

## 附录 A

## 矩阵论和范数

在阅读本附录之前,读者应熟悉很多数学工具的知识,包括特征值、特征向量和奇异值分解等;读者也应该熟练掌握各种向量、矩阵、信号和系统范数之间的区别,同时应该知道如何使用这些范数来度量性能。

### A.1 基本知识

设有复标量  $c = \alpha + j\beta$ , 其中  $\alpha = \operatorname{Re} c$  为实部,  $\beta = \operatorname{Im} c$  为虚部; 为了得到其幅值  $|c|$ , 复标量乘以其共轭  $\bar{c} = \alpha - j\beta$ , 并取其平方根得

$$|\bar{c}| = \sqrt{\bar{c}c} = \sqrt{\alpha^2 - j^2\beta^2} = \sqrt{\alpha^2 + \beta^2}$$

设  $A = (a_{ij}) \in \mathbb{C}^{n \times m}$ , 其中  $a_{ij} = \operatorname{Re} a_{ij} + j\operatorname{Im} a_{ij}$ ; 其共轭转置矩阵 (Hermit 伴随矩阵) 记作  $A^* \stackrel{\text{def}}{=} \bar{A}^T$ ; 矩阵的迹记作  $\operatorname{tr} A$  (对角线元素之和); 行列式记作  $\det A$ 。根据定义, 非奇异矩阵  $A$  的逆记作  $A^{-1}$ , 并满足  $A^{-1}A = AA^{-1} = I$ , 逆矩阵通过下式给出

$$A^{-1} = \frac{\operatorname{adj} A}{\det A} \quad (\text{A.1.1})$$

其中,  $\operatorname{adj} A$  称为  $A$  的伴随矩阵。同时还有

$$(AB)^T = B^T A^T, \quad (AB)^* = B^* A^* \quad (\text{A.1.2})$$

若假定逆矩阵都存在, 还有

$$(AB)^{-1} = B^{-1} A^{-1} \quad (\text{A.1.3})$$

若  $A^T = A$ , 则称  $A$  为对称阵; 若  $A^* = A$ , 则称  $A$  为 Hermit 阵。

如果一个矩阵的对称部分  $(A + A^*)$  的特征值全部为正, 或对于所有的非零向量  $x$  都有  $x^*(A + A^*)x > 0$ , 则称矩阵  $A$  是正定的。如果  $A$  是 Hermit 阵, 则正定条件就简化为  $x^*Ax > 0$ , 这可简单记作  $A > 0$ 。类似地, 对于 Hermit 阵  $A$ , 其半正定 ( $A \geq 0$ ) 的条件是  $x^*Ax \geq 0$ 。对于一个半正定矩阵  $A$ , 其矩阵平方根  $A^{1/2}$  满足  $A^{1/2}A^{1/2} = A$ 。类似可以定义矩阵的负定性和半负定性。

一个矩阵  $A$  的秩, 就是其列向量 (行向量) 独立元的最大个数, 记为  $\operatorname{rank}(A)$ 。

#### A.1.1 分块高斯消元法与 Schur 补

考虑分块实方阵

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad (\text{A.1.4})$$

希望利用分块高斯消元法使其变为上三角阵或下三角阵。为此,令

$$\begin{bmatrix} I & 0 \\ X & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ XA+C & XB+D \end{bmatrix}$$

如果  $A$  可逆,选择  $X = -CA^{-1}$ ,则

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & B \\ 0 & \Delta_A \end{bmatrix}$$

类似地利用右乘也能够得到

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ C & \Delta_A \end{bmatrix}$$

其中

$$\Delta_A = D - CA^{-1}B \quad (\text{A.1.5})$$

称为  $A$  在  $M$  中的 **Schur 补**。

进而,如果  $D$  可逆,则令

$$\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} \Delta_D & 0 \\ C & D \end{bmatrix}$$

或

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} = \begin{bmatrix} \Delta_D & B \\ 0 & D \end{bmatrix}$$

其中

$$\Delta_D = A - BD^{-1}C \quad (\text{A.1.6})$$

称为  $D$  在  $M$  中的 **Schur 补**。

Schur 补在矩阵代数中是非常有用的。利用矩阵乘法可以得到如下行列式计算公式

$$\begin{aligned} \det \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \det A \det(D - CA^{-1}B) = \det A \det \Delta_A \\ &= \det D \det(A - BD^{-1}C) = \det D \det \Delta_D \end{aligned} \quad (\text{A.1.7})$$

同样可以用于矩阵的三角分解,因为

$$\begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix}$$

或

$$\begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} = \begin{bmatrix} \Delta_D & 0 \\ 0 & D \end{bmatrix}$$

再考虑到

$$\begin{bmatrix} I & 0 \\ P & I \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -P & I \end{bmatrix}$$

则可得矩阵  $M$  的分解公式

$$\begin{aligned} M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} I & 0 \\ CA^{-1} & I \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & \Delta_A \end{bmatrix} \begin{bmatrix} I & A^{-1}B \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} I & BD^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \Delta_D & 0 \\ 0 & D \end{bmatrix} \begin{bmatrix} I & 0 \\ D^{-1}C & I \end{bmatrix} \end{aligned} \quad (\text{A.1.8})$$

上式也可以等价地写成

$$\begin{aligned} M = \begin{bmatrix} A & B \\ C & D \end{bmatrix} &= \begin{bmatrix} A \\ C \end{bmatrix} A^{-1} \begin{bmatrix} A & B \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \Delta_A \end{bmatrix} \\ &= \begin{bmatrix} B \\ D \end{bmatrix} D^{-1} \begin{bmatrix} C & D \end{bmatrix} + \begin{bmatrix} \Delta_D & 0 \\ 0 & 0 \end{bmatrix} \end{aligned} \quad (\text{A.1.9})$$

这个表达式也可以用来定义 Schur 补。

### A.1.2 矩阵求逆引理

如果式(A.1.4)的分块实方阵是可逆的,则根据式(A.1.8)可以写成

$$\begin{aligned} M^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I & -A^{-1}B \\ 0 & I \end{bmatrix} \begin{bmatrix} A^{-1} & 0 \\ 0 & \Delta_A^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ -CA^{-1} & I \end{bmatrix} \\ &= \begin{bmatrix} A^{-1} + A^{-1}B\Delta_A^{-1}CA^{-1} & -A^{-1}B\Delta_A^{-1} \\ -\Delta_A^{-1}CA^{-1} & \Delta_A^{-1} \end{bmatrix} \end{aligned} \quad (\text{A.1.10})$$

或者

$$\begin{aligned} M^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} &= \begin{bmatrix} I & 0 \\ -D^{-1}C & I \end{bmatrix} \begin{bmatrix} \Delta_D^{-1} & 0 \\ 0 & D^{-1} \end{bmatrix} \begin{bmatrix} I & -BD^{-1} \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} \Delta_D^{-1} & -\Delta_D^{-1}BD^{-1} \\ -D^{-1}C\Delta_D^{-1} & D^{-1} + D^{-1}C\Delta_D^{-1}BD^{-1} \end{bmatrix} \end{aligned} \quad (\text{A.1.11})$$

综合以上两式可以得出

$$\Delta_D^{-1} = A^{-1} + A^{-1}B\Delta_A^{-1}CA^{-1} \quad (\text{A.1.12})$$

$$\Delta_A^{-1} = D^{-1} + D^{-1}C\Delta_D^{-1}BD^{-1} \quad (\text{A.1.13})$$

**矩阵求逆引理** 设  $A \in \mathbb{R}^{n \times n}$ ,  $C \in \mathbb{R}^{m \times m}$  均可逆;  $B \in \mathbb{R}^{n \times m}$ ,  $D \in \mathbb{R}^{m \times n}$ , 则有如下恒等式

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[C^{-1} + DA^{-1}B]^{-1}DA^{-1} \quad (\text{A.1.14})$$

**证明** 考虑  $\Delta_A = D - CA^{-1}B$ ,  $\Delta_D = A - BD^{-1}C$ , 由式(A.1.12)和式(A.1.13)得

$$\Delta_D^{-1} = (A - BD^{-1}C)^{-1} = A^{-1} + A^{-1}B\Delta_A^{-1}CA^{-1} = A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1}$$

令  $D^{-1} = \tilde{C}$ , 则有

$$(A - B\tilde{C}C) = A^{-1} + A^{-1}B(\tilde{C}^{-1} - CA^{-1}B)^{-1}CA^{-1}$$

再令  $C = -D$ , 则有

$$(A + B\tilde{C}D) = A^{-1} - A^{-1}B(\tilde{C}^{-1} + DA^{-1}B)^{-1}DA^{-1}$$

再令  $\tilde{C} = C$ , 即可得式(A.1.14)。 ■

**推论 1** 式(A.1.14)还可以写成如下替换形式

$$[A + BCD]^{-1} = A^{-1} - A^{-1}B[I_m + CDA^{-1}B]^{-1}CDA^{-1} \quad (\text{A.1.15})$$

**推论 2** 如果  $C = I_m$ , 则式(A.1.14)变为

$$[A + BD]^{-1} = A^{-1} - A^{-1}B[I_m + DA^{-1}B]^{-1}DA^{-1} \quad (\text{A.1.16})$$

