Chapter 1

Linear Algebra

1.1 Vector Spaces

Definition 1.1 (Vector spaces)**.** A ***vector space*** $\mathcal{V}$ over a scalar field $\mathbb{F}$ is a set of elements called ***vectors***, together with two operations $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ and $\cdot : \mathbb{F} \times \mathcal{V} \to \mathcal{V}$, called the ***vector addition*** and ***scalar multiplication*** respectively, such that

1. $\boldsymbol{v} + \boldsymbol{w} = \boldsymbol{w} + \boldsymbol{v}$ for all $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$.
2. $(\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w} = \boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w})$ for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$.
3. There is a zero vector $\mathbf{0}$ such that $\boldsymbol{v} + \mathbf{0} = \boldsymbol{v}$ for all $\boldsymbol{v} \in \mathcal{V}$.
4. For every $\boldsymbol{v}$ in $\mathcal{V}$, there is a vector $\boldsymbol{w}$ such that $\boldsymbol{v} + \boldsymbol{w} = \mathbf{0}$.
5. $\alpha \cdot (\boldsymbol{v} + \boldsymbol{w}) = \alpha \cdot \boldsymbol{v} + \alpha \cdot \boldsymbol{w}$ for all $\alpha \in \mathbb{F}$ and $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$.
6. $\alpha \cdot (\beta \cdot \boldsymbol{v}) = (\alpha\beta) \cdot \boldsymbol{v}$ for all $\alpha, \beta \in \mathbb{F}$ and $\boldsymbol{v} \in \mathcal{V}$.
7. $(\alpha + \beta) \cdot \boldsymbol{v} = \alpha \cdot \boldsymbol{v} + \beta \cdot \boldsymbol{v}$ for all $\alpha, \beta \in \mathbb{F}$ and $\boldsymbol{v} \in \mathcal{V}$.
8. $1 \cdot \boldsymbol{v} = \boldsymbol{v}$ for all $\boldsymbol{v} \in \mathcal{V}$.

For notational convenience, we often drop the $\cdot$ and write $\alpha\boldsymbol{v}$ instead of $\alpha \cdot \boldsymbol{v}$.

Remark 1.2. In property 4 of the definition above, it is easy to see that for each $\boldsymbol{v}$, there is only one vector $\boldsymbol{w}$ such that $\boldsymbol{v} + \boldsymbol{w} = \mathbf{0}$. We often denote this $\boldsymbol{w}$ by $-\boldsymbol{v}$, and the vector substraction $- : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$ is then defined (or understood) as $\boldsymbol{v} - \boldsymbol{w} = \boldsymbol{v} + (-\boldsymbol{w})$.

Example 1.3. Let $\mathbb{F}$ be a scalar field. The space $\mathbb{F}^{\mathrm{n}}$ is the collection of n-tuple $\boldsymbol{v} = (\mathrm{v}_1, \mathrm{v}_2, \cdots, \mathrm{v}_\mathrm{n})$ with $\mathrm{v}_i \in \mathbb{F}$ with addition $+$ and scalar multiplication $\cdot$ defined by

$$(\mathrm{v}_1, \cdots, \mathrm{v}_\mathrm{n}) + (\mathrm{w}_1, \cdots, \mathrm{w}_\mathrm{n}) \equiv (\mathrm{v}_1 + \mathrm{w}_1, \cdots, \mathrm{v}_\mathrm{n} + \mathrm{w}_\mathrm{n}),$$
$$\alpha(\mathrm{v}_1, \cdots, \mathrm{v}_\mathrm{n}) \equiv (\alpha\mathrm{v}_1, \cdots, \alpha\mathrm{v}_\mathrm{n}).$$

Then $\mathbb{F}^{\mathrm{n}}$ is a vector space.

Example 1.4. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $\mathcal{V}$ be the collection of all $\mathbb{R}$-valued continuous functions on $[0, 1]$. The vector addition $+$ and scalar multiplication $\cdot$ is defined by

$$(f+g)(x) = f(x) + g(x) \qquad \forall\, f, g \in \mathcal{V},$$
$$(\alpha \cdot f)(x) = \alpha f(x) \qquad \forall\, f \in \mathcal{V}, \alpha \in \mathbb{F}.$$

Then $\mathcal{V}$ is a vector space, and is denoted by $\mathscr{C}([0, 1]; \mathbb{F})$. When the scalar field under consideration is clear, we simply use $\mathscr{C}([0, 1])$ to denote this vector space.

Definition 1.5 (Vector subspace)**.** Let $\mathcal{V}$ be a vector space over scalar field $\mathbb{F}$. A subset $\mathcal{W} \subseteq \mathcal{V}$ is called a vector subspace of $\mathcal{V}$ if itself is a vector space over $\mathbb{F}$.

1.1.1 The linear independence of vectors

Definition 1.6. Let $\mathcal{V}$ be a vector space over a scalar field $\mathbb{F}$. k vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\cdots$, $\boldsymbol{v}_k$ in $\mathcal{V}$ is said to be ***linearly dependent*** if there exists $(\alpha_1, \cdots, \alpha_k) \subseteq \mathbb{F}^k$, $(\alpha_1, \cdots, \alpha_k) \neq \mathbf{0}$ such that $\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_k \boldsymbol{v}_k = \mathbf{0}$. k vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\cdots$, $\boldsymbol{v}_k$ in $\mathcal{V}$ is said to be ***linearly independent*** if they are not linearly dependent. In other words, $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ are linearly independent if

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots + \alpha_k \boldsymbol{v}_k = \mathbf{0} \quad \Rightarrow \quad \alpha_1 = \alpha_2 = \cdots = \alpha_k = 0.$$

Example 1.7. The k vectors $\{1, x, x^2, \cdots, x^{k-1}\}$ are linearly independent in $\mathscr{C}([0, 1])$ for all $k \in \mathbb{N}$.

1.1.2 The dimension of a vector space

Definition 1.8. The ***dimension*** of a vector space $\mathcal{V}$ is the number of maximum linearly independent set in $\mathcal{V}$, and in such case $\mathcal{V}$ is called an n-dimensional vector space, where n the the dimension of $\mathcal{V}$. If for every number $\mathrm{n} \in \mathbb{N}$ there exists n linearly independent vectors in $\mathcal{V}$, the vector space $\mathcal{V}$ is said to be infinitely dimensional.

Example 1.9. The space $\mathbb{F}^{\mathrm{n}}$ is n-dimensional, and $\mathscr{C}([0,1])$ is infinitely dimensional (since $1, x, \cdots, x^{\mathrm{n}-1}$ are n linearly independent vectors in $\mathscr{C}([0,1])$).

1.1.3 Bases of a vector space

Definition 1.10 (Basis)**.** Let $\mathcal{V}$ be a vector space over $\mathbb{F}$. A set of vectors $\{\boldsymbol{v}_i\}_{i\in\mathcal{I}}$ in $\mathcal{V}$ is called a ***basis*** of $\mathcal{V}$ if for every $\boldsymbol{v} \in \mathcal{V}$, there exists a unique $\{\alpha_i\}_{i\in\mathcal{I}} \subseteq \mathbb{F}$ such that

$$\boldsymbol{v} = \sum_{\alpha\in\mathcal{I}} \alpha_i \boldsymbol{v}_i \,.$$

For a given basis $\mathcal{B} = \{\boldsymbol{v}_i\}_{i\in\mathcal{I}}$, the coefficients $\{\alpha_i\}_{i\in\mathcal{I}}$ given in the above relation is denoted by $[\boldsymbol{v}]_{\mathcal{B}}$.

Example 1.11 (Standard Basis of $\mathbb{F}^{\mathrm{n}}$)**.** Let $\mathrm{e}_i = (0,, \cdots, 0, 1, 0, \cdots, 0)$, where 1 locates at the i-th slot. Then the collection $\{\mathrm{e}_i\}_{i=1}^{\mathrm{n}}$ is a basis of the vector space $\mathbb{F}^{\mathrm{n}}$ over $\mathbb{F}$ since

$$(\alpha_1, \cdots, \alpha_{\mathrm{n}}) = \sum_{i=1}^{\mathrm{n}} \alpha_i \mathrm{e}_i \qquad \forall\, \alpha_i \in \mathbb{F}.$$

The collection $\{\mathrm{e}_i\}_{i=1}^{\mathrm{n}}$ is called the ***standard basis*** of $\mathbb{F}^{\mathrm{n}}$.

Example 1.12. Even though $\{1, x, \cdots, x^k, \cdots\}$ is a set of linearly independent vectors, it is not a basis of $\mathscr{C}([0,1])$. However, let $\mathscr{P}([0,1])$ be the collection of polynomials defined on $[0,1]$. Then $\mathscr{P}([0,1])$ is still a vector space, and $\{1, x, \cdots, x^k, \cdots\}$ is a basis of $\mathscr{P}([0,1])$.

1.2 Inner Products and Inner Product Spaces

Definition 1.13 (Inner product space)**.** Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A vector space $\mathcal{V}$ over a scalar field $\mathbb{F}$ with a bilinear form $(\cdot,\cdot) : \mathcal{V} \times \mathcal{V} \to \mathbb{F}$ is called an ***inner product space*** if the bilinear form satisfies

1. $(\boldsymbol{v}, \boldsymbol{v}) \geqslant 0$ for all $\boldsymbol{v} \in \mathcal{V}$.

2. $(\boldsymbol{v}, \boldsymbol{v}) = 0$ if and only if $\boldsymbol{v} = 0$.

3. $(\boldsymbol{v}, \boldsymbol{w}) = \overline{(\boldsymbol{w}, \boldsymbol{v})}$ for all $\boldsymbol{v}, w \in \mathcal{V}$, where the bar over the scalar $(\boldsymbol{w}, \boldsymbol{v})$ is the complex conjugate.

4. $(\boldsymbol{v} + \boldsymbol{w}, \boldsymbol{u}) = (\boldsymbol{v}, \boldsymbol{u}) + (\boldsymbol{w}, \boldsymbol{u})$ for all $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$.

5. $(\alpha \boldsymbol{v}, \boldsymbol{w}) = \alpha(\boldsymbol{v}, \boldsymbol{w})$ for all $\alpha \in \mathbb{F}$ and $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$.

The bilinear form $(\cdot, \cdot)$ is called an *inner product* on $\mathcal{V}$.

Example 1.14 (Standard Inner Product on $\mathbb{F}^{\mathrm{n}}$)**.** Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$, and $\mathbb{F}^{\mathrm{n}}$ be the vector space defined in Example 1.3. A special inner product on the vector space $\mathbb{F}^{\mathrm{n}}$ over $\mathbb{F}$, called the ***standard inner product*** on $\mathbb{F}^{\mathrm{n}}$, is defined by

$$(\boldsymbol{v}, \boldsymbol{w}) \equiv \sum_{i=1}^{\mathrm{n}} \mathrm{v}_i \overline{\mathrm{w}}_i \,,$$

where v_i and w_i are the i-th component of $\boldsymbol{v}$ and $\boldsymbol{w}$, respectively, and $\overline{\mathrm{w}}_i$ is the complex conjugate of w_i. We sometimes use $\boldsymbol{v} \cdot \boldsymbol{w}$ to denote $(\boldsymbol{v}, \boldsymbol{w})$.

Example 1.15. Let $\mathcal{V} = \mathscr{C}([0,1];\mathbb{R})$. Define

$$(f, g) = \int_0^1 f(x) g(x) dx \,.$$

Then $\big(\mathscr{C}([0,1];\mathbb{R}), (\cdot,\cdot)\big)$ is an inner product space. The norm induced by this inner product is given by

$$\|f\| = \left[\int_0^1 |f(x)|^2 dx\right]^{\frac{1}{2}},$$

and is called the L^2-norm.

Proposition 1.16. *Let $\mathcal{V}$ be an inner product space with inner product $(\cdot,\cdot)$. The inner product $(\cdot,\cdot)$ on $\mathcal{V}$ induces a norm defined by*

$$\|\boldsymbol{v}\| \equiv \sqrt{(\boldsymbol{v}, \boldsymbol{v})}$$

satisfying

1. $\|\boldsymbol{v}\| \geqslant 0$ *for all* $\boldsymbol{v} \in \mathcal{V}$.

2. $\|\boldsymbol{v}\| = 0$ *if and only if* $\boldsymbol{v} = 0$.

3. $\|\alpha \boldsymbol{v}\| = |\alpha| \|\boldsymbol{v}\|$ *for all* $\alpha \in \mathbb{F}$ *and* $\boldsymbol{v} \in \mathcal{V}$.

4. $\|\boldsymbol{v} + \boldsymbol{w}\| \leqslant \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$ *for all* $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$.

5. $|(\boldsymbol{v}, \boldsymbol{w})| \leqslant \|\boldsymbol{v}\| \|\boldsymbol{w}\|$ *for all* $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$.

Proof. Properties 1 through 3 are obvious. We focus on proving property 5 first, and as we will see, property 4 is a direct consequence of property 5.

