Contents

Chapter 1

基本矩陣分析

1.2 範數與特徵值

1.2.1 範數

Let X be a vector space over $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$.

Definition 1.2.1 (Vector norms)**.** Let $N : X \to \mathbb{R}^+$. Then N is a **(vector) norm** [範數], if

(i) $N(\alpha x) = |\alpha| N(x), \alpha \in \mathbb{K},$ for $x \in X$;

(ii) $N(x+y) \leq N(x) + N(y),$ for $x, y \in X$;

(iii) $N(x) = 0$ if and only if $x = 0$.

The usual notation is $\|x\| = N(x)$.

Definition 1.2.2 (p-norms; 1, 2, ∞ norms)**.** For any $p \geq 1$, we define the **p-norm** $\|\cdot\|_p : \mathbb{C}^n \to \mathbb{R}^+$ by

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}. \tag{1.2.1}$$

(Then it can be verified that it is a vector norm). Especially,

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \text{ (1-norm)}, \tag{1.2.2}$$

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2 \right)^{1/2} \quad \text{(2-norm = Euclidean-norm)}, \tag{1.2.3}$$

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \text{ (∞-norm = maximum norm)}. \tag{1.2.4}$$

Lemma 1.2.1 (Continuity of Norm)**.** $N(x)$ is a continuous function in the components $x_1, \ldots, x_n$ of x.

Proof.

$$|N(x) - N(y)| \leq N(x-y) \leq \sum_{j=1}^{n} |x_j - y_j| N(e_j) \leq \|x - y\|_\infty \sum_{j=1}^{n} N(e_j),$$

in which e_j is the j-th column of the identity matrix I_n. □

Lemma 1.2.2 (Equivalence of Norms)**.** Let N and M be any two norms on $\mathbb{C}^n$. Then there are constants $c_1, c_2 > 0$ such that

$$c_1 M(x) \leq N(x) \leq c_2 M(x), \text{ for all } x \in \mathbb{C}^n. \tag{1.2.5}$$

Definition 1.2.3 (Matrix Norm)**.** Let $\|\cdot\| : \mathbb{C}^{m\times n} \to \mathbb{R}^+$. Then $\|\cdot\|$ is a **matrix norm**, if

(H1) $\|\alpha A\| = |\alpha|\|A\|$;

(H2) $\|A+B\| \le \|A\| + \|B\|$;

(H3) $\|A\| = 0$ if and only if $A = 0$;

(H4) $\|AB\| \le \|A\|\|B\|$. (Required for some norms)

(H5) $\|Ax\| \le \|A\| \cdot \|x\|$ (for some assigned vector norm $\|\cdot\|$) (Required for some norms).

Definition 1.2.4 (Frobenius Norm)**.** The **Frobenius norm** $\|\cdot\|_F : \mathbb{C}^{n\times n} \to \mathbb{R}^+$ is defined by

$$\|A\|_F = \left(\sum_{i,j=1}^{n} |a_{i,j}|^2\right)^{1/2}. \tag{1.2.6}$$

(It can be verified that it is a matrix norm satisfying (H1)~(H5) in the definition of Matrix Norms.)

Proof.

- H4: By Cauchy-Schwartz inequality, we have

$$\begin{aligned}\|AB\|_F &= \left(\sum_{i,j}\left|\sum_k a_{ik}b_{kj}\right|^2\right)^{1/2}\\ &\le \left(\sum_{i,j}\left\{\sum_k |a_{ik}|^2\right\}\left\{\sum_k |b_{kj}|^2\right\}\right)^{1/2}\\ &= \left(\sum_i\sum_k |a_{ik}|^2\right)^{1/2}\left(\sum_j\sum_k |b_{kj}|^2\right)^{1/2}\\ &= \|A\|_F\|B\|_F.\end{aligned}$$

- H5: By Cauchy-Schwartz inequality, we have

$$\begin{aligned}\|Ax\|_2 &= \left(\sum_i\left|\sum_j a_{ij}x_j\right|^2\right)^{1/2}\\ &\le \left(\sum_i\left(\sum_j |a_{ij}|^2\right)\left(\sum_j |x_j|^2\right)\right)^{1/2}\\ &= \|A\|_F\|x\|_2.\end{aligned}$$

□

Remark 1. $\|I_n\|_F = \sqrt{n}$.

Definition 1.2.5 (Operator Norm)**.** Given a vector norm $\|\cdot\|$. An **associated (induced) matrix norm** is defined by

$$\|A\| = \sup_{x\neq 0}\frac{\|Ax\|}{\|x\|} = \max_{x\neq 0}\frac{\|Ax\|}{\|x\|}.$$

(It can be verified that it is a matrix norm satisfying (H1) ~(H5).)