### A.1.3 一些行列式恒等式

设  $A \in \mathbb{C}^{n \times n}$  为复方阵, 定义

$$\det A = \sum_{i=1}^n a_{ij}c_{ij} \quad (\text{沿列展开}) \quad (\text{A.1.17})$$

$$\det A = \sum_{j=1}^n a_{ij} c_{ij} \text{ (沿行展开)} \quad (\text{A. 1. 18})$$

其中,  $c_{ij}$  是指第  $ij$  个余子式, 即

$$c_{ij} = [\text{adj} A]_{ji} \stackrel{\text{def}}{=} (-1)^{i+j} A^{ij} \quad (\text{A. 1. 19})$$

此处  $A^{ij}$  是去掉矩阵的第  $i$  行和第  $j$  列构成的子阵。

如果  $\det A$  非零, 则矩阵是非奇异的。从定义出发, 可以直接得到  $\det A = \det A^T$ 。于是, 行列式其他一些恒等式如下:

(1) 设  $A_1, A_2$  是方阵, 而且具有相同维数, 则

$$\det(A_1 A_2) = \det(A_2 A_1) = \det A_1 \cdot \det A_2 \quad (\text{A. 1. 20})$$

(2) 设  $c$  为复标量,  $A \in \mathbb{C}^{n \times n}$ , 则

$$\det(cA) = c^n \det(A) \quad (\text{A. 1. 21})$$

(3) 设  $A$  为非奇异阵, 则

$$\det A^{-1} = \frac{1}{\det A} \quad (\text{A. 1. 22})$$

(4) 设  $A_1, A_2$  是适当维数的矩阵, 使得  $A_1 A_2$  以及  $A_2 A_1$  均为方阵 (但  $A_1, A_2$  本身不需要为方阵), 则

$$\det(I + A_1 A_2) = \det(I + A_2 A_1) \quad (\text{A. 1. 23})$$

这实际上是后面式 (A. 1. 25) 给出 Schur 公式的特例, 式 (A. 1. 23) 在控制领域是非常有用的, 原因是由其可以得出  $\det(I + GK) = \det(I + KG)$  的结论。

(5) 对于三角阵或分块三角阵的行列式, 就是对角块行列式的乘积, 即

$$\det \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix} = \det \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} = \det(A_{11}) \cdot \det(A_{22}) \quad (\text{A. 1. 24})$$

(6) Schur 公式 分块矩阵的行列式

$$\begin{aligned} \det \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} &= \det(A_{11}) \cdot \det(A_{22} - A_{21} A_{11}^{-1} A_{12}) \\ &= \det(A_{22}) \cdot \det(A_{11} - A_{12} A_{22}^{-1} A_{21}) \end{aligned} \quad (\text{A. 1. 25})$$

其中, 假定  $A_{11}$  和  $A_{22}$  是非奇异的。

**证明** 注意, 若  $A_{11}$  非奇异, 则  $A$  可以分解为

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ A_{21} A_{11}^{-1} & I \end{bmatrix} \begin{bmatrix} A_{11} & 0 \\ 0 & X \end{bmatrix} \begin{bmatrix} I & A_{11}^{-1} A_{12} \\ 0 & I \end{bmatrix} \quad (\text{A. 1. 26})$$

其中  $X = A_{22} - A_{21} A_{11}^{-1} A_{12}$ 。利用式 (A. 1. 20) 和式 (A. 1. 24) 计算行列式, 则式 (A. 1. 25) 的第一部分已经得到证明。类似地, 如果  $A_{22}$  是非奇异的, 则

$$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} I & A_{12} A_{22}^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} Y & 0 \\ 0 & A_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ A_{22}^{-1} A_{21} & I \end{bmatrix} \quad (\text{A. 1. 27})$$

其中  $Y = A_{11} - A_{12} A_{22}^{-1} A_{21}$ 。同样, 式 (A. 1. 25) 的后半部分也类似得出。 ■

## A.2 矩阵特征值和特征向量

**定义 A. 2. 1 (特征值和特征向量)** 设  $A \in \mathbb{C}^{n \times n}$  为复方阵, 特征值  $\lambda_i, i=1, 2, \dots, n$  是如下  $n$  阶特征方程的解



$$\det(A - \lambda I) = 0 \quad (\text{A. 2.1})$$

对应于特征值  $\lambda_i$  的(右)特征向量  $x_i \in \mathbb{C}^n$  就是如下方程的非零解

$$(A - \lambda_i I)x_i = 0 \Leftrightarrow Ax_i = \lambda_i x_i \quad (\text{A. 2.2})$$

相应的左特征向量  $y_i \in \mathbb{C}^n$  满足

$$y_i^* (A - \lambda_i I) = 0 \Leftrightarrow y_i^* A = \lambda_i y_i^* \quad (\text{A. 2.3})$$

当我们说特征向量时,一般指的是右特征向量。

**注释 1** 注意,如果  $x_i \in \mathbb{C}^n$  是  $A$  的一个特征向量,则对任意常数  $\alpha$ ,  $\alpha x_i$  也是一个特征向量。因此,通常把特征向量标准化,使其具有单位长度,即  $x_i^* x_i = 1$ 。

**注释 2** 由式(A. 2. 3)可以得到  $A^* y_i = \bar{\lambda}_i y_i$ , 因此可由  $A^*$  的右特征向量得到  $A$  的左特征向量  $y_i$ 。然而请注意,因为有特征值的共轭  $\bar{\lambda}_i$ , 所以对  $A$  与  $A^*$  来说,复特征值的顺序可能不同。

**注释 3** 特征值有时也称作特征增益。 $A$  的特征值集合也称为  $A$  的谱。 $A$  的特征值的最大绝对值就是  $A$  的谱半径

$$\rho(A) \stackrel{\text{def}}{=} \max_i |\lambda_i(A)| \quad (\text{A. 2.4})$$

互异特征值相对应的特征向量总是线性独立的。对于重特征值,情况可能不总是这样。也就是说,并非所有  $n \times n$  矩阵都有  $n$  个线性独立的特征向量。

如果矩阵  $A$  具有特征值  $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ , 相应的特征向量为  $\{t_1, t_2, \dots, t_n\}$ , 且令

$$T = [t_1, t_2, \dots, t_n], \quad \Lambda = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_n\} \quad (\text{A. 2.5})$$

则可以把式(A. 2. 2)写成  $AT = T\Lambda$  的形式。如果特征向量是线性独立的,使得  $T^{-1}$  存在,那么  $A$  可以“对角化”为如下形式

$$\Lambda = T^{-1}AT \quad (\text{A. 2.6})$$

如果特征值是互异的,该式总成立。而对其他情况如  $A = I$ , 可能也成立。对于互异的特征值,其左右特征向量是相互直交的,可进行尺度变换,使其相互正交

$$y_i^* = \begin{cases} 1, & \text{如果 } i=j \\ 0, & \text{如果 } i \neq j \end{cases}$$

然后,对矩阵  $A$  有如下对左右特征向量的二元展开或谱分解

$$A = \sum_{i=1}^n \lambda_i x_i y_i^* \quad (\text{A. 2.7})$$

**注释** 特征值不是互异(即有重特征值)的情况在理论上和计算上都更加复杂。幸运的是,从实际的观点来看,只要理解了互异特征值的情况就足够了。

### A. 2.1 特征值的特性

矩阵  $A$  的特征值有以下特性:

- (1) 矩阵  $A$  特征值的总和等于  $A$  的迹(对角线上元素之和):  $\text{tr}A = \sum_i \lambda_i$ ;
- (2) 矩阵  $A$  特征值的乘积等于  $A$  行列式的值:  $\det A = \prod_i \lambda_i$ ;
- (3) 上三角或下三角阵的特征值是其对角元素;
- (4) 实矩阵的特征值是实数或共轭复数对;

- (5)  $A$  和其转置  $A^T$  有相同的特征值(但一般情况下有不同的特征向量);
- (6) 当且仅当  $A$  的所有特征值非零, 其逆阵  $A^{-1}$  存在且其特征值为  $1/\lambda_1, \dots, 1/\lambda_n$ ;
- (7) 矩阵  $A+cI$  的特征值为  $\lambda_i+c$ ;
- (8) 当  $k$  为整数时, 矩阵  $cA^k$  的特征为  $c\lambda_i^k$ ;
- (9) 假定  $A$  为  $l \times m$  矩阵,  $B$  为  $m \times l$  矩阵, 则  $l \times l$  的矩阵  $AB$  和  $m \times m$  矩阵  $BA$  有相同的非零特征值; 更具体地说, 假设  $l > m$ , 那么矩阵  $AB$  和  $BA$  有  $m$  个相同的特征值, 而  $l-m$  个特征值为零;
- (10) 在相似变换下特征值保持不变, 即  $A$  和  $DAD^{-1}$  有相同的特征值;
- (11) 特征向量组成的矩阵  $T$  可对角化矩阵  $A$  (见式(A.2.6)) 和矩阵  $(I+A)^{-1}$  (证明:  $T^{-1}(I+A)^{-1}T = (T^{-1}(I+A)T)^{-1} = (I+A)^{-1}$ );
- (12) Gershorin 定理:  $n \times n$  矩阵  $A$  的特征值位于复平面内  $n$  个圆的并集上, 圆心是  $a_{ii}$ , 半径是  $r_i = \sum_{j \neq i} |a_{ij}|$  (第  $i$  行非对角元素之和), 则这些特征值也位于另外  $n$  个圆的并集上, 圆心仍是  $a_{ii}$ , 而半径是  $r'_i = \sum_{j \neq i} |a_{ji}|$  (第  $i$  列非对角元素之和);
- (13) Hermit 阵(因此还有对称阵)的特征值是实数;
- (14) 一个 Hermit 阵正定( $A > 0$ )的条件是, 当且仅当其所有特征值为正数。