Let $\alpha \in \mathbb{F}$ satisfy $\alpha(\boldsymbol{v}, \boldsymbol{w}) = |(\boldsymbol{v}, \boldsymbol{w})|$. Then $|\alpha| = 1$. For all $\lambda \in \mathbb{R}$,

$$\begin{aligned}(\lambda\alpha\boldsymbol{v} + \boldsymbol{w}, \lambda\alpha\boldsymbol{v} + \boldsymbol{w}) &= (\lambda\alpha\boldsymbol{v}, \lambda\alpha\boldsymbol{v}) + (\lambda\alpha\boldsymbol{v}, \boldsymbol{w}) + (\boldsymbol{w}, \lambda\alpha\boldsymbol{v}) + (\boldsymbol{w}, \boldsymbol{w}) \\ &= \lambda^2\|\boldsymbol{v}\|^2 + \lambda\alpha(\boldsymbol{v}, \boldsymbol{w}) + \overline{\lambda\alpha(\boldsymbol{v}, \boldsymbol{w})} + \|\boldsymbol{w}\|^2 \\ &= \lambda^2\|\boldsymbol{v}\|^2 + 2\lambda|(\boldsymbol{v}, \boldsymbol{w})| + \|\boldsymbol{w}\|^2 .\end{aligned}$$

Since the left-hand side of the quantity above is always non-negative for all $\lambda \in \mathbb{R}$, we must have

$$|(\boldsymbol{v}, \boldsymbol{w})|^2 - \|\boldsymbol{v}\|^2\|\boldsymbol{w}\|^2 \leqslant 0$$

which implies property 5. To prove property 4, we note that

$$\begin{aligned}\|\boldsymbol{v} + \boldsymbol{w}\| \leqslant \|\boldsymbol{v}\| + \|\boldsymbol{w}\| &\Leftrightarrow \|\boldsymbol{v} + \boldsymbol{w}\|^2 \leqslant (\|\boldsymbol{v}\| + \|\boldsymbol{w}\|)^2 \\ &\Leftrightarrow (\boldsymbol{v} + \boldsymbol{w}, \boldsymbol{v} + \boldsymbol{w}) \leqslant \|\boldsymbol{v}\|^2 + 2\|\boldsymbol{v}\|\|\boldsymbol{w}\| + \|\boldsymbol{w}\|^2 \\ &\Leftrightarrow \mathrm{Re}(\boldsymbol{v}, \boldsymbol{w}) \leqslant \|\boldsymbol{v}\|\|\boldsymbol{w}\|\end{aligned}$$

while the last inequality is valid because of property 5. □

Remark 1.17. The inequality in property 5 is called the ***Cauchy-Schwarz inequality***.

Definition 1.18. Let $\big(\mathcal{V}, (\cdot, \cdot)\big)$ be an inner product space. A basis $\mathcal{B}$ of $\mathcal{V}$ is called ***orthogonal*** if $\boldsymbol{u} \cdot \boldsymbol{v} = 0$ if $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{B}$ and $\boldsymbol{u} \neq \boldsymbol{v}$, and is called ***orthonormal*** if it is an orthogonal basis such that $\|\boldsymbol{v}\| = 1$ for all $\boldsymbol{v} \in \mathcal{B}$.

Definition 1.19 (Orthogoanl complement)**.** Let $\big(\mathcal{V}, (\cdot, \cdot)\big)$ be an inner product space over scalar field $\mathbb{F}$, and $\mathcal{W} \subseteq \mathcal{V}$ be a vector subspace of $\mathcal{V}$. The ***orthogonal complement*** of $\mathcal{W}$, denoted by $\mathcal{W}^{\perp}$, is the set

$$\mathcal{W}^{\perp} = \big\{\boldsymbol{v} \in \mathcal{V} \,\big|\, (\boldsymbol{v}, \boldsymbol{w}) = 0 \text{ for all } \boldsymbol{w} \in \mathcal{W}\big\} .$$

Proposition 1.20. *Let $\big(\mathcal{V}, (\cdot, \cdot)\big)$ be an inner product space over scalar field $\mathbb{F}$, and $\mathcal{W}$ be a vector subspace of $\mathcal{V}$. Then $\mathcal{W}^{\perp}$ is a vector subspace of $\mathcal{V}$.*

1.3 Normed Vector Spaces

The norm introduced in Proposition 1.16 is a good way of measure the magnitude of vectors. In general if a real-valued function can be used as a measurement of the magnitude of vectors if certain properties are satisfied.

Definition 1.21. Let $\mathcal{V}$ be a vector space over scalar field $\mathbb{F}$. A real-valued function $\|\cdot\|: \mathcal{V} \to \mathbb{R}$ is said to be a norm of $\mathcal{V}$ if

1. $\|\boldsymbol{v}\| \geqslant 0$ for all $\boldsymbol{v} \in \mathcal{V}$.

2. $\|\boldsymbol{v}\| = 0$ if and only if $\boldsymbol{v} = 0$.

3. $\|\alpha\boldsymbol{v}\| = |\alpha|\|\boldsymbol{v}\|$ for all $\boldsymbol{v} \in \mathcal{V}$ and $\alpha \in \mathbb{F}$.

4. $\|\boldsymbol{v} + \boldsymbol{w}\| \leqslant \|\boldsymbol{v}\| + \|\boldsymbol{w}\|$ for all $\boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}$.

The pair $(\mathcal{V}, \|\cdot\|)$ is called a normed vector space.

Example 1.22. Let $\mathcal{V} = \mathbb{F}^{\mathrm{n}}$, and $\|\cdot\|_p$ be defined by

$$\|\boldsymbol{x}\|_p = \begin{cases} \Big[\sum\limits_{i=1}^{\mathrm{n}} |x_i|^p\Big]^{\frac{1}{p}} & \text{if} \quad 1 \leqslant p < \infty\,, \\ \max\limits_{1 \leqslant i \leqslant \mathrm{n}} |x_i| & \text{if} \quad p = \infty\,, \end{cases}$$

where $\boldsymbol{x} = (x_1, \cdots, x_{\mathrm{n}})$. The function $\|\cdot\|_p$ is a norm of $\mathbb{F}^{\mathrm{n}}$, and is called the p-norm of $\mathbb{F}^{\mathrm{n}}$.

Theorem 1.23 (Hölder's inequality)**.** *Let $1 \leqslant p \leqslant \infty$. Then*

$$\big|(\boldsymbol{x}, \boldsymbol{y})\big| \leqslant \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_{p'} \qquad \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}^{\mathrm{n}}\,, \tag{1.1}$$

where $(\cdot,\cdot)$ is the standard inner product on $\mathbb{F}^{\mathrm{n}}$ and p' is the conjugate of p satisfying $\dfrac{1}{p} + \dfrac{1}{p'} = 1$.

Proof. Let $\boldsymbol{x} = (x_1, \cdots, x_{\mathrm{n}})$ and $\boldsymbol{y} = (y_1, \cdots, y_{\mathrm{n}})$ be given. Without loss of generality we can assume that $\boldsymbol{x} \neq \boldsymbol{0}$ and $\boldsymbol{y} \neq \boldsymbol{0}$. Define $\widetilde{\boldsymbol{x}} = \boldsymbol{x}/\|\boldsymbol{x}\|_p$ and $\widetilde{\boldsymbol{y}} = \boldsymbol{y}/\|\boldsymbol{y}\|_{p'}$. Then $\|\widetilde{\boldsymbol{x}}\|_p = 1$ and $\|\widetilde{\boldsymbol{y}}\|_{p'} = 1$. By Young's inequality

$$ab \leqslant \frac{1}{p}a^p + \frac{1}{p'}b^{p'} \qquad \forall\, a, b \geqslant 0\,,$$

we find that for $1 < p < \infty$,

$$\begin{aligned}\big|(\widetilde{\boldsymbol{x}}, \widetilde{\boldsymbol{y}})\big| &= \Big|\sum_{k=1}^{\mathrm{n}} \frac{x_k}{\|\boldsymbol{x}\|_p} \frac{y_k}{\|\boldsymbol{y}\|_{p'}}\Big| \leqslant \sum_{k=1}^{\mathrm{n}} \frac{|x_k|}{\|\boldsymbol{x}\|_p} \frac{|y_k|}{\|\boldsymbol{y}\|_{p'}} \leqslant \sum_{k=1}^{\mathrm{n}} \Big(\frac{1}{p} \frac{|x_k|^p}{\|\boldsymbol{x}\|_p^p} + \frac{1}{p'} \frac{|y_k|^{p'}}{\|\boldsymbol{y}\|_{p'}^{p'}}\Big) \\ &= \frac{1}{p\|\boldsymbol{x}\|_p^p} \sum_{k=1}^{\mathrm{n}} |x_k|^p + \frac{1}{p'\|\boldsymbol{y}\|_{p'}^{p'}} \sum_{k=1}^{\mathrm{n}} |y_k|^{p'} = \frac{\|\boldsymbol{x}\|_p^p}{p\|\boldsymbol{x}\|_p^p} + \frac{\|\boldsymbol{y}\|^{p'}}{p'\|\boldsymbol{y}\|_{p'}^{p'}} = 1\end{aligned}$$

which conclude the case for $1 < p < \infty$. The proof for the case that $p = 1$ or $p = \infty$ is trivial, and is left to the reader. □

Corollary 1.24 (Minkowski inequality)**.** *Let $1 \leqslant p \leqslant \infty$. Then*

$$\|\boldsymbol{x} + \boldsymbol{y}\|_p \leqslant \|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p \qquad \forall\, \boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}^{\mathrm{n}} .$$

Proof. We only prove the case that $1 < p < \infty$. First we note that

$$\begin{aligned}\|\boldsymbol{x} + \boldsymbol{y}\|_p^p &= \sum_{k=1}^{\mathrm{n}} |x_k + y_k|^p \leqslant \sum_{k=1}^{\mathrm{n}} |x_k + y_k|^{p-1}\big(|x_k| + |y_k|\big) \\ &= \sum_{k=1}^{\mathrm{n}} |x_k + y_k|^{p-1}|x_k| + \sum_{k=1}^{\mathrm{n}} |x_k + y_k|^{p-1}|y_k| .\end{aligned}$$

Let $\boldsymbol{u} = \big(|x_1|, |x_2|, \cdots, |x_{\mathrm{n}}|\big)$ and $\boldsymbol{v} = \big(|x_1+y_1|^{p-1}, |x_2+y_2|^{p-1}, \cdots, |x_{\mathrm{n}}+y_{\mathrm{n}}|^{p-1}\big)$. By Hölder's inequality,

$$\begin{aligned}\sum_{k=1}^{\mathrm{n}} |x_k + y_k|^{p-1}|x_k| &= (\boldsymbol{u}, \boldsymbol{v}) \leqslant \|\boldsymbol{u}\|_p \|\boldsymbol{v}\|_{p'} = \|\boldsymbol{x}\|_p \Big(\sum_{k=1}^{\mathrm{n}} |x_k + y_k|^{(p-1)p'}\Big)^{\frac{1}{p'}} \\ &= \|\boldsymbol{x}\|_p \Big(\sum_{k=1}^{\mathrm{n}} |x_k + y_k|^p\Big)^{\frac{p-1}{p}} = \|\boldsymbol{x}\|_p \|\boldsymbol{x} + \boldsymbol{y}\|_p^{p-1} .\end{aligned}$$

Similarly, we have $\sum\limits_{k=1}^{\mathrm{n}} |x_k + y_k|^{p-1}|y_k| \leqslant \|\boldsymbol{y}\|_p \|\boldsymbol{x} + \boldsymbol{y}\|_p^{p-1}$; thus

$$\|\boldsymbol{x} + \boldsymbol{y}\|_p^p \leqslant \big(\|\boldsymbol{x}\|_p + \|\boldsymbol{y}\|_p\big)\|\boldsymbol{x} + \boldsymbol{y}\|_p^{p-1}$$

which concludes the Minkowski inequality. □

Theorem 1.25. *Let $1 \leqslant p \leqslant \infty$, and p' be the conjugate of p; that is, $\dfrac{1}{p} + \dfrac{1}{p'} = 1$. Then*

$$\|\boldsymbol{x}\|_p = \sup_{\|\boldsymbol{y}\|_{p'}=1} \big|(\boldsymbol{x}, \boldsymbol{y})\big| \qquad \forall\, \boldsymbol{x} \in \mathbb{F}^{\mathrm{n}} .$$

Proof. By Hölder's inequality, it is clear that $\|\boldsymbol{x}\|_p \geqslant \sup_{\|\boldsymbol{y}\|_{p'}=1} \big|(\boldsymbol{x}, \boldsymbol{y})\big|$ for all $\boldsymbol{x} \in \mathbb{F}^{\mathrm{n}}$. On the other hand, note that $|x_k|^p = x_k \cdot \overline{x_k}|x_k|^{p-2}$; thus letting $y_k = \dfrac{\overline{x_k}|x_k|^{p-2}}{\|\boldsymbol{x}\|_p^{p-1}}$ we find that $\|\boldsymbol{y}\|_{p'} = 1$ which implies that

$$\big|(\boldsymbol{x}, \boldsymbol{y})\big| = \frac{1}{\|\boldsymbol{x}\|_p^{p-1}} \sum_{k=1}^{\mathrm{n}} |x_k|^p = \|\boldsymbol{x}\|_p$$

which implies that $\sup_{\|\boldsymbol{y}\|_{p'}=1} \big|(\boldsymbol{x}, \boldsymbol{y})\big| \geqslant \|\boldsymbol{x}\|_p$. □

Making use of Hölder's inequality (1.1) and the Riemann sum approximation of the Riemann integral, we can conclude the following

Theorem 1.26. *Let $1 \leqslant p \leqslant \infty$. If p' is the conjugate of p; that is, $\dfrac{1}{p} + \dfrac{1}{p'} = 1$, then*

$$\Big|\int_0^1 f(x)g(x)\,dx\Big| \leqslant \|f\|_p \|g\|_{p'} \qquad \forall\, f, g \in \mathscr{C}([0,1];\mathbb{R})\,,$$

where

$$\|f\|_p = \begin{cases} \Big(\displaystyle\int_0^1 |f(x)|^p dx\Big)^{\frac{1}{p}} & \text{if } 1 \leqslant p < \infty\,, \\ \displaystyle\max_{x\in[0,1]} |f(x)| & \text{if } p = \infty\,. \end{cases}$$

Remark 1.27. The Minkowski inequality implies that

$$\|f + g\|_p \leqslant \|f\|_p + \|g\|_p \qquad \forall\, f, g \in \mathscr{C}([0,1];\mathbb{R})\,.$$

In other words, the function $\|\cdot\|_p : \mathscr{C}([0,1];\mathbb{R}) \to \mathbb{R}$ is a norm on $\mathscr{C}([0,1];\mathbb{R})$, and is called the L^p-norm.