Remark 2. For any operator norm, $\|I_n\| = 1$.

Theorem 1.2.3 (1, 2, and ∞ Matrix Norms)**.**

$$\|A\|_1 = \sup_{x\neq 0} \frac{\|Ax\|_1}{\|x\|_1} = \max_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}| = \text{Max Absolute Column Sumns;} \tag{1.2.7}$$

$$\|A\|_\infty = \sup_{x\neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}| = \text{Max Absolute Row Sumns;} \tag{1.2.8}$$

$$\|A\|_2 = \sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \sqrt{\rho(A^*A)} = \text{Max Eigenvalues of } A^*A. \tag{1.2.9}$$

1.2.2 範數與特徵值的關係

Theorem 1.2.4 (Comparison of Spectral Radius and Matrix Norm (v1))**.**

(i) For any matrix norm $\|\cdot\|$ and matrix $A \in \mathbb{C}^{n\times n}$,

$$\rho(A) \leq \|A\|. \tag{1.2.10}$$

(ii) Given $A \in \mathbb{C}^{n\times n}$. Then, for any $\epsilon > 0$, there exists some matrix norm $\|\cdot\|\epsilon$ such that

$$\|\cdot\|_\epsilon \leq \rho(A) + \epsilon. \tag{1.2.11}$$

Theorem 1.2.5 (Convergence of A^n as $n\to\infty$)**.** Let $A \in \mathbb{C}^{n\times n}$. The following statements are equivalent:

(i) $\lim\limits_{m\to\infty} A^m = 0$;

(ii) $\lim\limits_{m\to\infty} A^m x = 0$ for all x;

(iii) $\rho(A) < 1$.

Theorem 1.2.6 (Comparison of Spectral Radius and Matrix Norm (v2))**.** For any operator norm $\|\cdot\|$,

$$\rho(A) = \lim_{k\to\infty} \|A^k\|^{1/k}. \tag{1.2.12}$$

Theorem 1.2.7 (Inverse of $I - A$ as An Power Series)**.** Let $A \in \mathbb{C}^{n\times n}$, and $\rho(A) < 1$. Then $(I-A)^{-1}$ exists and

$$(I-A)^{-1} = I + A + A^2 + \cdots. \tag{1.2.13}$$

Corollary 1.2.8 (Estimation of the Norm of the inverse of $I - A$)**.** If $\|A\| < 1$, then $(I-A)^{-1}$ exists and

$$\|(I-A)^{-1}\| \leq \frac{1}{1-\|A\|}. \tag{1.2.14}$$

Lemma 1.2.9. For any $A \in \mathbb{C}^{m\times n}$, and unitarycmatrices $U \in \mathbb{C}^{m\times m}, V \in \mathbb{C}^{n\times n}$, it holds

$$\begin{aligned} &\|UAV\|_F = \|A\|_F \text{ (by } \|UA\|_F = \sqrt{\|Ua_1\|_2^2 + \ldots + \|Ua_n\|_2^2}), \\ &\|UAV\|_2 = \|A\|_2. \end{aligned}$$

Theorem 1.2.10 (Singular Value Decomposition (SVD))**.** Let $A \in \mathbb{C}^{m\times n}$. Then there exist unitary matrices $U = [u_1, \ldots, u_m] \in \mathbb{C}^{m\times m}$ and $V = [v_1, \ldots, v_n] \in \mathbb{C}^{n\times n}$ such that

$$U^*AV = \mathrm{diag}(\sigma_1, \ldots, \sigma_p) = \Sigma_{m\times n}, \tag{1.2.15}$$

where $p = \min\{m, n\}$ and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$. (Here, σ_i denotes the i-th largest singular value of A). Equivalently,

$$A = U_{m\times m}\Sigma_{m\times n}V^*_{n\times n}. \tag{1.2.16}$$

It is called the **singular value decomposition (SVD)** of A.

Remark 3.

