*第十三章 若尔当典范形的讨论与应用

本章主要介绍若尔当典范形的应用, 包括矩阵函数的定义以及一些简单的矩阵方程的讨论. 也介绍了求若尔当典范形的转换矩阵的一些方法, 这是因为常微分方程里需要用到它. 若尔当典范形是线性代数的核心内容之一, 可是一般的高等代数课程在教到若尔当典范形时, 剩下的课时已经不多了, 往往匆匆教完这部分内容后, 课程也宣告结束. 学生没有时间去了解它的应用, 自然很快把它忘记了. 事实上, 一个理论只有在应用中才能体会到它的重要性, 才能被深刻在脑海里. 因此本章内容的学习可以加深对若尔当典范形的理解, 这对有志于报考研究生的读者特别有用, 因为学习了本章后将能开阔你的解题思路. 第四第五节并不直接用到典范形, 因为它们常被用到, 也放在这一章里. 供教师选用.

为简单起见, 本章的数域 K 总是取为复数域 $\mathbb{C}$.

§1 若尔当典范形的几何意义

上一章是从矩阵的角度研究它的相似典范形. 这一节是从线性变换的角度, 把矩阵看成线性变换关于某个基的矩阵, 然后讨论怎样选取合适的基使得对应的矩阵成为相似典范形. 这是因为相似的矩阵可以被看成同一个线性变换在不同基下的矩阵. 另一方面, 如果线性空间可以分解成线性变换 $\mathscr{A}$ 的不变子空间的直和, 那么 $\mathscr{A}$ 的矩阵可以相似于分块对角阵 (参见第七章命题 5.1). 因此我们首先讨论空间的分解问题.

以后如不特别说明, 总是假设 $\mathscr{A} \in \mathrm{End}(V)$ 是复数域上 n 维线性空间 V 的线性变换. $\mathscr{A}$ 关于 V 的基 $\eta_1, \cdots, \eta_n$ 的矩阵是 $A \in M_n(\mathbb{C})$.

定义 1.1 如果有复系数多项式

$$f(\lambda) = a_0\lambda^m + a_1\lambda^{m-1} + \cdots + a_{m-1}\lambda + a_m$$

使得

$$f(\mathscr{A}) = a_0\mathscr{A}^m + a_1\mathscr{A}^{m-1} + \cdots + a_{m-1}\mathscr{A} + a_m\mathscr{E} = 0,$$

这里 $\mathscr{E}$ 表示恒同线性变换, 则称 $f(\lambda)$ 是 $\mathscr{A}$ 的**零化多项式**. 次数最小的零化多项式称为 $\mathscr{A}$ 的**极小多项式**, 记为 $m_{\mathscr{A}}(\lambda)$.

显然 $\mathscr{A}$ 的极小多项式就是矩阵 A 的极小多项式. 第十二章 §6 得到的关于极小多项式的结果仍然有效.

命题 1.1 设 $\mathscr{A}$ 的极小多项式

$$m_{\mathscr{A}}(\lambda) = (\lambda-\lambda_1)^{k_1}(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_t)^{k_t},$$

其中 λ_i 各不相同, 则存在 V 的直和分解

$$V = W_{\lambda_1} \oplus W_{\lambda_2} \oplus \cdots \oplus W_{\lambda_t},$$

其中

$$W_{\lambda_i} = \{\alpha \in V \mid (\mathscr{A} - \lambda_i\mathscr{E})^{k_i}(\alpha) = 0\}, \quad i = 1, 2, \cdots, t,$$

是 $\mathscr{A}$ 的不变子空间, 称为属于特征值 λ_i 的**根子空间** (*root subspace*).

证明: 不难验证 W_{λ_i} 是线性子空间. 设 $\alpha \in W_{\lambda_i}$, 则

$$(\mathscr{A} - \lambda_i\mathscr{E})^{k_i}(\mathscr{A}(\alpha)) = \mathscr{A}((\mathscr{A} - \lambda_i\mathscr{E})^{k_i}(\alpha)) = 0,$$

说明 $\mathscr{A}(\alpha) \in W_{\lambda_i}$, 因此 W_{λ_i} 是 $\mathscr{A}$ 的不变子空间.