### A.2.2 系统状态矩阵的特征值

考虑用线性微分方程描述的系统

$$\dot{x}(t) = Ax(t) + Bu(t) \quad (\text{A.2.8})$$

除非  $A$  是对角阵, 这就是一个耦合的微分方程组。为简单起见, 假定矩阵  $A$  的特征向量  $t_i$  是线性独立的, 并构成矩阵  $T = [t_1, t_2, \dots, t_n]$ ; 引入新的状态向量  $z(t) = T^{-1}x(t)$ , 即  $x(t) = Tz(t)$ , 然后得到

$$T\dot{z}(t) = ATz(t) + Bu(t) \Leftrightarrow \dot{z}(t) = \Lambda z(t) + T^{-1}Bu(t) \quad (\text{A.2.9})$$

按照新的状态变量, 这是一个解耦的微分方程组。每一状态  $z_i(t)$  的自由解(即在  $u(t)=0$  条件下)是  $z_i(t) = z_{0i}e^{\lambda_i t}$ , 其中  $z_{0i}$  是  $t=0$  时刻的状态。如果  $\lambda_i$  为实数, 则当且仅当  $\lambda_i < 0$  时, 这个模态是稳定的(即当  $t \rightarrow \infty$  时  $z_i(t) \rightarrow 0$ )。如果  $\lambda_i = \text{Re}\lambda_i + j\text{Im}\lambda_i$  为复数, 则可看出  $e^{\lambda_i t} = e^{\text{Re}\lambda_i t} (\cos(\text{Im}\lambda_i t) + j\sin(\text{Im}\lambda_i t))$ , 当且仅当  $\text{Re}\lambda_i < 0$  时, 这个模态是稳定的(即当  $t \rightarrow \infty$  时  $z_i(t) \rightarrow 0$ )。事实上, 由于真实物理状态  $x = Tz$  当然是实数, 那么新的状态  $z_i$  是复数也无关紧要。因此, 对于一个线性系统, 当且仅当其状态矩阵  $A$  的所有特征值其实部都小于零, 即位于左半开平面, 则系统是稳定的。

## A.3 矩阵的 Kronecker 乘积

设  $A \in \mathbb{R}^{m \times m}$ ,  $B \in \mathbb{R}^{n \times n}$  是两个矩阵, 其特征值分别为

$$\lambda(A) = \{\alpha_1, \alpha_2, \dots, \alpha_m\}, \quad \lambda(B) = \{\beta_1, \beta_2, \dots, \beta_n\} \quad (\text{A.3.1})$$

则  $A$  与  $B$  的 Kronecker 乘积定义为

$$A \otimes B \stackrel{\text{def}}{=} \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mm}B \end{bmatrix}$$

**引理(Kronecker 乘积性质)** 给定矩阵  $A \in \mathbb{R}^{m \times m}$  和  $B \in \mathbb{R}^{n \times n}$ , 以及矩阵  $C \in \mathbb{R}^{m \times p}$ ,  $D \in \mathbb{R}^{n \times r}$ , 则如下事实成立:

- (1)  $(A \otimes B)(C \otimes D) = AC \otimes BD$ ;
- (2) 如果  $A$  和  $B$  是可逆的, 则  $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$ ;
- (3)  $(A \otimes B)$  的特征值是  $\lambda(A \otimes B) = \{\alpha_i \beta_j\}_{i=1,2,\dots,m; j=1,2,\dots,n}$ ;
- (4)  $\det(A \otimes B) = (\det A)^n (\det B)^m$ ;

(5) 对于任意可匹配维数的矩阵  $\{A, B, C, X\}$ , 如果  $C = AXB$ , 则有如下关系式:  $\text{vec}(C) = (B^T \otimes A) \text{vec}(X)$ , 其中  $\text{vec}(M)$  表示把矩阵  $M$  的各列向量排成单个列向量。

**证明** (1) 由定义可直接验证, 即

$$(A \otimes B)(C \otimes D) = \{q_{ij}\}_{m \times r} = \left\{ \sum_{k=1}^m a_{ik} c_{kj} BD \right\} = AC \otimes BD$$

(2) 根据(1)选择  $C = A^{-1}$ ,  $D = B^{-1}$ , 则有

$$(A \otimes B)(A^{-1} \otimes B^{-1}) = I \otimes I = I$$

(3) 选择  $C$  作为  $A$  的右特征向量, 即  $AC = [\alpha_1 \quad \alpha_2 \quad \cdots \quad \alpha_m]C$ ; 选择  $D$  作为  $B$  的左特征向量, 即  $BD = [\beta_1 \quad \beta_2 \quad \cdots \quad \beta_n]D$ , 根据(1)则有

$$(A \otimes B)(C \otimes D) = AC \otimes BD = [\alpha_1, \alpha_2, \dots, \alpha_m]C \otimes [\beta_1, \beta_2, \dots, \beta_n]D = \{\alpha_i \beta_j\}C \otimes D$$

从而有

$$\lambda(A \otimes B) = \{\alpha_i \beta_j\}_{i=1,2,\dots,m; j=1,2,\dots,n}$$

(4) 根据(3)有

$$\det(A \otimes B) = \prod_{i=1}^m \prod_{j=1}^n \alpha_i \beta_j$$

而由  $\det A = \prod_{i=1}^m \alpha_i$ ,  $\det B = \prod_{j=1}^n \beta_j$  推知  $(\det A)^n = \prod_{j=1}^n \prod_{i=1}^m \alpha_i$ ,  $(\det B)^m = \prod_{i=1}^m \prod_{j=1}^n \beta_j$ ; 从而有  $\det(A \otimes B) = (\det A)^n (\det B)^m$ 。

(5) 根据 Kronecker 乘积的定义, 以及两个列向量  $a, b$  满足  $\text{vec}(ab^T) = b \otimes a$  的事实, 有

$$\text{不妨令 } C = AXB = A \begin{bmatrix} x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_m^T \end{bmatrix} = \sum_{i=1}^m Ax_i b_i^T, \text{ 由此得到}$$

$$\text{vec}(C) = \text{vec} \left[ \sum_{i=1}^m Ax_i b_i^T \right] = \sum_{i=1}^m b_i \otimes Ax_i = \begin{bmatrix} \sum_{i=1}^m b_{i1} Ax_i \\ \sum_{i=1}^m b_{i2} Ax_i \\ \vdots \\ \sum_{i=1}^m b_{im} Ax_i \end{bmatrix}$$



同样通过计算得到

$$(B^T \otimes A) \text{vec}(X) = \begin{bmatrix} b_{11}A & b_{21}A & \cdots & b_{m1}A \\ b_{12}A & b_{22}A & \cdots & b_{m2}A \\ \vdots & \vdots & \ddots & \vdots \\ b_{1r}A & b_{mr}A & \cdots & b_{nr}A \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^m b_{i1}Ax_i \\ \sum_{i=1}^m b_{i2}Ax_i \\ \vdots \\ \sum_{i=1}^m b_{ir}Ax_i \end{bmatrix}$$

结论得证。 ■

## A.4 矩阵的 QR 分解

设  $A = [a_1, a_2, \dots, a_m] \in \mathbb{R}^{n \times m}$ , 且  $m \leq n, \text{rank} A = m$ ; 希望得到

$$A = QR \quad (\text{A.4.1})$$

其中  $R \in \mathbb{R}^{m \times m}$  是上三角阵, 而

$$Q = [q_1 \ q_2 \ \cdots \ q_m] \in \mathbb{R}^{n \times m} \quad (\text{A.4.2})$$

具有直交的列向量, 即  $\langle q_i, q_j \rangle = q_i^T q_j = \delta_{ij}, \forall i, j$ ; 而  $R$  的对角线上的元都是正的。

这个问题的本质就是把  $A$  中的列向量  $\{a_1, a_2, \dots, a_m\}$  转化为一组直交向量  $\{q_1, q_2, \dots, q_m\}$ , 而完成这一过程的经典方法就是 Gram-Schmidt 直交化程序。令

$$\begin{cases} q_1 = \frac{a_1}{\|a_1\|} \\ q_i = \frac{r_i}{\|r_i\|}, \quad r_i = a_i - \sum_{j=1}^{i-1} \langle a_i, q_j \rangle q_j, \quad i = 2, 3, \dots, m \end{cases} \quad (\text{A.4.3})$$

显然,  $\|q_i\| = 1, i = 1, 2, \dots, m$ , 且  $\langle q_i, q_j \rangle = 0, \forall i \neq j$ 。

又因为  $\text{rank} A = m$ , 则  $\{a_1, a_2, \dots, a_m\}$  线性独立, 于是有  $\|r_i\| > 0, i = 1, 2, \dots, m$ ; 而且式(A.4.3)可以写成

$$a_i = \sum_{j=1}^{i-1} \langle a_i, q_j \rangle q_j + \|r_i\| q_i$$

于是上三角阵  $R \in \mathbb{R}^{m \times m}$  可以写成

$$R = \begin{bmatrix} \|r_1\| & \langle a_2, q_1 \rangle & \cdots & \langle a_m, q_1 \rangle \\ 0 & \|r_2\| & \cdots & \langle a_m, q_2 \rangle \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \|r_m\| \end{bmatrix} \quad (\text{A.4.4})$$