(i) $U^*AV = \Sigma \Longrightarrow AV = U\Sigma$. So

$$A\begin{bmatrix}v_1 \mid v_2 \mid \cdots \mid v_n\end{bmatrix} = \begin{bmatrix}u_1 \mid u_2 \mid \cdots \mid u_m\end{bmatrix}\Sigma_{m\times n}.$$

- If $m < n$, then

$$\begin{aligned}
A\begin{bmatrix}v_1 \mid v_2 \mid \cdots \mid v_n\end{bmatrix} &= \begin{bmatrix}u_1 \mid u_2 \mid \cdots \mid u_m\end{bmatrix}\begin{bmatrix}\sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0\\ 0 & \sigma_2 & \ddots & \vdots & \vdots & & \vdots\\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots\\ 0 & \cdots & 0 & \sigma_m & 0 & \cdots & 0\end{bmatrix}\\
\Longleftrightarrow \begin{bmatrix}Av_1 \mid Av_2 \mid \cdots \mid Av_n\end{bmatrix} &= \begin{bmatrix}\sigma_1u_1 \mid \sigma_2u_2 \mid \cdots \mid \sigma_mu_m \mid 0 \mid \cdots \mid 0\end{bmatrix}\\
\Longleftrightarrow Av_i = \sigma u_i\ (i = 1,\dots,m) &\text{ and } Av_i = 0\ (i = m+1,\dots,n).
\end{aligned}$$

- If $m > n$, then

$$\begin{aligned}
A\begin{bmatrix}v_1 \mid v_2 \mid \cdots \mid v_n\end{bmatrix} &= \begin{bmatrix}u_1 \mid u_2 \mid \cdots \mid u_m\end{bmatrix}\begin{bmatrix}\sigma_1 & 0 & \cdots & 0\\ 0 & \sigma_2 & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \sigma_n\\ 0 & \cdots & \cdots & 0\\ \vdots & & & \vdots\\ 0 & \cdots & \cdots & 0\end{bmatrix}\\
\Longleftrightarrow \begin{bmatrix}Av_1 \mid Av_2 \mid \cdots \mid Av_n\end{bmatrix} &= \begin{bmatrix}\sigma_1u_1 \mid \sigma_2u_2 \mid \cdots \mid \sigma_nu_n\end{bmatrix}\\
\Longleftrightarrow Av_i = \sigma u_i\ (i = 1,\dots,n).
\end{aligned}$$

(ii) Let

$$\begin{aligned} f : \mathbb{C}^n &\to \mathbb{C}^m\\ x &\mapsto Ax. \end{aligned}$$

There exist orthonormal set $\{v_1, v_2, \cdots, v_n\}$ in $\mathbb{C}^n$, and orthonormal set $\{u_1, u_2, \cdots, u_m\}$ in $\mathbb{C}^m$ such that

- If $m < n$, then

$$Av_i = \sigma u_i\ (i = 1,\dots,m) \text{ and } Av_i = 0\ (i = m+1,\dots,n). \tag{1.2.17}$$

- If $m > n$, then

$$Av_i = \sigma u_i\ (i = 1,\dots,n). \tag{1.2.18}$$

(iii) $U^*A = \Sigma \Longrightarrow A = U\Sigma V^*$. So

- If $m < n$, then

$$A = \begin{bmatrix} u_1 \mid u_2 \mid \cdots \mid u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots & \vdots & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & \vdots \\ 0 & \cdots & 0 & \sigma_m & 0 & \cdots & 0 \end{bmatrix} \begin{bmatrix} v_1^* \\ v_2^* \\ \vdots \\ v_n^* \end{bmatrix}$$

$$\begin{aligned} &\iff A = \begin{bmatrix} \sigma_1 u_1 \mid \sigma_2 u_2 \mid \cdots \mid \sigma_m u_m \mid 0 \mid \cdots \mid 0 \end{bmatrix} \begin{bmatrix} v_1^* \\ v_2^* \\ \vdots \\ v_n^* \end{bmatrix} \\ &\iff A = \sigma_1 u_1 v_1^* + \cdots \sigma_1 u_m v_m^* + 0 v_{m+1}^* + \cdots + 0 v_n^* \\ &\iff A = \sigma_1 u_1 v_1^* + \cdots + \sigma_m u_m v_m^*. \end{aligned} \tag{1.2.19}$$

- If $m > n$, then

$$A = \begin{bmatrix} u_1 \mid u_2 \mid \cdots \mid u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n \\ 0 & \cdots & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \begin{bmatrix} v_1^* \\ v_2^* \\ \vdots \\ v_n^* \end{bmatrix}$$

$$\begin{aligned} &\iff \begin{bmatrix} Av_1 \mid Av_2 \mid \cdots \mid Av_n \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 \mid \sigma_2 u_2 \mid \cdots \mid \sigma_n u_n \end{bmatrix} \begin{bmatrix} v_1^* \\ v_2^* \\ \vdots \\ v_n^* \end{bmatrix} \\ &\iff A = \sigma_1 u_1 v_1^* + \cdots \sigma_m u_n v_n^*. \end{aligned} \tag{1.2.20}$$

We can summarize (1.2.19) and (1.2.20) as

$$A = \sigma_1 u_1 v_1^* + \cdots + \sigma_p u_p v_p^* \tag{1.2.21}$$

$$= \begin{bmatrix} u_1 \mid u_2 \mid \cdots \mid u_p \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_p \end{bmatrix} \begin{bmatrix} v_1^* \\ v_2^* \\ \vdots \\ v_p^* \end{bmatrix}, \tag{1.2.22}$$

where $p = \min\{m, n\}$.