任取 $\alpha \in V$. 由于

$$\frac{m_{\mathscr{A}}(\lambda)}{(\lambda-\lambda_1)^{k_1}}, \ \cdots, \ \frac{m_{\mathscr{A}}(\lambda)}{(\lambda-\lambda_t)^{k_t}}$$

是互素的多项式 (假设 λ_i 互不相同), 因此存在多项式 $u_1(\lambda), \cdots, u_t(\lambda)$ 使得 (第十章定理 3.8)

$$u_1(\lambda)\frac{m_{\mathscr{A}}(\lambda)}{(\lambda-\lambda_1)^{k_1}} + \cdots + u_t(\lambda)\frac{m_{\mathscr{A}}(\lambda)}{(\lambda-\lambda_t)^{k_t}} = 1.$$

用 $\mathscr{A}$ 代替 λ, 右端成为恒同变换, 作用在向量 α 上得到

$$\alpha = \alpha_1 + \alpha_2 + \cdots + \alpha_t,$$

其中

$$\alpha_i=u_i(\mathscr{A})(\mathscr{A}-\lambda_1\mathscr{E})^{k_1}\cdots(\mathscr{A}-\lambda_{i-1}\mathscr{E})^{k_{i-1}}(\mathscr{A}-\lambda_{i+1}\mathscr{E})^{k_{i+1}}\cdots(\mathscr{A}-\lambda_t\mathscr{E})^{k_t}(\alpha),$$

$$i=1,\cdots,t.$$

由于

$$(\mathscr{A}-\lambda_i\mathscr{E})^{k_i}(\alpha_i)=u_i(\mathscr{A})m_{\mathscr{A}}(\mathscr{A})(\alpha)=0,\quad i=1,\cdots,t,$$

所以 $\alpha_i\in W_{\lambda_i}$, 证明了 $V=W_{\lambda_1}+\cdots+W_{\lambda_t}$.

为证直和, 设 $\alpha\in W_{\lambda_1}\cap(W_{\lambda_2}+\cdots+W_{\lambda_t})$. 由于 $(\lambda-\lambda_1)^{k_1}$ 与 $(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_t)^{k_t}$ 互素, 存在 $v_1(\lambda),v_2(\lambda)$ 使得

$$v_1(\lambda)(\lambda-\lambda_1)^{k_1}+v_2(\lambda)(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_t)^{k_t}=1.$$

用 $\mathscr{A}$ 代替 λ, 作用在向量 α 上得到

$$\alpha=v_1(\mathscr{A})(\mathscr{A}-\lambda_1\mathscr{E})^{k_1}(\alpha)+v_2(\mathscr{A})(\mathscr{A}-\lambda_2\mathscr{E})^{k_2}\cdots(\mathscr{A}-\lambda_t\mathscr{E})^{k_t}(\alpha),$$

由于 $\alpha\in W_{\lambda_1}$, 得 $(\mathscr{A}-\lambda_1\mathscr{E})^{k_1}(\alpha)=0$, 又因 $\alpha\in W_{\lambda_2}+\cdots+W_{\lambda_t}$, 可得 $(\mathscr{A}-\lambda_2\mathscr{E})^{k_2}\cdots(\mathscr{A}-\lambda_t\mathscr{E})^{k_t}(\alpha)=0$, 因此 $\alpha=0$. 类似地可证 $W_{\lambda_i}\cap(W_{\lambda_1}+\cdots+W_{\lambda_{i-1}}+W_{\lambda_{i+1}}+\cdots+W_{\lambda_t})=0$. 所以 $V=W_{\lambda_1}\oplus\cdots\oplus W_{\lambda_t}$. □

从命题 1.1 中根子空间 W_{λ_i} 的定义可以看出, 这些子空间是被 $\mathscr{A}$ 唯一确定的. 但是命题 1.1 的分解还可以加细. 以得到与若尔当典范形对应的分解.

