第五章　矩阵函数及其微积分

引　言　怎样讨论矩阵的微积分?

在图像处理, 模式识别或移动通信等领域, 常需要利用特定的线性变换将高维向量压缩成低维向量或者将低维向量还原为高维向量, 并且使误差尽可能小. 描述此类问题的一个数学模型如下例:

例 5.0.1　设 $\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}^m$, 求半正交矩阵 $U \in \mathbb{R}^{m\times n}$, 使得

$$\|U\alpha - \beta\| \tag{5.0.1}$$

最小.

例 5.0.1 相当于求以矩阵 U 为自变量的函数 $J(U) = \|U\alpha - \beta\|$ 在约束条件 $U^TU = I$ 或 $UU^T = I$ 下的最小值点 (矩阵), 这样的优化问题具有普遍意义 (属于运筹学的研究领域). 比较一元或多元微分学可知, 解决此问题的一个可行办法是求函数 $J(U) = \|U\alpha - \beta\|$ 关于未知矩阵 U 的导数, 这就需要研究矩阵函数的微积分 (如果将 U 的元素都作为未知数列出, 则目标函数 $\|U\alpha - \beta\|$ 就是一个 mn 元函数, 可以使用多元微分学来研究此问题, 但那将是什么样的场景!). 我们在大学的许多课程和实践中对微积分的强大作用已经深有体会, 如果矩阵能与微积分相结合, 无疑将会产生更为巨大的作用. 那么如何才能将微积分引入到矩阵的研究中来呢? 比较数学分析或高等数学课程, 我们首先需要研究矩阵序列的收敛性, 这就需要计算两个矩阵之间的距离. 一旦有了距离概念, 就能够和数学分析或高等数学几乎完全平行地讨论矩阵序列的极限和矩阵函数的连续性, 进而讨论矩阵函数的微分学与积分学等理论. 我们在第一章内积空间中已经看到, 距离概念可以由长度或范数导出, 而长度或范数可以由内积导出, 因此研究矩阵函数的微积分实际上可以通过在矩阵空间中引入适当的内积后顺利进行. 但一般的无限维线性空间可能没有内积概念 (所有 $m \times n$ 阶函数矩阵构成的线性空间当然是无限维的), 因此我们将在本章第一节研究比内积导出的范数更为广泛的概念, 以使范数能够应用在更大的范围. 那么, 什么是矩阵的范数呢? 我们先看下面简单的例子:

例 5.0.2　设 x 是复数, 则当 $|x| < 1$ 时有

$$(1-x)^{-1} = 1 + x + x^2 + \cdots + x^m + \cdots \tag{5.0.2}$$

问题: 何时上式对于矩阵也成立, 即设 A 是 n 阶矩阵, 公式

$$(I-A)^{-1} = I + A + A^2 + \cdots + A^m + \cdots \tag{5.0.3}$$

何时成立? 即相当于 $|x| < 1$ 的条件是什么? 容易知道一个充分条件是 $\rho(A) < 1$. 但是矩阵的特征值及其谱半径都是不容易计算的, 我们能否改进这个条件? 答案是肯定的, 只需将 $\rho(A) < 1$ 换成 $\|A\| < 1$, 即 A 的范数小于 1, 任何一种范数即可! 因此, 矩阵的范数可以看作是实数的绝对值或者复数的模的推广, 是一种衡量矩阵 (包括向量) 大小的尺度.

另外, 我们对矩阵函数的微积分也不陌生, 比如三元函数 $f(x, y, z)$ 的梯度向量

$$\nabla f(x) = (\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}, \frac{\partial f}{\partial z}) \tag{5.0.4}$$

就是向量 (x,y,z) 的函数 $f(x,y,z)$ 关于向量 (x,y,z) 的导数 (向量).

第一节　向量与矩阵的范数

回忆内积空间中由内积导出的范数具有的特征, 我们在一般的实或复线性空间中引进下面的定义.

定义 5.1.1　设 $\mathbb{F}=\mathbb{R}$ 或 $\mathbb{C}$, V 为 $\mathbb{F}$ 上的一个线性空间. 如果 V 上的实 (向量) 函数 $\|\cdot\|$ 满足下列性质:

(1) 正定性: 对 $\forall x\in V$, $\|x\|\geq 0$ 且 $\|x\|=0 \iff x=0$;

(2) 齐次性: 对 $\forall k\in\mathbb{F}$, $x\in V$, 有

$$\|kx\|=|k|\,\|x\|;$$

(3) 三角不等式: 对 $\forall x,y\in V$, 有

$$\|x+y\|\leq\|x\|+\|y\|,$$

则称 V 是一个**赋范线性空间**或赋范空间, 记为 $(V,\|\cdot\|)$(常简记为 V, 如果范数是不需要强调的), 称 $\|x\|$ 是 V 中向量 x 的范数.

由齐次性立即可知: $\|0\|=0$, $\|-x\|=\|x\|$.

例 5.1.1　对实数域 $\mathbb{R}$ 而言, 普通的绝对值显然是 V 上的范数, 绝对值的 $a\,(a>0)$ 倍也是范数. 因此 $\mathbb{R}$ 上有无穷多范数. 请思考: $\mathbb{R}$ 上还有别的范数吗? 对复数域 $\mathbb{C}$ 考虑同样的问题.

例 5.1.2　设 $V=\mathbb{C}^n$ 或 $\mathbb{R}^n$, 则下列实值函数都是 $\mathbb{C}^n$(或 $\mathbb{R}^n$) 上的向量范数, 因而 V 为赋范空间:

(1)$\|x\|_\infty=\max\limits_{1\leq j\leq n}|x_j|$　　(**最大范数或 l_∞ 范数**或∞**- 范数**);

(2)$\|x\|_1=\sum\limits_{j=1}^{n}|x_j|$　　(**和范数**或l_1 **范数**或1**- 范数**);

(3)$\|x\|_2=\left(\sum\limits_{j=1}^{n}|x_j|^2\right)^{1/2}$　　(**欧几里得范数**或l_2 **范数**);

(4)$\|x\|_p=\left(\sum\limits_{j=1}^{n}|x_j|^p\right)^{1/p}$, $p\geq 1$　　(**Hölder 范数** 或l_p **范数**或p**- 范数**).

显然, 当 $n=1$ 时, 例5.1.2 中的所有范数都变成 $\mathbb{C}$ 或 $\mathbb{R}$ 上的普通范数 (模或绝对值).

容易看出, $\|\cdot\|_1$ 与 $\|\cdot\|_2$ 为Hölder 范数中取 $p=1$ 与 $p=2$ 的情形, 而 $\|\cdot\|_\infty$ 是Hölder 范数当 $p\to\infty$ 的极限情形. 直接验证可知 (见习题3), $\|\cdot\|_\infty$ 满足 定义5.1.1 的3个条件, 因而为 V 上的向量范数. l_p 范数 $(p\geq 1)$ 显然满足定义中的条件 (1) 和 (2). 为验证条件 (3), 只需应用下列**Minkowski 不等式**(证明见习题4):

$$\left(\sum_{j=1}^{n}|x_j+y_j|^p\right)^{1/p}\leq\left(\sum_{j=1}^{n}|x_j|^p\right)^{1/p}+\left(\sum_{j=1}^{n}|y_j|^p\right)^{1/p},\quad p\geq 1. \tag{5.1.1}$$

注 1. 定义 5.1.1 中的字母“l”是序列空间 (对照第二章第六节) 或 Lebesgue[49] 空间的统称. 确切地说, 数域 $\mathbb{F}$ 上的所有绝对收敛的无穷数列构成的线性空间称为 l^1 空间, 绝对平方收敛的无穷数列构成的线性空间称为 l^2 空间, 所有有界无穷数列构成的线性空间称为 l^∞ 空间, 以及 l^p 空间等等. 类似地, 可以定义绝对可积函数空间 L^1, 绝对平方可积函数空间 L^2, 以及 L^p, L^∞ 等等.

注 2. 当 $0 < p < 1$ 时, l^p 范数仍然满足向量范数的前两个条件, 但不满足三角不等式, 见习题 5.

注 3. 对于 $\mathbb{C}$ 或 $\mathbb{R}$ 上一般 n 维线性空间 V, 可以通过取 V 的一组基, 然后像 例 5.1.2 中一样定义 V 的范数.

注 4. 常将 1- 范数称为 Manhattan (曼哈顿)- 度量, 因为在赋范线性空间中可以由范数自然定义距离, 即

$$d(x, y) = \|x - y\|.$$

请读者在平面上或者空间中画出两点间的距离的示意图. 如果连接两点间的最短曲线称为线段, 请问 1- 范数下的线段是什么? ∞- 范数下的线段是什么?

例 5.1.3 (各种范数下的单位圆) 下面的图从左至右依次展示了 1- 范数, 普通范数 (欧几里得范数) 和 ∞- 范数下的平面上的单位圆 (1 维单位球面):

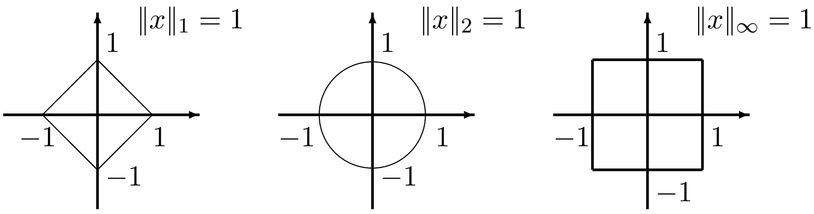

图 5.1.1

读者从此例可以看出, 在 1- 范数和 ∞- 范数下的单位圆按通常意义都是正方形! 一般地, 将赋范线性空间 V 中范数为 1 的向量的集合称为单位球面, 范数小于等于 1 的向量的集合称为单位球. 如果单位球面是多面体, 则称该范数是多面的. 因此 1- 范数和 ∞- 范数是多面的.(还有别的多面范数吗?)

请读者思考, 在 1- 范数和 ∞- 范数下的单位圆中的四个角是直角吗? 此时的角度与我们的常识一致吗?

矩阵是特殊的向量, 因此也可以定义矩阵的向量范数.

例 5.1.4 设 $\mathbb{F}^{n\times n}$ 为数域 $\mathbb{F}$ 上 n 阶矩阵所构成的线性空间, 对 $\forall A = (a_{ij})_{n\times n}$, 定义

$$\|A\|_F = \left(\sum_{i,j=1}^{n} |a_{ij}|^2\right)^{1/2} = \sqrt{\operatorname{tr}(A^*A)} \quad \text{(Frobenius 范数或 F- 范数)}$$

$$\|A\|_1 = \max_{1\le j\le n} \sum_{i=1}^{n} |a_{ij}| \quad \text{极大列和范数}$$

$$\|A\|_\infty = \max_{1\le i\le n} \sum_{j=1}^{n} |a_{ij}| \quad \text{极大行和范数}$$

则不难验证 (见习题 6), 它们都是 $\mathbb{F}^{n\times n}$ 中的范数, 因而 $\mathbb{F}^{n\times n}$ 成为赋范线性空间.

[49] Henri Léon Lebesgue(1875-1941), 法国数学家, 数学上有著名的 Lebesgue 积分.

注. 上例中的 1- 范数与 ∞- 范数与 例 5.1.2 中的 l_1- 范数和 l_∞- 范数并不一致, 其理由稍后将揭晓.

一个线性空间上有多少种范数? 如何由已知的范数构造新的范数? 下面的命题给出了部分答案, 证明见习题 7.

命题 5.1.1 (构造新范数) 设 $\|\cdot\|_\alpha$ 为 $\mathbb{F}^m(\mathbb{F}=\mathbb{C}$ 或 $\mathbb{R})$ 上的一种向量范数, $A\in\mathbb{F}^{m\times n}$ 是列满秩的矩阵. 对 $\forall x=(x_1,x_2,\cdots,x_n)^T\in\mathbb{F}^n$, 定义

$$\|x\|_\beta=\|Ax\|_\alpha,$$

则 $\|\cdot\|_\beta$ 为 $\mathbb{F}^n$ 的范数.

例 5.1.5 对 $(x,y)\in\mathbb{R}^2$, 规定 $\|(x,y)\|=\sqrt{x^2+2y^2-2xy}$, 则由 命题 5.1.1 可知这确是 $\mathbb{R}^2$ 中的一个向量范数 (为什么?). 此时的单位圆方程为 $x^2+2y^2-2xy=1$! 以我们熟悉的欧几里得度量来看, 这个"单位圆"是一个对称轴不是坐标轴的椭圆!

赋范线性空间中的单位球或单位球面具有重要的意义, 因为在几何上, 它们相当于实数轴上的单位闭区间或其端点, 相当于平面上的单位圆盘或单位圆周, 以及空间中的单位球或单位球面. 因此它们都是有界闭集 (或更精确地, 紧集), 从高等数学或数学分析课程中我们知道, 连续函数在有界闭集上一定有最大值和最小值. 研究赋范线性空间上的连续函数或变换 (算子) 的一个重要技巧就是设法将函数的定义域限制或转移到单位球或单位球面上.

例 5.1.6 按照矩阵的 1- 范数或 ∞- 范数, 所有基本矩阵 E_{ij}, 所有置换矩阵 (单位矩阵通过任意次交换行列的变换得到的矩阵) 等都在单位球面上. 按照 F- 范数, 所有基本矩阵仍在单位球面上, 但任何 $n\geq 2$ 阶置换矩阵的范数均为 $\sqrt{n}$, 从而都不在单位球面上.

从前面的讨论及上述例子我们知道, $\mathbb{F}$ ($\mathbb{F}=\mathbb{C}$ 或 $\mathbb{R}$) 上 n 维线性空间 V 中可以定义无穷多种向量范数. 那么, 这些向量范数之间有什么关系呢? 为回答此问题, 我们先引入一个定义.

定义 5.1.2 设 V 为线性空间 (有限维或无限维), $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是 V 中任意两种范数. 若存在常数 $C>0$, 使对 $\forall x\in V$, 都有

$$\frac{1}{C}\leq\frac{\|x\|_\alpha}{\|x\|_\beta}\leq C \tag{5.1.2}$$

则称 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是等价的.

例 5.1.7 在 $\mathbb{R}^2$ 中, 1- 范数, 欧几里得范数和 ∞- 范数都是等价的, 因为

$$(1/2)\|x\|_1\leq\|x\|_\infty\leq\|x\|_2\leq\|x\|_1\leq 2\|x\|_\infty.$$

于是可将不等式 (5.1.2) 中的常数 C 取为 2.

由定义可以直接得到下面的两种范数等价的简单刻画.

命题 5.1.2 设 V 为线性空间 (有限维或无限维), $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是 V 中的两种范数. 则 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 等价 $\iff$ 存在正的常数 C_1 与 C_2, 使得对 $\forall x\in V$, 都有

$$\|x\|_\alpha \le C_1\|x\|_\beta, \qquad \|x\|_\beta \le C_2\|x\|_\alpha, \tag{5.1.3}$$

下面的命题可以帮助我们更好地理解范数的等价, 其证明是直接的, 见习题8.

命题 5.1.3 设 $x_1, x_2, \cdots, x_n, \cdots$ 是线性空间 V(有限维或无限维) 中的 (向量) 序列, v 是 V 中某给定向量. 设 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是 V 的两个向量范数. 则 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 等价的充分必要条件是

$$\lim_{n\to\infty}\|x_n - v\|_\alpha = 0 \iff \lim_{n\to\infty}\|x_n - v\|_\beta = 0.$$

(此时称序列 $\{x_n\}$ 按范数收敛于 v.) 换句话说, 两个范数等价 $\iff$ 它们具有相同的敛散性.

引理 5.1.1 有限维线性空间的向量范数是向量坐标的连续函数.

证 设 V 是 n 维线性空间, $\alpha_1, \alpha_2, \cdots, \alpha_n$ 是 V 的一组基, 则对 $\forall x\in V$, x 可唯一地表示成

$$x = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_n\alpha_n.$$

显然 V 中任何一种范数 $\|\cdot\|$ 都是坐标 $x_1, x_2, \cdots, x_n$ 的函数, 故记

$$\|x\| = \varphi(x_1, x_2, \cdots, x_n).$$

要证对 $\forall y = y_1\alpha_1 + y_2\alpha_2 + \cdots + y_n\alpha_n \in V$, 如果每个 $|x_i - y_i| \to 0$, 则 $|\varphi(x_1, x_2, \cdots, x_n) - \varphi(y_1, y_2, \cdots, y_n)| \to 0$. 由三角不等式知

$$\begin{aligned}
&|\varphi(x_1, x_2, \cdots, x_n) - \varphi(y_1, y_2, \cdots, y_n)| = \\
&= |\,\|x\| - \|y\|\,| \le \|x - y\| = \|\sum_{i=1}^{n}(x_i - y_i)\alpha_i\| \\
&\le \sum_{i=1}^{n}|x_i - y_i|\,\|\alpha_i\| \le \sqrt{\sum_{i=1}^{n}\|\alpha_i\|^2}\sqrt{\sum_{i=1}^{n}|x_i - y_i|^2} \to 0
\end{aligned}$$

所以 $\|x\|$ 是坐标 $x_1, x_2, \cdots, x_n$ 的连续函数. □

定理 5.1.1 有限维线性空间中的任何两种向量范数都是等价的.

证 设 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是 V 中任意两种范数. 当 $x = 0$ 时, 不等式 (5.1.2) 式显然成立. 设 $x \ne 0$, 则 $\|x\|_\beta \ne 0$. 由 引理5.1.1, $\|x\|_\alpha$, $\|x\|_\beta$ 都是 $x_1, x_2, \cdots, x_n$ 的连续正函数, 因此

$$f(x_1, x_2, \cdots, x_n) = \frac{\|x\|_\alpha}{\|x\|_\beta}$$

也是 $x_1, x_2, \cdots, x_n$ 的连续函数. 考虑在 $\|\cdot\|_\beta$ 下的单位球面 $S = \{x\in V \mid \|x\|_\beta = 1\}$. 由于 S 为有界闭集, 且 S 上的点均不为零, 因此 f 在 S 上连续. 根据多元连续函数的性质, f 在 S 上有最大值 C_2 与最小值 $C_1 > 0$. 由于对 $\forall x \ne 0$, 都有 $x/\|x\|_\beta \in S$, 从而

$$C_1 \le \|\frac{x}{\|x\|_\beta}\|_\alpha = \frac{\|x\|_\alpha}{\|x\|_\beta} \le C_2.$$

即有 $C_1\|x\|_\beta \le \|x\|_\alpha \le C_2\|x\|_\beta$. $\square$

注. 在无限维的线性空间中, 两个向量范数是可以不等价的, 见下例.

例 5.1.8 设 $V=C[0,1]$ 是闭区间 $[0,1]$ 上全体实连续函数组成的无限维实线性空间. 则

$$\|f\|_\infty = \max_{0\le x\le 1} |f(x)| \tag{5.1.4}$$

与

$$\|f\|_1 = \int_0^1 |f(x)|dx \tag{5.1.5}$$

均是 V 中的范数. 它们等价吗? 考虑 V 中的函数列 (请读者画出这些函数的草图以便于理解)f_n, 其中 $f_1=1$, 而对每个 $n\ge 2$,

$$f_n = \begin{cases} 2nx, & 0\le x\le \frac{1}{2n}; \\ -2nx+2, & \frac{1}{2n} < x \le \frac{1}{n}; \\ 0, & x > \frac{1}{n}. \end{cases}$$

则易知 (见习题 9)

$$\lim_{n\to\infty} \|f_n\|_1 = 0$$

但

$$\lim_{n\to\infty} \|f_n\|_\infty = 1.$$

因此, 由 命题 5.1.3 知这两个范数不等价.

由于等价的范数导致相同的收敛性, 而函数的微积分学均由极限定义, 因此等价的范数将导致相同的微积分学. 粗略地说, 在有限维赋范线性空间中的微积分学本质上只有一种, 而在无限维赋范线性空间中则可以有不同的微积分学.

在 例 5.1.4 中将 $m\times n$ 矩阵看成 mn 维向量而定义了矩阵的向量范数. 但矩阵还有自身的特点, 即乘法. 因此有下面的定义.