## A.5 矩阵的奇异值分解

定义 A.5.1 一个(复)矩阵  $U$  如果满足

$$U^* = U^{-1} \quad (\text{A.5.1})$$



那么这个矩阵称为酉阵。

酉阵所有特征值的绝对值都等于1,而且所有的奇异值(由下面的定义看出)都为1。

**定义 A.5.2** 任一复矩阵  $A \in \mathbb{C}^{n \times m}$ , 可利用奇异值分解(SVD)将其分解为

$$A = U \Sigma V^* \quad (\text{A.5.2})$$

其中  $U \in \mathbb{C}^{n \times n}$ ,  $V \in \mathbb{C}^{m \times m}$  都是酉阵, 而  $\Sigma \in \mathbb{C}^{n \times m}$  包含一个对角阵  $\Sigma_1$ , 具有实的非负奇异值  $\sigma_i$  递减排列如下

$$\Sigma = \begin{bmatrix} \Sigma_1 \\ 0 \end{bmatrix}, \quad n \geq m \quad (\text{A.5.3})$$

或

$$\Sigma = [\Sigma_1 \quad 0], \quad m \geq n \quad (\text{A.5.4})$$

其中

$$\Sigma_1 = \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_k\}, \quad k = \min\{m, n\} \quad (\text{A.5.5})$$

且

$$\bar{\sigma} \equiv \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k \equiv \sigma \quad (\text{A.5.6})$$

酉阵  $U$  和  $V$  是  $A$  的列空间和行空间构成的直交基。 $V$  的列向量记为  $v_i$ , 称为右奇异向量或输入奇异向量;  $U$  的行向量记为  $u_i$ , 称为左奇异向量或输出奇异向量。定义  $\bar{u} \equiv u_1$ ,  $\bar{v} \equiv v_1$ ,  $\bar{u} \equiv u_k$ ,  $\bar{v} \equiv v_k$ 。

注意, 式(A.5.2)中的分解并不唯一。例如, 对于一个方阵, 其 SVD 选择为  $A = U' \Sigma V'^*$ , 其中  $U' = U S$ ,  $V' = V S$ ,  $S = \text{diag}\{e^{j\theta_i}\}$ , 而  $\theta_i$  为任意的实数; 然而奇异值  $\sigma_i$  却是唯一的。

奇异值就是  $AA^*$  或  $A^*A$  的  $k = \min\{m, n\}$  个最大特征值的正平方根, 即

$$\sigma_i(A) = \sqrt{\lambda_i(AA^*)} = \sqrt{\lambda_i(A^*A)} \quad (\text{A.5.7})$$

同时,  $U$  和  $V$  的列分别是  $AA^*$  和  $A^*A$  的单位特征向量。为了推导式(A.5.7), 有

$$AA^* = (U \Sigma V^*) (U \Sigma V^*)^* = (U \Sigma V^*) (V \Sigma^* U^*) = U \Sigma \Sigma^* U^* \quad (\text{A.5.8})$$

由于  $U$  是酉阵且满足  $U^* = U^{-1}$ , 则等价地有

$$(AA^*)U = U \Sigma \Sigma^* \quad (\text{A.5.9})$$

可以看出  $U$  是  $AA^*$  特征向量组成的矩阵。同样地,  $V$  是  $A^*A$  特征向量组成的矩阵。

**定义 A.5.3** 一个矩阵的秩, 等于该矩阵非零奇异值的个数。若  $A$  的秩为  $r$ , 且  $r < k = \min\{n, m\}$ , 则称矩阵  $A$  为亏秩矩阵。对  $i = r+1, \dots, k$ , 有奇异值  $\sigma_i$  为零。一个亏秩的方阵是奇异阵(非方阵总是奇异阵)。

给一个矩阵左乘或右乘一个非奇异矩阵, 它的秩保持不变。进而, 对于一个  $n \times m$  的矩阵  $A$  和  $m \times p$  的矩阵  $B$ , 它们乘积  $AB$  的秩可按如下定界(Sylvester 不等式)

$$\text{rank}(A) + \text{rank}(B) - m \leq \text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B)) \quad (\text{A.5.10})$$

一般来讲, 奇异值必须数值计算才能得到。然而对于  $2 \times 2$  矩阵, 很容易得到一个解析表达式。由式(A.5.7)知  $\sigma_i(A) = \sqrt{\lambda_i(A^*A)}$ , 引入

$$b \stackrel{\text{def}}{=} \text{tr}(A^T A) = \sum_{i,j} |a_{ij}|^2, \quad c \stackrel{\text{def}}{=} \det(A^T A)$$

矩阵特征值的总和就等于它的迹, 而乘积等于行列式的值, 所以有

$$\lambda_1 + \lambda_2 = b, \quad \lambda_1 \cdot \lambda_2 = c$$

对  $\lambda_1, \lambda_2$  求解, 利用  $\sigma_i(A) = \sqrt{\lambda_i(A^*A)}$  则有

$$\bar{\sigma}(A) = \sqrt{\frac{b + \sqrt{b^2 - 4c}}{2}}, \quad \sigma(A) = \sqrt{\frac{b - \sqrt{b^2 - 4c}}{2}} \quad (\text{A. 5.11})$$

由于有  $A = U\Sigma V^*$ , 假设  $A$  是一个  $m \times m$  的非奇异阵, 那么有

$$A^{-1} = V\Sigma^{-1}U^* \quad (\text{A. 5.12})$$

这就是  $A^{-1}$  的 SVD, 但奇异值的顺序却是反过来的。

奇异值可对特征值的大小定界, 即对矩阵  $A$  有

$$\sigma(A) \leq |\lambda_i(A)| \leq \bar{\sigma}(A) \quad (\text{A. 5.13})$$

由 SVD 的定义知, 下面的结论是显然的

$$\bar{\sigma}(A^*) = \bar{\sigma}(A), \quad \bar{\sigma}(A^T) = \bar{\sigma}(A) \quad (\text{A. 5.14})$$

不加证明地给出另一个重要结论:

$$\bar{\sigma}(AB) \leq \bar{\sigma}(A)\bar{\sigma}(B) \quad (\text{A. 5.15})$$

## A.6 矩阵的伪逆(广义逆)

考虑一个线性方程组

$$y = Ax \quad (\text{A. 6.1})$$

其中  $y \in \mathbb{C}^m$ ,  $A \in \mathbb{C}^{m \times n}$ 。式 (A. 6.1) 的一个最小二乘解, 就是求得一个  $x \in \mathbb{C}^n$ , 使得  $\|y - Ax\|_2$  最小化所有向量中的一个最小化向量。按照矩阵  $A$  的伪逆 (Moore-Penrose 广义逆), 这个解给定为

$$x = A^+ y \quad (\text{A. 6.2})$$

这个伪逆可以由  $A = U\Sigma V^*$  的 SVD 得到

$$A^+ = V_r \Sigma_r^{-1} U_r^* = \sum_{i=1}^r \frac{1}{\sigma_i(A)} v_i u_i^* \quad (\text{A. 6.3})$$

其中  $r$  是  $A$  的非零奇异值的个数, 则有

$$\bar{\sigma}(A) = \frac{1}{\sigma(A^+)} \quad (\text{A. 6.4})$$

注意, 任意矩阵  $A$ , 即使是奇异方阵或非方阵, 其伪逆  $A^+$  都存在; 而且伪逆满足

$$AA^+A = A \text{ 且 } A^+AA^+ = A^+ \quad (\text{A. 6.5})$$

注意以下情况 ( $A$  的秩为  $r$ ):

(1)  $r = n = m$ , 即  $A$  是非奇异阵。这种情况下  $A^+ = A^{-1}$  是逆阵;

(2)  $r = n \leq m$ , 即  $A$  是列满秩阵。这是“常见的最小二乘问题”, 希望对  $\|y - Ax\|_2$  最小化, 其解是

$$A^+ = (A^*A)^{-1}A^* \quad (\text{A. 6.6})$$

在这种情况下  $A^+A = I$ , 因此  $A^+$  是  $A$  的左逆矩阵。

(3)  $r = m \leq n$ , 即  $A$  是行满秩阵。在这种情况下, 我们对式 (A. 6.1) 有无数个解, 求取使  $\|x\|_2$  最小化的一个解, 可得到

$$A^+ = A^* (AA^*)^{-1} \quad (\text{A. 6.7})$$

在这种情况下  $AA^+ = I$ , 因此  $A^+$  是  $A$  的右逆矩阵。

(4)  $r < k = \min\{m, n\}$  (一般情况)。在这种情况下,  $A^*A$  和  $AA^*$  都是亏秩的; 必须利用式(A. 6. 3)得到伪逆矩阵。在这种情况下,  $A$  既没有左逆, 也没有右逆矩阵。

## A. 7 向量和矩阵范数

### A. 7.1 范数的一般概念

对一个向量、矩阵、信号或系统, 如果仅用一个数值来总体度量其大小将是十分有用的。为达此目的, 采用一种函数, 谓之范数。最常用的范数就是 Euclid 向量范数。

**定义 A. 7. 1**  $e$  (可以是向量、矩阵、信号或系统) 的范数定义为一个实数, 记作  $\|e\|$ , 满足如下列性质:

- (1) 非负性:  $\|e\| \geq 0$ ;
- (2) 绝对性:  $\|e\| = 0 \Leftrightarrow e = 0$  (对半范数有  $\|e\| = 0 \Leftarrow e = 0$ );
- (3) 齐次性:  $\|\alpha e\| = |\alpha| \cdot \|e\|$ , 对所有的复标量  $\alpha$  均成立;
- (4) 三角不等式

$$\|e_1 + e_2\| \leq \|e_1\| + \|e_2\| \quad (\text{A. 7. 1})$$

更精确地说,  $e$  是定义在复数域  $\mathbb{C}$  上向量空间 (包括有限维空间和无限维空间)  $V$  中的一个元素, 上面所列性质对于  $\forall e, e_1, e_2 \in V, \forall \alpha \in \mathbb{C}$  均成立。