1.4 Matrices

Definition 1.28 (Matrix)**.** Let $\mathbb{F}$ be a scalar field. The space $\mathbb{M}(\mathrm{m},\mathrm{n};\mathbb{F})$ is the collection of elements, called an *m*-by-*n* matrix or $\mathrm{m} \times \mathrm{n}$ matrix over $\mathbb{F}$, of the form

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

where $a_{ij} \in \mathbb{F}$ is called the (i,j)-th entry of A, and is denoted by $[\mathrm{A}]_{ij}$. We write $\mathrm{A} = [a_{ij}]_{1 \leqslant i \leqslant m; 1 \leqslant j \leqslant n}$ or simply $\mathrm{A} = [a_{ij}]_{\mathrm{m} \times \mathrm{n}}$ to denote that A is an $\mathrm{m} \times \mathrm{n}$ matrix whose (i,j)-th entry is a_{ij}. A is called a ***square matrix*** if $\mathrm{m} = \mathrm{n}$. The $1 \times \mathrm{m}$ matrix

$$a_{i*} = \begin{bmatrix} a_{i1} & a_{i2} & \cdots & a_{i\mathrm{n}} \end{bmatrix}$$

is called the i-th row of A, and the $m \times 1$ matrix

$$a_{*j} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{\mathrm{m}j} \end{bmatrix}$$

is called the j-th column of A.

Definition 1.29 (Matrix addition)**.** Let $\mathrm{A} = [a_{ij}]_{\mathrm{m} \times \mathrm{n}}$ and $\mathrm{B} = [b_{ij}]_{\mathrm{m} \times \mathrm{n}}$ be two $\mathrm{m} \times \mathrm{n}$ matrices over a scalar field $\mathbb{F}$. The sum of A and B, denoted by $\mathrm{A} + \mathrm{B}$, is another $\mathrm{m} \times \mathrm{n}$ matrix defined by $\mathrm{A} + \mathrm{B} = [a_{ij} + b_{ij}]_{\mathrm{m} \times \mathrm{n}}$ or more precisely,

$$\mathrm{A} + \mathrm{B} = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}.$$

Definition 1.30 (Scalar multiplication)**.** Let $\mathrm{A} = [a_{ij}]_{\mathrm{m} \times \mathrm{n}}$ be an $\mathrm{m} \times \mathrm{n}$ matrix over a scalar field $\mathbb{F}$, and $\alpha \in \mathbb{F}$. The scalar multiplication of α and A, denoted by $\alpha\mathrm{A}$, is an $\mathrm{m} \times \mathrm{n}$ matrix defined by $\alpha\mathrm{A} = [\alpha a_{ij}]_{\mathrm{m} \times \mathrm{n}}$ or more precisely,

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

Proposition 1.31. *The space $\mathbb{M}(m, n; \mathbb{F})$ is a vector space over $\mathbb{F}$ under the matrix addition and scalar multiplication defined in previous two definitions.*

Definition 1.32 (Matrix product)**.** Let $\mathrm{A} \in \mathbb{M}(m, n; \mathbb{F})$ and $\mathrm{B} \in \mathbb{M}(n, \ell; \mathbb{F})$ be two matrices over a scalar field $\mathbb{F}$. The matrix product of A and B, denoted by AB, is an $m \times \ell$ matrix given by $\mathrm{AB} = [c_{ij}]_{\mathrm{m} \times \mathrm{n}}$ with $c_{ij} = \sum\limits_{k=1}^{\mathrm{n}} a_{ik} b_{kj}$. In other words, the (i,j)-th entry of the product AB is the inner product of the i-th row of A and the j-th column of B.

Remark 1.33. The matrix product AB is only defined if the number of columns of A is the same as the number of rows of B. Therefore, even if AB is defined, BA might not make sense. When A and B are both $n \times \mathrm{n}$ square matrix, AB and BA are both defined; however, in general AB $\neq$ BA.

Remark 1.34. Let $\boldsymbol{v} \in \mathbb{F}^{\mathrm{n}}$ be a vector such that the k-th component of $\boldsymbol{v}$ is the same as the (i,k)-th entry of $\mathrm{A} \in \mathbb{M}(m,n;\mathbb{F})$, and $\boldsymbol{w} \in \mathbb{F}^{\mathrm{n}}$ be a vector such that the k-th component of $\boldsymbol{w}$ is the same as the (k,j)-th entry of $\mathrm{B} \in \mathbb{M}(n,\ell;\mathbb{F})$. Then the (i,j)-th entry of AB is simply the inner product of $\boldsymbol{v}$ and $\boldsymbol{w}$ in $\mathbb{F}^{\mathrm{n}}$.

Example 1.35. Let $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 2 & -1 & 1 \\ 3 & 0 & 2 \\ -1 & 1 & 0 \end{bmatrix}$. Then

$$AB = \begin{bmatrix} 0 & 1 & 1 \\ -4 & 1 & -2 \end{bmatrix}$$

but BA is not defined.

Proposition 1.36. *Let* $\mathrm{A} \in \mathbb{M}(\mathrm{m},\mathrm{n};\mathbb{F})$, $\mathrm{B} \in \mathbb{M}(\mathrm{n},\ell;\mathbb{F})$ *and* $\mathrm{C} \in \mathbb{M}(\ell,k;\mathbb{F})$*. Then*

$$\mathrm{A}(\mathrm{BC}) = (\mathrm{AB})\mathrm{C}\,.$$

Definition 1.37 (The range and the null space of matrices)**.** Let $\mathrm{A} \in \mathbb{M}(\mathrm{m},\mathrm{n};\mathbb{F})$. The ***range*** of A, denoted by $R(\mathrm{A})$, is the subset of $\mathbb{F}^{\mathrm{m}}$ given by

$$R(\mathrm{A}) = \left\{\mathrm{A}\boldsymbol{x} \in \mathbb{F}^{\mathrm{m}} \,\middle|\, \boldsymbol{x} \in \mathbb{F}^{\mathrm{n}}\right\},$$

and the ***null space*** of A, denoted by null(A), is the subset of $\mathbb{F}^{\mathrm{n}}$ given by

$$\mathrm{null}(\mathrm{A}) = \left\{\boldsymbol{x} \in \mathbb{F}^{\mathrm{n}} \,\middle|\, \mathrm{A}\boldsymbol{x} = \mathbf{0}\right\}.$$

Proposition 1.38. *Let* $\mathrm{A} \in \mathbb{M}(\mathrm{m},\mathrm{n};\mathbb{F})$*. Then* $R(\mathrm{A})$ *and* null(A) *are vector subspaces of* $\mathbb{F}^{\mathrm{n}}$ *and* $\mathbb{F}^{\mathrm{m}}$*, respectively.*

Definition 1.39 (Kronecker's delta)**.** The Kronecker delta is a function, denoted by δ, of two variables (usually positive integers) such that the function is 1 if the two variables are equal, and 0 otherwise. When the two variables are i and j, the value $\delta(i,j)$ is usually written as δ_{ij}; that is,

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j\,, \\ 1 & \text{if } i = j\,. \end{cases}$$

Definition 1.40 (Identity matrix)**.** The identity matrix of size n, denoted by $\mathrm{I_n}$, is the $\mathrm{n} \times \mathrm{n}$ square matrix with ones on the main diagonal and zeros elsewhere. In other words,

$$\mathrm{I_n} = [\delta_{ij}]_{\mathrm{n}\times\mathrm{n}}\,,$$

where δ_{ij} is the Kronecker delta.

When the size is clear from the context, $\mathrm{I_n}$ is sometimes denoted by I.

Definition 1.41 (Transpose)**.** Let $\mathrm{A} = [a_{ij}]_{\mathrm{m}\times\mathrm{n}}$ be a $\mathrm{m} \times \mathrm{n}$ matrix over scalar field $\mathbb{F}$. The transpose of A, denoted by $\mathrm{A^T}$, is the $\mathrm{n} \times \mathrm{m}$ matrix given by $[\mathrm{A^T}]_{ij} = a_{ji}$.

By the definition of product of matrices, we can easily derive the following two propositions.

Proposition 1.42. *Let* $\mathrm{A} \in \mathbb{M}(\mathrm{m}, \mathrm{n}; \mathbb{F})$ *and* $\mathrm{B} \in \mathbb{M}(\mathrm{n}, \ell; \mathbb{F})$*. Then* $(\mathrm{AB})^\mathrm{T} = \mathrm{B^T A^T}$*.*

Proposition 1.43. *Let* $\mathrm{A} = [a_{ij}]_{\mathrm{m}\times\mathrm{n}}$ *be a* $\mathrm{m} \times \mathrm{n}$ *matrix over scalar field* $\mathbb{F}$*, and* $(\cdot,\cdot)_{\mathbb{F}^\mathrm{n}}$ *and* $(\cdot,\cdot)_{\mathbb{F}^\mathrm{m}}$ *be the standard inner products on* $\mathbb{F}^\mathrm{n}$ *and* $\mathbb{F}^\mathrm{m}$*, respectively. Then*

$$(\mathrm{A}\boldsymbol{x}, \boldsymbol{y})_{\mathbb{F}^\mathrm{m}} = (\boldsymbol{x}, \overline{\mathrm{A^T}}\boldsymbol{y})_{\mathbb{F}^\mathrm{n}} \qquad \forall\, \boldsymbol{x} \in \mathbb{F}^\mathrm{n}, \boldsymbol{y} \in \mathbb{F}^\mathrm{m}\,.$$

Definition 1.44 (Rank and nullity of matrices)**.** The ***rank*** of a matrix A, denoted by rank(A), is the dimension of the vector space generated (or spanned) by its columns. The ***nullity*** of a matrix A, denoted by nullity(A), is the dimension of the null space of A.

Remark 1.45. The matrix $\overline{\mathrm{A^T}}$ is often called the ***conjugate transpose*** of the matrix A.

Remark 1.46. The rank defined above is also referred to the ***column rank***, and the ***row rank*** of a matrix is the dimension of the vector space spanned by its rows. One should immediately notice that the column rank of A equals the dimension of $R(\mathrm{A})$ and the row rank of A equalis the dimension of $R(\mathrm{A^T})$.

Theorem 1.47. *Let* $\mathrm{A} \in \mathbb{M}(\mathrm{m}, \mathrm{n}; \mathbb{F})$*. Then* $\mathrm{rank(A)} + \mathrm{nullity(A)} = \mathrm{n}$*.*

Proof. Without loss of generality, we assume that nulltiy(A) $= k < \mathrm{n}$, and $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ be a basis of null(A). Then there exists $\mathrm{n} - k$ vectors $\{\boldsymbol{v}_{k+1}, \cdots, \boldsymbol{v}_\mathrm{n}\}$ such that $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_\mathrm{n}\}$ is a basis of $\mathbb{F}^\mathrm{n}$. We conclude the theorem by showing that $\{\mathrm{A}\boldsymbol{v}_{k+1}, \cdots, \mathrm{A}\boldsymbol{v}_\mathrm{n}\}$ is a basis of $R(\mathrm{A})$.

First, we claim that $\{A\boldsymbol{v}_{k+1}, \cdots, A\boldsymbol{v}_n\}$ is a linearly independent set of vectors. To see this, suppose that $\alpha_{k+1}, \cdots, \alpha_n \in \mathbb{F}$ such that

$$\alpha_{k+1}A\boldsymbol{v}_{k+1} + \cdots + \alpha_n A\boldsymbol{v}_n = \mathbf{0}\,.$$

Then $A(\alpha_{k+1}\boldsymbol{v}_{k+1} + \cdots + \alpha_n\boldsymbol{v}_n) = \mathbf{0}$ which implies that $\alpha_{k+1}\boldsymbol{v}_{k+1} + \cdots + \alpha_n\boldsymbol{v}_n \in \text{null}(A)$. Since $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_k\}$ is a basis of null(A), there exist $\alpha_1, \cdots, \alpha_k \in \mathbb{F}$ such that

$$\alpha_1\boldsymbol{v}_1 + \alpha_k\boldsymbol{v}_k = \alpha_{k+1}\boldsymbol{v}_{k+1} + \cdots + \alpha_n\boldsymbol{v}_n\,.$$

By the linear independence of $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$, we must have $\alpha_1 = \cdots = \alpha_n = 0$ which shows the linear independence of $\{A\boldsymbol{v}_{k+1}, \cdots, A\boldsymbol{v}_n\}$.

Let $\boldsymbol{w} \in R(A)$. Then $\boldsymbol{w} = A\boldsymbol{v}$ for some $\boldsymbol{v} \in \mathbb{F}^n$. Since $\{\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n\}$ is a basis of $\mathbb{F}^n$, there exist $\beta_1, \cdots, \beta_n \in \mathbb{F}$ such that $\boldsymbol{v} = \beta_1\boldsymbol{v}_1 + \cdots + \beta_n\boldsymbol{v}_n$. As a consequence, by the fact that $A\boldsymbol{v}_j = \mathbf{0}$ for $1 \leqslant j \leqslant k$,

$$\boldsymbol{w} = A\boldsymbol{v} = A(\beta_1\boldsymbol{v}_1 + \cdots + \beta_n\boldsymbol{v}_n) = \beta_1 A\boldsymbol{v}_1 + \cdots \beta_n A\boldsymbol{v}_n = \beta_{k+1}A\boldsymbol{v}_{k+1} + \cdots + \beta_n A\boldsymbol{v}_n\,;$$

thus $\boldsymbol{w}$ can be written as a linear combination of $\{A\boldsymbol{v}_{k+1}, \cdots, A\boldsymbol{v}_n\}$. □

Theorem 1.48. *The rank of a matrix is the same as the rank of its transpose. In other words, for a given matrix the row rank equals the column rank.*

Proof. Let A be a m × n matrix, and $(\cdot,\cdot)_{\mathbb{F}^n}$, $(\cdot,\cdot)_{\mathbb{F}^m}$ be the standard inner products on $\mathbb{F}^n$, $\mathbb{F}^m$, respectively. Then Proposition 1.43 implies that

$$\begin{aligned}\boldsymbol{y} \in R(A)^\perp \;&\Leftrightarrow\; (\boldsymbol{y}, A\boldsymbol{x})_{\mathbb{F}^m} = 0 \text{ for all } \boldsymbol{x} \in \mathbb{F}^n \;\Leftrightarrow\; (\overline{A^T}\boldsymbol{y}, \boldsymbol{x})_{\mathbb{F}^n} = 0 \text{ for all } \boldsymbol{x} \in \mathbb{F}^n \\ &\Leftrightarrow\; \overline{A^T}\boldsymbol{y} = \mathbf{0} \;\Leftrightarrow\; \boldsymbol{y} \in \text{null}(\overline{A^T})\,.\end{aligned}$$

In other words, $R(A)^\perp = \text{null}(\overline{A^T})$. Since the column rank of A is the dimension of $R(A)$, we must have

$$\text{nullity}(A^T) = \text{nullity}(\overline{A^T}) = \dim\big(R(A)^\perp\big) = m - \text{the column rank of } A\,.$$

On the other hand, Theorem 1.47 implies that

$$\text{rank}(A^T) + \text{nullity}(A^T) = m\,;$$

thus the column rank of A is the same as the row rank of A. □

Definition 1.49. Let $\mathrm{A} \in \mathbb{M}(\mathrm{n}, \mathrm{n}; \mathbb{F})$ be a square matrix. A is said to be ***invertible*** if there exists $\mathrm{B} \in \mathbb{M}(\mathrm{n}, \mathrm{n}; \mathbb{F})$ such that $\mathrm{AB} = \mathrm{I_n}$. The matrix B is called the ***inverse matrix*** of A, and is usually denoted by A^{-1}.