定义 5.1.3 设 $\|\cdot\|$ 为 $\mathbb{F}^{n\times n}$ 上一个非负的实函数, 若 $\|\cdot\|$ 是 $\mathbb{F}^{n\times n}$ 的向量范数, 且对 $\forall A,B\in\mathbb{F}^{n\times n}$, 有

$$(\text{次乘性}) \qquad \|AB\| \le \|A\|\,\|B\| \tag{5.1.6}$$

则称 $\|\cdot\|$ 为 $\mathbb{F}^{n\times n}$ 上的一个**矩阵范数**.

注 1. (次乘性的合理性) 如果将矩阵范数定义中的次乘性的不等式反向或加强为“乘性”, 即 $\|AB\|\ge\|A\|\,\|B\|$, 则幂零矩阵的矩阵范数将是 0, 与正定性不符.

注 2. (次乘性的意义) 设 $\|A\|<1$, 则次乘性保证了 $\|A^k\|\to 0\,(k\to\infty)$, 因此矩阵范数的次乘性实际上保证了矩阵幂级数的敛散性的“合理性”.

例 5.1.9 例 5.1.4 中定义的矩阵的向量范数均满足次乘性条件 (5.1.6), 因此均是矩阵范数. 为此只需验证三种向量范数均满足次乘性. 此处仅对 F- 范数验证, 其余见习题 11.

设 $A=(a_{ij})_{n\times n}$, $B=(b_{ij})_{n\times n}$. 则

$$\begin{aligned}\|AB\|_F &= \sqrt{\sum_{i,j=1}^{n}|\sum_{k=1}^{n}a_{ik}b_{kj}|^2} \le \sqrt{\sum_{i,j=1}^{n}(\sum_{k=1}^{n}|a_{ik}|^2)(\sum_{k=1}^{n}|b_{kj}|^2)} \\ &= \sqrt{\sum_{i,j=1}^{n}|a_{ij}|^2}\sqrt{\sum_{i,j=1}^{n}|b_{ij}|^2} = \|A\|_F\,\|B\|_F,\end{aligned}$$

即不等式 (5.1.6) 成立. □

今后我们将把矩阵的 1- 范数, F- 范数和 ∞- 范数改记为矩阵的矩阵范数 $\|\!|\cdot|\!\|_1, \|\!|\cdot|\!\|_F, \|\!|\cdot|\!\|_\infty$ 等等.

例 5.1.10 由次乘性可知, $\|\!|A^k|\!\| \le \|\!|A|\!\|^k$. 因此, 如果 $\|\!|A|\!\| < 1$, 则数列 $\{\|\!|A^k|\!\|\}$ 收敛于 0.

例 5.1.11 (单位矩阵的范数) 在任何矩阵范数 $\|\!|\cdot|\!\|$ 之下, 单位矩阵的范数 $\|\!|I|\!\| \ge 1$. 这是因为由次乘性可知 $\|\!|I|\!\| = \|\!|I^2|\!\| \le \|\!|I|\!\|^2$. 请注意, 在向量范数之下, 单位矩阵的范数可以为任何正数!(为什么?)

矩阵范数的 定义 5.1.3 虽然考虑到了矩阵的乘法性质, 但还没有将矩阵和线性变换联系起来. 矩阵的“真正范数”应能同时体现矩阵的这二层含义, 或者说矩阵自身的范数应考虑到矩阵的乘法或者线性变换的复合. 考察下面的例子.

例 5.1.12 (线性变换的大小) 设 A 是 n 阶矩阵. 由第二章第六节线性变换与度量可知, Ax 的度量等于 x 的度量的 $|A|$ 倍. 用范数的语言, 即

$$\|Ax\|_\alpha = \|\!|A|\!\|\|x\|_\alpha. \tag{5.1.7}$$

因此, 矩阵 A 的范数应当定义为

$$\|\!|A|\!\| = \frac{\|Ax\|_\alpha}{\|x\|_\alpha}. \tag{5.1.8}$$

于是单位矩阵的范数为 1, 任何纯量矩阵 aI 的范数为 $|a|$, 这是合理的. 遗憾的是, 对任意矩阵 A, 公式 (5.1.8) 的右端可能不是一个常数, 而和向量 x 有关. 比如, 如果 A 是非零不可逆矩阵, 则存在非零向量 x 使得 $Ax=0$, 于是有 $\|\!|A|\!\|=0$, 这与正定性矛盾. 因此应该取公式 (5.1.8) 右端的最大值或者上确界, 即定义

$$\|\!|A|\!\| = \sup_{x\neq 0}\frac{\|Ax\|_\alpha}{\|x\|_\alpha}. \tag{5.1.9}$$

故有下述定义

定义 5.1.4 设 $\|\!|\cdot|\!\|$ 与 $\|\cdot\|$ 分别是 $\mathbb{F}^{n\times n}$ 及 $\mathbb{F}^n$ 的矩阵范数和向量范数. 若对 $\forall A\in\mathbb{F}^{n\times n}$ 及 $x\in\mathbb{F}^n$, 都有

$$\|Ax\| \le \|\!|A|\!\|\,\|x\| \tag{5.1.10}$$

则称矩阵范数 $\|\!|\cdot|\!\|$ 与向量范数 $\|\cdot\|$ 是相容的.

例 5.1.12 实际上是一个从向量范数出发构造与之相容的矩阵范数的方法, 我们将其写成下面的定理.

定理 5.1.2 设 $(V, \|\cdot\|)$ 是 n 维赋范线性空间, A 是任意 n 阶矩阵. 定义

$$\|\!|A|\!\|=\sup_{0\neq x\in\mathbb{F}^n}\frac{\|Ax\|}{\|x\|}\tag{5.1.11}$$

则上式定义了一个与向量范数 $\|\cdot\|$ 相容的矩阵范数, 称为由向量范数 $\|\cdot\|$ 诱导的矩阵范数或**算子范数**.

证 齐次性与相容性是显然的.

正定性. 对任意 $A\neq 0$, 有 $\|\!|A|\!\|>0$, 这是因为 V 有一组由单位向量构成的基, 因此如果 $Ax=0$ 对所有单位向量成立, 则必有 $A=0$. 故正定性得证.

为证三角不等式和次乘性, 请注意下式成立 (应用转移到单位球或单位球面的技巧, 见习题 15)

$$\|\!|A|\!\|=\max_{\|x\|=1}\|Ax\|=\sup_{0\neq x\in\mathbb{F}^n}\frac{\|Ax\|}{\|x\|}=\sup_{\|x\|\leq 1}\frac{\|Ax\|}{\|x\|}.\tag{5.1.12}$$

因赋范线性空间 $(\mathbb{F}^n, \|\cdot\|)$ 的单位闭球或单位球面皆为有界闭集, 而 $\|Ax\|$ 为 x 的连续函数, 故在单位球或单位球面上取得最大值, 所以公式 (5.1.12) 中的 "sup" 可以换为 "max".

对 $\forall\, A, B\in\mathbb{F}^{n\times n}$, 有

$$\begin{aligned}\|\!|A+B|\!\|&=\sup_{0\neq x\in\mathbb{F}^n}\frac{\|(A+B)x\|}{\|x\|}\leq\sup_{0\neq x\in\mathbb{F}^n}\frac{\|Ax\|+\|Bx\|}{\|x\|}\\&\leq\sup_{0\neq x\in\mathbb{F}^n}\frac{\|Ax\|}{\|x\|}+\sup_{0\neq x\in\mathbb{F}^n}\frac{\|Bx\|}{\|x\|}=\|\!|A|\!\|+\|\!|B|\!\|.\end{aligned}$$

所以三角不等式成立. 设 $\|ABx\|$ 在单位球面的最大值点为 y, 即 $\|\!|AB|\!\|=\|ABy\|$. 如果 $\|By\|=0$, 即 $By=0$, 则显然 $\|\!|AB|\!\|=0\leq\|\!|A|\!\|\,\|\!|B|\!\|$, 即次乘性成立; 如果 $\|By\|\neq 0$, 则

$$\begin{aligned}\|\!|AB|\!\|&=\frac{\|ABy\|}{\|By\|}\cdot\|By\|\leq\max_{\substack{0\neq x\in\mathbb{F}^n\\ \|Bx\|\neq 0}}\frac{\|A(Bx)\|}{\|Bx\|}\cdot\max_{\|x\|=1}\|Bx\|\\&\leq\sup_{0\neq x\in\mathbb{F}^n}\frac{\|Ax\|}{\|x\|}\cdot\|\!|B|\!\|=\|\!|A|\!\|\,\|\!|B|\!\|.\end{aligned}$$

即得次乘性. □

注 1. 可以证明 (见习题 13), 例 5.1.9 中给出的 矩阵的 1- 范数恰好是向量的 1- 范数的诱导范数, 而矩阵的 ∞- 范数恰好是向量的 ∞- 范数的诱导范数 (这是我们将矩阵的 1- 范数和 ∞- 范数没有仿照向量的相应范数定义的原因, 见注 2 和 例 5.1.4 后的注).

注 2. 如果仿照向量的 1- 范数与 ∞- 范数定义矩阵的 1- 范数和 ∞- 范数, 则它们不与向量的 1- 范数与 ∞- 范数相容, 实际上, 这样定义的 ∞- 范数不是矩阵范数, 见习题 14.

推论 5.1.1 (1) $\mathbb{F}^{n\times n}$ 上任意两种矩阵范数均等价;

(2) 对于 $\mathbb{F}^n$ 上每种向量范数, 都存在 $\mathbb{F}^{n\times n}$ 上与它相容的矩阵范数;

(3) 对于 $\mathbb{F}^{n\times n}$ 上每种矩阵范数, 都存在 $\mathbb{F}^n$ 上与它相容的向量范数.

证 只需证明 (3). 设 $\|\cdot\|$ 是 $\mathbb{F}^{n\times n}$ 上的一种矩阵范数, 对每个 $x\in\mathbb{F}^n$, 定义

$$\|x\| = \|xJ^T\| \tag{5.1.13}$$

其中 $J=\sum_{i=1}^n e_i$ 是全1向量矩阵 (实际上 $xJ^T=(x,x,\cdots,x)$ 是每列均为 x 的矩阵). 则公式 (5.1.13) 定义了一个与矩阵范数 $\|\cdot\|$ 相容的向量范数, 证明见习题 20. □

例 5.1.13 将公式 (5.1.13) 定义的向量范数称为矩阵范数 $\|\cdot\|$ 诱导的向量范数. 由矩阵的极大列和范数 $\|\cdot\|_1$ 诱导的向量范数实际上正是向量的1- 范数! 请读者思考: 矩阵的 ∞- 范数诱导的向量范数是 ∞- 范数吗? 矩阵的F- 范数诱导向量的2- 范数吗?

由矩阵和线性变换的对应关系可知, 矩阵范数自然诱导线性变换的范数. 另外, 也可以仿照矩阵范数来定义线性变换的范数, 故有下面的定义.

定义 5.1.5 设 U,V 是任意维实或复赋范线性空间, $\sigma\in\mathrm{Hom}(U,V)$. 定义

$$\|\sigma\| = \sup_{0\neq x\in U}\frac{\|\sigma(x)\|}{\|x\|} \tag{5.1.14}$$

则上式定义了线性空间 $\mathrm{Hom}(U,V)$ 上的一个与 U 中向量范数相容的向量范数, 即 $\|\sigma(x)\|\leq\|\sigma\|\|x\|$. 特别, 如果 $U=V$, 则此向量范数还是矩阵范数, 即满足次乘性 $\|\sigma\tau\|\leq\|\sigma\|\|\tau\|$(请验证!). 我们讨论的线性变换范数均指由公式 (5.1.14) 定义的范数, 它实际上是两个赋范线性空间 U 与 V 的范数诱导的算子范数.

注. 如果 U 是有限维赋范线性空间, 则由公式 (5.1.14) 定义的算子范数 $\|\sigma\|$ 一定有界 (为什么?), 此时称 σ 是**有界算子**(变换). 但若 U 是无限维的, 则此范数可能是无穷大, 此时称 σ 是无界算子.

例 5.1.14 考虑通常欧几里得范数下的平面上的正交投影变换 $\sigma:(x,y)^T\mapsto(x,0)^T$. 则 $\|\sigma\|=1$.

有了线性变换的范数概念, 就可以仿照函数的连续性来定义线性变换的连续性了.

定义 5.1.6 设 U,V 是实或复赋范线性空间 (有限维或无限维), $\sigma\in\mathrm{Hom}(U,V)$. 如果对任意给定的正数 $\varepsilon>0,\exists\delta>0$, 使得对任意 $x,y\in U$, $\|x-y\|<\delta$, 均有

$$\|\sigma(x)-\sigma(y)\|<\varepsilon,$$

则称 σ 是连续的.

定理 5.1.3 有限维赋范线性空间的线性变换均是连续的.

定理的证明只需注意到线性变换的范数是与向量范数相容的即可 (见习题24). 实际上, 我们在第二章即已知道线性变换实际上是若干个线性函数而已, 而每个线性函数都是连续的, 所以线性变换的连续性是自然的结论.

对无限维赋范线性空间, 定理5.1.3 不成立. 比如, 设 $V=\mathbb{R}[x]$ 的范数由公式 (5.1.4) 定

义 (见 例 5.1.8). 定义 $\sigma : \sum_{n\geq 0} a_n x^n \mapsto \sum_{n\geq 0} n a_n$. 则 $\sigma \in \mathrm{Hom}(V, \mathbb{R})$ 不是连续映射 (为什么?). 线性变换的连续性和有界性的关系见习题 15.

思考题

1. 在 $\mathbb{R}^2$ 中, 中心在原点的非等边矩形是否可以是单位圆? 中心在原点的正三角形与双曲线呢?

2. 三角不等式中的等号何时成立? 是否存在范数使得三角不等式总是等式?

3. 两个范数的乘积是否仍是范数? (和的情形见习题 18.)

4. 内积可以诱导范数. 哪些 p- 范数可以诱导内积, 即定义 $(x-y, x-y) = \|x-y\|^2$? 哪些不能?

5. 矩阵 A 与其共轭转置 A^* 的矩阵范数有何联系? 可逆矩阵与其逆矩阵的矩阵范数有何联系? 线性变换与其伴随变换的范数有何联系?

6. 矩阵范数中次乘性的等号何时成立? 是否存在矩阵范数使得次乘性中的等号永远成立?

7. 是否能够由一种矩阵范数定义一种不同于公式 (5.1.13) 的向量范数?

8. 能否在赋范线性空间中定义合理的角度? 研究 1- 范数和 ∞- 范数的单位圆中的几个角, 它们是直角吗?

第二节 矩阵序列与矩阵级数

本节我们讨论向量序列与矩阵序列的敛散性, 其方法与结论基本平行于数学分析和高等数学的相应概念, 请读者注意比较.

定义 5.2.1 设 $(V, \|\cdot\|)$ 是 n 维赋范线性空间, $x_1, x_2, \cdots, x_k, \cdots$ 是 V 的一个向量序列, α 是 V 的一个固定向量. 如果

$$\lim_{k\to\infty} \|x_k - \alpha\| = 0$$

则称向量序列 $x_1, x_2, \cdots, x_k, \cdots$ 在 $\|\cdot\|$ 收敛, 且 α 是该序列的极限, 记作

$$\lim_{k\to\infty} x_k = \alpha \quad 或 \quad x_k \to \alpha.$$

不收敛的向量序列称为发散的.

例 5.2.1 设 $V = \mathbb{R}$ 是普通赋范线性空间 (范数 = 绝对值), 则 定义 5.2.1 中的敛散性与高等数学中的相应概念完全一致.

设 $V = \mathbb{F}^n$ 是赋范线性空间, 则 定义 5.2.1 中的敛散性可以更具体地描述如下 (证明见习题 25):

引理 5.2.1 设 $\{x^{(k)}\}$ 是 $\mathbb{F}^n$ 中的向量序列, $x^{(k)} = (x_1^{(k)}, x_2^{(k)}, \cdots, x_n^{(k)})$, $k = 1, 2, \cdots$. 设 $x = (x_1, x_2, \cdots, x_n)$ 是 $\mathbb{F}^n$ 中的固定向量. 则 $\lim_{k\to\infty} x^{(k)} = x \Longleftrightarrow$

$$\lim_{k\to\infty} x_i^{(k)} = x_i, \quad i = 1, 2, \cdots, n.$$

从上面的引理可以看出, 向量序列的极限实际上是坐标序列的极限, 也就是说向量序列收敛 $\Longleftrightarrow$ 该序列的向量的每个坐标构成的数列 (共有 n 个) 均收敛. 所以, 向量序列的收敛可以称为“**按坐标收敛** ”. 另外, 此定义显然可以推广到所有无穷数列构成的无限维空间.

定理 5.2.1 设 $\{x^{(k)}\}$ 为 $\mathbb{F}^n$ 的向量序列, $\|\cdot\|$ 为 $\mathbb{F}^n$ 中的任一向量范数, 则

$$\lim_{k\to\infty} x^{(k)} = x \Longleftrightarrow \lim_{k\to\infty} \|x^{(k)} - x\| = 0.$$

证 由向量范数的等价性, 定理中的结论只要对一种向量范数成立, 则对任何一种向量范数都成立. 故就向量范数 $\|\cdot\|_\infty$ 来证明即可, 见习题24.

注. 定理5.2.1 可以等价地描述为向量序列按坐标收敛于向量 $x \iff$ 它**按范数收敛**于 x.

将 n 阶矩阵看作 n^2 维向量即可定义矩阵序列的敛散性. 为方便读者, 我们将其单独列出.

定义 5.2.2 设 $A_m = (a_{ij}^{(m)})_{n\times n} \in \mathbb{F}^{n\times n}$, $m = 1, 2, \cdots$. 若对任意的 i, j, 有 $\lim\limits_{m\to\infty} a_{ij}^{(m)} = a_{ij}$, $1 \le i, j \le n$, 则称矩阵序列 $\{A_m\}$ 收敛于矩阵 $A = (a_{ij})_{n\times n}$, 记为 $\lim\limits_{m\to\infty} A_m = A$. 否则, 称 $\{A_m\}$ 为发散的.

上述定义可称作矩阵序列**按元素收敛**或按坐标收敛.

由 定理5.2.1 可知, 矩阵序列按坐标收敛与按范数收敛等价, 即有下述推论

推论 5.2.1 设 $\{A_m\}$ 为 $\mathbb{F}^{n\times n}$ 的矩阵序列, $\|\cdot\|$ 为 $\mathbb{F}^{n\times n}$ 的任意向量范数, 则

$$\lim_{m\to\infty} A_m = A \iff \lim_{m\to\infty} \|A_m - A\| = 0.$$

注. 定理中的“向量范数”不必是“矩阵范数”.

例 5.2.2 设 $A_m = \begin{pmatrix} m\sin\frac{1}{m} & (1-\frac{1}{m})^{2m} \\ 1 & m^2(1-\cos\frac{1}{m}) \end{pmatrix}$, 则

$$\lim_{m\to\infty} A_m = \begin{pmatrix} 1 & e^{-2} \\ 1 & \frac{1}{2} \end{pmatrix}.$$

下面是矩阵序列极限的四则运算法则, 证明见习题25.

性质 5.2.1 (1) 若 $A_m \to A$, $B_m \to B$, $\{a_m\} \to a$, $\{b_m\} \to b$, $a_m, b_m, a, b \in \mathbb{F}$, 则

$$\lim_{m\to\infty} (a_m A_m + b_m B_m) = aA + bB, \qquad \lim_{m\to\infty} A_m B_m = AB.$$

特别地, 若 P 是可逆矩阵, 则

$$\lim_{m\to\infty} P^{-1} A_m P = P^{-1} A P.$$

(2) 若 $\lim\limits_{m\to\infty} A_m = A$, 则对 $\mathbb{F}^{n\times n}$ 中任意范数 $|\!|\!|\cdot|\!|\!|$, $|\!|\!|A_m|\!|\!|$ 有界.

(3) 若 $A_m \to A$, 且 A_m^{-1} 及 A^{-1} 存在, 则

$$\lim_{m\to\infty} A_m^{-1} = A^{-1}.$$

定义 5.2.3 设 $A \in \mathbb{F}^{n\times n}$, 则由 A 的幂可得一矩阵序列:

$$I,\ A,\ A^2,\ A^3, \cdots \tag{5.2.1}$$

若矩阵序列 (5.2.1) 收敛, 则称矩阵 A**幂收敛**.

