{21})a_{22} - (a_{12} + ca_{22})a_{21} \\ &= a_{11}a_{22} - a_{12}a_{21} \\ &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}. \end{aligned}$$

Suppose it is true for all $(n-1) \times (n-1)$ matrices. Consider the case $R_j + cR_i$. We divide the case into two subcases: (1) $i = 1$ and (2) $i \geq 2$. For $i = 1$, we have □

Theorem 4.6 (The Algorithm for Determinant). *Let $A = [a_{ij}]$ be an $n \times n$ matrix. If*

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{array}{c} \text{interschanging rows} \\ \sim \\ \text{adding a multiple of} \\ \text{one row to another} \end{array} \begin{bmatrix} a_1 & * & \cdots & * \\ 0 & a_2 & \cdots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_n \end{bmatrix}.$$

Then

$$\det A = (-1)^k a_1 a_2 \cdots a_n,$$

where k is the number times of interchanging two rows.

Example 4.2. The determinant

$$\begin{vmatrix} 1 & 2 & 4 \\ 2 & 3 & 5 \\ 3 & 4 & 8 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 4 \\ 0 & -1 & -3 \\ 0 & -2 & -4 \end{vmatrix} = - \begin{vmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \\ 0 & -2 & -4 \end{vmatrix} = - \begin{vmatrix} 1 & 2 & 4 \\ 0 & 1 & 3 \\ 0 & 0 & 2 \end{vmatrix} = -2.$$

Theorem 4.7. *Let A be an $n \times n$ matrix, and let C_{ij} be the (i,j)-cofactor of A, i.e.,*

$$C_{ij} = (-1)^{i+j} \det A_{ij}.$$

Then we have

$$\sum_{k=1}^{n} a_{ik} C_{jk} = a_{i1} C_{j1} + a_{i2} C_{j2} + \cdots + a_{in} C_{jn} = \begin{cases} \det A & \text{if } \ i = j \\ 0 & \text{if } \ i \neq j. \end{cases}$$

Proof. For $i = j$, it follows from the cofactor expansion formula. For $i \neq j$, let us verify for 3×3 matrices. For instance, for $i = 1$ and $j = 3$, we have

$$\begin{aligned} a_{11}C_{31} + a_{12}C_{32} + a_{13}C_{33} &= a_{11} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} - a_{12} \begin{vmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{vmatrix} + a_{13} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \\ &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \end{vmatrix} = 0. \end{aligned}$$

□

Definition 4.8. For an $n \times n$ matrix $A = [a_{ij}]$, the **classical adjoint matrix** of A is the $n \times n$ matrix $\operatorname{adj} A$ whose (i,j)-entry is the (j,i)-cofactor C_{ji}, that is,

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

Theorem 4.9. *For any $n \times n$ matrix A,*

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

Proof. We verify for the case of 3×3 matrices.

$$\begin{aligned} & \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} \\ = & \begin{bmatrix} \sum_{k=1}^3 a_{1k}C_{1k} & \sum_{k=1}^3 a_{1k}C_{2k} & \sum_{k=1}^3 a_{1k}C_{3k} \\ \sum_{k=1}^3 a_{2k}C_{1k} & \sum_{k=1}^3 a_{2k}C_{2k} & \sum_{k=1}^3 a_{2k}C_{3k} \\ \sum_{k=1}^3 a_{3k}C_{1k} & \sum_{k=1}^3 a_{3k}C_{2k} & \sum_{k=1}^3 a_{3k}C_{3k} \end{bmatrix} \\ = & \begin{bmatrix} \det A & 0 & 0 \\ 0 & \det A & 0 \\ 0 & 0 & \det A \end{bmatrix} \end{aligned}$$

□

Theorem 4.10. *An $n \times n$ matrix A is invertible if and only if* $\det A \neq 0$.

5 Elementary Matrices

Definition 5.1. The following $n \times n$ matrices are called **elementary matrices**.

$$E\big(c(i)\big) = \begin{array}{cl} \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & O & \\ & & 1 & & & & \\ & & & c & & & \\ & & & & 1 & & \\ & O & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} & \begin{array}{l} \\ \\ \\ i\text{th}, \quad c \neq 0; \\ \\ \\ \\ \end{array} \end{array}$$

$$E(i,j) = \begin{array}{cl} \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & O & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & O & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} & \begin{array}{l} \\ \\ i\text{th} \\ \\ j\text{th} \\ \\ \\ \end{array} \end{array} ;$$

$$E\big(j + c(i)\big) = \begin{array}{cl} \begin{bmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & 0 & & \\ & & \vdots & \ddots & \vdots & & \\ & & c & \cdots & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{bmatrix} & \begin{array}{l} \\ \\ i\text{th} \\ \\ j\text{th} \\ \\ \\ \end{array} \end{array} .$$