Proposition 1.50. *Let* $\mathrm{A} \in \mathbb{M}(\mathrm{n}, \mathrm{n}; \mathbb{F})$ *be invertible. Then* $rank(\mathrm{A}) = rank(\mathrm{A}^{-1}) = \mathrm{n}$.

Proof. Since $\mathrm{A}(\mathrm{A}^{-1}\boldsymbol{b}) = (\mathrm{AA}^{-1})\boldsymbol{b} = \boldsymbol{b}$ for all $\boldsymbol{b} \in \mathbb{F}^\mathrm{n}$, $R(\mathrm{A}) = \mathbb{F}^\mathrm{n}$ which implies that $\mathrm{rank}(\mathrm{A}) = \mathrm{n}$. We next show that $R(\mathrm{A}^{-1}) = \mathbb{F}^\mathrm{n}$. Denote A^{-1} by B, and let $\boldsymbol{b} \in \mathbb{F}^\mathrm{n}$. Then $\mathrm{B^T}(\mathrm{A^T}\boldsymbol{b}) = (\mathrm{B^T A^T})\boldsymbol{b} = \boldsymbol{b}$ since $\mathrm{B^T A^T} = (\mathrm{AB})^\mathrm{T} = \mathrm{I_n}$. This observation implies that $R(\mathrm{B^T}) = \mathbb{F}^\mathrm{n}$, and the theorem is then concluded by Theorem 1.48. ▫

Proposition 1.51. *Let* $\mathrm{A} \in \mathbb{M}(\mathrm{n}, \mathrm{n}; \mathbb{F})$ *be invertible. Then* $\mathrm{A}^{-1}\mathrm{A} = \mathrm{AA}^{-1} = \mathrm{I_n}$.

Proof. We show that for all $\boldsymbol{b} \in \mathbb{F}^\mathrm{n}$, $\mathrm{A}^{-1}\mathrm{A}\boldsymbol{b} = \boldsymbol{b}$. Since A is invertible, $\mathrm{rank}(\mathrm{A}^{-1}) = \mathrm{n}$; thus $R(\mathrm{A}^{-1}) = \mathbb{F}^\mathrm{n}$ which implies that for each $\boldsymbol{b} \in \mathbb{F}^\mathrm{n}$, there exists $\boldsymbol{x} \in \mathbb{F}$ such that $\mathrm{A}^{-1}\boldsymbol{x} = \boldsymbol{b}$. As a consequence,

$$(\mathrm{A}^{-1}\mathrm{A})\boldsymbol{b} = (\mathrm{A}^{-1}\mathrm{A})(\mathrm{A}^{-1}\boldsymbol{x}) = \mathrm{A}^{-1}(\mathrm{AA}^{-1})\boldsymbol{x} = \mathrm{A}^{-1}\boldsymbol{x} = \boldsymbol{b}\,.$$ ▫

1.4.1 Elementary Row Operations and Elementary Matrices

Definition 1.52 (Elementary row operations)**.** For an $\mathrm{n} \times \mathrm{m}$ matrix A, three types of elementary row operations can be performed on A:

1. The first type of row operation on A switches all matrix elements on the i-th row with their counterparts on j-th row.

2. The second type of row operation on A multiplies all elements on the i-th row by a non-zero scalar λ.

3. The third type of row operation on A adds j-th row multiplied by a scalar μ to the i-th row.

The elementary row operation on an $\mathrm{n} \times \mathrm{m}$ matrix A can be done by multiplying A by an $\mathrm{n} \times \mathrm{n}$ matrix, called an elementary matrix, on the left. The elementary matrices are defined in the following

Definition 1.53 (Elementary matrices)**.** An elementary matrix is a matrix which differs from the identity matrix by one single elementary row operation.

1. Switching the i_0-th and j_0-th rows of A, where $i_0 \neq j_0$, is done by left multiplied A by the matrix $\mathrm{E} = [e_{ij}]_{\mathrm{n}\times\mathrm{n}}$ given by

$$e_{ij} = \begin{cases} 1 & \text{if } (i,j) = (i_0, j_0) \text{ or } (i,j) = (j_0, i_0) \text{ or } i = j = k_0 \text{ for some } k_0 \neq i_0, j_0, \\ 0 & \text{otherwise,} \end{cases}$$

or in the matrix form,

$$\mathrm{E} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & & & & & & & & \vdots \\ \vdots & \ddots & 1 & \ddots & & & & & & & \vdots \\ \vdots & & 0 & 0 & 0 & & & 1 & & & \vdots \\ \vdots & & & \ddots & 1 & \ddots & & & & & \vdots \\ \vdots & & & & 0 & \ddots & 0 & & & & \vdots \\ \vdots & & & & & \ddots & 1 & \ddots & & & \vdots \\ \vdots & & & 1 & & & 0 & 0 & 0 & & \vdots \\ 0 & & & & & & & \ddots & 1 & \ddots & \vdots \\ 0 & & & & & & & & 0 & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ \leftarrow \text{the } i_0\text{-th row} \\ \\ \\ \\ \leftarrow \text{the } j_0\text{-th row} \\ \\ \\ \\ \end{matrix}$$

$\uparrow$ the i_0-th column $\uparrow$ the j_0-th column

2. Multiplying the k_0-th row of A by a non-zero scalar λ is done by left multiplied A by the matrix $\mathrm{E} = [e_{ij}]_{\mathrm{n}\times\mathrm{n}}$ given by

$$e_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ \lambda & \text{if } i = j = k_0, \\ 1 & \text{otherwise,} \end{cases}$$

or in the matrix form,

$$\mathrm{E} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & & & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & & & \vdots \\ \vdots & & 0 & 1 & 0 & & & & \vdots \\ \vdots & & & 0 & \lambda & 0 & & & \vdots \\ \vdots & & & & 0 & 1 & 0 & & \vdots \\ \vdots & & & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & & & 0 & 1 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \leftarrow \text{the } k_0\text{-th row} \\ \\ \\ \\ \\ \end{matrix}$$

3. Adding the j_0-th row of A multiplied by a scalar μ to the i_0-th row, where $i_0 \neq j_0$, is done by left multiplied A by the matrix $\mathrm{E} = [e_{ij}]_{\mathrm{n}\times\mathrm{n}}$ given by

$$e_{ij} = \begin{cases} 1 & \text{if } i = j, \\ \mu & \text{if } (i,j) = (i_0, j_0), \\ 0 & \text{otherwise,} \end{cases}$$

or in the matrix form,

$$\mathrm{E} = \begin{bmatrix} 1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 1 & 0 & & & & & & 0 \\ \vdots & \ddots & \ddots & \ddots & & & \mu & & 0 \\ \vdots & & \ddots & \ddots & \ddots & & & & 0 \\ \vdots & & & 0 & 1 & 0 & & & \vdots \\ \vdots & & & & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & & & 0 & 1 & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 & 1 \end{bmatrix} \begin{matrix} \\ \\ \leftarrow \text{the } i_0\text{-th row} \\ \\ \\ \\ \\ \\ \\ \end{matrix}$$

$$\uparrow$$
$$\text{the } j_0\text{-th column}$$

Proposition 1.54. *Every elementary matrix is invertible.*

Theorem 1.55. *Let* $\mathrm{A} \in \mathbb{M}(\mathrm{n}, \mathrm{n}; \mathbb{F})$ *be a square matrix. The following statements are equivalent:*

1. $R(\mathrm{A}) = \mathbb{F}^{\mathrm{n}}$.

2. $rank(\mathrm{A}) = \mathrm{n}$.

3. $\mathrm{A}\boldsymbol{x} = \boldsymbol{b}$ *has a unique solution* $\boldsymbol{x}$ *for all* $\boldsymbol{b} \in \mathbb{F}^{\mathrm{n}}$.

4. A *is invertible.*

5. $\mathrm{A} = \mathrm{E}_k \mathrm{E}_{k-1} \cdots \mathrm{E}_2 \mathrm{E}_1$ *for some elementary matrices* $\mathrm{E}_1, \cdots, \mathrm{E}_k$.

Proof. Note that by definition 1,2,3 are equivalent, and Proposition 1.50 shows that $4 \Rightarrow 2$. The implication from 3 to 4 is due to the fact that the map $\boldsymbol{b} \mapsto \boldsymbol{x}$, where $\boldsymbol{x}$ is the unique solution to $\mathrm{A}\boldsymbol{x} = \boldsymbol{b}$, is the inverse of A. Proposition 1.54 provides that $5 \Rightarrow 4$. That $3 \Rightarrow 5$ follows from that at most $\mathrm{n}(\mathrm{n}+1)$ elementary row operations has to be applied on A to reach the identity matrix. □

1.5 Determinants

In order to introduce the notion of the determinant of square matrices, we need to talk about permutations first. Note that there are many other ways of defining determinants, but it is quite elegant to use the notion of permutations, and we can derive a lot of useful results via this definition.

Definition 1.56 (Permutations)**.** A sequence $(k_1, k_2, \cdots, k_{\rm n})$ of positive integers not exceeding n, with the property that no two of the k_i are equal, is called a ***permutation of degree*** n. The collection of all permutations of degree n is denoted by $\mathbb{P}(\rm n)$.

A sequence $(k_1, k_2, \cdots, k_{\rm n})$ can be obtained from the sequence $(1, 2, \cdots, {\rm n})$ by a finite number of interchanges of pairs of elements. For example, if $k_1 \neq 1$, we can transpose 1 and k_1, obtaining $(k_1, \cdots, 1, \cdots)$. Proceeding in this way we shall arrive at the sequence $(k_1, k_2, \cdots, k_{\rm n})$ after n or less such interchanges of pairs.

In general, a permutation $(k_1, k_2, \cdots, k_{\rm n})$ can be expressed as

$$\tau_{(i_N, j_N)} \cdots \tau_{(i_2, j_2)} \tau_{(i_1, j_1)} (1, 2, \cdots, {\rm n}) = (k_1, k_2, \cdots, k_{\rm n}),$$

where $\tau_{(i,j)}$ is a "pair-interchange operator" which swaps the i-th and the j-th elements (of the object fed into), and N is the number of pair interchanges. We call such pair-interchange operators the permutation operator. Since $\tau_{(i,j)}$ is the inverse operator of itself, we also have

$$\tau_{(i_1, j_1)} \tau_{(i_2, j_2)} \cdots \tau_{(i_N, j_N)} (k_1, k_2, \cdots, k_{\rm n}) = (1, 2, \cdots, {\rm n}).$$

We remark here that the number of pair interchanges (from $(1, 2, \cdots, {\rm n})$ to $(k_1, k_2, \cdots, k_{\rm n})$) is not unique; nevertheless, if two processes of pair interchanges lead to the same permutation, then the numbers of interchanges differ by an even number. This leads to the following

Definition 1.57 (Even and odd permutations)**.** A permutation $(k_1, \cdots, k_{\rm n})$ is called an ***even*** (***odd***) ***permutation*** of degree n if the number required to interchange pairs of $(1, 2, \cdots, {\rm n})$ in order to obtain $(k_1, k_2, \cdots, k_{\rm n})$ is even (odd).

Example 1.58. If ${\rm n} = 3$, the permutation $(3, 1, 2)$ can be obtained by interchanging pairs of $(1, 2, 3)$ twice:

$$(1, 2, 3) \xrightarrow{\tau_{(1,3)}} (3, 2, 1) \xrightarrow{\tau_{(2,3)}} (3, 1, 2);$$

thus $(3,1,2)$ is an even permutation of $(1,2,3)$. On the other hand, $(1,3,2)$ is obtained by interchanging pairs of $(1,2,3)$ once:

$$(1,2,3) \xrightarrow{\tau_{(2,3)}} (1,3,2);$$

thus $(1,3,2)$ is an odd permutation of $(1,2,3)$.

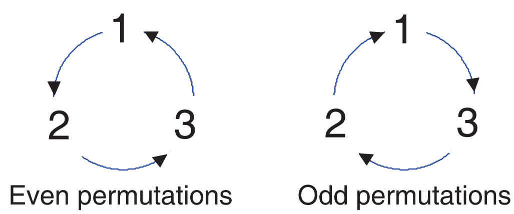

Figure 1.1: Even and odd permutations of degree 3

For n = 3, the even and odd permutations can also be viewed as the orientation of the permutation (k_1, k_2, k_3). To be more precise, if $(1,2,3)$ is arranged in a counter-clockwise orientation (see Figure 1.1), then an even permutation of degree 3 is a permutation in the counter-clockwise orientation, while an odd permutation of degree 3 is a permutation in the clockwise orientation. From figure 1.1, it is easy to see that $(3,1,2)$ is an even permutation of degree 3 and $(1,3,2)$ is an odd permutation of degree 3.