考虑四个不同对象的范数 (在四个不同向量空间的范数):

- (1)  $e$  是一个常向量;
- (2)  $e$  是一个常矩阵;
- (3)  $e$  是一个时变信号  $e(t)$ , 对每个固定时刻  $t$  都是一个常标量或常向量;
- (4)  $e$  是一个“系统”, 一个传递函数  $G(s)$  或冲激响应函数  $g(t)$ , 对每个固定的  $s$  或  $t$  是一个常标量或矩阵。

情况 1 和情况 2 涉及空间范数, 现在问题是如何对这些通道进行平均或求和; 情况 3 和情况 4 涉及函数范数或称时间范数, 是希望对时间或频率的函数进行平均或求和。注意, 前两个是有限维范数, 而后两个是无限维范数。

**注** 对于范数的记法。读者应该注意到, 在各种文献中关于范数的记法是不一致的, 所以应该仔细辨认以免混淆。首先, 尽管空间范数和时间范数有本质的区别, 但都采用同样的符号, 经常都使用  $\|\cdot\|$  这种记法。其次, 相同的记法经常用来表述完全不同的范数; 例如无限范数  $\|e\|_\infty$ , 如果  $e$  是一个常向量, 那么  $\|e\|_\infty$  是向量中的最大元; 如果  $e(t)$  是一个时域标量信号, 那么  $\|e\|_\infty$  就是时变函数  $|e(t)|$  的峰值; 如果  $E$  是一个常阵, 那么  $\|E\|_\infty$  是矩阵的最大元; 其他著作可能用  $\|E\|_\infty$  来表示最大的矩阵行之和。最后, 如果  $E(s)$  是一个稳定的真系统 (传递函数), 那么  $\|E\|_\infty$  就是  $\mathcal{H}_\infty$  范数, 即  $E$  是最大奇异值的峰值。

### A. 7.2 向量范数

设向量  $a = [a_1, a_2, \dots, a_m]^T \in V = \mathbb{C}^m$ , 对于  $p \geq 1$ , 定义其  $p$ -范数为

$$\|a\|_p \stackrel{\text{def}}{=} \left( \sum_{i=1}^m |a_i|^p \right)^{1/p} \quad (\text{A. 7. 2})$$



其中  $|a_i|$  表示复标量  $a_i$  的绝对值。

向量 1-范数  $\|a\|_1 = \sum_{i=1}^m |a_i|$  就是对向量中各元幅值求和的范数。

向量 2-范数  $\|a\|_2 = \left( \sum_{i=1}^m |a_i|^2 \right)^{1/2}$  就是 Euclid 范数, 相应于两点之间的最短距离, 具有如下性质

$$a^* a = \|a\|_2^2 \quad (\text{A. 7. 3})$$

其中  $a^* = \bar{a}^T$  表示向量  $a$  的复共轭转置。

向量  $a = [a_1, a_2, \dots, a_m]^T \in V = \mathbb{C}^m$ , 对于  $p = \infty$ , 其向量  $\infty$ -范数定义为

$$\|a\|_\infty = \|a\|_{\max} \stackrel{\text{def}}{=} \max_{1 \leq i \leq m} |a_i| \quad (\text{A. 7. 4})$$

就是最大范数, 是各元幅值的最大值。

不同向量范数仅乘以定常因子就可等价。例如, 对于  $a \in \mathbb{C}^m$ , 则

$$\|a\|_{\max} \leq \|a\|_2 \leq \sqrt{m} \|a\|_{\max} \quad (\text{A. 7. 5})$$

$$\|a\|_2 \leq \|a\|_1 \leq \sqrt{m} \|a\|_2 \quad (\text{A. 7. 6})$$

在图 A. 7. 1 中, 向量范数之间的区别用图形来说明。在  $m=2$  的情况下, 对  $\|a\|_p=1$  描绘了其轮廓线。

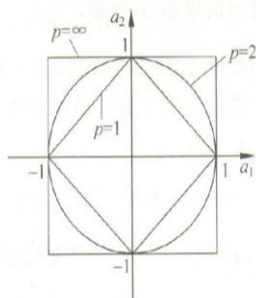


图 A. 7. 1 向量  $p$ -范数的轮廓:  $\|a\|_p=1$ , 对  $p=1, 2, \infty$

(数学上, 单位圆在  $\mathbb{R}^2$  上有三个不同的范数,)

### A. 7. 3 矩阵范数

定义 A. 7. 2 设  $A \in \mathbb{C}^{m \times n}$  为常阵, 其范数  $\|A\|$  称为矩阵范数, 除了满足定义 A. 7. 1 中的四个性质外, 还要满足乘法性质 (也称为一致性条件)

$$\|AB\| \leq \|A\| \cdot \|B\| \quad (\text{A. 7. 7})$$

如果存在矩阵范数满足 A. 7. 1 中的四个性质, 但不满足式 (A. 7. 7), 称其为广义矩阵范数。

首先检查作为向量  $p$  范数直接推广的三种范数:

(1) 求和矩阵范数。这是元素幅值之和

$$\|A\|_1 = \|A\|_{\text{sum}} \stackrel{\text{def}}{=} \sum_{i=1}^m \sum_{j=1}^n |a_{ij}| \quad (\text{A. 7. 8})$$

(2) **Frobenius 矩阵范数(或 Euclid 范数)**。这是元素幅值平方和的平方根

$$\|A\|_2 = \|A\|_F \stackrel{\text{def}}{=} \left( \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} = \sqrt{\text{tr}(A^* A)} \quad (\text{A.7.9})$$

(3) **最大元范数**。这是各元幅值的最大值,即

$$\|A\|_\infty = \|A\|_{\max} \stackrel{\text{def}}{=} \max_{i,j} |a_{ij}| \quad (\text{A.7.10})$$

由于式(A.7.10)不满足式(A.7.7),所以不是矩阵范数。但  $\sqrt{mn} \|A\|_\infty$  是一个矩阵范数。

上述三种范数有时也分别称为 1-范数、2-范数和  $\infty$ -范数,但不要和向量  $p$ -范数混淆。

(4) **诱导矩阵范数**。诱导矩阵范数在系统理论中是十分重要的,因为与系统中信号放大有很密切的联系。考虑方程

$$z = Aw \quad (\text{A.7.11})$$

可以认为  $w$  是系统的输入向量,而  $z$  是其输出向量,考虑按比率  $\|z\|/\|w\|$  定义矩阵  $A$  的“放大倍数”或“增益”。为此,对所有可能输入方向的最大化增益感兴趣,这将由下式定义的诱导范数给出

$$\|A\|_p \stackrel{\text{def}}{=} \max_{w \neq 0} \frac{\|Aw\|_p}{\|w\|_p} \quad (\text{A.7.12})$$

其中  $\|w\|_p = \left( \sum_i |w_i|^p \right)^{1/p}$  表示向量的  $p$ -范数。

换句话说,就是寻找向量  $w$  的方向,使得比率  $\|z\|_p/\|w\|_p$  最大化。因此,诱导范数给矩阵最大可能的“放大功率”。也经常采用下面等价的定义

$$\|A\|_p = \max_{\|w\|_p \leq 1} \|Aw\|_p = \max_{\|w\|_p = 1} \|Aw\|_p \quad (\text{A.7.13})$$

对于诱导的 1-范数、2-范数和  $\infty$ -范数,下面的恒等式成立

$$\|A\|_1 = \max_j \left( \sum_i |a_{ij}| \right) \quad (\text{最大行之和}) \quad (\text{A.7.14})$$

$$\|A\|_{\infty} = \max_i \left( \sum_j |a_{ij}| \right) \quad (\text{最大列之和}) \quad (\text{A.7.15})$$

$$\|A\|_2 = \sigma(A) = \sqrt{\rho(A^* A)} \quad (\text{奇异值或谱范数}) \quad (\text{A.7.16})$$

其中谱半径为  $\rho(A) = \max_i |\lambda_i(A)|$  就是矩阵  $A$  的最大特征值。注意,矩阵的诱导 2-范数等于其(最大)奇异值,经常称为谱范数。

**定理 A.7.1** 所有诱导范数  $\|A\|_p$  都是矩阵范数,因而满足乘法性质

$$\|AB\|_p \leq \|A\|_p \cdot \|B\|_p \quad (\text{A.7.17})$$

**证明** 对于一组串接的系统,复合系统可以表示为

$$z = Av, \quad v = Bw \Rightarrow z = ABw \quad (\text{A.7.18})$$

由诱导范数的定义知,首先引入  $v = Bw$ , 然后用  $\|v\|_p \neq 0$  乘以分子和分母,最后对包括  $w$  和  $v$  的项分别求最大化,则有

$$\|AB\|_p \stackrel{\text{def}}{=} \max_{w \neq 0} \frac{\|ABw\|_p}{\|w\|_p} = \max_{w \neq 0} \frac{\|Av\|_p}{\|w\|_p} \cdot \frac{\|Bw\|_p}{\|v\|_p} \leq \max_{v \neq 0} \frac{\|Av\|_p}{\|v\|_p} \cdot \max_{w \neq 0} \frac{\|Bw\|_p}{\|w\|_p}$$