Let I_n be the $n \times n$ identity matrix. We have

$$\begin{aligned} I_n &\overset{cR_i}{\to} E\big(c(i)\big), \\ I_n &\overset{R_i \leftrightarrow R_j}{\to} E(i,j), \\ I_n &\overset{R_j + cR_i}{\to} E\big(j + c(i)\big). \end{aligned}$$

Proposition 5.2. *Let A be an $m \times n$ matrix. Then*

(a) *$E\big(c(i)\big)A$ is the matrix obtained from A by multiplying c to the ith row.*

(b) *$E(i,j)A$ is the matrix obtained from A by interchanging the ith row and the jth row.*

(c) *$E\big(j + c(i)\big)A$ is the matrix obtained from A by adding the c multiple of the ith row to the jth row.*

Proposition 5.3. *Elementary matrices are invertible. Moreover,*

$$\begin{aligned} E\big(c(i)\big)^{-1} &= E(\tfrac{1}{c}(i)), \\ E(i,j)^{-1} &= (j,i), \\ E\big(j + c(i)\big)^{-1} &= E\big(j - c(i)\big). \end{aligned}$$

Theorem 5.4. *A square matrix A is invertible if and only if it can be written as product*

$$A = E_1 E_2 \cdots E_p,$$

of some elementary matrices E_1, E_2, $\ldots$, E_p.

Theorem 5.5. *A square matrix A is invertible if and only if* $\det A \neq 0$.

Proposition 5.6. *Let A and B $n \times n$ matrices. Then*

$$\det AB = \det A \det B.$$

Proof. Case 1: $\det A = 0$. Then A is not invertible by Theorem 4.10. We claim that AB is not invertible. Otherwise, if AB is invertible, then there is a matric C such that $(AB)C = I$; thus $A(BC) = I$; this means that A is invertible, contradictory to that $\det A \neq 0$. Again by Theorem 4.10, $\det(AB) = 0$. Therefore $\det AB = \det A \det B$.

Case 2: $\det A \neq 0$. It is easy to check directly that

$$\det EB = \det E \det B$$

for any elementary matrix E. Let A be written as $A = E_1E_2 \cdots E_k$ for some elementary matrices E_1, E_2, …, E_k. Then

$$\begin{aligned} \det AB &= \det(E_1E_2 \cdots E_kB) = (\det E_1)\det(E_2 \cdots E_kB) \\ &= (\det E_1)(\det E_2)\det(E_3 \cdots E_kB) = \cdots \\ &= \det E_1 \det E_2 \cdots \det E_k \det B = \det A \det B. \end{aligned}$$

□

6 LU-Decomposition

Theorem 6.1. *Let A be an $m \times n$ matrix. If A can be reduced to its row echelon form by elementary row operations without switching rows, then A admits a LU-decomposition, that is, there is a lower triangular $m \times m$ matrix L and an $m \times n$ upper triangular matrix U such that $A = LU$.*

Proof. Let $\rho_1, \rho_2, \ldots, \rho_k$ be a sequence of row operations such that

$$A \overset{\rho_1}{\sim} A_1 \overset{\rho_2}{\sim} \cdots \overset{\rho_k}{\sim} A_k = U,$$

where U is upper triangular. Let $E_1, E_2, \ldots, E_k$ be elementary matrices corresponding the row operations $\rho_1, \rho_2, \ldots, \rho_k$, respectively. Then $E_k \cdots E_2E_1A = U$. Let $L = E_1^{-1}E_2^{-1} \cdots E_k^{-1}$. Then L is a lower triangular and $A = LU$. □

Example 6.1.