(iv) For any $x \in \mathbb{C}^n$,

$$\begin{aligned} Ax &= (\sigma_1 u_1 v_1^* + \cdots + \sigma_p u_p v_p^*)x \\ &= \sigma_1 u_1 (v_1^* x) + \cdots \sigma_p u_p (v_p^* x) \\ &= \sigma_1 \langle v_1, x \rangle u_1 + \cdots \sigma_p \langle v_p, x \rangle u_p. \end{aligned}$$

(v) Note that

$$AA^* = (U\Sigma V^*)(U\Sigma V^*)^* = U(\Sigma\Sigma^*)U^* \in \mathbb{C}^{m \times m}, \tag{1.2.23}$$

where

$$\Sigma\Sigma^* = \begin{cases} \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_p^2 \end{bmatrix}_{m\times m} & \text{if } m < n, \\ \begin{bmatrix} \sigma_1^2 & & & & & & \\ & \sigma_2^2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_p^2 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}_{m\times m} & \text{if } m > n. \end{cases}$$

It implies that

$$\sigma_i^2,\ i = 1,\dots,p, \text{ are eigenvalues of } AA^* \text{ and } u_i,\ i = 1,\dots,m, \text{ are eigenvectors of } AA^*. \tag{1.2.24}$$

Similarly, it holds that

$$A^*A = (U\Sigma V^*)^*(U\Sigma V^*) = V(\Sigma^*\Sigma)V^* \in \mathbb{C}^{n\times n}, \tag{1.2.25}$$

where

$$\Sigma\Sigma^* = \begin{cases} \begin{bmatrix} \sigma_1^2 & & & & & & \\ & \sigma_2^2 & & & & & \\ & & \ddots & & & & \\ & & & \sigma_p^2 & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{bmatrix}_{n\times n} & \text{if } m < n, \\ \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_p^2 \end{bmatrix}_{n\times n} & \text{if } m > n. \end{cases}$$

It implies that

$$\sigma_i^2,\ i = 1,\dots,p, \text{ are eigenvalues of } A^*A \text{ and } v_i,\ i = 1,\dots,n, \text{ are eigenvectors of } A^*A. \tag{1.2.26}$$

(vi) Especially, $\sigma_1 = \|A\|_2 = \sqrt{A^*A}$.

Corollary 1.2.11 (2 matrix Norm v.s. Frobenius Norm)**.** For any $A \in \mathbb{C}^{n\times n}$,

$$\|A\|_2 \le \|A\|_F \le \sqrt{n}\|A\|_2. \tag{1.2.27}$$

1.2.3 不同範數之間的關係

Lemma 1.2.12 (Holder Inequality)**.** For any p, q satisfies $\frac{1}{p} + \frac{1}{q} = 1$, including the case $(p, q) = (1, \infty)$,

$$\int_S |f(x)g(x)|\, d\mu \le \left(\int_S |f(x)|^p\, d\mu\right)^{\frac{1}{p}} \cdot \left(\int_S |g(x)|^q\, d\mu\right)^{\frac{1}{q}}. \tag{1.2.28}$$

Corollary 1.2.13 (Generalization of Cauchy-Schwartz Inequality)**.** For any p, q satisfies $\frac{1}{p}+\frac{1}{q}=1$, including the case $(p,q)=(1,\infty)$,

$$\sum_{i=1}^{n}|x_i y_i| \leq \left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}} \cdot \left(\sum_{i=1}^{n}|y_i|^q\right)^{\frac{1}{q}}. \tag{1.2.29}$$

Theorem 1.2.14 (Jensen Inequality)**.** For any convex function ϕ, arbitrary function f, and set Ω satisfying $\mu(\Omega)=\int_\Omega d\mu = 1$, it holds

$$\phi\left(\int_\Omega f\, d\mu\right) \leq \int_\Omega (\phi \circ f)\, d\mu. \tag{1.2.30}$$

- We say a function ϕ is **convex** on a convex set X if

$$\phi(tx_1 + (1-t)x_2) \leq t\phi(x_1) + (1-t)\phi(x_2). \tag{1.2.31}$$

  for any $x_1, x_2 \in X$ and $t \in [0,1]$.