由诱导范数的定义得式(A.7.17)。

下面讨论乘法性质的含义。就矩阵范数而言,对任意维数的矩阵  $A$  和  $B$ ,只要乘积  $AB$  存在,乘法性质  $\|AB\| \leq \|A\| \cdot \|B\|$  总成立。尤其当选择  $A$  和  $B$  作为向量的时候也成立。从这种观测中可以得到:

(1) 选择  $B$  作为一个向量,即  $B=w$ ,那么对于任意矩阵范数,由式(A.7.17)得出

$$\|Aw\| \leq \|A\| \cdot \|w\| \quad (\text{A.7.19})$$

就说“矩阵范数  $\|A\|$  是与其相应的向量范数  $\|w\|$  是相容的”。很显然,由式(A.7.17)得出,任何诱导的矩阵  $p$ -范数,是与其相应的向量  $p$ -范数是相容的。同样地,Frobenius 范数和 2-范数是相容的(由于当  $w$  是一个向量,  $\|w\|_F = \|w\|_2$ )。

(2) 由式(A.7.19)还可以得出,对于任意矩阵范数有

$$\|A\| \geq \max_{w \neq 0} \frac{\|Aw\|}{\|w\|} \quad (\text{A.7.20})$$

注意,诱导范数的定义使得式(A.7.20)以等式成立。而性质  $\|A\|_F \geq \sigma(A)$  成立是因为  $\|w\|_F = \|w\|_2$ 。

(3) 同时选择  $A=z^*$  和  $B=w$  作为向量,然后利用 Frobenius 范数或式(A.7.12)中的诱导 2-范数(奇异值),得到 Cauchy-Schwarz 不等式

$$|z^* w| \leq \|z\|_2 \cdot \|w\|_2 \quad (\text{A.7.21})$$

其中  $z$  和  $w$  是同维的列向量,  $z^* w = \langle w, z \rangle$  是向量  $z$  和  $w$  的 Euclid 内积。

(4) 由内积还可以定义向量  $z$  和  $w$  之间的夹角  $\phi$

$$\phi = \arccos\left(\frac{|z^* w|}{\|z\|_2 \cdot \|w\|_2}\right) \quad (\text{A.7.22})$$

注意按此定义,  $\phi$  介于  $0^\circ \sim 90^\circ$ 。

#### A.7.4 谱半径

矩阵  $A$  的谱半径  $\rho(A)$  定义为其特征值幅值的最大值,即

$$\rho(A) = \max_i |\lambda_i(A)| \quad (\text{A.7.23})$$

这不是一个范数,因为它不满足范数的性质。

尽管谱半径不是一个范数,但却提供了任何矩阵范数的下界。

**定理 A.7.2** 对于任意的矩阵范数(尤其对于任意的诱导范数),下式成立

$$\rho(A) \leq \|A\| \quad (\text{A.7.24})$$

**证明** 由于  $\lambda_i(A)$  是  $A$  的特征值,则有  $At_i = \lambda_i t_i$ ,  $t_i$  表示特征向量。于是得到

$$|\lambda_i| \cdot \|t_i\| = \|\lambda_i t_i\| = \|At_i\| \leq \|A\| \cdot \|t_i\| \quad (\text{A.7.25})$$

(最后的不等式是根据式(A.7.17))。因此对于任意矩阵范数有  $|\lambda_i(A)| \leq \|A\|$ ,并且这对所有的特征值成立,结果得证。■

式(A.7.24)的一个简单物理解释是,特征值仅按某个方向(由特征向量给出)对矩阵增益进行度量,因而必然小于允许按任意方向对矩阵产生最大增益的度量,参看式(A.7.20)。

#### A.7.5 矩阵范数的一些关系式

矩阵  $A$  的各种范数是密切联系的,这一点可以从下面的不等式看出。假定  $A$  是一个  $m \times n$  的矩阵,那么



$$\bar{\sigma}(A) \leq \|A\|_F \leq \sqrt{\min(m, n)} \bar{\sigma}(A) \quad (\text{A. 7. 26})$$

$$\|A\|_{\max} \leq \bar{\sigma}(A) \leq \sqrt{mn} \|A\|_{\max} \quad (\text{A. 7. 27})$$

$$\bar{\sigma}(A) \leq \sqrt{\|A\|_1 \|A\|_{\infty}} \quad (\text{A. 7. 28})$$

$$\frac{1}{\sqrt{n}} \|A\|_{\infty} \leq \bar{\sigma}(A) \leq \sqrt{m} \|A\|_{\infty} \quad (\text{A. 7. 29})$$

$$\frac{1}{\sqrt{m}} \|A\|_1 \leq \bar{\sigma}(A) \leq \sqrt{n} \|A\|_1 \quad (\text{A. 7. 30})$$

$$\max\{\bar{\sigma}(A), \|A\|_F, \|A\|_1, \|A\|_{\infty}\} \leq \|A\|_{\text{sum}} \quad (\text{A. 7. 31})$$

所有的这些范数,除了  $\|A\|_{\max}$  之外都是矩阵范数,并且满足式(A. 7. 7)。

Frobenius 范数和最大奇异值(诱导 2-范数)的重要性质,就是关于酉变换保持不变的性质。也就是说,对于酉阵  $U_i$ ,满足  $U_i U_i^* = I$ ,则有

$$\|U_1 A U_2\|_F = \|A\|_F \quad (\text{A. 7. 32})$$

$$\bar{\sigma}(U_1 A U_2) = \bar{\sigma}(A) \quad (\text{A. 7. 33})$$

从矩阵  $A = U \Sigma V^*$  的 SVD 和式(A. 7. 32),得到 Frobenius 范数和奇异值  $\sigma_i(A)$  之间一个非常重要的关系式,即

$$\|A\|_F = \sqrt{\sum_i \sigma_i^2(A)} \quad (\text{A. 7. 34})$$

把 Perron-Frobenius 定理应用于方阵  $A$ ,则有

$$\max_D \|D A D^{-1}\|_1 = \max_D \|D A D^{-1}\|_{\infty} = \rho(|A|) \quad (\text{A. 7. 35})$$

$D$  是一个对角“尺度变换”矩阵,  $|A|$  表示矩阵  $A$  所有元被其幅值替换后的矩阵,而且  $\rho(|A|) = \max_i |\lambda_i(|A|)|$  是 Perron 根(Perron-Frobenius 特征值)。Perron 根大于或等于谱半径,即  $\rho(A) \leq \rho(|A|)$ 。

## A. 8 信号和系统范数

如前所述,涉及信号与系统的范数将是无限维范数。现在考虑信号与系统所处的时间域为  $T = T_c = \mathbb{R}$  或  $T = T_d = \mathbb{Z}$ ,前者称为连续时间域,后者称为离散时间域。时间域  $t$  可以转换为频率域  $s$ ,但本质仍然是无限维空间上的范数。

### A. 8. 1 信号范数

考虑随时间变化(或随频域变化)的信号  $e(t)$ ,  $t \in T_c$  的时间范数,与空间范数(向量和矩阵范数)相比较,发现选择时间范数会有很大区别。如图 A. 8. 1 中两个信号  $e_1(t)$  和  $e_2(t)$ ; 对于  $e_1(t)$ ,其  $\infty$ -范数(峰值)是 1,即  $\|e_1(\cdot)\|_{\infty} = 1$ ; 然而由于信号没有衰减,2-范数却是无限大,即  $\|e_1(\cdot)\|_2 = \infty$ 。对于  $e_2(t)$ ,反过来则是对的。

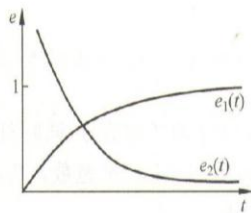
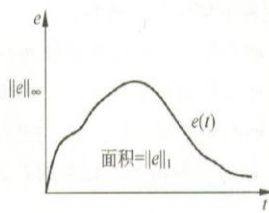
对于信号,可以用两步来计算范数:

(1) 在一个给定的时间或频率域上,采用向量范数(对于标量信号可简单地取绝对值)对所有通道“求和”;

(2) 采用时间范数对时间或频率“求和”。

回顾前面所述,向量范数在其仅有定常数差别的意义上是等价的。因此,无论在(1)步中采用什么范数都没有太大的影响,通常对向量和信号使用同样的  $p$ -范数。设有向量时间函数  $e(t)=[e_1(t), e_2(t), \dots, e_r(t)]^T, t \in T_c$ ; 或  $e(\cdot) \in L_p(\mathbb{R})$ , 其时间  $p$ -范数定义为

$$L_p \text{ 范数: } \|e(\cdot)\|_p = \left( \int_{-\infty}^{\infty} \sum_{i=1}^r |e_i(t)|^p dt \right)^{1/p} \quad (\text{A. 8. 1})$$

图 A. 8. 1 具有完全不同 2-范数和  $\infty$ -范数的信号图 A. 8. 2 信号的 1-范数和  $\infty$ -范数

下面的信号时间范数是常用的:

(1) 时间上的 1-范数(积分绝对误差,  $L_1$  范数,  $e(\cdot) \in L_1(\mathbb{R})$ )(1 维情况见图 A. 8. 2):

$$\|e(\cdot)\|_1 = \int_{-\infty}^{\infty} \sum_{i=1}^r |e_i(t)| dt \quad (\text{A. 8. 2})$$