7 Interpretation of Determinant

Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ be a 2×2 matrix. Let $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ be the standard basis of $\mathbb{R}^3$. Set

$$\boldsymbol{a}_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ 0 \end{bmatrix}, \ \boldsymbol{a}_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ 0 \end{bmatrix}.$$

Then the area of the parallelepiped spanned by the two vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ is the length of the cross product $\boldsymbol{a}_1 \times \boldsymbol{a}_2$, that is,

$$\begin{aligned} |\boldsymbol{a}_1 \times \boldsymbol{a}_2| &= |(a_{11}\boldsymbol{e}_1 + a_{21}\boldsymbol{e}_2) \times (a_{12}\boldsymbol{e}_1 + a_{22}\boldsymbol{e}_2)| \\ &= |a_{11}a_{22}\boldsymbol{e}_3 - a_{21}a_{22}\boldsymbol{e}_3| \\ &= |a_{11}a_{22} - a_{21}a_{22}|. \end{aligned}$$

Let $A = [a_{ij}] = [\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3]$ be a 3×3 matrix. The volume of the parallelotope spanned by the three vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\boldsymbol{a}_3$ is the length of the triple $(\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{a}_3$, that is,

$$\begin{aligned} |(\boldsymbol{a}_1 \times \boldsymbol{a}_2) \cdot \boldsymbol{a}_3| &= |\boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)| \\ &= |a_{11}\det A_{11} - a_{21}\det A_{21} + a_{31}\det A_{31}|. \end{aligned}$$

Example 7.1. The Vandermonde determinant is

$$\begin{aligned}
\begin{vmatrix} 1 & 1 & 1 \\ a_1 & a_2 & a_3 \\ a_1^2 & a_2^2 & a_3^2 \end{vmatrix} &= \begin{vmatrix} 1 & 1 & 1 \\ 0 & a_2 - a_1 & a_3 - a_1 \\ 0 & a_2^2 - a_2 a_1 & a_3^2 - a_3 a_1 \end{vmatrix} \\
&= \begin{vmatrix} a_2 - a_1 & a_3 - a_1 \\ a_2(a_2 - a_1) & a_3(a_3 - a_1) \end{vmatrix} \\
&= (a_3 - a_2)(a_3 - a_1)(a_2 - a_1).
\end{aligned}$$

$$\begin{vmatrix} 1 & 1 & 1 & \cdots & 1 \\ a_1 & a_2 & a_3 & \cdots & a_n \\ a_1^2 & a_2^2 & a_3^2 & \cdots & a_n^2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_1^{n-1} & a_2^{n-1} & a_3^{n-1} & \cdots & a_n^{n-1} \end{vmatrix} = \prod_{j>i} (a_j - a_i).$$

Example 7.2.

$$\begin{aligned}
\begin{vmatrix} 1 & -3 & 1 & -2 \\ 2 & -5 & -1 & -2 \\ 0 & -4 & 5 & 1 \\ -3 & 10 & -6 & 7 \end{vmatrix} &\overset{R_2 - 2R_1}{\underset{R_4 + 3R_1}{=}} \begin{vmatrix} 1 & -3 & 1 & -2 \\ 0 & 1 & -3 & 2 \\ 0 & -4 & 5 & 1 \\ 0 & 1 & -3 & 1 \end{vmatrix} \\
&\overset{R_3 + 4R_2}{\underset{R_4 - R_2}{=}} \begin{vmatrix} 1 & -3 & 1 & -2 \\ 0 & 1 & -3 & 2 \\ 0 & 0 & -7 & 9 \\ 0 & 0 & 0 & -1 \end{vmatrix} = 7.
\end{aligned}$$

Theorem 7.1 (Cramer's Rule). *Let $A = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n]$ be an $n \times n$ invertible matrix with the column vectors $\boldsymbol{a}_1$, $\boldsymbol{a}_2$, $\ldots$, $\boldsymbol{a}_n$. Then for any $\boldsymbol{b} \in \mathbb{R}^n$, the unique solution of the system*

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

is given by

$$x_i = \frac{\det A_i}{\det A}, \quad i = 1, 2, \ldots, n$$

where

$$A_i = [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{i-1}, \boldsymbol{b}, \boldsymbol{a}_{i+1}, \ldots, \boldsymbol{a}_n], \quad 1 \le i \le n.$$

Proof. If a vector $\boldsymbol{x}$ is the solution of the system, then

$$\begin{aligned}
A[\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{i-1}, \boldsymbol{x}, \boldsymbol{e}_{i+1}, \ldots, \boldsymbol{e}_n] &= [A\boldsymbol{e}_1, \ldots, A\boldsymbol{e}_{i-1}, A\boldsymbol{x}, A\boldsymbol{e}_{i+1}, \ldots, A\boldsymbol{e}_n] \\
&= [\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{i-1}, \boldsymbol{b}, \boldsymbol{a}_{i+1}, \ldots, \boldsymbol{a}_n].
\end{aligned}$$

Taking determinant of both sides, we have

$$(\det A) x_i = \det A_i, \quad 1 \le i \le n.$$

□