由矩阵序列极限的四则运算法则立即可得 (证明见习题29) 下述幂收敛的等价条件.

命题 5.2.1 设矩阵 A 与 B 相似, 则 A 幂收敛 $\iff$ B 幂收敛. 特别地, 矩阵 A 幂收敛 $\iff$ 其 Jordan 标准形幂收敛.

例 5.2.3 (1) 对角矩阵幂收敛 $\iff$ 其对角元素为 1 或绝对值小于 1;

(2) 设 $J = J_n(\lambda)$ 是特征值为 λ 的 n 阶 Jordan 块, 则 (其中约定, 若 $s > k$, 则 $C_k^s = 0$)

$$J^m = \begin{pmatrix} \lambda^m & C_m^1\lambda^{m-1} & C_m^2\lambda^{m-2} & \cdots & C_m^{n-1}\lambda^{m-n+1} \\ & \lambda^m & C_m^1\lambda^{m-1} & \cdots & C_m^{n-2}\lambda^{m-n+2} \\ & & \lambda^m & \cdots & C_m^{n-3}\lambda^{m-n+3} \\ & & & \ddots & \vdots \\ & & & & \lambda^m \end{pmatrix}. \tag{5.2.2}$$

因此 J 幂收敛 $\iff$ $|\lambda| < 1$ 或 $\lambda = 1$ 且 $n = 1$.

我们将上面的讨论总结为下面的定理.

定理 5.2.2 设 $A \in \mathbb{C}^{n\times n}$. 则 A 幂收敛 $\iff$ A 的任一特征值 λ 满足: $|\lambda| \le 1$, 并且, 若 $|\lambda| = 1$, 则 $\lambda = 1$ 且对角线元素为 1 的 Jordan 块都是一阶的.

注. "对角线元素为 1 的 Jordan 块都是一阶的" 的含义是特征值 1 的代数重数等于几何重数. 等价地, 最小多项式中的因式 $x - 1$ 的次数为 1.

推论 5.2.2 设 $S_m = \sum_{k=0}^m A^k$, 则 $\{S_m\}$ 收敛 $\iff$ $\lim\limits_{k\to\infty} A^k = 0$.

为了给出矩阵幂收敛的一个范数条件, 我们先讨论矩阵特征值与矩阵范数的一个基本关系.

定理 5.2.3 设 $A \in \mathbb{C}^{n\times n}$, $\|\!|\cdot|\!\|$ 为 $\mathbb{C}^{n\times n}$ 的任意一种矩阵范数, 则 $\rho(A) \le \|\!|A|\!\|$.

证 本章习题19实际上是对本定理的直接证明. 下面给出另一个利用幂收敛的具有启发性的有趣证明, 请读者体会之. 作矩阵

$$B = \frac{1}{\|\!|A|\!\| + \varepsilon} A,$$

其中 ε 为任意正实数. 则

$$\|\!|B|\!\| = \frac{1}{\|\!|A|\!\| + \varepsilon} \|\!|A|\!\| < 1.$$

于是, $\lim\limits_{m\to\infty} \|\!|B^m|\!\| = 0$, 由 推论 5.2.1, $\lim\limits_{m\to\infty} B^m = 0$. 于是由 定理 5.2.2, B 的所有特征值的模都小于或等于 1. 即

$$\frac{1}{\|\!|A|\!\| + \varepsilon} |\lambda| \le 1,$$

其中 λ 为 A 的任一特征值. 于是, $|\lambda| \le \|\!|A|\!\| + \varepsilon$. 因为 ε 为任意正实数, 所以 $|\lambda| \le \|\!|A|\!\|$. □

推论 5.2.3 (Neumann[50] 引理) 设矩阵 A 的某个矩阵范数小于 1, 则 A 幂收敛, $I-A$ 可逆且

$$(I-A)^{-1}=I+A+A^2+\cdots+A^m+\cdots=\sum_{k=0}^{\infty}A^k. \tag{5.2.3}$$

例 5.2.4 设两个离散随机变量的联合分布矩阵是方阵, 则该矩阵幂收敛. 这是因为该矩阵的所有元素之和为 1, 因此其 1- 范数必定小于 1 或者该矩阵仅有一个元素为 1 其余均为 0, 因此也幂收敛.

以下我们讨论矩阵的幂级数.

定义 5.2.4 设 $\{A_k\,|\,k=0,1,2,\cdots\}$ 为一个矩阵序列. 对 $n\geq 0$, 令 $S_n=\sum_{k=0}^{n}A_k$. 若 $\lim\limits_{n\to\infty}S_n=S$, 则称矩阵级数 $\sum_{k=0}^{\infty}A_k$ 收敛于 S, 记作 $\sum_{k=0}^{\infty}A_k=S$. 否则, 称 $\sum_{k=0}^{\infty}A_k$ 发散.

矩阵级数有着与普通数项级数相类似的性质.

性质 5.2.2 (1) 若 $\sum_{k=0}^{\infty}A_k$ 收敛, 则 $\lim\limits_{k\to\infty}A_k=0$;

(2) 若 $\sum_{k=0}^{\infty}A_k=A$, $\sum_{k=0}^{\infty}B_k=B, a\in\mathbb{F}$, 则

$$\sum_{k=0}^{\infty}(A_k+B_k)=A+B,\quad \sum_{k=0}^{\infty}aA_k=aA.$$

下面我们讨论矩阵 A 的幂级数 $\sum_{k=0}^{\infty}a_kA^k$ 的收敛问题.

从形式上看, $\sum_{k=0}^{\infty}a_kA^k$ 可以看成是在函数级数 $\sum_{k=0}^{\infty}a_kt^k$ 中用 A 代替 t 得到. 因此将 $\sum_{k=0}^{\infty}a_kA^k$ 的收敛性与 $\sum_{k=0}^{\infty}a_kt^k$ 的收敛性联系起来是很自然的.

引理 5.2.2 设 J 为对角线元素为 λ 的 n 阶 Jordan 块, $f(t)=\sum_{k=0}^{\infty}a_kt^k$ 是收敛半径为 r 的幂级数. 则当 $|\lambda|<r$ 时, 矩阵幂级数 $\sum_{k=0}^{\infty}a_kJ^k$ 是收敛的, 且其和为矩阵

$$\begin{pmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2}f''(\lambda) & \cdots & \frac{1}{(n-1)!}f^{(n-1)}(\lambda) \\ & f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(n-2)!}f^{(n-2)}(\lambda) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda) \\ & & & & f(\lambda) \end{pmatrix}.$$

[50] Carl Gottfried Neumann (1832-1925), 德国数学家, 等式 (5.2.3) 中的算子幂级数称为 Neumann 级数.

证 令 $S_m=\sum_{k=0}^m a_kJ^k$, $S_m(\lambda)=\sum_{k=0}^m a_k\lambda^k$. 由公式 (5.2.2) 知,

$$J^k=\begin{pmatrix} \lambda^k & C_k^1\lambda^{k-1} & C_k^2\lambda^{k-2} & \cdots & C_k^{n-1}\lambda^{k-n+1} \\ & \lambda^k & C_k^1\lambda^{k-1} & \cdots & C_k^{n-2}\lambda^{k-n+2} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & C_k^1\lambda^{k-1} \\ & & & & \lambda^k \end{pmatrix},$$

其中, 若 $s>k$, 则 $C_k^s=0$. 所以

$$S_m = \begin{pmatrix} \sum\limits_{k=0}^m a_k\lambda^k & \sum\limits_{k=0}^m a_kC_k^1\lambda^{k-1} & \sum\limits_{k=0}^m a_kC_k^2\lambda^{k-2} & \cdots & \sum\limits_{k=0}^m a_kC_k^{n-1}\lambda^{k-n+1} \\ & \sum\limits_{k=0}^m a_k\lambda^k & \sum\limits_{k=0}^m a_kC_k^1\lambda^{k-1} & \cdots & \sum\limits_{k=0}^m a_kC_k^{n-2}\lambda^{k-n+2} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & \sum\limits_{k=0}^m a_kC_k^1\lambda^{k-1} \\ & & & & \sum\limits_{k=0}^m a_k\lambda^k \end{pmatrix}$$

$$= \begin{pmatrix} S_m(\lambda) & S_m'(\lambda) & \frac{1}{2!}S_m''(\lambda) & \cdots & \frac{1}{(n-1)!}S_m^{(n-1)}(\lambda) \\ & S_m(\lambda) & S_m'(\lambda) & \cdots & \frac{1}{(n-2)!}S_m^{(n-2)}(\lambda) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & S_m'(\lambda) \\ & & & & S_m(\lambda) \end{pmatrix}.$$

因 $S_m(t)=\sum_{k=0}^m a_kt^k$ 的收敛半径为 r, 且 $|\lambda|<r$, 所以 $S_m(\lambda),S_m'(\lambda),S_m''(\lambda),\cdots,S_m^{(n-1)}(\lambda)$ 皆收敛, 且 $\lim\limits_{m\to\infty}S_m(\lambda)=f(\lambda)$, $\lim\limits_{m\to\infty}S_m^{(k)}(\lambda)=f^{(k)}(\lambda)$, $k=1,2,\cdots,n-1$. 因此,

$$\lim_{m\to\infty}S_m=\sum_{k=0}^{\infty}a_kJ^k=\begin{pmatrix} f(\lambda) & f'(\lambda) & \frac{1}{2!}f''(\lambda) & \cdots & \frac{1}{(n-1)!}f^{(n-1)}(\lambda) \\ & f(\lambda) & f'(\lambda) & \cdots & \frac{1}{(n-2)!}f^{(n-2)}(\lambda) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda) \\ & & & & f(\lambda) \end{pmatrix}.$$

例 5.2.5 设 $f(t)=\sum_{k=0}^\infty(\frac{t}{3})^k$, 求 $f(J)$, 其中

$$J=\begin{pmatrix} 2 & 1 & & \\ & 2 & 1 & \\ & & 2 & 1 \\ & & & 2 \end{pmatrix}.$$

解 $f(t)=(1-\frac{t}{3})^{-1}$, 其收敛半径为3. 因 J 的特征值2落在 $f(t)$ 的收敛域内, 所以 $f(J)=\sum_{k=0}^{\infty}(\frac{J}{3})^k$ 是收敛的且

$$f(J)=\begin{pmatrix} f(2) & f'(2) & \frac{1}{2!}f''(2) & \frac{1}{3!}f'''(2) \\ & f(2) & f'(2) & \frac{1}{2!}f''(2) \\ & & f(2) & f'(2) \\ & & & f(2) \end{pmatrix}.$$

因

$$f'(t)=\frac{1}{3}(1-\frac{t}{3})^{-2},\ f''(t)=\frac{2}{9}(1-\frac{t}{3})^{-3},\ f'''(t)=\frac{2}{9}(1-\frac{t}{3})^{-4},$$

所以

$$f(2)=3,\quad f'(2)=3,\quad f''(2)=6,\quad f'''(2)=18.$$

因此

$$f(J)=\begin{pmatrix} 3 & 3 & 3 & 3 \\ & 3 & 3 & 3 \\ & & 3 & 3 \\ & & & 3 \end{pmatrix}.$$

由 引理 5.2.2 便可得到下面的关于矩阵幂级数收敛的基本定理.

定理 5.2.4 (Lagrange[51] -Sylvester 定理) 设 $f(t)=\sum_{k=0}^{\infty}a_kt^k$, 它的收敛半径为 r. 设矩阵 A 的 Jordan 标准形为

$$J=\begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{pmatrix},\quad J_i=\begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{pmatrix},\quad 1\le i\le s,$$

其变换矩阵为 P, 即 $A=PJP^{-1}$. 若对所有的 $i=1,2,\cdots,s$, 都有 $|\lambda_i|<r$, 则矩阵幂级数 $\sum_{k=0}^{\infty}a_kA^k$ 收敛, 其和为

$$f(A)=\sum_{k=0}^{\infty}a_kA^k=P\begin{pmatrix} f(J_1) & & & \\ & f(J_2) & & \\ & & \ddots & \\ & & & f(J_s) \end{pmatrix}P^{-1},$$

其中

$$f(J_i)=\begin{pmatrix} f(\lambda_i) & f'(\lambda_i) & \frac{1}{2!}f''(\lambda_i) & \cdots & \frac{1}{(n_i-1)!}f^{(n_i-1)}(\lambda_i) \\ & f(\lambda_i) & f'(\lambda_i) & \cdots & \frac{1}{(n_i-2)!}f^{(n_i-2)}(\lambda_i) \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & f'(\lambda_i) \\ & & & & f(\lambda_i) \end{pmatrix},\quad i=1,2,\cdots,s.$$

[51] Joseph Louis Lagrange(1736-1813), 著名法国数学家和天文学家, 出生于 Turin (现意大利), 对数学的众多分支以及经典力学和天体力学有杰出贡献.

例 5.2.6 已知 $f(t)=2-t+2t^3$, 求 $f(A)$, 其中

$$A=\begin{pmatrix} 1 & 1 & & & \\ & 1 & & & \\ & & 2 & 1 & \\ & & & 2 & 1 \\ & & & & 2 \end{pmatrix}.$$

解 由于 $f(t)$ 次数较低, 读者可以尝试直接计算本题. 但注意多项式是最简单的幂级数, 因此利用 定理5.2.4 更为简洁. 因 $f(1)=3$, $f'(1)=5$, $f(2)=16$, $f'(2)=23$, $f''(2)=24$, 所以

$$f(A)=\begin{pmatrix} 3 & 5 & & & \\ & 3 & & & \\ & & 16 & 23 & 12 \\ & & & 16 & 23 \\ & & & & 16 \end{pmatrix}.$$

思考题

1. 若 $\lim\limits_{n\to\infty} A_nB_n$ 存在, 是否 $\lim\limits_{n\to\infty} A_n$, $\lim\limits_{n\to\infty} B_n$ 一定存在? 为什么?
2. 设 A,B 均幂收敛, $A+B, AB$ 幂收敛吗?

第三节 矩阵函数的导数与积分

有了前面的Lagrange-Sylvester定理, 我们可以像复变函数论那样, 利用矩阵幂级数来定义矩阵函数. 下列结果是复变函数论中已知的结论 ($z\in\mathbb{C}$):

$$e^z=\sum_{m=0}^{\infty}\frac{z^m}{m!},$$

$$\sin z=\sum_{m=1}^{\infty}(-1)^{m-1}\frac{z^{2m-1}}{(2m-1)!},$$

$$\cos z=1+\sum_{m=1}^{\infty}(-1)^{m}\frac{z^{2m}}{(2m)!}.$$

上面三个幂级数在复平面上都是收敛的. 因而由Lagrange-Sylvester定理, 下列各矩阵幂级数

$$\sum_{m=0}^{\infty}\frac{A^m}{m!},\quad \sum_{m=1}^{\infty}(-1)^{m-1}\frac{A^{2m-1}}{(2m-1)!},\quad I+\sum_{m=1}^{\infty}(-1)^{m}\frac{A^{2m}}{(2m)!},$$

都收敛. 它们的和 (矩阵) 分别用记号 e^A, $\sin A$, $\cos A$ 来表示, 并分别称为方阵 A 的指数函数, 正弦函数及余弦函数. 同样地, 由

$$\ln(1+z)=\sum_{m=1}^{\infty}(-1)^{m-1}\frac{z^m}{m},\quad |z|<1,$$

$$(1+z)^a=1+\sum_{m=1}^{\infty}\frac{a(a-1)\cdots(a-m+1)}{m!}z^m,\quad |z|<1,\quad a\text{ 为任意实数},$$

可定义方阵函数

$$\ln(I+A)=\sum_{m=1}^{\infty}(-1)^{m-1}\frac{A^m}{m},\quad \rho(A)<1,$$

$$(I+A)^a=I+\sum_{m=1}^{\infty}\frac{a(a-1)\cdots(a-m+1)}{m!}A^m,\quad \rho(A)<1,$$

这里 $\rho(A)$ 为 A 的谱半径.

由上述几个矩阵函数的定义可得, $e^{\lambda I}=e^{\lambda}I$, $\sin(\lambda I)=(\sin\lambda)I$, $\cos(\lambda I)=(\cos\lambda)I$.

根据 Lagrange-Sylvester 定理, 我们将矩阵函数 e^A 与 $\sin A$ 在 n 阶 Jordan 块处的函数值 (矩阵) 写成以下的命题, 以方便读者计算或引用, 其余几个矩阵函数可类似计算, 此处略去.

命题 5.3.1 设 J 是特征值为 λ 的 n 阶 Jordan 块, 则

$$e^A=e^{\lambda}\begin{pmatrix}1 & 1 & \frac{1}{2!} & \frac{1}{3!} & \cdots & \frac{1}{(n-1)!}\\ & 1 & 1 & \frac{1}{2!} & \cdots & \frac{1}{(n-2)!}\\ & & \ddots & \ddots & \ddots & \vdots\\ & & & 1 & 1 & \frac{1}{2!}\\ & & & & 1 & 1\\ & & & & & 1\end{pmatrix} \tag{5.3.1}$$

$$\sin A=\begin{pmatrix}\sin\lambda & \cos\lambda & -\frac{\sin\lambda}{2!} & -\frac{\cos\lambda}{3!} & \cdots & \frac{\sin[(2\pi)/(n-1)+\lambda]}{(n-1)!}\\ & \sin\lambda & \cos\lambda & -\frac{\sin\lambda}{2!} & \cdots & \frac{\sin[(2\pi)/(n-2)+\lambda]}{(n-2)!}\\ & & \ddots & \ddots & \ddots & \vdots\\ & & & \sin\lambda & \cos\lambda & -\frac{\sin\lambda}{2!}\\ & & & & \sin\lambda & \cos\lambda\\ & & & & & \sin\lambda\end{pmatrix} \tag{5.3.2}$$

例 5.3.1 设 $A=\begin{pmatrix}1 & 1\\ 0 & 2\end{pmatrix}$, 求 e^A, $\sin A$, $\cos A$.

解 A 有特征值 $\lambda_1=1, \lambda_2=2$. 因此 A 与对角矩阵 $\begin{pmatrix}1 & 0\\ 0 & 2\end{pmatrix}$ 相似. 计算得对应于 λ_1, λ_2 的特征向量为

$$\alpha_1=\begin{pmatrix}1\\ 0\end{pmatrix},\quad \alpha_2=\begin{pmatrix}1\\ 1\end{pmatrix}.$$

令 $P = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, 则 $A = P\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}P^{-1}$. 于是

$$\begin{aligned}
e^A &= P\begin{pmatrix} e & 0 \\ 0 & e^2 \end{pmatrix}P^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} e & 0 \\ 0 & e^2 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e & -e+e^2 \\ 0 & e^2 \end{pmatrix}, \\
\sin A &= P\begin{pmatrix} \sin 1 & 0 \\ 0 & \sin 2 \end{pmatrix}P^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \sin 1 & 0 \\ 0 & \sin 2 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \\
&= \begin{pmatrix} \sin 1 & -\sin 1 + \sin 2 \\ 0 & \sin 2 \end{pmatrix}, \\
\cos A &= P\begin{pmatrix} \cos 1 & 0 \\ 0 & \cos 2 \end{pmatrix}P^{-1} = \begin{pmatrix} \cos 1 & -\cos 1 + \cos 2 \\ 0 & \cos 2 \end{pmatrix}.
\end{aligned}$$

例 5.3.2 设 n 阶矩阵 A 的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$. 则 e^A, $\sin A$, $\cos A$ 的特征值分别为 $e^{\lambda_1}, e^{\lambda_2}, \cdots, e^{\lambda_n}$; $\sin\lambda_1, \sin\lambda_2, \cdots, \sin\lambda_n$; $\cos\lambda_1, \cos\lambda_2, \cdots, \cos\lambda_n$.

单纯矩阵的幂级数可由其谱分解方便地得到, 即有下述定理 (证明见习题 33)

定理 5.3.1 设单纯矩阵 A 的谱分解为

$$A = \sum_{i=1}^{s} \lambda_i P_i.$$

设幂级数 $f(t)$ 的收敛半径 $r > \rho(A)$. 则

$$f(A) = \sum_{i=1}^{s} f(\lambda_i) P_i \tag{5.3.3}$$

例 5.3.3 在第四章例 4.2.3 中我们已经知道正规矩阵

$$A = \begin{pmatrix} 8 & 4 & -1 \\ 4 & -7 & 4 \\ -1 & 4 & 8 \end{pmatrix}$$

的谱分解为 $A = 9P_1 - 9P_2$, 因此 $\sin A = (\sin 9)(P_1 - P_2)$.