(2) 时间上的 2-范数(二次范数, 积分平方误差, 信号的“能量”,  $L_2$  范数,  $e(\cdot) \in L_2(\mathbb{R})$ ):

$$\|e(\cdot)\|_2 = \left( \int_{-\infty}^{\infty} \sum_{i=1}^r |e_i(t)|^2 dt \right)^{1/2} \quad (\text{A. 8. 3})$$

(3) 时间上的  $\infty$ -范数(时域峰值,  $L_\infty$  范数,  $e(\cdot) \in L_\infty(\mathbb{R})$ )(1 维情况见图 A. 8. 2):

$$\|e(\cdot)\|_\infty = \sup_{t \in \mathbb{R}} \left[ \max_{i \in \{1, 2, \dots, r\}} |e_i(t)| \right] \quad (\text{A. 8. 4})$$

对于定义在离散时间点上的函数  $e(t)=[e_1(t), e_2(t), \dots, e_r(t)]^T, t \in T_d$ ; 或  $e(\cdot) \in l_p(\mathbb{Z})$ , 同样可以定义下面的信号时间范数:

$$l_p \text{ 范数: } \|e(\cdot)\|_p = \left( \sum_{t=-\infty}^{\infty} \sum_{i=1}^r |e_i(t)|^p \right)^{1/p} \quad (\text{A. 8. 5})$$

(1) 时间上的 1-范数(求和绝对误差,  $l_1$  范数,  $e(\cdot) \in l_1(\mathbb{Z})$ )

$$\|e(\cdot)\|_1 = \sum_{t=-\infty}^{\infty} \sum_{i=1}^r |e_i(t)| \quad (\text{A. 8. 6})$$

(2) 时间上的 2-范数(二次范数, 求和平方误差, 信号的“能量”,  $l_2$  范数,  $e(\cdot) \in l_2(\mathbb{Z})$ )

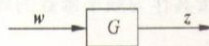
$$\|e(\cdot)\|_2 = \left( \sum_{t=-\infty}^{\infty} \sum_{i=1}^r |e_i(t)|^2 \right)^{1/2} \quad (\text{A. 8. 7})$$

(3) 时间上的  $\infty$ -范数(时域峰值,  $l_\infty$  范数,  $e(\cdot) \in l_\infty(\mathbb{Z})$ )

$$\|e(\cdot)\|_\infty = \sup_{t \in \mathbb{Z}} \left[ \max_{i \in \{1, 2, \dots, r\}} |e_i(t)| \right] \quad (\text{A. 8. 8})$$

## A. 8. 2 系统范数的概念

考虑图 A. 8. 3 中的系统, 具有稳定的  $m \times r$  维传递函数矩阵  $G(s)$  或冲激响应矩阵  $g(t)$ 。为了评估其性能需要回答一个问题: 给定关于容许输入信号  $w(t)$  的信息; 输出  $z(t)$  能有多大? 为了回



A. 8. 3 系统 G

答这个问题，必须计算有关的系统范数。

系统范数是系统级的范数，反映的是系统的传递特性。通常用如下 2-范数来评估输出信号

$$\|z(\cdot)\|_2 = \left( \int_{-\infty}^{\infty} \sum_{j=1}^m |z_j(t)|^2 dt \right)^{1/2} \quad (\text{A. 8. 9})$$

将考虑三种不同的输入选择：

(1)  $w(t)$  是一个单位冲激序列；

(2)  $w(t)$  是满足  $\|w(\cdot)\|_2 = 1$  的任意信号；

(3)  $w(t)$  是满足  $\|w(\cdot)\|_2 = 1$ ，但当  $t \geq 0$  时  $w(t) = 0$  的任意信号，同时只测量  $t \geq 0$  时的  $z(t)$ 。在这三种情况下相关的系统范数分别是  $\mathcal{H}_2$ 、 $\mathcal{H}_\infty$  和 Hankel 范数。

### A. 8. 3 $\mathcal{H}_2$ 范数

考虑一个严格真系统  $G(s)$ ，在空间上用(矩阵)Frobenius 范数并在频域上积分来定义  $\mathcal{H}_2$  范数，即

$$\|G(\cdot)\|_2 \stackrel{\text{def}}{=} \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \text{tr}(G(j\omega))^* G(j\omega) d\omega} \quad (\text{A. 8. 10})$$

$\|G(j\omega)\|_F^2 = \sum_{ij} |G_{ij}(j\omega)|^2$

由此看到  $G(s)$  必须是严格真的，否则  $\mathcal{H}_2$  范数为无穷大。对  $\mathcal{H}_2$  范数也可给出另外一种解释，按 Parseval 定理，式(A. 8. 10)等价于冲激响应的  $\mathcal{H}_2$  范数

$$\|G(\cdot)\|_2 = \|g(\cdot)\|_2 \stackrel{\text{def}}{=} \sqrt{\int_0^{\infty} \text{tr}(g^T(\tau)g(\tau)) d\tau} \quad (\text{A. 8. 11})$$

$\|g(\tau)\|_F^2 = \sum_{ij} |g_{ij}(\tau)|^2$

**注释 1** 需要注意的是，当  $\omega$  或  $\tau$  的值给定时， $G(j\omega)$  和  $g(\tau)$  都是常阵。

**注释 2** 可以改变式(A. 8. 11)中的积分与求和的次序得到

$$\|G(\cdot)\|_2 = \|g(\cdot)\|_2 = \sqrt{\sum_{ij} \int_0^{\infty} |g_{ij}(\tau)|^2 d\tau} \quad (\text{A. 8. 12})$$

此处  $g_{ij}(t)$  是冲激响应矩阵  $g(t)$  的第  $ij$  个元素。由此可见， $\mathcal{H}_2$  范数可解释为一个接一个地将单位冲激  $\delta_j(t)$  作用于每一个输入产生的输出(在将冲激加于下一个输入前，需先让输出稳定到零)。

总结起来对  $\mathcal{H}_2$  范数有如下的确定性性能解释

$$\|G(\cdot)\|_2 = \max_{w(t)=\text{单位冲激}} \|z(\cdot)\|_2 \quad (\text{A. 8. 13})$$

用最优控制中的二次判据(LQG)，对  $\mathcal{H}_2$  范数可给出一个随机理论的解释，这时所量测的是白噪声激励下输出的均方根期望值(rms)。

### A. 8. 4 $\mathcal{H}_\infty$ 范数

考虑一个真线性稳定系统  $G(s)$ ，在空间上(矩阵情况下)利用奇异值(诱导的 2-范数)，并按其作为频率函数的峰值来定义  $\mathcal{H}_\infty$  范数，即

$$\|G(\cdot)\|_\infty \stackrel{\text{def}}{=} \max_{\omega} \sigma(G(j\omega)) \quad (\text{A. 8. 14})$$

就性能而言，由式(A. 8. 14)可以看出，该范数是传递函数“幅值”的峰值。



然而,  $\mathcal{H}_\infty$  范数也有几种时域性能解释。首先, 它是对任意频率正弦输入最差情况下的增益。当  $t \rightarrow \infty$ , 设  $z(\omega)$  表示系统在正弦输入  $w(\omega)$  持续作用下的响应, 所以有  $z(\omega) = G(j\omega)w(\omega)$ 。对于给定频率  $\omega$ , 放大倍数(增益)  $\|z(\omega)\|_2 / \|w(\omega)\|_2$  依赖于  $w(\omega)$  的方向, 而在最不利方向上的增益由最大奇异值给出

$$\sigma(G(j\omega)) = \max_{w(\omega) \neq 0} \frac{\|z(\omega)\|_2}{\|w(\omega)\|_2}$$

增益也依赖于频率, 而在最不利的频率上的增益则由  $\mathcal{H}_\infty$  范数给出

$$\|G(\cdot)\|_\infty = \max_{\omega} \max_{w(\omega) \neq 0} \frac{\|z(\omega)\|_2}{\|w(\omega)\|_2} = \max_{\|w(\omega)\|_2=1} \|z(\omega)\|_2 \quad (\text{A. 8. 15})$$

其次, 由后面的表 A. 8. 2 可以看出,  $\mathcal{H}_\infty$  范数等于任何时域信号的诱导(最坏情况)2-范数:

$$\|G(\cdot)\|_\infty = \max_{w(t) \neq 0} \frac{\|z(\cdot)\|_2}{\|w(\cdot)\|_2} = \max_{\|w(\cdot)\|_2=1} \|z(\cdot)\|_2 \quad (\text{A. 8. 16})$$

式(A. 8. 16)之所以成立, 是因为最坏的输入信号  $w(t)$  是频率为  $\omega^*$  的正弦信号, 而方向是由最大增益  $\sigma(G(j\omega^*))$  给出。

再次,  $\mathcal{H}_\infty$  范数等于诱导的功率(rms)范数; 最后, 这也可以解释为用随机信号期望值表述的一个诱导范数。所有这些解释都使  $\mathcal{H}_\infty$  范数在工程应用中极为有用。

### A. 8. 5 $\mathcal{H}_2$ 与 $\mathcal{H}_\infty$ 范数之间的差别

为了理解在  $\mathcal{H}_2$  与  $\mathcal{H}_\infty$  范数之间的差别, 由式(A. 8. 16)注意到, 可以把 Frobenius 范数写成用奇异值表述的形式, 此时有

$$\|G(\cdot)\|_2 = \sqrt{\frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_i \sigma_i^2(G(j\omega)) d\omega} \quad (\text{A. 8. 17})$$