设矩阵 A 的初等因子组是

$$(\lambda-\lambda_1)^{k_1},\ \cdots,\ (\lambda-\lambda_s)^{k_s},$$

其中 λ_i 可以重复. 相应的若尔当典范形为

$$J=T^{-1}AT=\operatorname{diag}(J_{k_1}(\lambda_1),\cdots,J_{k_s}(\lambda_s)),\quad T\in GL(n,\mathbb{C}).$$

用可逆矩阵 T 作基变换:

$$(\eta_1'\ \eta_2'\cdots\eta_n')=(\eta_1\ \eta_2\cdots\eta_n)T,\tag{1.1}$$

那么线性变换 $\mathscr{A}$ 关于基 $\eta_1',\cdots,\eta_n'$ 的矩阵就是若尔当典范形 J. 为了讨论方便, 把基 $\eta_1',\cdots,\eta_n'$ 按对应的若尔当块重新命名为

$$\eta_{11}',\cdots,\eta_{1k_1}',\eta_{21}',\cdots,\eta_{2k_2}',\cdots,\eta_{s1}',\cdots,\eta_{sk_s}',$$

则根据线性变换矩阵的定义, 有

$$\left.\begin{aligned}&\mathscr{A}(\eta'_{i1}) = \lambda_i\eta'_{i1},\\&\mathscr{A}(\eta'_{i2}) = \lambda_i\eta'_{i2} + \eta'_{i1},\\&\cdots\cdots\cdots\cdots\cdots\cdots\\&\mathscr{A}(\eta'_{ij}) = \lambda_i\eta'_{ij} + \eta'_{i.j-1},\\&\cdots\cdots\cdots\cdots\cdots\cdots\\&\mathscr{A}(\eta'_{ik_i}) = \lambda_i\eta'_{ik_i} + \eta'_{i.k_i-1},\end{aligned}\right. \qquad i = 1,\cdots,s, \tag{1.2}$$

上式可以被改写成

$$\begin{aligned}&\eta'_{i,k_i-1} = (\mathscr{A} - \lambda_i\mathscr{E})\eta'_{ik_i},\\&\eta'_{i,k_i-2} = (\mathscr{A} - \lambda_i\mathscr{E})\eta'_{i,k_i-1} = (\mathscr{A} - \lambda_i\mathscr{E})^2\eta'_{ik_i},\\&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\&\eta'_{i,j} = (\mathscr{A} - \lambda_i\mathscr{E})\eta'_{i,j+1} = (\mathscr{A} - \lambda_i\mathscr{E})^{k_i-j}\eta'_{ik_i},\\&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\&\eta'_{i1} = (\mathscr{A} - \lambda_i\mathscr{E})\eta'_{i,2} = (\mathscr{A} - \lambda_i\mathscr{E})^{k_i-1}\eta'_{ik_i},\\&(\mathscr{A} - \lambda_i\mathscr{E})^{k_i}\eta'_{ik_i} = 0.\end{aligned} \tag{1.3}$$

从 (1.3) 式可以看出 $\eta'_{i1},\cdots,\eta'_{ik_i}$ 被 η'_{ik_i} 完全确定. 令

$$V_i = L(\eta'_{i1},\cdots,\eta'_{ik_i}),\quad i = 1,\cdots,s.$$

V_i 是 $\mathscr{A}$ 的不变子空间, 它被 η'_{ik_i} 完全确定. 由 V_i 的定义可得直和分解式

$$V = V_1 \oplus \cdots \oplus V_s.$$

不难看出, 所有对应于同一个特征值 c 的 V_i 之和就是根子空间 W_c. 以下的定理 1.2 实际上是第十二章 §5 的结果的另一种表述.

定理 1.2 *存在 V 到 $\mathscr{A}$ 的不变子空间的直和分解*

$$V = V_1 \oplus \cdots \oplus V_s, \tag{1.4}$$

使得 $\mathscr{A}$ 限制在每个 V_i 上的矩阵都是若尔当块. □

一般说来, 定理 1.2 的分解不是唯一的.