下面我们给出指数函数 e^A 的一些基本性质.

命题 5.3.2 (1) 若 $AB = BA$, 则 $e^A e^B = e^{A+B} = e^B e^A$;
(2) $(e^A)^{-1} = e^{-A}$;
(3) $|\,e^A\,| = e^{\operatorname{tr} A}$.

证 (1) 因 $AB = BA$, 所以二项式定理成立, 故有

$$(A+B)^m = \sum_{k=0}^{m} C_m^k A^{m-k} B^k.$$

所以

$$e^{A+B}=\sum_{m=0}^{\infty}\frac{1}{m!}\sum_{k=0}^{m}C_m^kA^{m-k}B^k$$
$$=\sum_{m=0}^{\infty}\sum_{k=0}^{m}\frac{1}{k!(m-k)!}A^{m-k}B^k$$
$$=\left(\sum_{k=0}^{\infty}\frac{A^k}{k!}\right)\left(\sum_{j=0}^{\infty}\frac{B^j}{j!}\right)=e^Ae^B.$$

(2) 在 (1) 中令 $B=-A$, 则得 $e^Ae^{-A}=I$, 所以

$$(e^A)^{-1}=e^{-A}.$$

(3) 设 A 的特征值为 $\lambda_1,\lambda_2,...,\lambda_n$, 则 e^A 的特征值为 $e^{\lambda_1},e^{\lambda_2},...,e^{\lambda_n}$, 因此 $|e^A|=e^{\lambda_1}e^{\lambda_2}\cdots e^{\lambda_n}=e^{\lambda_1+\lambda_2+\cdots+\lambda_n}=e^{\operatorname{tr}A}$. □

对于矩阵正弦函数和余弦函数, 我们可推出如下结论 (证明见习题 34).

命题 5.3.3 (1) (Euler[52] 公式) $e^{\mathrm{i}A}=\cos A+\mathrm{i}\sin A$,

$$\cos A=\frac{1}{2}(e^{\mathrm{i}A}+e^{-\mathrm{i}A}),$$
$$\sin A=\frac{1}{2\mathrm{i}}(e^{\mathrm{i}A}-e^{-\mathrm{i}A}),$$
$$\cos(-A)=\cos A,\quad \sin(-A)=-\sin A.$$

(2) 若 $AB=BA$, 则

$$\cos(A+B)=\cos A\cos B-\sin A\sin B,$$
$$\sin(A+B)=\sin A\cos B+\cos A\sin B.$$

注. 上述公式与实数或复数的情形略有差异, 即涉及到两个矩阵乘积的公式需要交换性方能成立. 请参考本节的思考题.

定义 5.3.1 若 $m\times n$ 矩阵 A 的每个元素 a_{ij} 都是变量 t 的函数, 则称 A 为函数矩阵或矩阵函数, 记为 $A(t)=(a_{ij}(t))_{m\times n}$.

仿照数学分析或高等数学课程对函数的极限, 连续性, 导数与积分的定义, 我们将按照元素或分量来定义矩阵函数的相应概念. 于是有如下定义.

定义 5.3.2 设 $A(t)=(a_{ij}(t))_{m\times n}$. 若对 $\forall\, a_{ij}(t)$, $1\le i\le m$, $1\le j\le n$, 都有 $\lim\limits_{t\to t_0}a_{ij}(t)=a_{ij}$(其中 $a_{ij}\in\mathbb{C}$), 则称函数矩阵 $A(t)$ 在 t_0 点处的极限为 $A=(a_{ij})_{m\times n}$.

性质 5.3.1 设 $\lim\limits_{t\to t_0}A(t)=A,\lim\limits_{t\to t_0}B(t)=B$.

(1) 若 $A(t)$, $B(t)$ 是同类型的矩阵, 则

$$\lim_{t\to t_0}(A(t)+B(t))=A+B.$$

[52] Leonhard Euler(1707-1783), 著名瑞士数学家, 物理学家, 一生大部分时间在俄国和德国, 对数学的众多分支以及力学, 流体动力学, 光学和天文学有重要贡献, 第 2002 号小行星以其名字命名. 公式 $e^{\mathrm{i}A}=\cos A+\mathrm{i}\sin A$ 对任意复数和复数矩阵均成立, $e^{\mathrm{i}\pi}+1=0$ 被称为世界第一公式.

(2) 若 $A(t)$, $B(t)$ 分别为 $m\times n$, $n\times s$ 矩阵, 则

$$\lim_{t\to t_0}(A(t)B(t))=\lim_{t\to t_0}A(t)\lim_{t\to t_0}B(t)=AB.$$

(3) 设 k 为常数, 则

$$\lim_{t\to t_0}(kA(t))=kA=k\lim_{t\to t_0}A(t).$$

像 定义5.3.2 一样, 可以通过函数矩阵 $A(t)=(a_{ij}(t))_{m\times n}$ 的每一个元素在一点或某一区间内的连续性, 可微性和可积性来分别定义 $A(t)$ 在一点或某一区间内连续, 可微和可积.

若 $A(t)$ 可微, 其导数定义如下:

$$A'(t)=(a'_{ij}(t))_{m\times n}.$$

若 $A(t)$ 可积, 定义:

$$\int_a^b A(t)\,\mathrm{d}\,t=\left(\int_a^b a_{ij}(t)\,\mathrm{d}\,t\right)_{m\times n}.$$

性质 5.3.2 (1) $(aA(t)+bB(t))'=aA'(t)+bB'(t)$;
(2) $(A(t)B(t))'=A'(t)B(t)+A(t)B'(t)$;
(3) $\left(\int_a^t A(s)\mathrm{d}\,s\right)'=A(t)$;
(4) $\int_a^t A'(s)\mathrm{d}\,s=A(t)-A(a)$;
(5) $\int_a^t BA(s)\mathrm{d}\,s=B\int_a^t A(s)\mathrm{d}\,s$, $\int_a^t A(s)B\mathrm{d}\,s=(\int_a^t A(s)\mathrm{d}\,s)B$, B 为常数矩阵.

命题 5.3.4 (1) $\frac{\mathrm{d}e^{At}}{\mathrm{d}t}=Ae^{At}$;
(2) $\frac{\mathrm{d}\sin At}{\mathrm{d}t}=A\cos At$, $\frac{\mathrm{d}\cos At}{\mathrm{d}t}=-A\sin At$.
特别地, $\frac{\mathrm{d}e^{At}}{\mathrm{d}t}|_{t=0}=A$; $\frac{\mathrm{d}\sin At}{\mathrm{d}t}|_{t=0}=A$.

证 只证 (1), 其余证明类似, 见习题34. 显然可设 A 是 Jordan 块 $J_n(\lambda)=\lambda I+N$. 由 命题5.3.2(1) 知, $e^{Jt}=e^{\lambda t}e^{Nt}$, 故

$$\frac{\mathrm{d}e^{Jt}}{\mathrm{d}t}=\lambda e^{\lambda t}e^{Nt}+e^{\lambda t}\frac{\mathrm{d}e^{Nt}}{\mathrm{d}t}.$$

因此只需验证 $\frac{\mathrm{d}e^{Nt}}{\mathrm{d}t}=Ne^{Nt}$ 即可. □

推论 5.3.1 对任意方阵 A 有

$$\int_{t_0}^t Ae^{As}\mathrm{d}s=e^{At}-e^{At_0} \tag{5.3.4}$$

$$\int_{t_0}^{t} A\sin As\mathrm{d}s = \cos At_0 - \cos At \tag{5.3.5}$$

$$\int_{t_0}^{t} A\cos As\mathrm{d}s = \sin At - \sin At_0 \tag{5.3.6}$$

特别, 若 A 可逆, 则

$$\int_{t_0}^{t} e^{As}\mathrm{d}s = A^{-1}(e^{At} - e^{At_0}).$$

例 5.3.4 设向量 $x = x(t) = (x_1(t), x_2(t), \cdots, x_n(t))^T$ 及对称矩阵 $A = A(t) = (A_{ij}(t))_{n\times n}$ 都是可微的, 求二次型 x^TAx 关于变量 t 的导数.

解

$$\begin{aligned}(x^TAx)' &= (x^T)'Ax + x^T(Ax)' \\ &= (x^T)'Ax + x^TA'x + x^TAx'.\end{aligned}$$

又

$$((x^T)'Ax)^T = x^TA^Tx' = x^TAx'.$$

而 $(x^T)'Ax$ 为一阶矩阵, 所以 $(x^T)'Ax = x^TAx'$. 于是, 我们得到

$$(x^TAx)' = x^TA'x + 2x^TAx'.$$

我们将在下节详细讨论矩阵函数及其微积分的计算.

思考题

1. $e^Ae^B = e^Be^A$ 成立的可能性有多大? 更一般地, 设 $f(x)$ 是一个幂级数, 则 $f(A)f(B) = f(B)f(A)$ 成立的可能性如何? 一般地, 如何比较与 A 可交换的矩阵的数量 (当然是无穷多个) 和与 A 不可交换的矩阵的数量?
2. 试举例说明矩阵 e^Ae^B, e^Be^A 与 e^{A+B} 可以两两不等. 又, 如果 $e^Ae^B = e^Be^A$, 是否有 $e^Ae^B = e^{A+B}$?
3. 矩阵的勾股定理是否成立, 即是否有 $\cos^2 A + \sin^2 A = I$?
4. 公式 $(A(t)^2)' = 2A(t)A'(t)$ 正确吗?
5. 设 $A(t)$ 可逆, 如何计算 $(A(t)^{-1})'$? 又 $A'(t)$ 是否可逆?
6. 设 $A(t)$ 是正交矩阵, 问 $A'(t)$ 还是正交矩阵吗?
7. 例 5.4.2 表明, 即使 A 不可逆, 积分 $\int_{t_0}^{t} e^{As}\mathrm{d}s$ 仍然有意义. 应如何计算?

第四节 矩阵函数的计算

本节我们将利用 Lagrange-Sylvester 定理讨论矩阵函数及其微积分的具体计算.

定理 5.4.1 (1) 设

$$J = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda \end{pmatrix}_{s\times s}, \quad 则 \quad e^{Jt} = e^{\lambda t}\begin{pmatrix} 1 & t & \frac{t^2}{2!} & \cdots & \frac{t^{s-1}}{(s-1)!} \\ & 1 & t & \cdots & \frac{t^{s-2}}{(s-2)!} \\ & & \ddots & \ddots & \vdots \\ & & & \ddots & t \\ & & & & 1 \end{pmatrix}.$$

(2) 设 $A=PJP^{-1}$ 且 A 的 Jordan 标准形为

$$J=J_1\oplus J_2\oplus\cdots\oplus J_s,$$

则

$$e^{At}=P(e^{J_1t}\oplus e^{J_2t}\oplus\cdots\oplus e^{J_st})P^{-1}.$$

证 只证 (1), 因为 (2) 是 (1) 的直接推论. 因为 $J=\lambda I+N$, 其中

$$N=\begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & 0 \end{pmatrix},$$

所以

$$\begin{aligned} e^{Jt}&=e^{\lambda tI+tN}=e^{\lambda t}e^{tN}=e^{\lambda t}\sum_{k=0}^{\infty}\frac{(tN)^k}{k!}\\ &=e^{\lambda t}\sum_{k=0}^{s-1}\frac{t^kN^k}{k!}. \end{aligned}$$

由于 N^k 仅第 k 条上对角线的元素为 1, 其余元素均为 0, 所以定理成立. □

例 5.4.1 求 e^{Jt}, 其中

$$J=\begin{pmatrix} J_1 & \\ & J_2 \end{pmatrix},\quad J_1=\begin{pmatrix} a & 1 & \\ & a & 1 \\ & & a \end{pmatrix},\quad J_2=\begin{pmatrix} b & 1 \\ & b \end{pmatrix}.$$

解 由 定理 5.4.1, 有

$$e^{J_1t}=e^{at}\begin{pmatrix} 1 & t & \frac{t^2}{2!} \\ & 1 & t \\ & & 1 \end{pmatrix},\qquad e^{J_2t}=e^{bt}\begin{pmatrix} 1 & t \\ & 1 \end{pmatrix}.$$

所以,

$$e^{Jt}=\begin{pmatrix} e^{J_1t} & \\ & e^{J_2t} \end{pmatrix}=\begin{pmatrix} e^{at} & te^{at} & \frac{t^2e^{at}}{2!} & & \\ & e^{at} & te^{at} & & \\ & & e^{at} & & \\ & & & e^{bt} & te^{bt} \\ & & & & e^{bt} \end{pmatrix}.$$

例 5.4.2 设 N 是 3 阶幂零 Jordan 块, 则

$$\int_0^t e^{Ns}\mathrm{d}s=\begin{pmatrix} t & \frac{t^2}{2!} & \frac{t^3}{3!} \\ & t & \frac{t^2}{2!} \\ & & t \end{pmatrix}.$$

设 A 为 n 阶方阵. 若已知 A 的最小多项式为

$$m(\lambda)=\lambda^m+a_1\lambda^{m-1}+\cdots+a_{s-1}\lambda+a_m,$$

则矩阵 A 的任何次幂都可由 $I,\, A,\, A^2,\cdots,A^{m-1}$ 的线性组合表示. 因此, 一个由矩阵幂级数定义的矩阵函数 $f(A)$, 可以通过一个次数不超过 $m-1$ 的 A 的多项式 $g(A)$ 来表示. 我们要找的这个多项式 $g(t)$ 有什么特点呢? 回忆两个 n 次多项式相等 $\iff$ 它们在 $n+1$ 个点处的值相同, 因此为确定 $g(A)$, 只需要研究一些特殊点处的值即可. 根据 Lagrange-Sylvester 定理, 只需要研究 $g(t)$ 以及 $g(t)$ 的适当阶数的导数在 A 的所有特征值处的值即可. 因此有下述定义.

定义 5.4.1 设方阵 A 的最小多项式为

$$m(\lambda)=(\lambda-\lambda_1)^{k_1}(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_s)^{k_s}, \tag{5.4.1}$$

其中 $\lambda_1,\,\lambda_2,\cdots,\,\lambda_s$ 互不相同. 对 $f(t)=\sum_{k=0}^{\infty}a_kt^k$, 称

$$\begin{array}{ll}\{ & f(\lambda_1),\, f'(\lambda_1),\cdots,\, f^{(k_1-1)}(\lambda_1),\\ & f(\lambda_2),\, f'(\lambda_2),\cdots,\, f^{(k_2-1)}(\lambda_2),\\ & \cdots\cdots\cdots\cdots,\\ & f(\lambda_s),\, f'(\lambda_s),\cdots,\, f^{(k_s-1)}(\lambda_s) \quad \}\end{array} \tag{5.4.2}$$

为函数 $f(t)$ 在矩阵 A 的谱上的数值.

定理 5.4.2 (Sylvester 矩阵定理) 设 $f(t)=\sum_{k=0}^{\infty}a_kt^k,\, g(t)=\sum_{k=0}^{\infty}b_kt^k$. 则 $f(A)=g(A) \iff f(t)$ 与 $g(t)$ 在 A 的谱上的数值相等, 即, $f(\lambda_i)=g(\lambda_i),\, f'(\lambda_i)=g'(\lambda_i),\cdots,\, f^{(k_i-1)}(\lambda_i)=g^{(k_i-1)}(\lambda_i),\, i=1,2,\cdots,s$.

证 设 A 的 Jordan 标准形为 $P^{-1}AP=J=J_1\oplus\cdots\oplus J_s$, 则由 $f(A)=g(A) \iff f(J)=g(J) \iff f(J_i)=g(J_i),\forall i$, 再对照 Lagrange-Sylvester 定理中的每个块即可, 详见习题 40. □

推论 5.4.1 设矩阵 A 可对角化, 则 $f(A)=g(A) \iff f(\lambda)=g(\lambda),\forall\lambda\in\sigma(A)$.

例 5.4.3 已知矩阵 A 的最小多项式是 $m(\lambda)=\lambda^2(\lambda-\pi)$. 证明 $\sin A=A-\frac{1}{\pi}A^2$.

证 设 $f(t)=\sin t,\, g(t)=t-\frac{1}{\pi}t^2$. 由 定理 5.4.2, 只要证明 $f(t)$ 与 $g(t)$ 在 A 的谱上的数值相同, 则 $f(A)=g(A)$. 直接计算可得,

$$\begin{aligned}&f(0)=\sin 0=0=g(0),\quad f'(0)=\cos 0=1=g'(0),\\&f(\pi)=\sin\pi=0=g(\pi).\end{aligned}$$

因此 $\sin A=f(A)=g(A)=A-\frac{1}{\pi}A^2$. □

例 5.4.4 求 e^{At}, 其中

$$A=\begin{pmatrix}2&1&4\\0&2&0\\0&3&1\end{pmatrix}.$$

解 A 的特征多项式为

$$|\lambda I-A|=\begin{vmatrix}\lambda-2 & -1 & -4\\ 0 & \lambda-2 & 0\\ 0 & -3 & \lambda-1\end{vmatrix}=(\lambda-1)(\lambda-2)^2.$$

直接验证便知, $(x-1)(x-2)$ 不是 A 的零化多项式, 所以 A 的最小多项式为

$$m(x)=(x-1)(x-2)^2.$$

令 $f(\lambda)=e^{\lambda t}$(在这里把 t 看成常数), 则 $e^{At}=f(A)$. 设 $g(\lambda)=a_0+a_1\lambda+a_2\lambda^2$(其中 a_0, a_1, a_2 为待定系数). 要求 $f(\lambda)$ 与 $g(\lambda)$ 在 A 的谱上的数值相等, 即

$$\begin{cases}e^t=f(1)=g(1)=a_0+a_1+a_2,\\ e^{2t}=f(2)=g(2)=a_0+2a_1+4a_2,\\ te^{2t}=f'(2)=g'(2)=a_1+4a_2.\end{cases}$$

解上述方程得

$$a_0=4e^t-3e^{2t}+2te^{2t},\ a_1=-4e^t+4e^{2t}-3te^{2t},\ a_2=e^t-e^{2t}+te^{2t}.$$

因而得到

$$e^{At}=\begin{pmatrix}e^{2t} & 12e^t-12e^{2t}+13te^{2t} & -4e^t+4e^{2t}\\ 0 & e^{2t} & 0\\ 0 & -3e^t+3e^{2t} & e^t\end{pmatrix}.$$

在上述例子中我们采用的是待定系数法. 下面我们介绍一种在实际应用中经常采用的Lagrange插值法. 下面的引理是熟知的 (见习题41).

引理 5.4.1 (Lagrange插值公式) 设 $f(x)$ 是 n 次多项式, $a_i, i=1,2,\cdots,n+1$ 是不同的数. 对每个 $i=1,2,\cdots,n+1$, 令

$$L_i(x)=\frac{(x-a_1)\cdots(x-a_{i-1})(x-a_{i+1})\cdots(x-a_{n+1})}{(a_i-a_1)\cdots(a_i-a_{i-1})(a_i-a_{i+1})\cdots(a_i-a_{n+1})} \tag{5.4.3}$$

则

$$f(x)=\sum_{i=1}^{n+1}f(a_i)L_i(x).$$

公式 (5.4.3) 中的 n 次多项式 $L_i(x)$ 称为Lagrange插值多项式. 于是有

命题 5.4.1 设方阵 A 的最小多项式 $m(\lambda)=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_m)$ 无重根, $f(t)$ 为任一收敛半径 $r>\rho(A)$ 的幂级数. 则

$$f(A)=\sum_{i=1}^{m}f(\lambda_i)L_i(A),$$

其中 $L_i(x),\quad i=1,2,\cdots,m$ 是Lagrange插值多项式.

例 5.4.5 设 $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, 求 e^{At}.