由此看到, 最小化  $\mathcal{H}_\infty$  范数, 相当于最小化最大奇异值的峰值(“最坏方向、最坏频率”), 而最小化  $\mathcal{H}_2$  范数则相当于最小化所有频率上的所有奇异值的平方和(“平均方向、平均频率”)。总结起来有:

- (1)  $\mathcal{H}_\infty$  范数: “压低最大奇异值的峰值”;
- (2)  $\mathcal{H}_2$  范数: “压低全部”(所有频率上的所有奇异值)。

为什么  $\mathcal{H}_\infty$  范数应用如此之广? 在鲁棒控制中采用  $\mathcal{H}_\infty$  范数, 主要是因为它便于表示非结构化模型不确定性, 也因为其满足相乘特性

$$\|A(\cdot)B(\cdot)\|_\infty \leq \|A(\cdot)\|_\infty \cdot \|B(\cdot)\|_\infty \quad (\text{A. 8. 18})$$

这些结论可由式(A. 8. 16)直接得出, 它表明  $\mathcal{H}_\infty$  范数是一种诱导范数。

$\mathcal{H}_2$  范数有何不妥之处?  $\mathcal{H}_2$  范数有许多良好的数学与数值计算方面的特性, 并且其最小化有重要的工程意义。然而,  $\mathcal{H}_2$  范数不是一种诱导范数, 并且不满足相乘特性。这意味着无法通过计算单个组件的  $\mathcal{H}_2$  范数, 来判断其串联(级联)后会如何表现。

### A. 8. 6 Hankel 范数

对于一个稳定的系统  $G(s)$ , 当施加一个延续到  $t=0$  的输入  $w(t)$ , 然后量测  $t>0$  时的输出  $z(t)$ , 并选择  $w(t)$  最大化两个信号 2-范数的比值, 即定义系统的 Hankel 范数

$$\|G(\cdot)\|_H \stackrel{\text{def}}{=} \max_{w(t)} \frac{\sqrt{\int_0^\infty \|z(\tau)\|_2^2 d\tau}}{\sqrt{\int_{-\infty}^0 \|w(\tau)\|_2^2 d\tau}} \quad (\text{A. 8. 19})$$

这个 Hankel 范数是一种从过去输入到未来输出的诱导范数。可以证明, Hankel 范数等于

$$\|G(\cdot)\|_H = \sqrt{\rho(PQ)} \quad (\text{A. 8. 20})$$

其中  $\rho$  是谱半径(最大特征值的绝对值);  $P$  是系统  $(A, B, C)$  能控性 Gram 阵  $P \stackrel{\text{def}}{=} \int_0^\infty e^{A\tau} B B^T e^{A^T \tau} d\tau$ ,  $Q$  是系统  $(A, B, C)$  能观性 Gram 阵  $Q \stackrel{\text{def}}{=} \int_0^\infty e^{A^T \tau} C^T C e^{A\tau} d\tau$ 。之所以称其为 Hankel 范数, 是因为矩阵  $PQ$  具有 Hankel 矩阵的特殊结构(沿“错位”对角线有相同元素)。对应的 Hankel 奇异值是  $PQ$  特征值的正平方根, 即

$$\sigma_i = \sqrt{\lambda_i(PQ)} \quad (\text{A. 8. 21})$$

Hankel 范数与  $\mathcal{H}_\infty$  范数有密切关系, 从而有

$$\|G(\cdot)\|_H \equiv \sigma_1 \leq \|G(\cdot)\|_\infty \leq 2 \sum_{i=1}^n \sigma_i \quad (\text{A. 8. 22})$$

这样, Hankel 范数总是小于(或等于)  $\mathcal{H}_\infty$  范数。

### A. 8. 7 各种系统范数的信号解释

考虑一个系统  $G$ , 输入为  $d$  而输出为  $e$

$$e = Gd \quad (\text{A. 8. 23})$$

为了得到好的性能, 可能对于任何容许的输入信号  $d$ , 要求输出信号  $e$  尽可能“小”, 因此需要规定:

- (1) 什么样的  $d$  是容许的( $d$  属于哪个集合)?
  - (2) 所描述的“小”究竟是什么意思(应该对  $e$  使用什么范数)?
- 一些可能的输入信号集合是:

(1)  $d(t)$  由冲激函数  $\delta(t)$  组成, 在状态产生阶跃变化时, 这就引入了 LQ 目标函数, 并建立了  $\mathcal{H}_2$  范数;

- (2)  $d(t)$  是一个零均值的白噪声过程;
- (3) 按固定的频率  $d(t) = \sin(\omega t)$ , 适用  $t = -\infty$  (相对应于稳态的正弦响应);
- (4)  $d(t)$  是由所有容许频率正弦函数构成的集合;
- (5)  $d(t)$  是能量有界的, 即  $\|d(\cdot)\|_2 \leq 1$ ;
- (6)  $d(t)$  是幅值有界的, 即  $\|d(\cdot)\|_\infty \leq 1$ 。

前 3 个输入信号集是对规定的信号, 而后 3 个是有界范数的输入信号类。亟须解决的物理问题, 就是确定哪一个输入类是合理的。

对输出信号的度量可以考虑如下范数:

- (1) 1-范数  $\|e(\cdot)\|_1$ ;
- (2) 2-范数(能量)  $\|e(\cdot)\|_2$ ;
- (3)  $\infty$ -范数(幅值的峰值)  $\|e(\cdot)\|_\infty$ 。

采用其他范数也是可能的, 但是必须再次强调, 是由工程问题决定哪一个范数最合适。现在, 将分别考虑按输入类的定义和输出范数来确定究竟选择哪一个系统范数。也就是说,

需要用合适的系统增益来测试系统的性能。SISO 系统的结果如表 A. 8. 1 所示, 其中  $d(t)$  和  $e(t)$  是标量信号。在这些表示中,  $G(s)$  是传递函数,  $g(t)$  则为相应的冲激响应。注意, 尤其是

$$\mathcal{H}_\infty \text{ 范数: } \|G(\cdot)\|_\infty \stackrel{\text{def}}{=} \max_{d(t)} \bar{\sigma}(G(j\omega)) = \max_{d(t)} \frac{\|e(\cdot)\|_2}{\|d(\cdot)\|_2} \quad (\text{A. 8. 24})$$

以及

$$L_1 \text{ 范数: } \|g(\cdot)\|_1 \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} g(t) dt = \max_{d(t)} \frac{\|e(\cdot)\|_\infty}{\|d(\cdot)\|_\infty} \quad (\text{A. 8. 25})$$

(其中右边的两个等式不是根据定义, 而是泛函分析的重要结果)。我们从表 A. 8. 1 和表 A. 8. 2 可以看出,  $\mathcal{H}_2$  范数和  $\mathcal{H}_\infty$  范数出现在很多位置上, 这对其在控制中的广泛应用奠定了基础。除此以外, 如果认为  $d(t)$  是具有任意频率正弦函数的集合, 并且使用 2-范数来度量输出, 那么结果是用  $\mathcal{H}_\infty$  范数。同样地, 如果输入是白噪声并且用 2-范数来度量输出, 则结果是用  $\mathcal{H}_2$  范数。

表 A. 8. 1 对于 2 个规定的输入信号和 2 个不同的输出信号范数产生的系统范数

	$d(t) = \delta(t)$	$d(t) = \sin(\omega t)$
$\ e(\cdot)\ _2$	$\ G(\cdot)\ _2$	$\infty$ (常用)
$\ e(\cdot)\ _\infty$	$\ g(\cdot)\ _\infty$	$\bar{\sigma}(G(j\omega))$

表 A. 8. 2 对于 2 组不同的范数有界输入信号和 2 个不同输出范数的系统范数  
(对角线上的元都是诱导范数)

	$\ d(\cdot)\ _2$	$\ d(\cdot)\ _\infty$
$\ e(\cdot)\ _2$	$\ G(\cdot)\ _\infty$	$\infty$
$\ e(\cdot)\ _\infty$	$\ G(\cdot)\ _2$	$\ g(\cdot)\ _1$

表 A. 8. 1 和表 A. 8. 2 中的结果通过采用合适的矩阵和向量范数, 可以推广到 MIMO 系统。特别地, 如果对于  $\mathcal{H}_\infty$  范数采用  $\|G(\cdot)\|_\infty = \max_{\omega} \bar{\sigma}(G(j\omega))$ , 对于  $L_1$  范数采用  $\|g(\cdot)\|_1 = \max_i \|g_i(\cdot)\|_1$ , 其中  $g_i(t)$  表示冲激响应矩阵的第  $i$  行, 则表 A. 8. 2 对角线上的诱导范数可以推广到 MIMO 情况。事实上,  $\mathcal{H}_\infty$  范数和  $L_1$  范数都是诱导范数, 使其适合在鲁棒性分析中应用, 比如说在小增益定理中就用到。由下面一个真标量系统的定界可以看出, 这两个范数是紧密联系的

$$\|G(\cdot)\|_\infty \leq \|g(\cdot)\|_1 \leq (2n+1) \cdot \|G(\cdot)\|_\infty \quad (\text{A. 8. 26})$$

这里  $n$  是最小实现的状态维数。对于一个多变量  $m \times n$  系统, 有如下的推广

$$\|G(\cdot)\|_\infty \leq \sqrt{m} \cdot \|g(\cdot)\|_1 \leq \sqrt{mn} \cdot (2n+1) \cdot \|G(\cdot)\|_\infty \quad (\text{A. 8. 27})$$