变换矩阵的求法

最后我们讨论如何确定变换矩阵T, 使得$T^{-1}AT$成为若尔当典范形. 第一种方法是利用子空间分解, 由子空间的基合成空间V的基. 从定理 1.2 前面的讨论可知确定T相当于找出一个与分解(1.4)相对应的基$\eta'_{11},\cdots,\eta'_{sk_s}$. 而确定这个基的关键是找出向量$\eta'_{ik_i}$, 它们满足条件

$$(\mathscr{A}-\lambda_i\mathscr{E})^{k_i}(\eta'_{ik_i})=0,\quad (\mathscr{A}-\lambda_i\mathscr{E})^{k_i-1}(\eta'_{ik_i})\neq 0. \tag{1.5}$$

因此在确定了矩阵A的初等因子组后, 可利用条件(1.5)确定向量η'_{ik_i}, 当具有相同特征值的初等因子不止一个时, 应按次数由高到低的次序, 先确定次数k_i高的η'_{ik_i}, 再用公式(1.3)算出其余的基向量$\eta'_{i1},\cdots,\eta'_{i,k_i-1}$. 在求次数较低的初等因子对应的基向量$\eta'_{jk_j}$时, 应在满足条件(1.5)的向量中选取与已经求出的基向量线性无关的向量. 最后把这些向量的坐标作为列向量即可写出矩阵T.

例 1.1 已知矩阵

$$A=\begin{pmatrix}4&-1&2\\-9&4&-6\\-9&3&-5\end{pmatrix}$$

的初等因子是$\lambda-1$, $(\lambda-1)^2$. 求变换矩阵T使$T^{-1}AT$是若尔当典范形.

解: 因为两个初等因子对应同一个特征值1, 因此先求与次数较高的初等因子对应的η'_{22}. 它需要满足(1.5), 但因极小多项式$m_{\mathscr{A}}(\lambda)=(\lambda-1)^2$, 故$(A-E)^2=0$, 只剩下一个条件:

$$(A-E)\eta'_{22}=\begin{pmatrix}3&-1&2\\-9&3&-6\\-9&3&-6\end{pmatrix}\eta'_{22}\neq 0.$$

不妨取$\eta'_{22}=(1,0,0)^{\mathrm{T}}$. 算出$\eta'_{21}=(A-E)\eta'_{22}=(3,-9,-9)^{\mathrm{T}}$.

η'_{11}应该满足

$$(A-E)\eta'_{11}=\begin{pmatrix}3&-1&2\\-9&3&-6\\-9&3&-6\end{pmatrix}\eta'_{11}=0,$$

而且应与η'_{21}线性无关. 可取$\eta'_{11}=(1,3,0)^{\mathrm{T}}$. 因此

$$T=\begin{pmatrix}1&3&1\\3&-9&0\\0&-9&0\end{pmatrix}.$$

$$T^{-1}AT = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}. \quad \square$$

当同一个特征值对应的初等因子很多时, 计算将会相当复杂.

第二种方法是直接解方程. 把 $T^{-1}AT = J$ 化成 $AT = TJ$, 以矩阵 T 的 n^2 个元素作为未知量, 得到一个线性方程组, 然后求解. 其中会有不少自由未知量. 这种方法只能用于 n 很小的情形.

第三种方法是利用多项式矩阵的初等变换先把特征矩阵 $\lambda E - A$ 化成对角形 (不一定化成正规形):

$$P_1(\lambda)(\lambda E - A)Q_1(\lambda) = D(\lambda),$$

并记录下变换矩阵 $Q_1(\lambda)$. 从对角矩阵 $D(\lambda)$ 就能得到矩阵 A 的初等因子组 (第十二章命题 4.3). 然后写出相应的若尔当典范形 J. 再通过初等变换把 J 的特征矩阵也化成 $D(\lambda)$:

$$P_2(\lambda)(\lambda E - J)Q_2(\lambda) = D(\lambda).$$

于是

$$\lambda E - J = [P_2(\lambda)^{-1}P_1(\lambda)](\lambda E - A)[Q_1(\lambda)Q_2(\lambda)^{-1}].$$