解 因为 $|\lambda I - A| = \begin{vmatrix} \lambda - 1 & -2 \\ -2 & \lambda - 1 \end{vmatrix} = (\lambda+1)(\lambda-3)$, 所以 A 的特征值为 $\lambda_1 = -1$, $\lambda_2 = 3$, 因而其最小多项式为

$$m(x) = (x+1)(x-3).$$

令 $f(\lambda) = e^{\lambda t}$, 则按 Lagrange 插值公式, 有

$$f(A) = f(\lambda_1)L_1(A) + f(\lambda_2)L_2(A),$$

其中

$$L_1(A) = \frac{A - \lambda_2 I}{\lambda_1 - \lambda_2} = -\frac{1}{4}\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix}, \quad L_2(A) = \frac{A - \lambda_1 I}{\lambda_2 - \lambda_1} = \frac{1}{4}\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}.$$

又 $f(\lambda_1) = e^{-t}$, $f(\lambda_2) = e^{3t}$. 所以

$$\begin{aligned} e^{At} = f(A) &= e^{-t} \cdot (-\frac{1}{4})\begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix} + e^{3t}\frac{1}{4}\begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \\ &= \frac{1}{2}\begin{pmatrix} e^{-t} + e^{3t} & -e^{-t} + e^{3t} \\ -e^{-t} + e^{3t} & e^{-t} + e^{3t} \end{pmatrix}. \end{aligned}$$

如果矩阵 A 的最小多项式有重根, 则 Lagrange 插值公式称为 Lagrange-Sylvester 插值公式, 我们仅将其列出, 有兴趣的读者可以尝试证明之, 见习题 44.

命题 5.4.2 (Lagrange-Sylvester 插值公式) 设 n 阶方阵 A 的最小多项式为

$$m(\lambda) = (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_s)^{k_s},$$

其中 $\sum_{i=1}^{s} k_i = m \le n$. 则

$$f(A) = \sum_{i=1}^{s} \varphi_i(A) \left(a_{i1} I + a_{i2}(A - \lambda_i I) + \cdots + a_{ik_i}(A - \lambda_i I)^{k_i - 1} \right),$$

其中

$$\begin{aligned} &\varphi_i(A) = (A - \lambda_1 I)^{k_1} \cdots (A - \lambda_{i-1} I)^{k_{i-1}}(A - \lambda_{i+1} I)^{k_{i+1}} \cdots (A - \lambda_s I)^{k_s}, \quad 1 \le i \le s, \\ &a_{ij} = \frac{1}{(j-1)!} \frac{\mathrm{d}^{j-1}}{\mathrm{d}\,\lambda^{j-1}} \left((\lambda - \lambda_i)^{k_i} \frac{f(\lambda)}{m(\lambda)} \right) \Bigg|_{\lambda = \lambda_i}, \quad 1 \le i \le s;\ j = 1, 2, \cdots, k_i. \end{aligned}$$

第五节 * 自变量为矩阵的函数的导数及应用

本章前面关于矩阵函数的导数定义只涉及到一个未知量, 本质上是一元函数的导数. 正如本章引言所述, 我们更需要关心多个变量的情形, 比如若 例 5.3.4 中的向量 x 与矩阵 A 中的自变量均为二元的, 如何计算二次型 $x^T A x$ 的导数? 为此, 先考察下面的例子.

例 5.5.1 若 A 是 n 阶实常数矩阵, 试求实二次型 x^TAx 在单位球面 $x^Tx=1$ 上的最大值与最小值.

作 Lagrange 辅助函数 $L=x^TAx+\lambda(1-x^Tx)$, 其梯度$\nabla(L)$ 的第 i 个分量为

$$2\sum_{j=1}^{n}a_{ij}x_j-2\lambda x_i.$$

因此 L 的极大值点需要满足条件 $2\sum\limits_{j=1}^{n}a_{ij}x_j-2\lambda x_i=0, 1\le i\le n$, 此恰好是 $Ax=\lambda x$. 因此, 二次型 x^TAx 在单位球面 $x^Tx=1$ 上的最大值与最小值均在单位特征向量处取得, 并且最大值与最小值就是最大特征值与最小特征值! 由此我们也知道了, 求条件极值的 Lagrange 乘子法中的乘子 λ 在此处恰好就是特征值.

上面例子中函数 $L=f(x)=f(x_1,x_2,\cdots,x_n)$ 的梯度

$$\nabla(f)=(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n}) \tag{5.5.1}$$

给出了一个多元函数的“向量值”导数, 即若 $f(x)=f(x_1,x_2,\cdots,x_n)$ 是一个 n 元函数, 其梯度可以看作是其导数 $\frac{\partial f}{\partial x}$, 于是我们引入以下定义.

定义 5.5.1 设 f(或 $f(x)=f(x_1,x_2,\cdots,x_n)$) 是 n 元 (实或复) 可微函数. 称 f 的梯度向量 $\nabla(f)$ 为其**导数向量**即下面的行函数向量

$$\frac{\partial f}{\partial x}=(\frac{\partial f}{\partial x_1},\frac{\partial f}{\partial x_2},\cdots,\frac{\partial f}{\partial x_n}) \tag{5.5.2}$$

在不致混淆的情况下, 导数向量简称为导数.

更一般地, 我们需要定义多元函数向量 (即映射) 或函数矩阵的导数. 为简单起见, 我们假定所涉及的函数的定义域是整个空间 (实际上只要是一个适当的子集即可) 且都是可微的.

定义 5.5.2 设 $f=(f_1,f_2,\cdots,f_m)^T:\mathbb{F}^n\longrightarrow\mathbb{F}^m$ 是一个 n 元可微映射 (写成列向量), 即每个 n 元函数 $f_i, 1\le i\le m$ 均为可微函数, 则映射 f 的导数定义为如下的 $m\times n$ 矩阵

$$\frac{\partial(f_1,f_2,\cdots,f_m)}{\partial(x_1,x_2,\cdots,x_n)}=\begin{pmatrix}\frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n}\\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n}\end{pmatrix} \tag{5.5.3}$$

矩阵 $\frac{\partial(f_1,f_2,\cdots,f_m)}{\partial(x_1,x_2,\cdots,x_n)}$ 称为映射 (或函数向量) f 的 Jacobian 矩阵, 常简记为 $J(f)$. 如果 $J(f)$ 是方阵, 则其行列式称为 Jacobian 行列式.

注意, 如果映射 f 是行向量 $(f_1,f_2,\cdots,f_m)$, 则定义其导数为公式 (5.5.3) 中矩阵的转置. (由于一部分人将梯度向量看成是列向量, 另一部分人则将其看成是行向量, 因此用行向量记号还是用列向量记号常常导致混乱, 务必特别仔细. 沿用高等数学的习惯, 我们将梯度向量看成是行向量.) 今后, 我们将把任何函数或映射 f 的导数记为 $J(f)$, 而 $\nabla(f)$ 将意味着 f 是一个多元函数.

例 5.5.2 设 x 是 n 元 (未知) 列向量, A 是 $m\times n$ 阶常数矩阵, 则
(1)$J(x)=\frac{\partial x}{\partial x}=I_n=\frac{\partial x^T}{\partial x}=J(x^T)$;
(2)$J(Ax)=\frac{\partial Ax}{\partial x}=A,\ \ J(x^TA^T)=\frac{\partial x^TA^T}{\partial x}=A^T$.

例 5.5.3 设 $f:(x,y)^T\mapsto(a_1x+b_1y,a_2x+b_2y)^T$ 是 $\mathbb{F}^2$ 上的线性变换, 则其导数为

$$J(f)=\begin{pmatrix}a_1 & b_1\\ a_2 & b_2\end{pmatrix}.$$

这恰好是线性变换 f 在标准基下的矩阵! 一般地, 如果 $f=(f_1,f_2,\cdots,f_m)^T:\mathbb{F}^n\longrightarrow\mathbb{F}^m$ 是一个线性变换, 设 f 在标准基 - 标准基下的矩阵为 $A_{m\times n}$, 则 f 的导数 (矩阵) 正是 A 本身! (证明见习题 54.)

例 5.5.4 设 $f(x)=f(x_1,x_2,\cdots,x_n)$ 是 n 元二次可微函数, 则其导数 (= 梯度)$J(f)=\frac{\partial f}{\partial x}$ 的导数 $J(J(f))$ 称为 f 的 Hessian [53] 矩阵, 记为 $H(f)$, 即有

$$H(f)=J(J(f))=\begin{pmatrix}\frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2\partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n\partial x_1}\\ \frac{\partial^2 f}{\partial x_1\partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n\partial x_2}\\ \cdots & \cdots & \cdots & \cdots\\ \frac{\partial^2 f}{\partial x_1\partial x_n} & \frac{\partial^2 f}{\partial x_2\partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2}\end{pmatrix} \tag{5.5.4}$$

由多元函数的微分学可知, 如果 f 是二次连续可微的, 则其 Hessian 矩阵 $H(f)$ 是对称矩阵. 函数 $f(x)$ 的 Hessian 矩阵 $H(f)$ 可以看作是 $f(x)$ 的二阶导数.

例 5.5.5 设 $f(x,y)=x^2+2xy+y^2-x$, 则其导数为 $J(f)=(2x+2y-1,2x+2y)$, 其 Hessian 矩阵为

$$H(f)=\begin{pmatrix}2 & 2\\ 2 & 2\end{pmatrix}.$$

例 5.5.6 设 $f=f(x,y)=(x^2+2xy+y^2-x,x^2+2xy+y^2+y)^T$ 是二元多项式映射 (即 f 的两个分量 f_1,f_2 均为二元多项式), 则其 Jacobian 矩阵 $J(f)$ 为

$$J(f)=\begin{pmatrix}2x+2y-1 & 2x+2y\\ 2x+2y & 2x+2y+1\end{pmatrix},$$

易知其 Jacobian 行列式 $|J(f)|=1$. 数学上未解决的著名猜想 – **Jacobian 猜想**(1939 年) 是说如果一个 $\mathbb{C}^n$ 到其自身的多项式映射 f 的 Jacobian 行列式 $|J(f)|=1$, 则 f 一定有逆 (必定也是多项式映射). 线性变换是最简单的多项式映射, Jacobian 猜想对线性映射成立 (见习题 48). 目前二次多项式映射的 Jacobian 猜想已由 S.Wang [54] 于 1980 年所证明. 比如, 本题中的映射是可逆的, 见习题 48.

[53] Ludwig Otto Hesse(1811-1874), 德国数学家, Carl Gustav Jacob Jacobi 的学生.
[54] Stuart Sui-Sheng Wang(王穗生, 1946-), 华裔美籍数学家, 现美国 Oakland 大学教授.

多元函数的导数 (向量) 的四则运算法则与普通函数的导数法则基本类似 (注意区分行向量与列向量), 我们列出其中如下部分.

命题 5.5.1 设 f,g 均为 n 元向量 x 的可微函数, $h(x)$ 是 n 元行映射, $p(x),q(x)$ 是 m 维行映射, $a,b\in\mathbb{F}$, 则

(1) (线性法则)$J(af+bg)=aJ(f)+bJ(g)$;

(2) (乘法公式)$J(pq^T)=pJ(q)+qJ(p)$;

(3) (除法公式)$J(f/g)=(gJ(f)-fJ(g))/g^2$;

(4) (链法则) $J(f(h))=\frac{\partial f(h)}{\partial h}J(h^T)$.

证 我们只证明链法则 (4). 记 $x=(x_1,x_2,\cdots,x_n)^T,h=(h_1,h_2,\cdots,h_n)$. 则由多元函数的链法则可知

$$\frac{\partial f(h(x))}{\partial x}=\nabla(f(h_1,h_2,\cdots,h_n))=(\sum_{i=1}^{n}\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_1},\sum_{i=1}^{n}\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_2},\cdots,\sum_{i=1}^{n}\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_n}).$$

另一方面, 由

$$\frac{\partial f(h)}{\partial h}\frac{\partial h^T}{\partial x}=(\frac{\partial f}{\partial h_1},\frac{\partial f}{\partial h_2},\cdots,\frac{\partial f}{\partial h_n})J(h^T)=(\sum_{i=1}^{n}\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_1},\sum_{i=1}^{n}\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_2},\cdots,\sum_{i=1}^{n}\frac{\partial f}{\partial h_i}\frac{\partial h_i}{\partial x_n}).$$

故链法则成立. □

注. 通常我们将多元复合函数 $f(h)$ 中的中间变量 h 写成行向量的形式即 $h=(h_1,h_2,\cdots,h_n)$, 因此要求链法则中的第二个因子是 h^T 的导数而不是 h 的导数.

例 5.5.7 设 $f=f(u,v)$ 是二元函数, $g(x,y)=(x^2,ax+y^2)$ 为二元映射. 则 $f(g)=f(x^2,ax+y^2)$, 由链法则可得

$$J(f(g))=\frac{\partial f(g)}{\partial(x,y)}=J(f)\frac{\partial g^T}{\partial(x,y)}=(\frac{\partial f}{\partial u},\frac{\partial f}{\partial v})\begin{pmatrix}2x&0\\a&2y\end{pmatrix}=(2x\frac{\partial f}{\partial u}+a\frac{\partial f}{\partial v},2y\frac{\partial f}{\partial v}),$$

此与多元微分学的结果一致.

注. 如果 $f=(f_1,f_2,\cdots,f_m)^T$ 与 $g=(g_1,g_2,\cdots,g_m)^T$ 是两个 m 维列映射, 则乘法公式 $J(f^Tg)$ 的表达应该与 命题 5.5.1(2) 中的乘法公式略有差别但本质一致, 见习题 49.

现在可以得到求二次型 x^Tx 的导数的正确公式了 (请对照直接计算的结果, 更一般的结果见习题 50):

$$\frac{\partial x^Tx}{\partial x}=x^T\frac{\partial x}{\partial x}+x^T\frac{\partial x^T}{\partial x}=2x^T \tag{5.5.5}$$

再来讨论 例 5.5.1 中的条件极值问题. 直接对 Lagrange 辅助函数 L 关于向量 x 求导可得

$$J(L)=J(x^TAx)+J(\lambda(1-x^Tx))=x^TJ(Ax)+(Ax)^TJ(x^T)-\lambda J(x^Tx)=2x^TA-2\lambda x^T,$$

因此其极值点必须满足条件 $x^TA=\lambda x^T$, 即 x^T 是 A 的左特征向量或 x 是 A^T 的右特征向量.

总结以上的讨论可知, 如果矩阵函数是一元函数, 则其导数运算法则与一元函数的导数运算法则完全一致, 而若矩阵函数是多元函数, 则其导数本质上是对自变量向量求导 (注意我们始终是对列向量 x 求导, 而没有讨论对行向量 x^T 求导), 因此其运算法则与函数的导数运算法则有非常大的差别.

例 5.5.8 (坐标变换) 本题之前我们一直在对向量 $x=(x_1,x_2,\cdots,x_n)^T$ 求导, 假如我们对向量 $y=(x_1,x_1+x_2,\cdots,x_1+x_n)^T$ 求导会得到什么结果? 这相当于对复合函数 $f(g)$ 求导, 其中 f 是普通的 n 元函数, 而 $y=g(x)=(x_1,x_1+x_2,\cdots,x_1+x_n)$ 是 n 元映射 (相当于一个坐标变换). 具体计算见习题 51.

我们知道, 和一个方阵关联的最重要的两个数值是其行列式和迹. 这两个数值实际上都是从 $\mathbb{F}^{n\times n}$ 到 $\mathbb{F}$ 的多元 (n^2 元) 函数, 我们现在研究它们的导数. $\mathbb{F}^{n\times n}$ 中的一个未知矩阵可以写成 $X=(x_{ij})$, 我们将其行列式 $|X|$ 改记为 $D(X)$, 而其迹仍记为 $\mathrm{tr}(X)$ 或 $\mathrm{tr}X$, 称为迹函数. 由于 $D(X)$ 与 $\mathrm{tr}(X)$ 关于诸变量 x_{ij} 的导数 $\nabla(D(X))=J(D(X))$ 与 $\mathrm{tr}(X)$ 均为 n^2 维的行向量, 不便于表示, 因此我们将它们都排列成矩阵的形式如下

$$J(D(X))=\frac{\partial D(X)}{\partial X}=\begin{pmatrix} \frac{\partial D(X)}{\partial x_{11}} & \frac{\partial D(X)}{\partial x_{12}} & \cdots & \frac{\partial D(X)}{\partial x_{1n}} \\ \frac{\partial D(X)}{\partial x_{21}} & \frac{\partial D(X)}{\partial x_{22}} & \cdots & \frac{\partial D(X)}{\partial x_{2n}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial D(X)}{\partial x_{n1}} & \frac{\partial D(X)}{\partial x_{n2}} & \cdots & \frac{\partial D(X)}{\partial x_{nn}} \end{pmatrix} \tag{5.5.6}$$

与

$$J(\mathrm{tr}(X))=\frac{\partial \mathrm{tr}(X)}{\partial X}==\begin{pmatrix} \frac{\partial \mathrm{tr}(X)}{\partial x_{11}} & \frac{\partial \mathrm{tr}(X)}{\partial x_{12}} & \cdots & \frac{\partial \mathrm{tr}(X)}{\partial x_{1n}} \\ \frac{\partial \mathrm{tr}(X)}{\partial x_{21}} & \frac{\partial \mathrm{tr}(X)}{\partial x_{22}} & \cdots & \frac{\partial \mathrm{tr}(X)}{\partial x_{2n}} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial \mathrm{tr}(X)}{\partial x_{n1}} & \frac{\partial \mathrm{tr}(X)}{\partial x_{n2}} & \cdots & \frac{\partial \mathrm{tr}(X)}{\partial x_{nn}} \end{pmatrix} \tag{5.5.7}$$

一般将式 (5.5.6) 与式 (5.5.7) 分别称为行列式与迹函数的梯度矩阵. 下面的命题表明, 这两个貌似复杂的矩阵实际上很简单 (证明见习题 52).

命题 5.5.2 设 X 是 n 阶矩阵, 则其行列式的梯度矩阵是其伴随矩阵的转置, 而迹函数的导数是单位矩阵, 即

$$J(D(X))=(\mathrm{adj}X)^T, \qquad\qquad J(\mathrm{tr}(X))=I_n \tag{5.5.8}$$

例 5.5.9 设 $X=\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}$, 则 $D(X)=x_{11}x_{22}-x_{21}x_{12}$, 因此

$$J(D(X))=\begin{pmatrix} x_{22} & -x_{21} \\ -x_{12} & x_{11} \end{pmatrix}=(\mathrm{adj}X)^T.$$

以上对矩阵的迹函数 $\mathrm{tr}(\cdot)$ 的考察相对简单, 更有趣的问题是研究矩阵 AXB(及类似矩阵) 的迹, 其中 A,B 为给定的矩阵, 即求 $J(\mathrm{tr}(AXB))=\frac{\partial \mathrm{tr}(AXB)}{\partial X}$, 见习题 53.

例 5.5.10 设 $x = r\sin\theta\cos\phi, y = r\sin\theta\sin\phi, z = r\cos\theta$. 则

$$J = \frac{\partial(x,y,z)}{\partial(r,\theta,\phi)} = \begin{pmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ r\cos\theta\cos\phi & r\cos\theta\sin\phi & -r\sin\theta \\ -r\sin\theta\sin\phi & r\sin\theta\cos\phi & 0 \end{pmatrix},$$

因此 $|J| = r^2\sin\theta$ 正是球面坐标下的积分系数.

设 n 元函数 $f(x) = f(x_1, x_2, \cdots, x_n)$ 的导数为 $J(f)$, 我们已经看到 $f(x)$ 的二阶导数即其 Hessian 矩阵 $H(f)$. 自然应该将矩阵函数 $H(f)$ 关于向量 $(x_1, x_2, \cdots, x_n)$ 的导数定义为函数 $f(x)$ 关于向量 $(x_1, x_2, \cdots, x_n)$ 的三阶导数. 那么, 我们应该如何定义 n 阶矩阵函数 $H(f)$ 关于向量 $(x_1, x_2, \cdots, x_n)$ 的导数呢? 更一般地, 设 $f(X)$ 是 $p\times q$ 阶矩阵 X 的 $m\times n$ 阶矩阵函数, 如何定义 $f(X)$ 关于 X 的导数呢? 仿照向量函数对向量的导数, 一个可行的办法是将矩阵看作向量, 即利用将矩阵的向量化变换 vec, 然后按照向量函数对向量的导数计算导数 (矩阵) 即可.