Corollary 1.2.15.

(i) For any convex function ϕ and arbitrary function f defined on the interval I with $|I|=1$,

$$\phi\left(\int_I f(x)\, dx\right) \leq \int_I \phi(f(x))\, dx. \tag{1.2.32}$$

Especially,

$$\left(\int_I f(x)\, dx\right)^2 \leq \int_I f^2(x)\, dx. \tag{1.2.33}$$

(ii) For each fixed $n \in \mathbb{N}$, let $\Omega = \{1, 2, \cdots, n\}$ and $\mu(i) = \frac{1}{n}$, $i = 1, \cdots, n$. Then, for any convex function ϕ and arbitrary function f,

$$\phi\left(\frac{1}{n}\sum_{i=1}^{n} f(i)\right) \leq \frac{1}{n}\sum_{i=1}^{n}\phi(f(i)). \tag{1.2.34}$$

Theorem 1.2.16 (Inequality between Different Norms)**.**

(i) For any $q \geq p$,

$$1 \leq \frac{\|x\|_p}{\|x\|_q} \leq n^{\frac{1}{p}-\frac{1}{q}},\ \forall x \in \mathbb{C}^n. \tag{1.2.35}$$

- The proof of the second inequality proof needs (1.2.34).

In partucular, for any $p > 0$,

$$1 \leq \frac{\|x\|_p}{\|x\|_\infty} \leq n^{\frac{1}{p}},\ \forall x \in \mathbb{C}^n. \tag{1.2.36}$$

(ii) For any $\frac{1}{p}+\frac{1}{q}=1$,

$$|y^* x| \leq \|x\|_p \|y\|_q,\ \forall x, y \in \mathbb{C}^n. \tag{1.2.37}$$

Moreover,

$$\max\{|y^*x| : \|y\|_q = 1\} = \|x\|_p,\ \forall x \in \mathbb{C}^n, \tag{1.2.38}$$
$$\max\{|y^*x| : \|x\|_p = 1\} = \|y\|_q,\ \forall y \in \mathbb{C}^n. \tag{1.2.39}$$

- The proof of the second inequality proof needs (1.2.29).

(iii) For any $p \geq 1$, and matrix $A = \begin{bmatrix} a_1 \mid a_2 \mid \cdots \mid a_n \end{bmatrix} = \begin{bmatrix} a_{ij} \end{bmatrix} \in \mathbb{C}^{m \times n}$,

$$\max_{1 \leq j \leq n} \|a_j\|_p \leq \|A\|_p \leq n^{\frac{p-1}{p}} \max_{1 \leq j \leq n} \|a_j\|_p, \tag{1.2.40}$$

$$\max_{i,j} |a_{ij}| \leq \|A\|_p \leq n^{\frac{p-1}{p}} m^{\frac{1}{p}} \max_{i,j} |a_{ij}|. \tag{1.2.41}$$

(iv) For any $p \geq 1$, and matrix $A \in \mathbb{C}^{m \times n}$,

$$m^{\frac{1-p}{p}} \|A\|_1 \leq \|A\|_p \leq n^{\frac{p-1}{p}} \|A\|_1. \tag{1.2.42}$$

(v) For any $\frac{1}{p} + \frac{1}{q} = 1$, and matrix $A \in \mathbb{C}^{m \times n}$,

$$\|A\|_p = \|A^*\|_q. \tag{1.2.43}$$

(vi) For any $p \geq 1$, and matrix $A \in \mathbb{C}^{m \times n}$,

$$n^{-\frac{1}{p}} \|A\|_\infty \leq \|A\|_p \leq m^{\frac{1}{p}} \|A\|_\infty. \tag{1.2.44}$$

(vii) For any $\frac{1}{p} + \frac{1}{q} = 1$, and matrix $A \in \mathbb{C}^{m \times n}$,

$$\|A\|_2 \leq \sqrt{\|A\|_p \|A\|_q}. \tag{1.2.45}$$

(viii) For any $q > p \geq 1$, and matrix $A \in \mathbb{C}^{m \times n}$,

$$n^{\frac{1}{q} - \frac{1}{p}} \|A\|_q \leq \|A\|_p \leq m^{\frac{1}{p} - \frac{1}{q}} \|A\|_q. \tag{1.2.46}$$