再求出 $Q_1(\lambda)Q_2(\lambda)^{-1}$ 被 $\lambda E - J$ 右除所得到的余式 $Q_0 \in M_n(\mathbb{C})$:

$$Q_1(\lambda)Q_2(\lambda)^{-1} = U(\lambda)(\lambda E - J) + Q_0.$$

根据第十二章引理 3.2, Q_0 也可用以下方法求得: 把 $Q_1(\lambda)Q_2(\lambda)^{-1}$ 展开成矩阵多项式:

$$Q_1(\lambda)Q_2(\lambda)^{-1} = C_0\lambda^k + C_1\lambda^{k-1} + \cdots + C_k, \quad C_0, \cdots, C_k \in M_n(\mathbb{C}).$$

则

$$Q_0 = C_0J^k + C_1J^{k-1} + \cdots + C_k.$$

于是根据第十二章定理 3.3 的证明, 有

$$Q_0^{-1}AQ_0 = J.$$

至于如何将变换矩阵记录下来, 我们在第四章 §8 已经遇到过这种情形. 下面的例子介绍了具体的方法, 请读者自己解释这样做的理由.

例 1.2 求矩阵

$$A = \begin{pmatrix} 4 & -1 & 2 \\ -9 & 4 & -6 \\ -9 & 3 & -5 \end{pmatrix}$$

的若尔当典范形及变换矩阵.

解: 先作初等变换把特征矩阵 $\lambda E - A$ 化为对角形 (中间过程略, 下同):

$$\begin{pmatrix} \lambda-4 & 1 & -2 \\ 9 & \lambda-4 & 6 \\ 9 & -3 & \lambda+5 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda-1 & 0 \\ 0 & 0 & (\lambda-1)^2 \\ 0 & 0 & 1 \\ 1 & 2 & -\lambda-2 \\ 0 & 1 & -3 \end{pmatrix}.$$

上面的计算可表示为

$$\begin{pmatrix} P_1(\lambda) & 0 \\ 0 & E \end{pmatrix} \begin{pmatrix} \lambda E - A \\ E \end{pmatrix} Q_1(\lambda) = \begin{pmatrix} P_1(\lambda)(\lambda E - A)Q_1(\lambda) \\ Q_1(\lambda) \end{pmatrix}.$$

因此 A 的初等因子是 $\lambda-1$, $(\lambda-1)^2$. 其若尔当典范形为

$$J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}.$$

再作初等变换把特征矩阵 $\lambda E - J$ 化为与上面相同的对角形:

$$\begin{pmatrix} \lambda-1 & 0 & 0 \\ 0 & \lambda-1 & -1 \\ 0 & 0 & \lambda-1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda-1 & 0 \\ 0 & 0 & (\lambda-1)^2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & \lambda-1 \end{pmatrix}.$$

上面的计算可表示为

$$\begin{pmatrix} P_2(\lambda) & 0 \\ 0 & E \end{pmatrix} \begin{pmatrix} \lambda E - J \\ E \end{pmatrix} Q_2(\lambda) = \begin{pmatrix} P_2(\lambda)(\lambda E - J)Q_2(\lambda) \\ Q_2(\lambda) \end{pmatrix}.$$

为求 $Q_1(\lambda)Q_2(\lambda)^{-1}$, 我们使用以下的分块矩阵乘法:

$$\begin{pmatrix} Q_2(\lambda) \\ Q_1(\lambda) \end{pmatrix} Q_2(\lambda)^{-1} = \begin{pmatrix} E \\ Q_1(\lambda)Q_2(\lambda)^{-1} \end{pmatrix}.$$

具体计算如下 (注意只能做初等列变换, 不能做行变换):

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & \lambda-1 \\ 0 & 0 & 1 \\ 1 & 2 & -\lambda-2 \\ 0 & 1 & -3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 2 & -2\lambda-1 & 1 \\ 1 & -3 & 0 \end{pmatrix}.$$

于是

$$Q_1(\lambda)Q_2(\lambda)^{-1} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix}\lambda + \begin{pmatrix} 0 & 1 & 0 \\ 2 & -1 & 1 \\ 1 & -3 & 0 \end{pmatrix}.$$