定义 5.5.3 设 $f : \mathbb{F}^{p\times q} \longrightarrow \mathbb{F}^{m\times n}$ 是一个 $p\times q$ 元矩阵函数, 即

$$f(X) = \begin{pmatrix} f_{11}(X) & f_{12}(X) & \cdots & f_{1n}(X) \\ f_{21}(X) & f_{22}(X) & \cdots & f_{2n}(X) \\ \cdots & \cdots & \cdots & \cdots \\ f_{m1}(X) & f_{m2}(X) & \cdots & f_{mn}(X) \end{pmatrix},$$

其中 $X = (x_{ij})_{p\times q}$ 是自变量矩阵, 则 $f(X)$ 的导数定义为 $mp\times nq$ 阶矩阵

$$J(f) = \frac{\partial f(X)}{\partial X} = \frac{\partial \mathrm{vec}(f(X))}{\partial \mathrm{vec}X} \tag{5.5.9}$$

例 5.5.11 设 $f : \mathbb{F}^{2\times 2} \longrightarrow \mathbb{F}^{2\times 2}$ 是 2×2 元恒等矩阵函数, 即 $f(X) = X$, 则

$$J(f) = \frac{\partial \mathrm{vec}(f(X))}{\partial \mathrm{vec}X} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = I_4!$$

这是自然的, 因为我们总是将自变 (向) 量 x 视为列向量 (即使 n 元函数 $f(x_1, x_2, \cdots, x_n)$ 中的自变量写成了行的形式, 我们依然认为 $f(x)$ 是列向量 $x = (x_1, x_2, \cdots, x_n)^T$ 的函数, 这等于将函数 $f(x)$ 看作是从 $\mathbb{R}^n$ 到 $\mathbb{R}$ 的映射), 而 $\partial\mathrm{vec}X$ 规定自变量是按照列顺序排列的.

下面的命题阐述矩阵函数 (我们没有区分矩阵函数与函数矩阵这两个概念) 的导数的一些性质, 证明见习题 53.

命题 5.5.3 (1) 设 $f = f(x) = (f_{ij}(x))_{m\times p}$ 与 $g = g(x) = (g_{ij}(x))_{p\times q}$ 是两个 n 元函数矩阵, 则

$$J(f(x)g(x)) = (g(x)^T \otimes I_m)J(f(x)) + (I_q \otimes f(x))J(g(x)) \tag{5.5.10}$$

特别地,

$$J(xx^T) = x\otimes I_n + I_n \otimes x \tag{5.5.11}$$

(2) 设 $A\in\mathbb{C}^{m\times n}, B\in\mathbb{C}^{p\times q}, X\in\mathbb{C}^{n\times p}$, 则

$$J(AXB)=\frac{\partial AXB}{\partial X}=B^T\otimes A; J((AXB)^T)=\frac{\partial B^TX^TA^T}{\partial X}=B\otimes A^T=(J(AXB))^T.$$

(3) 设 $A\in\mathbb{C}^{m\times n}, X\in\mathbb{C}^{n\times m}$, 则

$$\frac{\partial \mathrm{tr}(AX)}{\partial X}=\frac{\partial \mathrm{tr}(XA)}{\partial X}=A^T.$$

(4) 设 $A\in\mathbb{C}^{n\times n}, X\in\mathbb{C}^{m\times n}$, 则

$$\frac{\partial \mathrm{tr}(XAX^T)}{\partial X}=X(A+A^T).$$

(5) 设 $\alpha,\beta\in\mathbb{C}^{n\times 1}, X\in\mathbb{C}^{m\times n}$, 则

$$\frac{\partial(\alpha^TX^TX\beta)}{\partial X}=X(\alpha\beta^T+\beta\alpha^T).$$

(6) 设 $\alpha,\beta\in\mathbb{C}^{m\times 1}, X\in\mathbb{C}^{m\times n}$, 则

$$\frac{\partial(\alpha^TXX^T\beta)}{\partial X}=(\alpha\beta^T+\beta\alpha^T)X.$$

(7) 设 $X\in\mathbb{C}^{m\times n}, f=f(X)=(f_{ij}(X))_{p\times q}$ 与 $g=g(X)=(g_{ij}(X))_{m\times n}$ 是两个函数矩阵, 则

$$\frac{\partial(f(g(X)))}{\partial X}=\frac{\partial(f(g))}{\partial g}\frac{\partial(g)}{\partial X}.$$

利用矩阵函数的微分学求解矩阵函数的极值问题, 有下述与多元微分学类似的两个结论 (证明略去, 读者可参考 [16] 第五章):

定理 5.5.1 (无约束极值的必要条件) 设 $X\in\mathbb{R}^{m\times n}, f(X)$ 是关于矩阵变量 X 的二阶连续可导函数.

(1) 若 X_0 是 $f(X)$ 的一个极值点, 则 $J(f)(X_0)=\frac{\partial f(X)}{\partial X}|_{X=X_0}=0$; 换句话说, $f(X)$ 的极值点均满足方程

$$J(f)=0\text{ 或 }\frac{\partial f(X)}{\partial X}=0.$$

(2) 若 $f(X)$ 还是关于矩阵变量 X 的凸函数, 则 $f(X)$ 的极值点必是最值点.

定理 5.5.2 (约束极值的必要条件) 设 $X\in\mathbb{R}^{m\times n}, f(X)$ 是关于矩阵变量 X 的二阶连续可导函数. 设 X_0 是 $f(X)$ 在约束条件 $c(X)=0$ 下的一个条件极值点, $L(X,\lambda)=f(X)-\lambda c(X)$ 是 Lagrange 辅助函数. 则存在数 $\lambda_0\in\mathbb{R}$, 使得

$$J(L)_X(X_0,\lambda_0)=\frac{\partial L(X,\lambda)}{\partial X}|_{(X_0,\lambda_0)}=0, c(X_0)=0.$$

我们现在解决本章引言 例 5.0.1 中的问题, 其背景如下.

考虑 K 个用户的移动通信的 CDMA 系统, 设第 k 个用户的信号幅值为 A_k, 特定时刻发射的比特信号为 $b_k\,(=\pm 1)$. 则描述该系统的一个简化数学模型为:

$$y = RAb + n \tag{5.5.12}$$

其中 y 是通信基站的接收信号向量, K 阶对称矩阵 R 是 (用户扩频波形的) 自相关矩阵, $b=(b_1,\cdots,b_k)^T$, K 阶列向量 n 是方差为 σ^2 的噪声向量 (一般假定 n 是加性白 Gauss 噪声且与用户信号不相关). 问题: 如何设计第 k 个用户的检测器 m_k(即使用 $\hat{b}_k = \mathrm{sgn}(m_k^T y)$ 作为第 k 个用户发射信号的估计), 使得所有 K 个用户的检测器

$$M = (m_1, \cdots, m_K) \tag{5.5.13}$$

具有**最小均方误差**? 即使得以矩阵 M 为自变量的目标函数

$$J(M) = E\{\|b - My\|_2^2\} \tag{5.5.14}$$

最小化, 其中 $\|\bullet\|$ 是向量的欧几里得范数.

利用数学期望与矩阵的迹的性质可以证明 (此处略去) 目标函数 $J(M)$ 具有下述表达

$$J(M) = \mathrm{tr}(I) + \mathrm{tr}(M(RA^2R + \sigma^2 R)M^T) - \mathrm{tr}(ARM^T) - \mathrm{tr}(MRA) \tag{5.5.15}$$

由 定理 5.5.1, 矩阵函数 $J(M)$ 的最小值点 M 必须满足方程

$$\frac{\partial J(M)}{\partial M} = 0 \tag{5.5.16}$$

由 命题 5.5.3 可知 (细节见习题 46, 注意 R 是对称矩阵)

$$\frac{\partial J(M)}{\partial M} = 2M(RA^2R + \sigma^2 R) - 2AR \tag{5.5.17}$$

于是方程 (5.5.16) 的解 M 必须满足方程

$$M(RA^2R + \sigma^2 R) = AR \tag{5.5.18}$$

特别, 若 R 可逆, 则可得下述唯一的最优盲多用户检测器

$$M = A(RA^2 + \sigma^2 I)^{-1}.$$

上述方法称为**盲多用户检测**(英文缩写为 BMUD), 是以矩阵为自变量的函数的微分学的前沿应用之一.

思考题

1. 设 $\alpha \in \mathbb{C}^{m\times 1}, \beta \in \mathbb{C}^{n\times 1}$, 则 $J(\alpha^T X \beta) = \frac{\partial \alpha^T X \beta}{\partial X} =?$
2. 设 $\alpha \in \mathbb{C}^{m\times 1}, \beta \in \mathbb{C}^{n\times 1}$, 则 $J(\alpha^T X^T \beta) = \frac{\partial \alpha^T X^T \beta}{\partial X} =?$
3. 设 X 是方阵, $J(X^2) =?$
4. 如果定义 $J(X) = \frac{\partial X}{\partial \mathrm{rvec} X}$, 将得到何种结果? 是否还有其它方式?
5. 试比较隐函数存在定理与 Jacobian 猜想.

第六节　应用 I：线性常微分方程

本节我们介绍矩阵函数的微积分在现代控制理论中的应用.

设有一个多输入 - 多输出连续时间系统, 如下图所示:

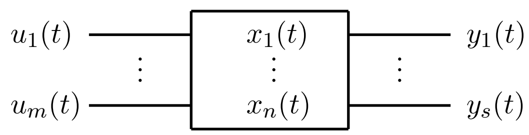

图 5.6.1

其中 $u_1(t), \cdots, u_m(t)$ 是 m 个输入, $y_1(t), \cdots, y_s(t)$ 是 s 个输出, $x_1(t), \cdots, x_n(t)$ 是系统的 n 个状态变量. 记

$$x = x(t) = (x_1(t), \cdots, x_n(t))^T, u = u(t) = (u_1(t), \cdots, u_m(t))^T, y = y(t) = (y_1(t), \cdots, y_s(t))^T,$$

分别称为系统的状态向量, 输入向量 (或控制向量), 输出向量(或观测向量). 描述该系统的最简单的数学模型为下面的**状态方程**

$$\begin{cases} x'(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \tag{5.6.1}$$

其中 $A \in \mathbb{C}^{n\times n}$ 称为系数矩阵或**系统矩阵**, $B \in \mathbb{C}^{n\times m}$ 为输入矩阵 (或**控制矩阵**), $C \in \mathbb{C}^{s\times n}$ 为输出矩阵 (或**观测矩阵**), $D \in \mathbb{C}^{s\times m}$ 为联系矩阵. 由于联系矩阵 D 只涉及输入和输出信号的静态传递, 而和系统状态变量无关, 故常令 $D = 0$. 我们仅限于讨论 A, B, C 均为常数矩阵的系统, 这样的系统称为**定常线性系统**, 一般简称为系统 (A, B, C). 方程组 (5.6.1) 中的第一个方程称为系统的状态方程, 第二个方程称为系统的**输出方程**.

由线性微分方程理论可知, 根据系统在某一时刻 t_0 的状态, 即可描述整个系统在任何时刻 $t \geq t_0$ 的状态. 换句话说, 如果给定初始条件 $x(t_0) = (x_1(t_0), \cdots, x_n(t_0))$, 则方程组 (5.6.1) 的解完全确定. 此称为定解问题. 定解问题的解决过程一般分为两个步骤, 即先求方程组 (5.6.1) 对应的齐次方程组的解, 再利用常数变易法或其他办法求方程组 (5.6.1) 的一个特解.

例 5.6.1 当 $n = 1$ 时, 定解问题涉及的齐次微分方程就是普通的线性常微分方程 $x'(t) = ax, x(t)|_{t=t_0} = x(t_0)$. 利用分离变量法直接积分即可得该定解问题的解为 $x = x(t_0)e^{a(t-t_0)}$. 此时相应的非齐次微分方程为 $x'(t) = ax + f(t), x(t)|_{t=t_0} = x(t_0)$, 其中 $f(t)$ 为已知函数. 利用熟知的常数变易法技巧可求得其解为

$$x = x(t_0)e^{a(t-t_0)} + \int_{t_0}^{t} e^{a(t-s)} f(s)\,\mathrm{d}\,s.$$

一般地, 定解问题涉及的齐次方程组为

$$\begin{cases} x'(t) = Ax \\ x(t)|_{t=t_0} = x(t_0), \end{cases} \tag{5.6.2}$$

其中 $x = x(t) = (x_1(t), x_2(t), \cdots, x_n(t))^T$ 为未知向量函数, $x(t_0) = (x_1(t_0), x_2(t_0), \cdots, x_n(t_0))^T$ 为给定的 n 维向量, A 为给定的 n 阶常数方阵.

当 $n \geq 2$ 时, 由于无法分离变量, 定解问题 (5.6.2) 的解不能由 例 5.6.1 给出的方法求出, 但可利用 Taylor 级数求解, 其框架如下:

由 $x'(t) = Ax$ 可知

$$x''(t) = A^2 x, \cdots, x^{(k)}(t) = A^k x, \cdots \tag{5.6.3}$$

故

$$x'(t_0)=Ax(t_0),x''(t_0)=A^2x(t_0),\cdots,x^{(k)}(t_0)=A^kx(t_0),\cdots \tag{5.6.4}$$

将 $x(t)$ 在 $t=t_0$ 处展开成 Taylor 级数可得

$$\begin{aligned} x &= x(t_0)+(t-t_0)x'(t_0)+\tfrac{(t-t_0)^2}{2!}x''(t_0)+\cdots+\tfrac{(t-t_0)^k}{k!}x^{(k)}(t_0)+\cdots \\ &= x(t_0)+(t-t_0)Ax(t_0)+\tfrac{(t-t_0)^2}{2!}A^2x(t_0)+\cdots+\tfrac{(t-t_0)^k}{k!}A^kx(t_0)+\cdots \\ &= e^{A(t-t_0)}x(t_0) \end{aligned} \tag{5.6.5}$$

由公式 (5.6.5) 给出的解还是唯一的 (见习题 55), 故有下述定理

定理 5.6.1 定解问题 (5.6.2) 有唯一解, 且其解为

$$x(t)=e^{A(t-t_0)}x(t_0).$$

例 5.6.2 求常系数线性齐次微分方程组

$$\begin{cases} x'(t)=Ax(t), \\ x(t)|_{t=0}=x(0), \end{cases}$$

的解, 其中

$$A=\begin{pmatrix} 2 & 2 & -1 \\ -1 & -1 & 1 \\ -1 & -2 & 2 \end{pmatrix},\ x(0)=\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

解 由 定理 5.6.1, 方程组的解为 $x(t)=e^{At}x(0)$. 下面求 e^{At}. 直接计算可得 A 的 Jordan 标准形及变换矩阵分别为

$$J=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},\quad P=\begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 1 & -1 & 0 \end{pmatrix}.$$

从而,

$$\begin{aligned} x(t) &= e^{At}x(0)=Pe^{Jt}P^{-1}x(0) \\ &= \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \\ 1 & -1 & 0 \end{pmatrix}\begin{pmatrix} e^t & 0 & 0 \\ 0 & e^t & te^t \\ 0 & 0 & e^t \end{pmatrix}\begin{pmatrix} 0 & -1 & 1 \\ 0 & -1 & 0 \\ 1 & 2 & -1 \end{pmatrix}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 2t+1 \\ -2t+1 \\ -2t+1 \end{pmatrix}e^t. \end{aligned}$$

设 $A=(a_{ij})_{n\times n}$ 与 $B=(b_{ij})_{n\times m}$ 是常数矩阵, $u(t)=(u_1(t),u_2(t),...,u_m(t))^T$ 是已知向量函数, $x(t)=(x_1(t),x_2(t),...,x_n(t))^T$ 是未知向量函数. 下面考虑定解问题所涉及的线性常系数非齐次微分方程组

$$\begin{cases} x'(t)=Ax(t)+Bu(t) \\ x(t)|_{t=t_0}=x(t_0) \end{cases} \tag{5.6.6}$$

由线性微分方程组的理论可知, 方程组 (5.6.6) 的通解是其相应的齐次方程组的通解加上它自身的一个特解, 而该特解可以通过 Lagrange 常数变易法求得. 此处介绍另一个方法即积分因子法, 即给方程组 (5.6.6) 的两端同乘以 e^{-At}, 得

$$e^{-At}\frac{\mathrm{d}x(t)}{\mathrm{d}t}-e^{-At}Ax(t)=e^{-At}Bu(t). \tag{5.6.7}$$

注意上式的左端为 $(e^{-At}x(t))'$, 因而有下述定理 (细节见习题 58)

定理 5.6.2 方程组 (5.6.6) 有唯一解, 且其解为

$$x(t)=e^{A(t-t_0)}x(t_0)+\int_{t_0}^{t}e^{A(t-s)}Bu(s)\,\mathrm{d}s.$$

注. 积分因子法等价于作变量代换 $x=e^{At}y$ 或 $y=e^{-At}x$.

例 5.6.3 求定解问题

$$\begin{cases} x'(t)=Ax(t)+u(t)\\ x(t)|_{t=0}=(1,1,1)^T \end{cases}$$

的解, 其中

$$A=\begin{pmatrix} 3 & -1 & 1\\ 2 & 0 & -1\\ 1 & -1 & 2 \end{pmatrix},\ u(t)=(0,0,e^{2t})^T.$$

解 由 定理 5.6.2, 方程组的解为

$$x(t)=e^{At}x(0)+\int_0^t e^{A(t-s)}u(s)\,\mathrm{d}s.$$

首先求 e^{At}. 由于

$$|\lambda I-A|=\begin{vmatrix} \lambda-3 & 1 & -1\\ -2 & \lambda & 1\\ -1 & 1 & \lambda-2 \end{vmatrix}=\lambda(\lambda-2)(\lambda-3).$$

故 A 有 3 个不同的特征值, 因此 A 与对角矩阵相似. 与特征值 $\lambda_1=0$, $\lambda_2=2$, $\lambda_3=3$ 相应的三个线性无关的特征向量为

$$\alpha_1=(1,5,2)^T,\quad \alpha_2=(1,1,0)^T,\quad \alpha_3=(2,1,1)^T.$$

于是

$$P=\begin{pmatrix} 1 & 1 & 2\\ 5 & 1 & 1\\ 2 & 0 & 1 \end{pmatrix},\quad P^{-1}=-\frac{1}{6}\begin{pmatrix} 1 & -1 & -1\\ -3 & -3 & 9\\ -2 & 2 & -4 \end{pmatrix}.$$

因而

$$
\begin{aligned}
e^{At}x(0) &= P\begin{pmatrix} 1 & & \\ & e^{2t} & \\ & & e^{3t}\end{pmatrix}P^{-1}x(0)\\
&= \begin{pmatrix} 1 & 1 & 2\\ 5 & 1 & 1\\ 2 & 0 & 1\end{pmatrix}\begin{pmatrix} 1 & & \\ & e^{2t} & \\ & & e^{3t}\end{pmatrix}\cdot(-\frac{1}{6})\begin{pmatrix} 1 & -1 & -1\\ -3 & -3 & 9\\ -2 & 2 & -4\end{pmatrix}\begin{pmatrix} 1\\ 1\\ 1\end{pmatrix}\\
&= -\frac{1}{6}\begin{pmatrix} -1+3e^{2t}-8e^{3t}\\ -5+3e^{2t}-4e^{3t}\\ -2-4e^{3t}\end{pmatrix}.
\end{aligned}
$$

下面计算积分 $\int_0^t e^{A(t-s)}u(s)\,\mathrm{d}\,s$. 直接计算得

$$
e^{A(t-s)}u(s) = -\frac{1}{6}\begin{pmatrix} -e^{2s}+9e^{2t}-8e^{3t-s}\\ -5e^{2s}+9e^{2t}-4e^{3t-s}\\ -2e^{2s}-4e^{3t-s}\end{pmatrix}.
$$

$$
\int_0^t e^{A(t-s)}u(s)\,\mathrm{d}\,s = -\frac{1}{6}\begin{pmatrix} \frac{1}{2}+(9t+\frac{15}{2})e^{2t}-8e^{3t}\\ \frac{5}{2}+(9t+\frac{3}{2})e^{2t}-4e^{3t}\\ 1+3e^{2t}-4e^{3t}\end{pmatrix}.
$$