Definition 1.59 (The permutation symbol)**.** The permutation symbol $\varepsilon_{k_1 k_2 \cdots k_n}$ is a function of permutations of degree n defined by

$$\varepsilon_{k_1 k_2 \cdots k_n} = \begin{cases} 1 & \text{if } (k_1, k_2, \cdots, k_n) \text{ is an even permutation of degree n,} \\ -1 & \text{if } (k_1, k_2, \cdots, k_n) \text{ is an odd permutation of degree n.} \end{cases}$$

Remark 1.60. One can extend the domain the permutation symbol to all the sequence $(k_1, k_2, \cdots, k_n)$ by defining that $\varepsilon_{k_1 k_2 \cdots k_n} = 0$ if $(k_1, k_2, \cdots, k_n)$ is not a permutation of degree n.

Definition 1.61 (Determinants)**.** Given an n × n matrix $A = [a_{ij}]$, the determinants of A, denoted by det(A), is defined by

$$\det(A) = \sum_{(k_1, \cdots, k_n) \in \mathbb{P}(n)} \varepsilon_{k_1 k_2 \cdots k_n} \prod_{\ell=1}^{n} a_{\ell k_\ell}.$$

We note that the product $\prod_{\ell=1}^{n} a_{\ell k_\ell}$ in the definition of the determinant is formed by multiplying n-elements which appears exactly once in each row and column.

Proposition 1.62. *Let* E *be an elementary matrix. Then*

1. $\det(E) \neq 0$.

2. $\det(E) = \det(E^T)$.

3. *If* A *is an* $n \times n$ *matrix, then* $\det(EA) = \det(E)\det(A)$.

The proof of the proposition above is not difficult, and is left as an exercise.

Corollary 1.63. *Let* $\boldsymbol{v}_1, \cdots, \boldsymbol{v}_n \in \mathbb{R}^n$ *be (column) vectors,* $c \in \mathbb{R}$*, and*

$$\begin{aligned} A &= \left[\boldsymbol{v}_1 \vdots \cdots \vdots \boldsymbol{v}_n\right], \\ B &= \left[\boldsymbol{v}_1 \vdots \cdots \vdots \boldsymbol{v}_{j-1} \vdots \lambda \boldsymbol{v}_j \vdots \boldsymbol{v}_{j+1} \vdots \cdots \vdots \boldsymbol{v}_n\right], \\ C &= \left[\boldsymbol{v}_1 \vdots \cdots \vdots \boldsymbol{v}_{j-1} \vdots \boldsymbol{v}_j + \mu \boldsymbol{v}_i \vdots \boldsymbol{v}_{j+1} \vdots \cdots \vdots \boldsymbol{v}_n\right] \quad \text{for some } i \neq j\,. \end{aligned}$$

Then $\det(B) = \lambda \det(A)$*, and* $\det(C) = \det(A)$.

Proof. The corollary is easily concluded since $B = E_1 A$ and $C = E_2 A$ for some elementary matrices E_1 and E_2 with $\det(E_1) = c$ and $\det(E_2) = 1$. ▫

Corollary 1.64. *Let* A *be an* $n \times n$ *matrix. Then* A *is invertible if and only if* $\det(A) \neq 0$.

Proof. ($\Rightarrow$) Since A is invertible, Theorem 1.55 implies that

$$A = E_k E_{k-1} \cdots E_2 E_1$$

for some elementary matrices $E_1, \cdots, E_k$, and this corollary follows from Proposition 1.62.

($\Leftarrow$) Note that A is invertible if and only if $\operatorname{rank}(A) = \operatorname{rank}(A^T) = n$. Therefore, if A is not invertible, the row vectors of A are linearly dependent; thus there exists a non-zero vectors $(\alpha_1, \cdots, \alpha_n) \in \mathbb{F}^n$ such that

$$\alpha_1 \boldsymbol{v}_1 + \alpha_2 \boldsymbol{v}_2 + \cdots \alpha_n \boldsymbol{v}_n = \mathbf{0}\,,$$

where $A^T = \left[\boldsymbol{v}_1 \vdots \cdots \vdots \boldsymbol{v}_n\right]$. Suppose that $\alpha_j \neq 0$. Then

$$\boldsymbol{v}_j = \beta_1 \boldsymbol{v}_1 + \cdots \beta_{j-1} \boldsymbol{v}_{j-1} + \beta_{j+1} \boldsymbol{v}_{j+1} + \cdots \beta_n \boldsymbol{v}_n\,;$$

thus after applying $(\mathrm{n}-1)$-times elementary row operations of the third type (adding some multiple of certain row to another row) on A we reach a matrix whose j-th row is a zero (row) vector. Thereofre, for some elementary matrices $\mathrm{E}_1, \cdots, \mathrm{E}_{\mathrm{n}-1}$ we have

$$\det(\mathrm{E}_{\mathrm{n}-1} \cdots \mathrm{E}_1 \mathrm{A}) = 0$$

which implies that $\det(\mathrm{A}) = 0$. □

Corollary 1.65. *Let* A *be an* $\mathrm{n}\times\mathrm{n}$ *matrix. Then the determinant of* A *and* A^{T}*, the transpose of* A*, are the same; that is,*

$$\det(\mathrm{A}) = \det(\mathrm{A}^{\mathrm{T}}).$$

Proof. If A is not invertible, then A^{T} is not invertible either because of Theorem 1.48. Therefore, $\det(\mathrm{A}) = 0 = \det(\mathrm{A}^{\mathrm{T}})$.

Now suppose that A is invertible. Then Theorem 1.55 implies that

$$\mathrm{A} = \mathrm{E}_k \mathrm{E}_{k-1} \cdots \mathrm{E}_2 \mathrm{E}_1$$

for some elementary matrices $\mathrm{E}_1, \cdots, \mathrm{E}_k$. Since all $\mathrm{E}_j^{\mathrm{T}}$'s are also elementary matrices, by Proposition 1.62 we conclude that

$$\begin{aligned}\det(\mathrm{A}^{\mathrm{T}}) &= \det(\mathrm{E}_1^{\mathrm{T}} \cdots \mathrm{E}_k^{\mathrm{T}}) = \det(\mathrm{E}_1^{\mathrm{T}}) \cdots \det(\mathrm{E}_k^{\mathrm{T}}) \\ &= \det(\mathrm{E}_k^{\mathrm{T}}) \cdots \det(\mathrm{E}_1^{\mathrm{T}}) \\ &= \det(\mathrm{E}_k) \cdots \det(\mathrm{E}_1) = \det(\mathrm{E}_k \cdots \mathrm{E}_1) = \det(\mathrm{A}) .\end{aligned}$$ □

Corollary 1.66. *Let* A, B *be* $\mathrm{n} \times \mathrm{n}$ *matrices. Then* $\det(\mathrm{AB}) = \det(\mathrm{A})\det(\mathrm{B})$.

Proof. If A is not invertible, then AB is not invertible either; thus in this case $\det(\mathrm{A})\det(\mathrm{B}) = 0 = \det(\mathrm{AB})$.

Now suppose that A is invertible. Then Theorem 1.55 implies that

$$\mathrm{A} = \mathrm{E}_k \mathrm{E}_{k-1} \cdots \mathrm{E}_2 \mathrm{E}_1$$

for some elementary matrices $\mathrm{E}_1, \cdots, \mathrm{E}_k$. As a consequence, Proposition 1.62 implies that

$$\begin{aligned}\det(\mathrm{AB}) &= \det(\mathrm{E}_k \cdots \mathrm{E}_1 \mathrm{B}) = \det(\mathrm{E}_k)\det(\mathrm{E}_{k-1} \cdots \mathrm{E}_1 \mathrm{B}) \\ &= \cdots = \det(\mathrm{E}_k) \cdots \det(\mathrm{E}_1)\det(\mathrm{B}) \\ &= \det(\mathrm{E}_k \cdots \mathrm{E}_1)\det(\mathrm{B}) = \det(\mathrm{A})\det(\mathrm{B}) .\end{aligned}$$ □

Definition 1.67 (Minor, Cofactor, and Adjoint matrices)**.** Let A be an $\mathrm{n} \times \mathrm{n}$ matrix, and $\mathrm{A}(\hat{i}, \hat{j})$ be the $(\mathrm{n}-1) \times (\mathrm{n}-1)$ matrix obtained by eliminating the i-th row and j-th column of A; that is,

$$\mathrm{A}(\hat{i}, \hat{j}) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1(j-1)} & a_{1(j+1)} & \cdots & a_{1\mathrm{n}} \\ \vdots & \ddots & & \vdots & \vdots & & \vdots \\ a_{(i-1)1} & a_{(i-1)2} & \cdots & a_{(i-1)(j-1)} & a_{(i-1)(j+1)} & \cdots & a_{(i-1)\mathrm{n}} \\ a_{(i+1)1} & a_{(i+1)2} & \cdots & a_{(i+1)(j-1)} & a_{(i+1)(j+1)} & \cdots & a_{(i+1)\mathrm{n}} \\ \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ a_{\mathrm{n}1} & a_{\mathrm{n}2} & \cdots & a_{\mathrm{n}(j-1)} & a_{\mathrm{n}(j+1)} & \cdots & a_{\mathrm{nn}} \end{bmatrix}.$$

The (i, j)-th ***minor*** of A is the determinant of $A(\hat{i}, \hat{j})$, and the (i, j)-th ***cofactor***, is the (i, j)-th minor of A multiplied by $(-1)^{i+j}$. The ***adjoint matrix*** of A, denoted by Adj(A), is the transpose of the cofactor matrix; that is,

$$\big[\mathrm{Adj}(\mathrm{A})\big]_{ij} = (-1)^{i+j} \det\big(\mathrm{A}(\hat{j}, \hat{i})\big)\,.$$

Example 1.68. Let $\mathrm{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & -1 & 2 \\ 0 & 2 & -1 \end{bmatrix}$. Then the minor matrix of A is $\begin{bmatrix} -3 & -3 & 6 \\ -8 & -1 & 2 \\ 7 & -7 & -7 \end{bmatrix}$, the cofactor matrix of A is $\begin{bmatrix} -3 & 3 & 6 \\ 8 & -1 & -2 \\ 7 & 7 & -7 \end{bmatrix}$, and the adjoint matrix of A is $\begin{bmatrix} -3 & 8 & 7 \\ 3 & -1 & 7 \\ 6 & -2 & -7 \end{bmatrix}$.

The following lemma provides a way of computing the minors of a matrix.

Lemma 1.69. *Let* A *be an* $\mathrm{n} \times \mathrm{n}$ *matrix. Then*

$$\det\big(\mathrm{A}(\hat{i}, \hat{j})\big) = (-1)^{i+j} \sum_{(k_1, \cdots, k_\mathrm{n}) \in \mathbb{P}(\mathrm{n}),\, k_i = j} \varepsilon_{k_1 k_2 \cdots k_\mathrm{n}} \prod_{\substack{1 \leqslant \ell \leqslant \mathrm{n} \\ \ell \neq i}} a_{\ell k_\ell}.$$

Proof. Fix $(i, j) \in \{1, 2, \cdots, \mathrm{n}\} \times \{1, 2, \cdots, \mathrm{n}\}$. The matrix $\mathrm{A}(\hat{i}, \hat{j})$ is given by $\mathrm{A}(\hat{i}, \hat{j}) = [b_{\alpha\beta}]$, where $\alpha, \beta = 1, 2, \cdots, \mathrm{n}-1$, and

$$b_{\alpha\beta} = \begin{cases} a_{\alpha\beta} & \text{if } \alpha < i \text{ and } \beta < j, \\ a_{(\alpha+1)\beta} & \text{if } \alpha > i \text{ and } \beta < j, \\ a_{\alpha(\beta+1)} & \text{if } \alpha < i \text{ and } \beta > j, \\ a_{(\alpha+1)(\beta+1)} & \text{if } \alpha > i \text{ and } \beta > j. \end{cases}$$

Each permutation $(\sigma_1, \sigma_2, \cdots, \sigma_{\mathrm{n}-1})$ of degree $\mathrm{n}-1$ corresponds a unique permutation $(k_1, k_2, \cdots, k_\mathrm{n})$ of degree n such that

1. $k_i = j$;

2. for each $\tau \in \{1, \cdots, i-1\}$ and $\iota \in \{i, i+1, \cdots, \mathrm{n}-1\}$,

$$k_\tau = \begin{cases} \sigma_\tau & \text{if } \sigma_\tau < j\,, \\ \sigma_\tau + 1 & \text{if } \sigma_\tau \geqslant j\,, \end{cases} \qquad \text{and} \qquad k_{\iota+1} = \begin{cases} \sigma_\iota & \text{if } \sigma_\iota < j\,, \\ \sigma_\iota + 1 & \text{if } \sigma_\iota \geqslant j\,. \end{cases}$$

We now determine the sign of $\varepsilon_{\sigma_1\sigma_2\cdots\sigma_{\mathrm{n}-1}}$ and $\varepsilon_{k_1k_2\cdots k_{\mathrm{n}}}$. Note that if a process of pair interchanges of the permutation $(\sigma_1, \sigma_2, \cdots, \sigma_{\mathrm{n}-1})$ leads to $(1, 2, \cdots, \mathrm{n}-1)$, then similar process of pair interchanges of the permutation $(k_1, k_2, \cdots, k_{i-1}, j, k_{i+1}, \cdots, k_{\mathrm{n}})$, by leaving the i-th slot fixed, leads to the permutation of degree n

$$\begin{cases} (1, 2, \cdots, j-1, j+1, \cdots, i-1, j, i, \cdots, \mathrm{n}) & \text{if } \ i > j, \\ (1, 2, \cdots, i-1, j, i, \cdots, j-1, j+1, \cdots, \mathrm{n}) & \text{if } \ i < j, \\ (1, 2, \cdots, \mathrm{n}) & \text{if } \ j = i. \end{cases}$$