用 J 从右边代入 λ 即得

$$Q_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 0 \\ 2 & -1 & 1 \\ 1 & -3 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 & 0 \\ 2 & -3 & -1 \\ 1 & -3 & 0 \end{pmatrix}.$$

经验证确实有

$$Q_0^{-1}AQ_0 = J. \quad \square$$

这三种方法各有长短. 总之初等因子的次数越高, 具有同一个特征值的初等因子越多, 计算也越复杂, 哪个方法都一样. 当然最省力的是请计算机算, 可参见第十二章 §5 的上机实验.

在解线性常微分方程时会用到若尔当典范形的变换矩阵. 又如在求一个矩阵的高次方幂时, 借助若尔当典范形可大大减少计算量.

例 1.3 设

$$A = \begin{pmatrix} -1 & -4 \\ 1 & 3 \end{pmatrix},$$

计算 A^{100}.

解: 先求出 A 的初等因子是 $(\lambda-1)^2$. 因此若尔当典范形是 $J = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. 因 $A-E = \begin{pmatrix} -2 & -4 \\ 1 & 2 \end{pmatrix}$, 可取 $\eta_2' = (1,0)^{\mathrm{T}}$, $\eta_1' = (A-E)\eta_2' =$

$(-2,1)^{\mathrm{T}}$. 因此变换矩阵是 $T=\begin{pmatrix}-2 & 1\\ 1 & 0\end{pmatrix}$. 所以

$$\begin{aligned}A^{100}&=(TJT^{-1})^{100}=TJ^{100}T^{-1}\\&=\begin{pmatrix}-2 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1 & 100\\ 0 & 1\end{pmatrix}\begin{pmatrix}0 & 1\\ 1 & 2\end{pmatrix}=\begin{pmatrix}-199 & -400\\ 100 & 201\end{pmatrix}.\end{aligned}$$ □

习　题 13–1

1. 对下列矩阵 A, 求变换矩阵 T, 使 $T^{-1}AT$ 为若尔当典范形:

(1) $\begin{pmatrix}1 & -3 & 4\\ 4 & -7 & 8\\ 6 & -7 & 7\end{pmatrix}$;　(2) $\begin{pmatrix}3 & -2 & 1\\ 2 & -2 & 2\\ 3 & -6 & 5\end{pmatrix}$;

(3) $\begin{pmatrix}3 & -4 & 0 & 2\\ 4 & -5 & -2 & 4\\ 0 & 0 & 3 & -2\\ 0 & 0 & 2 & -1\end{pmatrix}$;　(4) $\begin{pmatrix}1 & -3 & 0 & 3\\ -2 & -6 & 0 & 13\\ 0 & -3 & 1 & 3\\ -1 & -4 & 0 & 8\end{pmatrix}$.

2. 证明: 每个复方阵 A 可分解为 $A=B+C$, 其中 B 为可对角化矩阵, C 为幂零阵, 且 $BC=CB$.

3. 设 J 是特征值为 1 的 n 阶若尔当块, 试求使 $g(J)$ 相似于 J 的多项式 $g(\lambda)$ 应满足的充分必要条件.

*__4.__ 证明: 每个复方阵可分解为两个复对称矩阵的乘积, 并且其中的一个是可逆的.

§ 2　矩阵函数

这一节要讨论的问题是如何把复函数 $f(\lambda)$ 推广成矩阵变量的函数 $f(A)$? 当 $f(\lambda)$ 是复系数多项式时, 这个问题早已解决. 现在要研究一般复函数的情形.

首先考虑若尔当块 $J=J_k(c)$ 的情形. 则 $J=cE_k+H$, 这里

$$H=\begin{pmatrix}0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \ddots & \vdots\\ 0 & 0 & 0 & \ddots & 1\\ 0 & 0 & 0 & \cdots & 0\end{pmatrix},$$

H 具有以下性质: $H^k = 0$, $H^{k-1} \neq 0$. 我们把某个方幂等于 0 的矩阵称为**幂零矩阵** (*nilpotent matrix*), 因此 H 是一个 k 阶幂零矩阵.