因此方程组的解为

$$
x(t) = -\frac{1}{6}\begin{pmatrix} -\frac{1}{2}+(9t+\frac{21}{2})e^{2t}-16e^{3t}\\ -\frac{5}{2}+(9t+\frac{9}{2})e^{2t}-8e^{3t}\\ -1+3e^{2t}-8e^{3t}\end{pmatrix}.
$$

高阶常系数齐次微分方程的定解问题可以转化为线性微分方程组来求解. 考虑下面的定解问题:

$$
\begin{cases} y^{(n)}+a_1y^{(n-1)}+a_2y^{(n-2)}+\cdots+a_ny=0,\\ y^{(i)}(t)|_{t=0}=y_0^{(i)}, \quad 0\le i\le n-1. \end{cases} \tag{5.6.8}
$$

令

$$
\begin{cases} x_1 = y,\\ x_2 = y' = x_1',\\ x_3 = y^{(2)} = x_2',\\ \cdots\cdots\cdots\\ x_n = y^{(n-1)} = x_{n-1}', \end{cases}
$$

则

$$
\begin{cases} x_1' &= & x_2,\\ x_2' &= & x_3,\\ \cdots & \cdots & \cdots\\ x_{n-1}' &= & x_n,\\ x_n' &= & -a_nx_1-a_{n-1}x_2-\cdots-a_1x_n. \end{cases}
$$

令

$$
x(t) = (x_1(t), x_2(t), \cdots, x_n(t))^T,\ x(0) = (x_1(0), x_2(0), \cdots, x_n(0))^T = (y_0, y_0', \cdots, y_0^{(n-1)})^T.
$$

则高阶常微分方程 (5.6.8) 等价于一阶微分方程组

$$\begin{cases} x'(t) = Ax(t) \\ x(t)|_{t=0} = x(0) \end{cases} \tag{5.6.9}$$

其中

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{pmatrix}. \tag{5.6.10}$$

可以看出, A 恰好是方程 (5.6.8) 的特征多项式 $r^n + a_1 r^{n-1} + a_2 r^{n-2} + \cdots + a_n$ 的友矩阵. 由 定理 5.6.1, 方程组 (5.6.9) 的解为 $x(t) = e^{At}x(0)$. 而方程 (5.6.8) 的解是方程组 (5.6.9) 的解的第一个分量, 所以定解问题 (5.6.8) 的解为

$$\begin{aligned} y &= (1, 0, \cdots, 0)x(t) = (1, 0, \cdots, 0)e^{At}x(0) \\ &= (1, 0, \cdots, 0)e^{At}\begin{pmatrix} y_0 \\ y_0' \\ \vdots \\ y_0^{(n-1)} \end{pmatrix}. \end{aligned}$$

相应于方程 (5.6.8) 的 n 阶常系数非齐次线性方程的定解问题:

$$\begin{cases} y^{(n)} + a_1 y^{(n-1)} + a_2 y^{(n-2)} + \cdots + a_n y = u(t), \\ y^{(i)}(t)|_{t=0} = y_0^{(i)}, \qquad 0 \le i \le n-1, \end{cases} \tag{5.6.11}$$

可作相应的讨论. 容易推出定解问题 (5.6.11) 的解是方程组

$$\begin{cases} x'(t) = Ax(t) + Bu(t), \\ x(t)|_{t=0} = x(0), \end{cases}$$

的解的第一个分量, 其中

$$\begin{aligned} &x(t) = (x_1(t), x_2(t), \cdots, x_n(t))^T = (y, y', \cdots, y^{(n-1)})^T, \\ &x(0) = (x_1(0), x_2(0), \cdots, x_n(0))^T = (y_0, y_0', \cdots, y_0^{(n-1)})^T, \\ &A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{pmatrix}, \qquad B = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}. \end{aligned}$$

由 定理 5.6.2 知, 定解问题 (5.6.11) 的解为

$$y(t) = (1, 0, \cdots, 0)\left(e^{At}x(0) + \int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s\right).$$

例 5.6.4 求下述三阶常微分方程的解 $y(t)$:

$$\begin{cases} y^{(3)} - 3y^{(2)} - 6y' + 8y = u(t), \\ (y(0), y'(0), y''(0)) = (1, 0, 1). \end{cases}$$

解 令 $x(t) = (x_1(t), x_2(t), x_3(t))^T$, 其中 $x_1(t) = y(t)$, $x_2(t) = y'(t)$, $x_3(t) = y''(t)$. 则 $x(0) = (1, 0, 1)^T$.

由前面的讨论知,

$$y(t) = (1, 0, 0)\left(e^{At}x(0) + \int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s\right),$$

其中

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -8 & 6 & 3 \end{pmatrix}, \quad B = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

A 的特征多项式为 $|\lambda I - A| = \lambda^3 - 3\lambda^2 - 6\lambda + 8 = (\lambda - 1)(\lambda + 2)(\lambda - 4)$. 所以 A 的 Jordan 标准形为

$$J = \begin{pmatrix} 1 & & \\ & -2 & \\ & & 4 \end{pmatrix}, \quad \text{其变换矩阵为} \quad P = \begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 4 \\ 1 & 4 & 16 \end{pmatrix}.$$

因此

$$\begin{aligned} e^{At} &= P\begin{pmatrix} e^t & & \\ & e^{-2t} & \\ & & e^{4t} \end{pmatrix}P^{-1} \\ &= \begin{pmatrix} 1 & 1 & 1 \\ 1 & -2 & 4 \\ 1 & 4 & 16 \end{pmatrix}\begin{pmatrix} e^t & & \\ & e^{-2t} & \\ & & e^{4t} \end{pmatrix} \cdot \frac{1}{18}\begin{pmatrix} 16 & 4 & -2 \\ 4 & -5 & 1 \\ -2 & 1 & 1 \end{pmatrix} \\ &= \frac{1}{18}\begin{pmatrix} 16e^t + 4e^{-2t} - 2e^{4t} & 4e^t - 5e^{-2t} + e^{4t} & -2e^t + e^{-2t} + e^{4t} \\ 16e^t - 8e^{-2t} - 8e^{4t} & 4e^t + 10e^{-2t} + 4e^{4t} & -2e^t - 2e^{-2t} + 4e^{4t} \\ 16e^t + 16e^{-2t} - 32e^{4t} & 4e^t - 20e^{-2t} + 16e^{4t} & -2e^t + 4e^{-2t} + 16e^{4t} \end{pmatrix}, \end{aligned}$$

$$e^{A(t-s)}Bu(s) = \frac{1}{18}u(s)\begin{pmatrix} -2e^{t-s} + e^{-2(t-s)} + e^{4(t-s)} \\ -2e^{t-s} - 2e^{-2(t-s)} + 4e^{4(t-s)} \\ -2e^{t-s} + 4e^{-2(t-s)} + 16e^{4(t-s)} \end{pmatrix}.$$

所以

$$\begin{aligned} y(t) &= (1, 0, 0)\left(e^{At}x(0) + \int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s\right) \\ &= \frac{1}{18}(14e^t + 5e^{-2t} - e^{4t}) - \frac{1}{9}e^t\int_0^t e^{-s}u(s)\,\mathrm{d}\,s + \frac{1}{18}e^{-2t}\int_0^t e^{2s}u(s)\,\mathrm{d}\,s \\ &\qquad + \frac{2}{9}e^{4t}\int_0^t e^{-4s}u(s)\,\mathrm{d}\,s. \end{aligned}$$

注 1. 如果仅求 n 阶齐次方程 (5.6.8) 的解, 则不必使用本节介绍的办法, 因为转化为矩阵形式仍然需要求特征方程的根, 而由线性微分方程的理论, 一旦知道特征方程的根, 则方

程 (5.6.8) 的通解就已经知道了.

注 2. 实际计算定常线性系统 (5.6.1) 的解的常用方法是 Laplace 变换.

第七节 应用 II：线性系统的可控性与可测性

本节我们利用上节的理论研究定常线性系统 (5.6.1) 的可控性与可观测性.

要控制一个定常系统 (5.6.1), 需要了解系统的状态 $x(t)$. 但在一般情况下, 不能直接测量到它 (比如人造卫星系统或深水探测仪等), 因此必须通过得到的观测向量 $y(t)$ 反过来判断 $x(t)$. 能否通过观测向量 $y(t)$ 确定出系统的全部状态, 这便是所谓的可观测性问题. 掌握了系统的状态后, 能否控制它使其达到预期的目的, 便是所谓的可控制性问题.

定义 5.7.1 对一个定常线性系统, 若在某个有限时间区间 $[0,t_1]$ 内存在着输入 $u(t)\,(0\le t\le t_1)$, 能使系统从初始状态 $x(0)$ 转移到 $x(t)=0$, 则称此状态 $x(0)$ 是可控的. 若任意初始状态 $x(0)$ 都是可控的, 则称此系统是**完全可控**或**完全能控**的.

注. 按照自然的理解, 所谓“完全能控”应该是指, 对任何给定的系统初始状态 $x(0)$, 均能由一定的控制信号 (即输入信号) 使得系统在有限时间内变成任意事先指定的状态 $x(t)$, 而不仅仅是使系统状态变为 0. 实际上, 可以证明这两种定义是等价的 (但显然使用 $x(t)=0$ 更为简便), 见习题 62.

例 5.7.1 设 $n=1$, 则定常系统实际上是一个一阶线性微分方程 $x'(t)=ax(t)+bu(t)$. 显然该系统完全可控 $\iff b\neq 0$.

例 5.7.2 设定常系统的状态方程为

$$x'(t)=\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix}x(t)+\begin{pmatrix}a\\ b\end{pmatrix}u(t),$$

则该系统完全能控 $\iff ab\neq 0$.

定理 5.7.1 系统 (A,B,C) 完全能控 $\iff$ n 阶 Hermite 矩阵

$$W(0,t_1)=\int_0^{t_1} e^{-tA}BB^*e^{-tA^*}\,\mathrm{d}\,t \tag{5.7.1}$$

为非奇异矩阵.

证 充分性: 设 $W(0,t_1)$ 非奇异. 由上一节的 定理 5.6.2 知,

$$x(t_1)=e^{At_1}x(0)+\int_0^{t_1} e^{A(t_1-t)}Bu(t)\,\mathrm{d}\,t. \tag{5.7.2}$$

令

$$u(t)=-B^*e^{-A^*t}W(0,t_1)^{-1}x(0). \tag{5.7.3}$$

将 $u(t)$ 代入 (5.7.2) 得

$$\begin{aligned}x(t_1) &= e^{At_1}x(0) - \textstyle\int_0^{t_1} e^{A(t_1-t)}BB^*e^{-A^*t}W(0,t_1)^{-1}x(0)\,\mathrm{d}t \\ &= e^{At_1}x(0) - e^{At_1}\left(\textstyle\int_0^{t_1} e^{-At}BB^*e^{-A^*t}\,\mathrm{d}t\right)W(0,t_1)^{-1}x(0) \\ &= e^{At_1}x(0) - e^{At_1}W(0,t_1)W(0,t_1)^{-1}x(0) = 0.\end{aligned}$$

这说明在控制输入 $u(t)$ 作用下, 能使系统 $x(0)$ 转移到 $x(t_1)=0$. 由于 $x(0)$ 的任意性, 因此系统是完全能控的.

必要性: 设有向量 $\alpha=(a_1,a_2,\cdots,a_n)^T$ 满足

$$W(0,t_1)\alpha = 0. \tag{5.7.4}$$

则 $\alpha^*W(0,t_1)\alpha=0$. 因而

$$\int_0^{t_1}\alpha^*(e^{-At}BB^*e^{-A^*t})\alpha\,\mathrm{d}t = 0.$$

因此

$$\alpha^*e^{-At}B = 0, \quad 0\le t\le t_1. \tag{5.7.5}$$

由于系统是完全能控的, 所以存在某个输入 $u(t)$ 使得 $x(t_1)=0$. 故由公式 (5.7.2), 有

$$e^{At_1}x(0) + \int_0^{t_1} e^{A(t_1-t)}Bu(t)\,\mathrm{d}t = 0.$$

所以

$$x(0) = -e^{-At_1}\int_0^{t_1} e^{A(t_1-t)}Bu(t)\,\mathrm{d}t = -\int_0^{t_1} e^{-At}Bu(t)\,\mathrm{d}t.$$

因而由等式 (5.7.5),

$$\alpha^*x(0) = -\int_0^{t_1}\alpha^*e^{-At}Bu(t)\,\mathrm{d}t = 0.$$

由 $x(0)$ 的任意性知 $\alpha=0$, 与假设 $\alpha\neq 0$ 矛盾. 故 $W(0,t_1)$ 是非奇异的. □

由于 定理5.7.1 涉及到矩阵函数的积分, 所以在实际应用中常使用下面的简单判别准则.

定理 5.7.2 系统 (A,B,C) 完全能控 $\Longleftrightarrow$ 矩阵

$$W = (B, AB, A^2B, \cdots, A^{n-1}B)$$

的秩为 n. 一般将 $n\times nm$ 阶矩阵 W 称为**能控性矩阵**.

证 系统 (A,B,C) 完全能控 $\Longleftrightarrow$ 对任意 $x(0)$, 关于未知向量 $u(t)$ 的方程 (5.7.2)(其中 $x(t_1)=0$) 有解, 即方程

$$\int_0^{t_1} e^{A(t_1-t)}Bu(t)\,\mathrm{d}t = -e^{At_1}x(0) \tag{5.7.6}$$

对于任意 $x(0)$ 总有解. 因为矩阵 e^{At_1} 总可逆, 故方程 (5.7.6) 可以化简为

$$\int_0^{t_1} e^{-At}Bu(t)\,\mathrm{d}t = -x(0) \tag{5.7.7}$$

由 Cayley-Hamilton 定理可知, n 阶矩阵 e^{-At} 可表示为

$$e^{-At}=\sum_{i=0}^{n-1}a_i(t)A^i,$$

因此方程 (5.7.7) 又可写为

$$\sum_{i=0}^{n-1}(A^iB)\int_0^{t_1}(a_i(t)u(t))\,\mathrm{d}\,t=-x(0) \tag{5.7.8}$$

记

$$z_i(t)^T=\int_0^{t_1}a_i(t)u(t)\,\mathrm{d}\,t,\ 0\le i\le n-1.$$

令 $z=(z_0(t),z_1(t),\cdots,z_{n-1}(t))^T$, 则等式 (5.7.8) 为

$$(B,AB,A^2B,\cdots,A^{n-1}B)z=-x(0) \tag{5.7.9}$$

于是系统 (A,B,C) 完全能控 $\iff$ 方程组 (5.7.9) 的系数矩阵的秩是 n. □

思考. 上面证明中的最后一句话为什么成立? 注意此时向量 z 的维数并不是未知数的个数, 那么, 方程 (5.7.9) 中未知数的个数是多少呢?

由 定理 5.7.2 可知, 如果控制矩阵 B 是行满秩的, 则系统一定是可控的.

为了更好地理解连续型的定常线性系统的可控性 (以及可测性), 以下我们对照离散型的定常线性系统. 类似于连续型的定常系统 (5.6.1), 离散型的定常系统的一个简单数学模型是如下的 (矩阵) 差分方程:

$$\begin{cases} x(k+1)=Ax(k)+Bu(k), & k\ge 0\\ y(k)=Cx(k), & k\ge 0\end{cases} \tag{5.7.10}$$

其中 $x(k),u(k),y(k)$ 分别看作是连续型的定常线性系统的状态变量, 控制信号和观测信号的离散化, 而离散系统 (5.7.10) 中的两个方程可以分别看作是连续型的定常线性系统的状态方程与输出方程的离散化. 离散系统 (5.7.10) 也简记为 (A,B,C). 利用迭代法可知, 方程 (5.7.10) 可以写为

$$x(k)=A^kx(0)+\sum_{i=0}^{k-1}A^{k-1-i}Bu(i),k\ge 1 \tag{5.7.11}$$

特别地,

$$x(n)=A^nx(0)+(B,AB,\cdots,A^{n-1}B)(u(n-1),u(n-2),\cdots,u(0))^T \tag{5.7.12}$$

上式正是 定理 5.7.2 中能控性矩阵的来历. 显然, 离散型的定常系统 (A,B,C) 与连续型的定常系统 (A,B,C) 具有完全相同的可控制性.

例 5.7.3 已知

$$A=\begin{pmatrix}2&-2&1\\1&1&0\\-1&0&-1\end{pmatrix},\quad B=\begin{pmatrix}2&1\\1&0\\-1&-1\end{pmatrix}.$$

试判断系统 (A,B,C) 是否完全能控.

解 由条件知, 系统 (A,B,C) 具有三个状态变量, 两个输入. 下面用 定理 5.7.2 来判断.

$$W=\begin{pmatrix} 2 & 1 & 1 & 1 & -5 & 0 \\ 1 & 0 & 3 & 1 & 4 & 2 \\ -1 & -1 & -1 & -2 & 0 & -1 \end{pmatrix} \to \begin{pmatrix} 0 & 1 & -5 & -1 & -13 & -4 \\ 1 & 0 & 3 & 1 & 4 & 2 \\ 0 & -1 & 2 & -1 & 4 & 1 \end{pmatrix}$$
$$\to \begin{pmatrix} 0 & 1 & -5 & -1 & -13 & -4 \\ 1 & 0 & 3 & 1 & 4 & 2 \\ 0 & 0 & -3 & -2 & -9 & -3 \end{pmatrix}.$$

所以 W 的秩为 3, 由此知系统是完全能控的.

例 5.7.4 如果定常系统是单输入系统, 则控制矩阵 B 为 n 维列向量, 因此系统的能控性矩阵 W 是一个 n 阶方阵. 由 定理 5.7.2, 此时系统可控 $\iff$ 能控性矩阵 W 是可逆矩阵. 这就给出可控性的一个直观解释: 即控制信号可以使系统的任何两个状态相互转化, 换言之, 系统的变化是一个可逆的过程.

如果 B 的列数大于 1, 则能控性矩阵 W 具有较多的列, 因此利用 W 的秩判断系统的可控性时, 应注意避免做过多无效计算, 可以先计算矩阵 B 的秩, 然后再计算矩阵 (B,AB) 的秩, 等等, 因为计算 A 的高次幂往往较计算较低次幂大为困难.

下面我们讨论定常系统的可观测性.

定义 5.7.2 对于一个定常线性系统, 若在有限时间区间 $[0,t_1]$ 内能通过观测系统的输出 $y(t)$ 而唯一地确定初始状态 $x(0)$, 则称此系统是**完全能观测** 或**完全可观测**的或完全可测的.

注意上面定义中的输出 $y(t)$ 是指时间区间 $[0,t_1]$ 内的所有输出, 因此所谓连续型的定常系统的可观测性是用一个时间段内的所有 (当然是无限多个) 观测信号 (即系统的反馈) 来反推系统的一个 (初始) 状态. 这是合理的, 因为观测矩阵 C 一般来说是不可逆的, 因此不能指望用一个反馈信号就能确定系统的初始状态.

例 5.7.5 若观测矩阵 C 是可逆矩阵, 则定常系统显然是完全可观测的.

例 5.7.6 设定常系统的系统方程为

$$\begin{cases} x'(t)=\begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} x(t)+Bu(t) \\ y(t)=(a,b)x(t) \end{cases},$$

则该系统完全可测 $\iff$ 观测矩阵 $C=(a,b)\neq 0$ 且 $\lambda_1\neq\lambda_2$. 证明见习题 65.

定理 5.7.3 系统 (A,B,C) 完全能观测 $\iff$ n 阶 Hermite 矩阵

$$M(0,t_1)=\int_0^{t_1} e^{A^*t}C^*Ce^{At}\,\mathrm{d}\,t$$

为非奇异矩阵.

证 充分性: $y(t)=Cx(t)=Ce^{At}x(0)+C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s$. 所以

$$Ce^{At}x(0)=y(t)-C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s.$$

上式两边左乘 $e^{A^*t}C^*$, 并从 0 到 t_1 积分, 得

$$M(0,t_1)x(0)=\int_0^{t_1}\left(y(t)-C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s\right)\mathrm{d}\,t.$$

由于 $M(0,t_1)$ 非奇异, 所以

$$x(0)=M(0,t_1)^{-1}\int_0^{t_1}\left(y(t)-C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s\right)\mathrm{d}\,t.$$

这说明 $x(0)$ 是唯一确定的, 因而系统是完全能观测的.