For the case that $i \neq j$, another $|i-j|$-times of pair interchanges leads to $(1, 2, \cdots, \mathrm{n})$. To be more precise, suppose that $i > j$. We first interchange the $(i-2)$-th and the $(i-1)$-th components, and then interchange that $(i-3)$-th and the $(i-2)$-th components, and so on. After $(i-j)$-times of pair interchanges, we reach $(1, 2, \cdots, \mathrm{n})$. Symbolically,

$$\begin{array}{c} (1, 2, \cdots, j-1, j+1, \cdots, i-\overleftrightarrow{1, j}, i, \cdots, \mathrm{n}) \\ \downarrow \tau_{(i\,-\,2,\, i\,-\,1)} \\ (1, 2, \cdots, j-1, j+1, \cdots, i-\overleftrightarrow{2, j}, i-1, \cdots, \mathrm{n}) \\ \downarrow \tau_{(i\,-\,3,\, i\,-\,2)} \\ (1, 2, \cdots, j-1, j+1, \cdots, i-\overleftrightarrow{3, j}, i-2 \cdots, \mathrm{n}) \\ \downarrow \\ \vdots \\ \downarrow \\ (1, 2, \cdots, \mathrm{n}). \end{array}$$

Similar argument applies to the case $i < j$; thus

$$\varepsilon_{\sigma_1\sigma_2\cdots\sigma_{\mathrm{n}-1}} = (-1)^{|i-j|}\varepsilon_{k_1k_2\cdots k_{\mathrm{n}}} = (-1)^{i+j}\varepsilon_{k_1k_2\cdots k_{\mathrm{n}}}.$$

As a consequence,

$$\begin{aligned} \det\big(\mathrm{A}(\hat{i}, \hat{j})\big) &= \sum_{(\sigma_1, \sigma_2, \cdots, \sigma_{\mathrm{n}-1}) \in \mathbb{P}(\mathrm{n}-1)} \varepsilon_{\sigma_1\sigma_2\cdots\sigma_{\mathrm{n}-1}} \prod_{\tau=1}^{\mathrm{n}-1} b_{\tau\sigma_\tau} \\ &= (-1)^{i+j} \sum_{(k_1, \cdots, k_{\mathrm{n}}) \in \mathbb{P}(\mathrm{n}),\, k_i = j} \varepsilon_{k_1k_2\cdots k_{\mathrm{n}}} \prod_{\substack{1 \leqslant \ell \leqslant \mathrm{n} \\ \ell \neq i}} a_{\ell k_\ell}. \end{aligned}$$

□

Theorem 1.70. *Let* A *be an* n × n *matrix. Then*

$$\mathrm{Adj}(A)A = A\mathrm{Adj}(A) = \det(A)I_n.$$

Proof. Let $A = [a_{ij}]$. By definition of matrix multiplications,

$$\begin{aligned}\big(\mathrm{Adj}(A)A\big)_{ij} &= \sum_{m=1}^{n} \big(\mathrm{Adj}(A)\big)_{im} a_{mj} = \sum_{m=1}^{n} \Big[\sum_{(k_1,\cdots,k_n)\in\mathbb{P}(n),\, k_m=i} \varepsilon_{k_1k_2\cdots k_n} \prod_{\substack{1\leqslant \ell\leqslant n\\ \ell\neq m}} a_{\ell k_\ell}\Big] a_{mj} \\ &= \begin{cases} \displaystyle\sum_{(k_1,\cdots,k_n)\in\mathbb{P}(n)} \varepsilon_{k_1k_2\cdots k_n} \prod_{\ell=1}^{n} a_{\ell k_\ell} & \text{if } i=j, \\ 0 & \text{if } i\neq j. \end{cases}\end{aligned}$$

The conclusion then follows from the definition of the determinant. □

Corollary 1.71. *Let* $A = [a_{ij}]$ *be an* n × n *matrix, and* $C = [c_{ij}]$ *be the adjoint matrix of* A*. Then*

$$\det(A) = \sum_{j=1}^{n} a_{ij}c_{ji} = \sum_{j=1}^{n} a_{ji}c_{ij} \quad \forall\, 1\leqslant i\leqslant n\,.$$

Corollary 1.72. *Let* A *be an* n × n *matrix and* $\det(A)\neq 0$*. Then the matrix* $\dfrac{\mathrm{Adj}(A)}{\det(A)}$ *is the inverse matrix of* A*, or equivalently,*

$$\mathrm{Adj}(A) = \det(A)A^{-1}. \tag{1.2}$$

1.5.1 Variations of determinants

Let δ be an operator satisfying the "product rule" $\delta(fg) = f\delta g + (\delta f)g$. Typically δ will be differential operators. By the definition of the determinant,

$$\begin{aligned}\delta \det(A) &= \sum_{(k_1,\cdots,k_n)\in\mathbb{P}(n)} \varepsilon_{k_1k_2\cdots k_n}\delta \prod_{\ell=1}^{n} a_{\ell k_\ell} \\ &= \sum_{i=1}^{n}\Big[\sum_{(k_1,\cdots,k_n)\in\mathbb{P}(n)} \varepsilon_{k_1k_2\cdots k_n}\delta a_{ik_i} \prod_{\substack{1\leqslant \ell\leqslant n\\ \ell\neq i}} a_{\ell k_\ell}\Big] \\ &= \sum_{i,j=1}^{n}\Big[\sum_{(k_1,\cdots,k_n)\in\mathbb{P}(n),\, k_i=j} \varepsilon_{k_1k_2\cdots k_n}\delta a_{ik_i} \prod_{\substack{1\leqslant \ell\leqslant n\\ \ell\neq i}} a_{\ell k_\ell}\Big] \\ &= \sum_{i,j=1}^{n} (-1)^{i+j} \det\big(A(\hat{i},\hat{j})\big)\delta a_{ij}\,.\end{aligned}$$

Therefore, we obtain the following

Theorem 1.73. *Let* A *be an* n×n *matrix, and* δ *be an operator satisfying* $\delta(fg) = f\delta g + (\delta f)g$ *whenever the product makes sense. Then*

$$\delta \det(\mathrm{A}) = \mathrm{tr}\big(\mathrm{Adj}(\mathrm{A})\delta\mathrm{A}\big)\,, \tag{1.3}$$

where $\delta\mathrm{A} \equiv [\delta a_{ij}]_{\mathrm{n}\times\mathrm{n}}$ *if* $\mathrm{A} = [a_{ij}]_{\mathrm{n}\times\mathrm{n}}$. *In particular, if* A *is invertible,*

$$\delta \det(\mathrm{A}) = \det(\mathrm{A})\mathrm{tr}(\big(\mathrm{A}^{-1}\delta\mathrm{A}\big)\,.$$

Example 1.74. Let $\mathrm{A}(x) = \begin{bmatrix} f(x) & g(x) \\ h(x) & k(x) \end{bmatrix}$ and $\delta = \dfrac{d}{dx}$. Then

$$\begin{aligned} \delta \det(\mathrm{A}) &= \mathrm{tr}\Big(\begin{bmatrix} k & -g \\ -h & f \end{bmatrix} \begin{bmatrix} f' & g' \\ h' & k' \end{bmatrix} \Big) = \mathrm{tr}\Big(\begin{bmatrix} kf' - gh' & kg' - gk' \\ -hf' + fh' & -hg' + fk' \end{bmatrix} \Big) \\ &= kf' - gh' - hg' + fk'\,. \end{aligned}$$

1.6 Bounded Linear Maps

Definition 1.75 (Linear map)**.** Let $\mathcal{V}$ and $\mathcal{W}$ be two vector spaces over a scalar field $\mathbb{F}$. A map $L : \mathcal{V} \to \mathcal{W}$ is called a ***linear map*** from $\mathcal{V}$ into $\mathcal{W}$ if

$$L(\alpha\boldsymbol{v} + \boldsymbol{w}) = \alpha L(\boldsymbol{v}) + L(\boldsymbol{w}) \qquad \forall\, \alpha \in \mathbb{F} \text{ and } \boldsymbol{v}, \boldsymbol{w} \in \mathcal{V}.$$

For notational convenience, we often write $L\boldsymbol{v}$ instead of $L(\boldsymbol{v})$. When $\mathcal{V}$ and $\mathcal{W}$ are finite dimensional, linear maps (from $\mathcal{V}$ into $\mathcal{W}$) are sometimes called ***linear transformations*** (from $\mathcal{V}$ into $\mathcal{W}$).

Let $L_1, L_2 : \mathcal{V} \to \mathcal{W}$ be two linear maps, and $\alpha \in \mathbb{F}$ be a scalar. It is easy to see that $\alpha L_1 + L_2 : \mathcal{V} \to \mathcal{W}$ is also a linear map. This is equivalent to say that the collection of linear maps is a vector space, and this induces the following

Definition 1.76. The vector space $\mathscr{L}(\mathcal{V}, \mathcal{W})$ is the collection of linear maps from $\mathcal{V}$ to $\mathcal{W}$.

Definition 1.77 (Boundedness of linear maps)**.** Let $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ and $(\mathcal{W}, \|\cdot\|_{\mathcal{W}})$ be two normed vector spaces over a scalar field $\mathbb{F}$. A linear map $L : \mathcal{V} \to \mathcal{W}$ is said to be bounded if the number

$$\|L\|_{\mathscr{B}(\mathcal{V},\mathcal{W})} \equiv \sup_{\|\boldsymbol{v}\|_{\mathcal{V}}=1} \|L\boldsymbol{v}\|_{\mathcal{W}} = \sup_{\boldsymbol{v}\neq 0} \frac{\|L\boldsymbol{v}\|_{\mathcal{W}}}{\|\boldsymbol{v}\|_{\mathcal{V}}} \tag{1.4}$$

is finite. The collection of all bounded linear map from $\mathcal{V}$ to $\mathcal{W}$ is denoted by $\mathscr{B}(\mathcal{V}, \mathcal{W})$, and $\mathscr{B}(\mathcal{V}, \mathcal{V})$ is also denoted by $\mathscr{B}(\mathcal{V})$ for simplicity.

Remark 1.78. When the domain $\mathcal{V}$ and the target $\mathcal{W}$ under consideration are clear, we use $\|\cdot\|$ instead of $\|\cdot\|_{\mathscr{B}(\mathcal{V},\mathcal{W})}$ to simplify the notation of operator norm.

Remark 1.79. If $\mathcal{V}$ is finite dimensional, then $\mathscr{L}(\mathcal{V},\mathcal{W}) = \mathscr{B}(\mathcal{V},\mathcal{W})$.

Proposition 1.80. *Let $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ and $(\mathcal{W}, \|\cdot\|_{\mathcal{W}})$ be two normed vector spaces over a scalar field $\mathbb{F}$. Then $\big(\mathscr{B}(\mathcal{V},\mathcal{W}), \|\cdot\|\big)$ with $\|\cdot\|$ defined by* (1.4) *is a normed vector space.* (*Therefore, $\|\cdot\|$ is also called an operator norm*).

Definition 1.81 (Dual space)**.** Let $(\mathcal{V}, \|\cdot\|)$ be a normed vector space over field $\mathbb{F}$. An element in $\mathscr{B}(\mathcal{V},\mathbb{F})$ is called a bounded linear functional on $\mathcal{V}$, and the space $\big(\mathscr{B}(\mathcal{V},\mathbb{F}), \|\cdot\|_{\mathscr{B}(\mathcal{V},\mathbb{F})}\big)$ is called the dual space of $(\mathcal{V}, \|\cdot\|)$, and is usually denoted by $\mathcal{V}'$.

Definition 1.82. Let $(\mathcal{V}, \|\cdot\|_{\mathcal{V}})$ and $(\mathcal{W}, \|\cdot\|_{\mathcal{W}})$ be two normed vector spaces over a scalar field $\mathbb{F}$, and $L \in \mathscr{B}(\mathcal{V},\mathcal{W})$. The collection of all elements $\boldsymbol{v} \in \mathcal{V}$ such that $L\boldsymbol{v} = \boldsymbol{0}$ is called the kernel (or the null space) of L and is denoted by $\ker(L)$ or $\mathrm{Null}(L)$. In other words,

$$\ker(L) = \big\{\boldsymbol{v} \in \mathcal{V} \,\big|\, L\boldsymbol{v} = \boldsymbol{0}\big\}\,.$$

Theorem 1.83 (Riesz Representation Theorem)**.** *Let $(\mathcal{V}, (\cdot,\cdot)_{\mathcal{V}})$ be an inner product space, and $f : \mathcal{V} \to \mathbb{R}$ be a bounded linear map. Then there exists a unique $\boldsymbol{w} \in \mathcal{V}$ such that $f(\boldsymbol{v}) = (\boldsymbol{v}, \boldsymbol{w})_{\mathcal{V}}$ for all $\boldsymbol{v} \in \mathcal{V}$.*

Proof. The uniqueness for such a vector $\boldsymbol{w}$ is simply due to the fact that there is no non-trivial vector which is orthogonal to itself.

Now we show the existence of $\boldsymbol{w}$. If $f(\boldsymbol{v}) = 0$ for all $\boldsymbol{v} \in \mathcal{V}$, then $\boldsymbol{w} = \boldsymbol{0}$ does the job. Now suppose that $\ker(f) \subsetneq \mathcal{V}$. Then there exists $\boldsymbol{u} \in \ker(f)^{\perp}$ such that $\|\boldsymbol{u}\|_{\mathcal{V}} = 1$.