如果 $f(\lambda)$ 在 c 附近有幂级数展开式

$$f(c+h) = f(c) + f'(c)h + \frac{f''(c)}{2!}h^2 + \cdots,$$

我们很自然地想到定义 $f(\lambda)$ 在若尔当块 $J = cE + H$ 的值为

$$f(J) \overset{\text{def}}{=} f(c)E + f'(c)H + \frac{f''(c)}{2!}H^2 + \cdots + \frac{f^{(k-1)}(c)}{(k-1)!}H^{k-1}$$
$$= \begin{pmatrix} f(c) & f'(c) & \dfrac{f''(c)}{2!} & \dfrac{f'''(c)}{3!} & \cdots & \dfrac{f^{(k-1)}(c)}{(k-1)!} \\ 0 & f(c) & f'(c) & \dfrac{f''(c)}{2!} & \cdots & \dfrac{f^{(k-2)}(c)}{(k-2)!} \\ 0 & 0 & f(c) & f'(c) & \cdots & \dfrac{f^{(k-3)}(c)}{(k-3)!} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ddots & f'(c) \\ 0 & 0 & 0 & 0 & \cdots & f(c) \end{pmatrix}. \tag{2.1}$$

定义 2.1 设矩阵 $A \in M_n(\mathbb{C})$ 的特征值是 $\lambda_1, \cdots, \lambda_s$, 它的若尔当典范形 $J = \operatorname{diag}(J_1, \cdots, J_t) = T^{-1}AT$. 如果复函数 $f(\lambda)$ 在各特征值处都有幂级数展开式, 则定义 A 的**矩阵函数** (*matrix function*) 为

$$f(A) = T\operatorname{diag}(f(J_1), \cdots, f(J_t))T^{-1}. \tag{2.2}$$

可以证明矩阵函数的定义与变换矩阵 T 的取法无关, 因比较繁杂而略去. 仔细观察公式 (2.1) 可以发现, 对于若尔当块 $J_k(c)$, 定义 $f(J_k(c))$ 只用到了 $f(c), f'(c), \cdots, f^{(k-1)}(c)$ 的值. 如果矩阵 A 的极小多项式

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

那么为了确定矩阵函数 $f(A)$, 需要用到

$$f(\lambda_1), f'(\lambda_1), \cdots, f^{(k_1-1)}(\lambda_1), \cdots, f(\lambda_s), f'(\lambda_s), \cdots, f^{(k_s-1)}(\lambda_s). \tag{2.3}$$

从 (2.3) 可以看出, A 的所有特征值以及每个特征值在极小多项式中的重数在确定 A 的矩阵函数时起着关键性的作用. 因此把 A 的所有特征值以及每个特征值在极小多项式中的重数称为矩阵 A 的**谱** (*spectrum*),

记为 Λ_A. 把 (2.3) 中的数称为函数 $f(\lambda)$ 在 A 的谱上的值, 记为 $f(\Lambda_A)$. 这样我们就得到了以下命题.

命题 2.1 函数 $f(\lambda)$ 在矩阵 A 处的矩阵函数 $f(A)$ 被 f 在谱上的值 $f(\Lambda_A)$ 唯一确定. □

公式 (2.1) 和 (2.2) 定义的矩阵函数 $f(A)$ 可以被改写成

$$f(A)=\sum_{i=1}^{s}\sum_{j=0}^{k_i-1}f^{(j)}(\lambda_i)Z_{ij},$$

这里的 Z_{ij} 是被矩阵 A 唯一确定的复矩阵, 与函数 $f(\lambda)$ 无关. Z_{ij} 共有 $\deg m_A(\lambda)$ 个, 它们的作用类似于拉格朗日插值多项式中的基本插值多项式 $l_i(x)$ (参见第十章例 7.1). 我们可以通过选取简单的多项式函数 $f(\lambda)$ 把 Z_{ij} 确定下来. 以后计算 A 的矩阵函数就很方便了.