必要性: 设系统是完全能观测的. 则满足

$$y(t)=Ce^{At}x(0)+C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s \tag{5.7.13}$$

的 $x(0)$ 是唯一确定的. 若 $M(0,t_1)$ 奇异, 则存在非零的 $\alpha\in\mathbb{C}^n$ 满足

$$M(0,t_1)\alpha=0.$$

因而 $\alpha^*M(0,t_1)\alpha=0$. 于是

$$\int_0^{t_1}\alpha^*e^{A^*t}C^*Ce^{At}\alpha\,\mathrm{d}\,t=0.$$

由此得到

$$y(t)=Ce^{At}(x(0)+\alpha)+C\int_0^t e^{A(t-s)}Bu(s)\,\mathrm{d}\,s.$$

这说明若 $x(0)$ 满足等式 (5.7.13), 则 $x(0)+\alpha$ 也满足 (5.7.13). 因而, 由 $\alpha\neq 0$ 知, $x(0)$ 不是唯一确定的, 矛盾. 所以 $M(0,t_1)$ 非奇异. □

类似于可控性的情形, 离散型的定常系统 (A,B,C) 与连续型的定常系统 (A,B,C) 具有完全相同的可测性的下列简单判别准则, 证明见习题66.

定理 5.7.4 系统 (A,B,C) 完全能观测 $\Longleftrightarrow$ 矩阵

$$M=\begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1}\end{pmatrix}$$

的秩为 n. 矩阵 M 称为能观测性矩阵.

例 5.7.7 设某系统的状态方程与输出方程为

$$\begin{cases} x'(t)=\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}x(t)+\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}u(t), \\ y(t)=\begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}x(t). \end{cases}$$

试判断这系统的能控性与能观测性.

解 由于控制矩阵 B 是可逆的, 所以系统是完全能控的.

系统的能观测性矩阵为

$$M=\begin{pmatrix} C \\ CA \end{pmatrix}=\begin{pmatrix} 1 & -1 \\ 1 & -1 \\ 0 & 2 \\ 0 & 2 \end{pmatrix}.$$

所以 M 的秩为 2, 说明系统是完全能观测的.

本节最后, 我们讨论定常系统的状态方程的化简问题.

定常系统的状态变量可以有不同的选取, 状态变量的所有线性组合构成一个线性空间. 因此, 选择适当的状态变量实际上就是选择该线性空间一组适当的基. 利用不同的状态变量之间的线性关系可以有效地简化系统的系统方程.

设定常系统的系统矩阵 A 的 Jordan 标准形为 $J=PAP^{-1}$, 其中 P 为变换矩阵. 现令 $z(t)=P^{-1}x(t)$ 或 $x(t)=Pz(t)$, 则定常系统 (A,B,C) 可以简化为

$$\begin{cases} z'(t)=Jz(t)+P^{-1}Bu(t) \\ y(t)=CPz(t) \end{cases} \tag{5.7.14}$$

利用简化的系统方程 (5.7.14) 可以较为方便地判断系统的测控性, 见下面的例子.

例 5.7.8 对单输入定常系统 (A,B,C), 如果 A 的特征值均不相同, 则系统可控 $\Longleftrightarrow$ 控制矩阵 $P^{-1}B$ 无零行, 而系统可测 $\Longleftrightarrow$ 观测矩阵 CP 无零列, 详见习题 63 与 66.

思考题

1. 在 例 5.7.8 中, 能否用 A 为单纯矩阵代替条件"A 的特征值均不相同"?

2. 定常系统 (A,B,C) 与其简化的系统 (5.7.14) 的可控性与可测性是否完全一致?

习 题 五

1. 设 $\|\cdot\|$ 是酉空间 $\mathbb{C}^n$ 的向量范数, 证明向量范数的下列基本性质:

(1) 零向量的范数为零;

(2) 当 x 是非零向量时: $\left\|\dfrac{x}{\|x\|}\right\|=1$;

(3) $\|-x\|=\|x\|$;

(4) $|\,\|x\|-\|y\|\,|\le\|x-y\|$.

2. 证明: 若 $x\in\mathbb{C}^n$, 则

(1) $\|x\|_2\le\|x\|_1\le\sqrt{n}\|x\|_2$; (2) $\|x\|_\infty\le\|x\|_1\le n\|x\|_\infty$; (3) $\|x\|_\infty\le\|x\|_2\le\sqrt{n}\|x\|_\infty$.

3. (1) 试构造 $\mathbb{R}^2$ 上的一个向量范数, 使得该范数不是任何 p- 范数;

(2) 画出你构造的范数的单位圆;

(3) 试对 $\mathbb{R}^3$ 做 (1) 与 (2), 并比较你的单位球与1- 范数和 ∞- 范数的单位球;

(4) 证明当 $0 < p < 1$ 时, l_p 范数仍然满足向量范数的前两个条件, 但不满足三角不等式. 在平面上画出 $p = 1/2, 3/2$ 时的单位圆, 并就 $p < 1$ 与 $p \geq 1$ 的一般情形作比较.

4. 证明 Minkowski 不等式 (5.1.1).

5. (1) 证明由内积诱导的向量范数满足**平行四边形恒等式** 或**极化恒等式**

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2).$$

(2) 解释上式的意义;

(3) 证明: 如果一个向量范数满足平行四边形恒等式, 则该范数一定是由某内积诱导的范数;

(4) 由 (3) 的结论判断哪些 l_p 范数是由内积诱导的, 并给出一个由内积诱导的新范数.

6. 验证 例 5.1.4.

7. 证明 命题 5.1.1 中的正定性与齐次性.

8. 证明 命题 5.1.3.

9. 证明 例 5.1.8 中的两个范数不等价.

10. 证明赋范线性空间中的单位球均为凸集, 即若 x, y 属于单位球, 则 $\alpha x + \beta y$ 也属于单位球, 其中 α, β 为正数且 $\alpha + \beta = 1$. 对照习题 5, 解释这种现象.

11. 验证矩阵的极大列和范数与极大行和范数均满足次乘性.

12. 设矩阵 A 的 F- 范数等于 a, U 是酉矩阵, 问 AU 与 UA 的 F- 范数各是多少? 请总结你的计算.

13. 证明矩阵的 1- 范数, 2- 范数和 ∞- 范数分别是向量的 1- 范数, 2- 范数和 ∞- 范数的诱导范数 (因此与之相容).

14. 证明: (1) 矩阵仿照向量的 1- 范数是矩阵范数, 但与向量的 1- 范数不相容, 试求与其相容的向量范数;

(2) 矩阵仿照向量的 ∞- 范数是向量范数但不是矩阵范数.

15. 证明公式 (5.1.12).

16. (1) 证明 $\|A\|_2 = (\rho(A^*A))^{1/2}$ 定义了一个矩阵范数, 称为 A 的**谱范数** ;

(2) 试求一个与矩阵的谱范数相容的向量范数;

(3) 证明若 A 是正规矩阵, 则 A 的谱范数就是其谱半径 $\rho(A)$;

(4) 设 V 是由全体 Hermite 矩阵构成的复线性空间, 证明谱半径给出 V 上的一个向量范数. 该范数是矩阵范数吗?

17. 试构造两种矩阵范数使得一个矩阵 A 的两种范数分别为 2 与 1/3. 能否使所有非零矩阵的两种范数之积等于 1?

18. (1) 证明向量范数的代数性质: 有限种向量范数的任意正线性组合仍是向量范数;

(2) 设 $\|\cdot\|_\alpha$ 与 $\|\cdot\|_\beta$ 是两种向量范数或矩阵范数, $p > 0$. 判断

$$[(\|\cdot\|_\alpha)^p + (\|\cdot\|_\beta)^p]^{1/p}$$

是否为向量范数或矩阵范数?

(3) 判断矩阵范数是否有与向量范数相同的代数性质 (1)?

19. 利用特征值的定义直接证明矩阵 A 的谱半径不超过矩阵 A 的任何一种矩阵范数. 此结论可以换成矩阵的任何一种向量范数吗?

20. 证明公式 (5.1.13) 定义了一个与矩阵范数 $\|\cdot\|$ 相容的向量范数.

21. 设 T 为正交矩阵, 又 $A \in \mathbb{R}^{n\times n}$. 证明:

(1) $\|T\|_2 = 1$;

(2) $\|A\|_2 = \|TA\|_2$;

(3) 试解释上面的两个结果.

22. 设 A, B 为 n 阶矩阵, 其中 A 可逆而 B 不可逆, 设 $\|\cdot\|$ 是任何一种矩阵范数. 定义 A 的条件数 $Cond(A) = \|A\|\|A^{-1}\|$. 证明: $\|A - B\| \geq 1/\|A^{-1}\|$. 解释这个结果.

23. (奇异值与矩阵的范数) 设 $A \in \mathbb{C}^{m\times n}$, $\sigma_1, \sigma_2, \cdots, \sigma_n$ 是 A 的全部奇异值. 证明:
(1) $Cond(A) = \sigma_1(A)/\sigma_n(A)$, 其中 $\sigma_1(A)$ 与 $\sigma_n(A)$ 分别是 A 的最大和最小奇异值.(参考第四章例 4.5.5.)
(2) $\|A\|_F = \left(\sum_{i=1}^r \sigma_i^2\right)^{1/2} = (\mathrm{tr}(A^*A))^{1/2}$;
(3) $\|A\|_2 = \sigma_{\max}(A)$.

24. (1) 证明 定理 5.1.3;
(2) 设 U, V 是任意赋范线性空间 (不必有限维), $\sigma \in \mathrm{Hom}(U, V)$. 证明: σ 连续 $\iff$ σ 有界.

25. 证明 引理 5.2.1 与 定理 5.2.1.

26. 设 $A_k = \begin{pmatrix} \frac{1}{k^2} & \frac{k^2+k}{k^2+1} \\ 2 & (1-\frac{2}{k})^k \end{pmatrix}$, 求 $\lim\limits_{k\to\infty} A_k$.

27. 设 $\lim_{k\to\infty} A_k = A$.
(1) 如果 A_k 均为正定矩阵, 问 A 有何特点?
(2) 如果 A_k 均为正规矩阵, 问 A 有何特点?
(3) 如果 A_k 均为可逆矩阵, 问 A 有何特点?

28. 若 $\lim\limits_{n\to\infty} A^n = B$, 则 B 为幂等矩阵.

29. 证明 命题 5.2.1.

30. 设 $A = \begin{pmatrix} 2 & -\frac{1}{2} \\ 2 & 0 \end{pmatrix}$, 求 $\displaystyle\sum_{k=0}^{\infty} \frac{A^k}{2^k}$.

31. 设 $A = \begin{pmatrix} -0.6 & 1 & 0.8 \\ 0 & 0.2 & 0 \\ -0.6 & 1 & 0.8 \end{pmatrix}$. 试判断 A 是否幂收敛.

32. (1) 设 $A = \begin{pmatrix} 0 & 0 \\ 1 & -2 \end{pmatrix}$, 求 e^A, $\sin A$, $\cos A$;
(2) 已知 $J = \begin{pmatrix} -2 & & & \\ & 1 & 1 & \\ & & 1 & \\ & & & 2 \end{pmatrix}$, 求 e^J, $\sin J$, $\cos J$.

33. (1) 证明 定理 5.3.1;
(2) 利用 (1) 求第四章习题 1 中所有正规矩阵的指数函数, 正弦函数和余弦函数.

34. 证明 命题 5.3.3 与 命题 5.3.4(2).

35. 对下列方阵 A, 求矩阵函数 e^{At}:

$$(1)\, A = \begin{pmatrix} 2 & -2 & 3 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{pmatrix},\quad (2)\, A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -8 & -12 & -6 \end{pmatrix},\quad (3)\, A = \begin{pmatrix} -2 & 1 & 3 \\ 0 & -3 & 0 \\ 0 & 2 & -2 \end{pmatrix}.$$

36. 求下列两类矩阵的矩阵函数: $\cos A$, $\sin A$, e^A:
(1)A 为幂等矩阵;
(2)A 为对合矩阵 (即 $A^2 = I$).

37. 设函数矩阵 $A(t) = \begin{pmatrix} \sin t & \cos t & t \\ \frac{\sin t}{t} & e^t & t^2 \\ 1 & 0 & t^3 \end{pmatrix}$, 其中 $t \neq 0$. 计算 $\lim\limits_{t\to 0} A(t)$, $\dfrac{\mathrm{d}}{\mathrm{d}\,t}A(t)$, $\dfrac{\mathrm{d}^2}{\mathrm{d}\,t^2}A(t)$.

38. 设函数矩阵 $A(t)=\begin{pmatrix} e^{2t} & te^{t} & t^2 \\ e^{-t} & 2e^{2t} & 0 \\ 3t & 0 & 0 \end{pmatrix}$, 计算 $\int_0^1 A(t)\,\mathrm{d}\,t$ 和 $\dfrac{\mathrm{d}}{\mathrm{d}\,t}\int_0^{t^2} A(s)\,\mathrm{d}\,s$.

39. 证明: (1) 若 A 为实反对称矩阵, 则 e^A 为正交矩阵;
(2) 若 A 为 Hermite 阵, 则 $e^{\mathrm{i}A}$ 为酉矩阵.

40. 详细证明 定理 5.4.2.

41. 证明 引理 5.4.1 即 Lagrange 插值公式并利用线性空间的直和分解理论解释之.

42. (1) 设 $J_n(\lambda)$ 是一个 n 阶 Jordan 块, 求 $\sin Jt, \cos Jt$;
(2) 对任意 n 阶矩阵 A, 导出 $\sin At$ 与 $\cos At$ 的一般表达式.

43. (1) 设 $A=\begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{pmatrix}$, 求 $e^{At}, \sin At, \cos At$;
(2) 设 $A=\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, 求 $e^{At}, \sin At, \cos At$.

44. 证明 命题 5.4.2 即 Lagrange-Sylvester 插值公式.(提示: 研究商 $\frac{f(x)}{m_A(x)}$ 或利用极限研究无重根的 Lagrange 插值公式.)

45. 设 N 是 n 阶幂零块, 验证 $\frac{\mathrm{d}e^{Nt}}{\mathrm{d}t}=Ne^{Nt}$ 并计算 $\int_0^t e^{Ns}\mathrm{d}s$.

46. (1) 设 $A=\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$, 计算积分 $\int_0^t e^{As}\mathrm{d}s$;
(2) 设 $A=\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, 计算 e^A 与 e^{At};
(3) 设 $A^2=A$, 计算 e^{At} 与 $\int_0^t e^{As}\mathrm{d}s$.

47. 设 $A^2-A+2I=0$, 计算 e^{At} 与 $\int_0^t e^{As}\mathrm{d}s$.

48. (1) 证明 Jacobian 猜想对线性映射成立;
(2) 证明 Jacobian 猜想的逆命题成立;
(3) 试求 例 5.5.6 中多项式映射的逆映射. 试对次数不超过 2 (即每个分量的次数不超过 2) 的二元多项式映射证明 Jacobian 猜想. 如何将你的证明推广到 n 元?

49. 如果 $f=(f_1(x),\cdots,f_n(x))^T$ 与 $g=(g_1(x),\cdots,g_n(x))^T$ 是两个 n 元列映射, 计算 $J(f^Tg)$ 并与乘法公式 (命题 5.5.1(2)) 比较.

50. 设 $x=(x_1,x_2,\cdots,x_n)^T, A$ 是 n 阶矩阵, 求 n 元函数 $(\|x-Ax\|_2)^2$ 的导数与 Hessian 矩阵.

51. 计算 例 5.5.8 的复合函数的导数并与多元函数的链法则比较.

52. 证明 命题 5.5.2.

53. (1) 证明 命题 5.5.3, 并以此推出 $J(\mathrm{tr}(AX))$ 与 $J(\mathrm{tr}(XB))$;
(2) 验证公式 (5.5.17).

54. 计算一般线性变换的导数, 证明 例 5.5.3 的结论.

55. 验证公式 (5.6.5) 给出定解问题 (5.6.2) 的解且该解是唯一的.

56. 求下列微分方程组的通解：

$$(1)\ x'(t)=\begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix}x(t);\qquad (2)\ x'(t)=\begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & -1 \\ 0 & 1 & 1 \end{pmatrix}x(t).$$

57. 求下列微分方程组 $x'(t)=Ax(t)$ 满足初始条件 $x(0)$ 的解：

$$(1)\ A=\begin{pmatrix}1&12\\3&1\end{pmatrix},\ x(0)=\begin{pmatrix}0\\1\end{pmatrix};\qquad (2)\ A=\begin{pmatrix}1&2&1\\1&-1&1\\2&0&1\end{pmatrix},\ x(0)=\begin{pmatrix}1\\0\\0\end{pmatrix}.$$

58. 分别利用积分因子法和常数变易法详细证明 定理 5.6.2.

59. (1) 求解微分方程组

$$x'(t)=\begin{pmatrix}1&0&0\\2&1&-2\\3&2&1\end{pmatrix}x(t)+\begin{pmatrix}1\\0\\0\end{pmatrix}e^{2t}.$$

(2) 求 $x'(t)=Ax(t)+Bu(t)$ 满足初始条件 $x(0)$ 的解, 其中

$$A=\begin{pmatrix}-6&1&0\\-11&0&1\\-6&0&0\end{pmatrix},\ B=\begin{pmatrix}2\\6\\2\end{pmatrix},\ u(t)=1,\ x(0)=\begin{pmatrix}1\\0\\-1\end{pmatrix}.$$

60. 求方程 $y'''+6y''+11y'+6y=e^{-t}$ 满足 $y(0)=y'(0)=y''(0)=0$ 的解.

61. (1) 证明微分方程 $x'(t)=Ax(t)+\gamma e^{at}$ 有形如 $x(t)=\beta e^{at}$ 的解 $\iff$ $(\alpha I-A)\beta=\gamma$, 其中 β,γ 都是 n 维向量, $a\in\mathbb{C}$;

(2) 解 $x'(t)=Ax(t)+e^{2t}C$, 其中

$$A=\begin{pmatrix}3&1\\2&2\end{pmatrix},\ C=\begin{pmatrix}-1\\-1\end{pmatrix},\ x(0)=\begin{pmatrix}0\\1\end{pmatrix}.$$

62. 证明 定义 5.7.1 后的注.(提示: 利用 定理 5.7.1.)

63. (1) 设定常系统的系统矩阵 A 是对角矩阵, 试给出该系统可控性的一个判断准则;

(2) 设定常系统的系统矩阵 A 是一个 Jordan 块, 试给出该系统可控性的一个判断准则;

(3) 设单输入定常系统的系统矩阵 A 是 Frobenius 标准形 (即 (5.6.10) 中的矩阵), 而其控制矩阵 B 为标准向量 e_n, 证明该系统是可控的. 这样的定常系统称为可控标准形.

64. 根据你在 63 题中给出的判断准则, 研究下面几个定常系统的可控性.

(1)$A=\begin{pmatrix}2&0\\0&1\end{pmatrix}$, $B=(1,1)^T$; (2)$A=\begin{pmatrix}2&0\\0&1\end{pmatrix}$, $B=(1,0)^T$;

(3)$A=\begin{pmatrix}a&0\\0&b\end{pmatrix}$, $B=(c,d)^T$; (4)$A=\begin{pmatrix}a&1\\0&a\end{pmatrix}$, $B=(c,d)^T$.

65. 证明 例 5.7.6 的结论.

66. 证明 定理 5.7.4.

67. (1) 设定常系统的系统矩阵 A 是对角矩阵, 试给出该系统可测性的一个判断准则;

(2) 设定常系统的系统矩阵 A 是一个 Jordan 块, 试给出该系统可测性的一个判断准则;

(3) 设定常系统的系统矩阵 A 是 Frobenius 标准形的转置矩阵, 而其输出矩阵 C 为标准行向量 e_n^T, 证明该系统是可测的. 这样的定常系统称为可测标准形.

68. 根据你在 67 题中给出的判断准则, 研究下面几个定常系统的可测性.

(1)$A=\begin{pmatrix}a&1\\0&a\end{pmatrix}$, $C=(c,d)^T$; (2)$A=\begin{pmatrix}a&1&0\\0&a&1\\0&0&a\end{pmatrix}$, $C=(c,d,f)^T$.