For $\boldsymbol{v} \in \mathcal{V}$, consider the vector $\boldsymbol{y} = f(\boldsymbol{v})\boldsymbol{u} - f(\boldsymbol{u})\boldsymbol{v}$. Then $\boldsymbol{y} \in \ker(f)$; thus $\boldsymbol{y} \cdot \boldsymbol{u} = 0$. Therefore,

$$0 = f(\boldsymbol{v})\|\boldsymbol{u}\|_{\mathcal{V}}^2 - f(\boldsymbol{u})(\boldsymbol{v}, \boldsymbol{u})_{\mathcal{V}} = f(\boldsymbol{v}) - (\boldsymbol{v}, \overline{f(\boldsymbol{u})}\boldsymbol{u})_{\mathcal{V}}$$

which implies that $f(\boldsymbol{v}) = (\boldsymbol{v}, \boldsymbol{w})_{\mathcal{V}}$ with $\boldsymbol{w} = \overline{f(\boldsymbol{u})}\boldsymbol{u}$. ▫

By the Riesz representation theorem, we conclude the following

Theorem 1.84. *Let $(\mathcal{V}, (\cdot,\cdot)_{\mathcal{V}})$ and $(\mathcal{W}, (\cdot,\cdot)_{\mathcal{W}})$ be two inner product spaces. Then for all $L \in \mathscr{B}(\mathcal{V},\mathcal{W})$, there exists a unique $L^* \in \mathscr{B}(\mathcal{W},\mathcal{V})$ such that*

$$(L\boldsymbol{v}, \boldsymbol{w})_{\mathcal{W}} = (\boldsymbol{v}, L^*\boldsymbol{w})_{\mathcal{V}} \qquad \forall\, \boldsymbol{v} \in \mathcal{V}, \boldsymbol{w} \in \mathcal{W}\,.$$

Definition 1.85 (Dual operator)**.** Let $\mathcal{V}$ and $\mathcal{W}$ be two inner product spaces, and $L : \mathcal{V} \to \mathcal{W}$ be a bounded linear map. The ***dual operator*** of L, denoted by L^*, is the unique linear map from $\mathcal{W}$ into $\mathcal{V}$ satisfying

$$(L\boldsymbol{v}, \boldsymbol{w})_{\mathcal{W}} = (\boldsymbol{v}, L^*\boldsymbol{w})_{\mathcal{V}} \qquad \forall\, \boldsymbol{v} \in \mathcal{V}, \boldsymbol{w} \in \mathcal{W},$$

where $(\cdot,\cdot)_{\mathcal{V}}$ and $(\cdot,\cdot)_{\mathcal{W}}$ are inner products on $\mathcal{V}$ and $\mathcal{W}$, respectively.

Definition 1.86 (Symmetry of linear maps)**.** An linear map $L \in \mathscr{B}(\mathcal{H})$ is said to be ***symmetric*** if $L = L^*$.

The last part of this section contributes to the following theorem which states that *every bounded linear maps near by* (*measured by the operator norm*) *an invertible bounded linear map is also invertible*.

Theorem 1.87. *Let* $\mathrm{GL}(n)$ *be the set of all invertible linear maps on* $(\mathbb{R}^n, \|\cdot\|_2)$*; that is,*

$$\mathrm{GL}(n) = \big\{ L \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n) \,\big|\, L \text{ is one-to-one (and onto)} \big\} .$$

1. *If* $L \in \mathrm{GL}(n)$ *and* $K \in \mathscr{B}(\mathbb{R}^n, \mathbb{R}^n)$ *satisfying* $\|K - L\|\|L^{-1}\| < 1$, *then* $K \in \mathrm{GL}(n)$.

2. *The mapping* $L \mapsto L^{-1}$ *is continuous on* $\mathrm{GL}(n)$*; that is,*

$$\forall\, \varepsilon > 0\,, \exists\, \delta > 0 \ni \|K^{-1} - L^{-1}\| < \varepsilon \quad \textit{whenever} \quad \|K - L\| < \delta\,.$$

Proof. 1. Let $\|L^{-1}\| = \dfrac{1}{\alpha}$ and $\|K - L\| = \beta$. Then $\beta < \alpha$; thus for every $x \in \mathbb{R}^n$,

$$\begin{aligned}\alpha\|x\|_{\mathbb{R}^n} = \alpha\|L^{-1}Lx\|_{\mathbb{R}^n} &\leqslant \alpha\|L^{-1}\|\|Lx\|_{\mathbb{R}^n} = \|Lx\|_{\mathbb{R}^n} \leqslant \|(L-K)x\|_{\mathbb{R}^n} + \|Kx\|_{\mathbb{R}^n} \\ &\leqslant \beta\|x\|_{\mathbb{R}^n} + \|Kx\|_{\mathbb{R}^n}\,.\end{aligned}$$

As a consequence, $(\alpha - \beta)\|x\|_{\mathbb{R}^n} \leqslant \|Kx\|_{\mathbb{R}^n}$ and this implies that $K : \mathbb{R}^n \to \mathbb{R}^n$ is one-to-one hence invertible.

2. Let $L \in \mathrm{GL}(n)$ and $\varepsilon > 0$ be given. Choose $\delta = \min\Big\{\dfrac{1}{2\|L^{-1}\|}, \dfrac{\varepsilon}{2\|L^{-1}\|^2}\Big\}$. If $\|K - L\| < \delta$, then $K \in \mathrm{GL}(n)$. Since $L^{-1} - K^{-1} = K^{-1}(K - L)L^{-1}$, we find that if $\|K - L\| < \delta$,

$$\|K^{-1}\| - \|L^{-1}\| \leqslant \|K^{-1} - L^{-1}\| \leqslant \|K^{-1}\|\|K - L\|\|L^{-1}\| < \frac{1}{2}\|K^{-1}\|$$

which implies that $\|K^{-1}\| < 2\|L^{-1}\|$. Therefore, if $\|K - L\| < \delta$,

$$\|L^{-1} - K^{-1}\| \leqslant \|K^{-1}\|\|K - L\|\|L^{-1}\| < 2\|L^{-1}\|^2\delta < \varepsilon\,. \qquad \square$$

1.6.1 Matrix norms

Each $\mathrm{m} \times \mathrm{n}$ matrix $\mathrm{A} \in \mathbb{M}(\mathrm{m}, \mathrm{n}; \mathbb{F})$ induces a linear map $L : \mathbb{F}^{\mathrm{n}} \to \mathbb{F}^{\mathrm{m}}$ in a natural way: let $\mathrm{A} = [a_{ij}]_{\mathrm{m}\times\mathrm{n}}$ be a $\mathrm{m} \times \mathrm{n}$ matrix, $\mathcal{B} = \{\mathrm{e}_j\}_{j=1}^{\mathrm{n}}$ and $\widetilde{\mathcal{B}} = \{\widetilde{\mathrm{e}}_k\}_{k=1}^{\mathrm{m}}$ be the standard basis of $\mathbb{F}^{\mathrm{n}}$ and $\mathbb{F}^{\mathrm{m}}$, respectively. We define the linear map $L : \mathbb{F}^{\mathrm{n}} \to \mathbb{F}^{\mathrm{m}}$ by

$$Lx = \sum_{i=1}^{\mathrm{m}} \sum_{j=1}^{\mathrm{n}} a_{ij} x_j \widetilde{\mathrm{e}}_i \in \mathbb{F}^{\mathrm{m}}, \quad \text{where} \quad x = \sum_{j=1}^{\mathrm{n}} x_j \mathrm{e}_j \in \mathbb{F}^{\mathrm{n}},$$

or equivalently, $[Lx]_{\widetilde{\mathcal{B}}} = \mathrm{A}[x]_{\mathcal{B}}$. The linear map L is called the ***linear map induced by the matrix*** A.

By matrix norms it means the operator norm of the induced linear map. However, as introduced in Section 1.6, the operator norm of a linear map depends on the norms equipped on the vector spaces. In particular, we have introduced p-norm on $\mathbb{F}^{\mathrm{n}}$, and we have the following

Definition 1.88. Let $A \in \mathbb{M}(\mathrm{m}, \mathrm{n}; \mathbb{F})$ with induced linear map $L : \mathbb{F}^{\mathrm{n}} \to \mathbb{F}^{\mathrm{m}}$. The p-norm of A, denoted by $\|A\|_p$, is the operator norm of $L : (\mathbb{F}^{\mathrm{n}}, \|\cdot\|_p) \to (\mathbb{F}^{\mathrm{m}}, \|\cdot\|_p)$ given by

$$\|A\|_p = \sup_{\|x\|_p=1} \|Lx\|_p = \sup_{x\neq 0} \frac{\|Lx\|_p}{\|x\|_p}.$$

Remark 1.89. We can also choose different p in the domain and the co-domain. In other words, the (p, q)-norm of $A \in \mathbb{M}(\mathrm{m}, \mathrm{n}, \mathbb{F})$ is the operator norm of the induced linear map $L : (\mathbb{F}^{\mathrm{n}}, \|\cdot\|_p) \to (\mathbb{F}^{\mathrm{m}}, \|\cdot\|_q)$ given by

$$\|A\|_{(p,q)} = \sup_{\|x\|_p=1} \|Lx\|_q = \sup_{x\neq 0} \frac{\|Lx\|_q}{\|x\|_p}.$$

From now on, for notational simplicity we use Ax to denote $[Lx]_{\widetilde{\mathcal{B}}}$ if $\widetilde{\mathcal{B}}$ is the standard basis of the co-domain.

Example 1.90. Consider the case $p = 1$ and $p = \infty$, respectively.

1. $p = \infty$: $\|A\|_\infty = \sup_{\|x\|_\infty=1} \|Ax\|_\infty = \max\Big\{ \sum_{j=1}^{\mathrm{m}} |a_{1j}|, \sum_{j=1}^{\mathrm{m}} |a_{2j}|, \cdots \sum_{j=1}^{\mathrm{m}} |a_{nj}| \Big\}$.

   Reason: Let $x = (x_1, x_2, \cdots, x_{\mathrm{n}})^{\mathrm{T}}$ and $A = \big[a_{ij}\big]_{n\times\mathrm{m}}$. Then

$$Ax = \begin{bmatrix} a_{11}x_1 + \cdots + a_{1m}x_m \\ a_{21}x_1 + \cdots + a_{2m}x_m \\ \vdots \\ a_{n1}x_1 + \cdots + a_{nm}x_m \end{bmatrix}$$

Assume $\max_{1\leqslant i\leqslant n}\sum_{j=1}^{m}|a_{ij}| = \sum_{j=1}^{m}|a_{kj}|$ for some $1 \leqslant k \leqslant n$. Let

$$x = (\mathrm{sgn}(a_{k1}), \mathrm{sgn}(a_{k2}), \cdots, \mathrm{sgn}(a_{kn}))\ .$$

Then $\|x\|_\infty = 1$, and $\|Ax\|_\infty = \sum_{j=1}^{m}|a_{kj}|$.

On the other hand, if $\|x\|_\infty = 1$, then

$$|a_{i1}x_1 + a_{i2}x_2 + \cdots a_{im}x_m| \leqslant \sum_{j=1}^{\mathrm{m}}|a_{ij}| \leqslant \max\Big\{\sum_{j=1}^{\mathrm{m}}|a_{1j}|, \sum_{j=1}^{\mathrm{m}}|a_{2j}|, \cdots \sum_{j=1}^{\mathrm{m}}|a_{nj}|\Big\};$$

thus $\|A\|_\infty = \max\Big\{\sum_{j=1}^{\mathrm{m}}|a_{1j}|, \sum_{j=1}^{\mathrm{m}}|a_{2j}|, \cdots \sum_{j=1}^{\mathrm{m}}|a_{nj}|\Big\}$. In other words, $\|A\|_\infty$ is the largest sum of the absolute value of row entries.

2. $p = 1$: $\|A\|_1 = \max\Big\{\sum_{i=1}^{n}|a_{i1}|, \sum_{i=1}^{n}|a_{i2}|, \cdots, \sum_{i=1}^{n}|a_{im}|\Big\}$.

Reason: Let $(\cdot,\cdot)$ denote the inner product in $\mathbb{R}^{\mathrm{m}}$. Then for $x \in \mathbb{R}^{\mathrm{n}}$ with $\|x\|_1 = 1$, by Hölder's inequality (1.1) and Theorem 1.25 we have

$$\begin{aligned}\|Ax\|_1 &= \sup_{\|y\|_\infty=1}(Ax, y) = \sup_{\|y\|_\infty=1}(x, A^{\mathrm{T}}y) \leqslant \sup_{\|y\|_\infty=1}\|x\|_1\|A^{\mathrm{T}}y\|_\infty \\ &= \sup_{\|y\|_\infty=1}\|A^{\mathrm{T}}y\|_\infty = \|A^{\mathrm{T}}\|_\infty\,;\end{aligned}$$

thus $\|A\|_1 = \sup_{\|x\|_1=1}\|Ax\|_1 \leqslant \|A^{\mathrm{T}}\|_\infty$. Similarly, if $y \in \mathbb{R}^{\mathrm{m}}$ and $\|y\|_\infty = 1$, then

$$\begin{aligned}\|A^{\mathrm{T}}y\|_\infty &= \sup_{\|x\|_1=1}(x, A^{\mathrm{T}}y) = \sup_{\|x\|_1=1}(Ax, y) \leqslant \sup_{\|x\|_1=1}\|Ax\|_1\|y\|_\infty \\ &= \sup_{\|x\|_1=1}\|Ax\|_1 = \|A\|_1\end{aligned}$$

which implies that $\|A^{\mathrm{T}}\|_\infty = \sup_{\|y\|_\infty=1}\|A^{\mathrm{T}}y\|_\infty \leqslant \|A\|_1$. As a consequence,

$$\|A\|_1 = \|A^{\mathrm{T}}\|_\infty = \max\Big\{\sum_{i=1}^{n}|a_{i1}|, \sum_{i=1}^{n}|a_{i2}|, \cdots, \sum_{i=1}^{n}|a_{im}|\Big\}.$$

1.7 Representation of Linear Transformations

In Section 1.6.1, we see that any $\mathrm{m} \times \mathrm{n}$ matrix is associated with a linear map. On the other hand, suppose that $\mathcal{V}$ is a n-dimensional vector space with basis $\mathcal{B} = \{\boldsymbol{v}_j\}_{j=1}^{\mathrm{n}}$, and $\mathcal{W}$ is a m-dimensional vector space with basis $\widetilde{\mathcal{B}} = \{\boldsymbol{w}_i\}_{i=1}^{\mathrm{m}}$. Define $\mathbf{V} = \big[\boldsymbol{v}_1 \vdots \cdots \vdots \boldsymbol{v}_{\mathrm{n}}\big]$ and $\mathbf{W} = \big[\boldsymbol{w}_1 \vdots \cdots \vdots \boldsymbol{w}_{\mathrm{m}}\big]$, and let $L \in \mathscr{L}(\mathcal{V}, \mathcal{W})$. Since $L\boldsymbol{v}_j \in \mathcal{W}$, for each $1 \leqslant j \leqslant \mathrm{n}$ we can write $L\boldsymbol{v}_j = \sum\limits_{i=1}^{\mathrm{m}} a_{ij}\boldsymbol{w}_i$ for some coefficients a_{ij}. Moreover, if $\boldsymbol{u} \in \mathcal{V}$, then

$$\boldsymbol{u} = \sum_{j=1}^{\mathrm{n}} c_j \boldsymbol{v}_j \qquad \text{or} \qquad \boldsymbol{c} = [\boldsymbol{u}]_{\mathcal{B}} \qquad \text{or} \qquad \boldsymbol{u} = \mathbf{V}\boldsymbol{c}\,,$$

and by the linearity of L,

$$L\boldsymbol{u} = L\Big(\sum_{j=1}^{\mathrm{n}} c_j\boldsymbol{v}_j\Big) = \sum_{j=1}^{\mathrm{n}} c_j L\boldsymbol{v}_j = \sum_{j=1}^{\mathrm{n}}\sum_{i=1}^{\mathrm{m}} c_j a_{ij}\boldsymbol{w}_i = \sum_{i=1}^{\mathrm{m}}\Big(\sum_{j=1}^{\mathrm{n}} a_{ij}c_j\Big)\boldsymbol{w}_i\,.$$

Let $b_i = \sum\limits_{j=1}^{\mathrm{n}} a_{ij}c_j$, and $\boldsymbol{b} = [b_1, \cdots, b_{\mathrm{m}}]^{\mathrm{T}}$. Then

$$[L\boldsymbol{u}]_{\widetilde{\mathcal{B}}} = \boldsymbol{b} = A\boldsymbol{c} = A[\boldsymbol{u}]_{\mathcal{B}}\,.$$

The discussion above induces the following

Definition 1.91. Let $\mathcal{V}, \mathcal{W}$ be two vector spaces, $\dim(\mathcal{V}) = \mathrm{n}$ and $\dim(\mathcal{W}) = \mathrm{m}$, and $\mathcal{B}, \widetilde{\mathcal{B}}$ are basis of $\mathcal{V}, \mathcal{W}$, respectively. For $L \in \mathscr{L}(\mathcal{V}, \mathcal{W})$, the ***matrix representation*** of L relative to bases $\mathcal{B}$ and $\widetilde{\mathcal{B}}$, denoted by $[L]_{\widetilde{\mathcal{B}},\mathcal{B}}$, is the matrix satisfying

$$[L\boldsymbol{u}]_{\widetilde{\mathcal{B}}} = [L]_{\widetilde{\mathcal{B}},\mathcal{B}}[\boldsymbol{u}]_{\mathcal{B}} \qquad \forall\, \boldsymbol{u} \in \mathcal{V}\,.$$

If $L \in \mathscr{L}(\mathcal{V}, \mathcal{V})$, we simply use $[L]_{\mathcal{B}}$ to denote $[L]_{\mathcal{B},\mathcal{B}}$.

1.8 Matrix Diagonalization

Definition 1.92 (Eigenvalues and Eigenvectors)**.** Let $\mathcal{V}$ be a finite dimensional vector spaces over a scalar field $\mathbb{F}$, and $L \in \mathscr{B}(\mathcal{V})$. A scalar $\lambda \in \mathbb{F}$ is said to be an ***eigenvalue*** of L if there is a non-zero vector $\boldsymbol{v} \in \mathcal{V}$ such that $L\boldsymbol{v} = \lambda\boldsymbol{v}$. The collection of all eigenvalues of L is denoted by $\sigma(L)$.

For an eigenvalue $\lambda \in \mathbb{F}$ of L, a non-zero vector $\boldsymbol{v} \in \mathcal{V}$ satisfying $L\boldsymbol{v} = \lambda\boldsymbol{v}$ is called an ***eigenvector*** associated with the eigenvalue λ, and the collection of all $\boldsymbol{v} \in \mathcal{V}$ such that $L\boldsymbol{v} = \lambda\boldsymbol{v}$ is called the ***eigenspace*** associated with λ.

Let $\dim(\mathcal{V}) = \mathrm{n}$ and $\mathcal{B}$ be a basis of $\mathcal{V}$. Then if $\lambda \in \mathbb{F}$ is an eigenvalue of $L \in \mathscr{B}(\mathcal{V})$, there exists non-zero vector $\boldsymbol{v} \in \mathcal{V}$ such that

$$[L]_{\mathcal{B}}[\boldsymbol{v}]_{\mathcal{V}} = [L\boldsymbol{v}]_{\mathcal{B}} = \lambda[\boldsymbol{v}]_{\mathcal{B}} ;$$

thus the matrix representation $[L]_{\mathcal{B}}$ of L satisfies that $[L]_{\mathcal{B}} - \lambda\mathrm{I}_{\mathrm{n}}$ is singular (not invertible). Therefore, $\det([L]_{\mathcal{B}} - \lambda\mathrm{I}_{\mathrm{n}}) = 0$ which motivates the following

Definition 1.93. Let $A \in \mathbb{M}(\mathrm{n}, \mathrm{n}; \mathbb{F})$ be a $\mathrm{n} \times \mathrm{n}$ matrix over scalar field $\mathbb{F}$. An ***eigenvalue*** of A is a scalar $\lambda \in \mathbb{F}$ such that $\det(A - \lambda\mathrm{I}_{\mathrm{n}}) = 0$.

Theorem 1.94. *Let $L \in \mathscr{B}(\mathbb{F}^{\mathrm{n}})$ be symmetric. Then $\sigma(L) \subseteq \mathbb{R}$.*

Proof. Let $\lambda \in \sigma(L)$, and $\boldsymbol{v}$ be an eigenvector associated with λ. Then

$$\lambda(\boldsymbol{v}, \boldsymbol{v})_{\mathbb{F}^{\mathrm{n}}} = (\lambda\boldsymbol{v}, \boldsymbol{v})_{\mathbb{F}^{\mathrm{n}}} = (L\boldsymbol{v}, \boldsymbol{v})_{\mathbb{F}^{\mathrm{n}}} = (\boldsymbol{v}, L^*\boldsymbol{v})_{\mathbb{F}^{\mathrm{n}}} = (\boldsymbol{v}, L\boldsymbol{v})_{\mathbb{F}^{\mathrm{n}}} = (\boldsymbol{v}, \lambda\boldsymbol{v}) = \bar{\lambda}(\boldsymbol{v}, \boldsymbol{v})_{\mathbb{F}^{\mathrm{n}}}$$

which implies that $\lambda \in \mathbb{R}$. □

Lemma 1.95. *Let $L \in \mathscr{B}(\mathbb{F}^{\mathrm{n}})$ be symmetric, and $(\cdot, \cdot)_{\mathbb{F}^{\mathrm{n}}}$ be the standard inner product on $\mathbb{F}^{\mathrm{n}}$. Then the two numbers*

$$m \equiv \inf_{\|u\|_{\mathbb{F}^{\mathrm{n}}}=1} (Lu, u)_{\mathbb{F}^{\mathrm{n}}} \quad \textit{and} \quad M \equiv \sup_{\|u\|_{\mathbb{F}^{\mathrm{n}}}=1} (Lu, u)_{\mathbb{F}^{\mathrm{n}}}$$

belong to $\sigma(L)$.

Proof. Suppose that $M \notin \sigma(L)$. Let $[u, v] = (Mu - Lu, v)_{\mathbb{F}^{\mathrm{n}}}$. Then $[\cdot, \cdot]$ is an inner product on $\mathbb{F}^{\mathrm{n}}$; thus the Cauchy-Schwarz inequality (Proposition 1.16) implies that

$$\big|[u, v]\big| \leqslant \big|[u, u]\big|^{1/2}\big|[v, v]\big|^{1/2} .$$

By Theorem 1.25, we find that

$$\begin{aligned} \|Mu - Lu\|_{\mathbb{F}^{\mathrm{n}}} &= \sup_{\|v\|_{\mathbb{F}^{\mathrm{n}}}=1} \big|(Mu - Lu, v)_{\mathbb{F}^{\mathrm{n}}}\big| = \sup_{\|v\|_{\mathbb{F}^{\mathrm{n}}}=1} \big|[u, v]\big| \leqslant \sup_{\|v\|_{\mathbb{F}^{\mathrm{n}}}=1} \big|[u, u]\big|^{1/2}\big|[v, v]\big|^{1/2} \\ &\leqslant (M - m)^{1/2}(Mu - Lu, u)_{\mathbb{F}^{\mathrm{n}}}^{1/2} \qquad \forall\, u \in \mathbb{F}^{\mathrm{n}} , \end{aligned} \tag{1.5}$$

where we use the fact that $\sup_{\|v\|_{\mathbb{F}^{\mathrm{n}}}=1} \big|[v, v]\big|^{1/2} = (M - m)^{1/2}$ to conclude the last inequality.

Let $\mathcal{B}$ be the standard basis of $\mathbb{F}^n$, and $\{\boldsymbol{u}_k\}_{k=1}^\infty$ be a sequence of vectors in $\mathbb{F}^n$ such that $\|\boldsymbol{u}_k\|_{\mathbb{F}^n} = 1$, and $\lim\limits_{k\to\infty}(L\boldsymbol{u}_k, \boldsymbol{u}_k)_{\mathbb{F}^n} = M$. Then (1.5) implies $\|M\boldsymbol{u}_k - L\boldsymbol{u}_k\|_{\mathbb{F}^n} \to 0$ as $k \to \infty$. Since $M \notin \sigma(L)$, $M\mathrm{I}_n - [L]_{\mathcal{B}}$ is invertible; thus

$$[\boldsymbol{u}_k]_{\mathcal{B}} = (M\mathrm{I}_n - [L]_{\mathcal{B}})^{-1}(M[\boldsymbol{u}_k]_{\mathcal{B}} - [L]_{\mathcal{B}}[\boldsymbol{u}_k]_{\mathcal{B}}) \to \mathbf{0} \quad \text{in} \quad \mathbb{F}^n$$

which contradicts to $\|u_k\|_{\mathbb{F}^n} = 1$ for all $k \in \mathbb{N}$. Hence $M \in \sigma(L)$. Similarly, $m \in \sigma(L)$. □

Definition 1.96 (Diagonalizable linear maps)**.** Let $\mathcal{V}$ be a finite dimensional vector spaces over a scalar field $\mathbb{F}$. A linear map $L : \mathcal{V} \to \mathcal{V}$ is said to be ***diagonalizable*** if there is a basis $\mathcal{B}$ of $\mathcal{V}$ such that each $\boldsymbol{v} \in \mathcal{B}$ is an eigenvector of L.

Theorem 1.97. *Let $L \in \mathscr{B}(\mathbb{R}^n)$ be symmetric. Then there exists an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of L.*

Example 1.98 (The 2-norm of matrices)**.** Let $(\cdot,\cdot)_{\mathbb{R}^k}$ denote the inner product in Euclidean space $\mathbb{R}^k$, and $A \in \mathbb{M}(m, n; \mathbb{R})$. Since $A^{\mathrm{T}}A$ is a symmetric n × n matrix, it is diagonalizable by an orthonormal matrix P; that is, $A^{\mathrm{T}}A = P\Lambda P^{\mathrm{T}}$ for some orthonormal n × n matrix P and diagonal n × n matrix $\Lambda = [\lambda_i\delta_{ij}]$. Therefore,

$$\|Ax\|_2^2 = (Ax, Ax)_{\mathbb{R}^m} = (x, A^{\mathrm{T}}Ax)_{\mathbb{R}^n} = (x, P\Lambda P^{\mathrm{T}}x)_{\mathbb{R}^n} = (P^{\mathrm{T}}x, \Lambda P^{\mathrm{T}}x)_{\mathbb{R}^n}$$

which implies that

$$\begin{aligned}\sup_{\|x\|_2=1}\|Ax\|_2^2 &= \sup_{\|x\|_2=1}(P^{\mathrm{T}}x, \Lambda P^{\mathrm{T}}x)_{\mathbb{R}^n} = \sup_{\|y\|_2=1}(y, \Lambda y)_{\mathbb{R}^n} \quad (\text{Let } y = P^{\mathrm{T}}x, \text{ then } \|y\|_2 = 1)\\ &= \sup_{\|y\|_2=1}(\lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2)\\ &= \max\{\lambda_1, \cdots, \lambda_n\} = \text{maximum eigenvalue of } A^{\mathrm{T}}A.\end{aligned}$$

As a consequence, $\|A\|_2 = \sqrt{\text{maximum eigenvalue of } A^{\mathrm{T}}A}$.

1.9 The Einstein Summation Convention

In mathematics, especially in applications of linear algebra to physics, the Einstein summation convention is a notational convention that implies summation over a set of indexed terms in a formula, thus achieving notational brevity. According to this convention, when an index variable appears twice in a single term it implies summation of that term over all the values of the index. For example, with this convention, the inner product $\boldsymbol{u} \cdot \boldsymbol{v}$ of two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^{\mathrm{n}}$, where $\boldsymbol{u} = (u_1, \cdots, u_{\mathrm{n}})$ and $\boldsymbol{v} = (v_1, \cdots, v_{\mathrm{n}})$, can be expressed as $u_i v_i$, and the i-th component of the cross product $\boldsymbol{u} \times \boldsymbol{v}$ of two vectors $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^3$ can be expressed as $\varepsilon_{ijk} u^j v^k$.

In this book, we make a further convention that repeated Latin indices are summed from 1 to n, and repeated Greek indices are summed from 1 to n − 1, where n is the space dimension. In other words, we use the symbol $f_i g_i$ to denote the sum $\sum_{i=1}^{\mathrm{n}} f_i g_i$, and the symbol $f_\alpha g_\alpha$ to denote the sum $\sum_{i=1}^{\mathrm{n}-1} f_\alpha g_\alpha$. Starting from the next Chapter, we use such summation convention for notational simplicity.