例 2.1 已知矩阵

$$A=\begin{pmatrix}1&9&-4\\4&-24&11\\10&-66&30\end{pmatrix},$$

的初等因子是 $\lambda-3$, $(\lambda-2)^2$, 求矩阵函数 A^5 及 $\exp(A)$.

解: A 的极小多项式是 $m_A(\lambda)=(\lambda-3)(\lambda-2)^2$. 因此

$$f(A)=f(3)Z_{10}+f(2)Z_{20}+f'(2)Z_{21}.$$

分别取 $f(\lambda)=1$, λ 及 λ^2, 得到

$$\begin{cases}E=Z_{10}+Z_{20},\\A=3Z_{10}+2Z_{20}+Z_{21},\\A^2=9Z_{10}+4Z_{20}+4Z_{21}.\end{cases}$$

解得

$$Z_{10}=4E-4A+A^2=\begin{pmatrix}-3&21&-9\\2&-14&6\\6&-42&18\end{pmatrix},$$

$$Z_{20}=-3E+4A-A^2=\begin{pmatrix}4&-21&9\\-2&15&-6\\-6&42&-17\end{pmatrix},$$

$$Z_{21}=-6E+5A-A^2=\begin{pmatrix}2&-12&5\\2&-12&5\\4&-24&10\end{pmatrix}.$$

因此

$$A^5=3^5Z_{10}+2^5Z_{20}+5\cdot 2^4Z_{21}=\begin{pmatrix}-441&3471&-1499\\582&-3882&1666\\1586&-10782&4630\end{pmatrix},$$

$$\begin{aligned}\exp(A)&=\mathrm{e}^3Z_{10}+\mathrm{e}^2Z_{20}+\mathrm{e}^2Z_{21}\\&=\begin{pmatrix}-3\mathrm{e}^3+6\mathrm{e}^2&21\mathrm{e}^3-33\mathrm{e}^2&-9\mathrm{e}^3+14\mathrm{e}^2\\2\mathrm{e}^3&-14\mathrm{e}^3+3\mathrm{e}^2&6\mathrm{e}^3-\mathrm{e}^2\\6\mathrm{e}^3-2\mathrm{e}^2&-42\mathrm{e}^3+18\mathrm{e}^2&18\mathrm{e}^3-7\mathrm{e}^2\end{pmatrix}.\quad\square\end{aligned}$$

最后, 我们讨论矩阵函数的初等因子. 从公式 (2.2) 可以知道, $f(A)$ 的初等因子组是各个块 $f(J_i)$ 的初等因子组的保留重复因子的并. 因此问题归结为确定若尔当块的矩阵函数的初等因子.

定理 2.2 若尔当块的矩阵函数 $f(J_k(c))$ 的初等因子组有以下几种情况:

(1) 当 $k=1$ 或 $k>1$, $f'(c)\neq 0$ 时, $f(J_k(c))$ 只有一个初等因子 $(\lambda-f(c))^k$;

(2) 当 $k>1$, $f'(c)=\cdots=f^{(h-1)}(c)=0$, $f^{(h)}(c)\neq 0\,(1<h<k)$ 时, $f(J_k(c))$ 的初等因子组由 $h-r$ 个 $(\lambda-f(c))^q$, r 个 $(\lambda-f(c))^{q+1}$ 构成, 这里 $k=hq+r$, $0\leqslant r<h$;

(3) 当 $k>1$, $f'(c)=\cdots=f^{(k-1)}(c)=0$ 时, $f(J_k(c))$ 的初等因子组由 k 个 $\lambda-f(c)$ 构成.

证明: 矩阵 $f(J_k(c))$ 如 (2.1) 式所示. 这是一个上三角形矩阵, 它只有一个特征值 $f(c)$. 因此 $k=1$ 的结论是显然的. 当 $k>1$, $f'(c)\neq 0$ 时, 矩阵 $f(J_k(c))-f(c)E$ 的秩等于 $k-1$, 所以 $f(J_k(c))$ 的若尔当块只能有一个, 也就是说它的唯一的初等因子是 $(\lambda-c)^k$. 这样就证明了 (1). 结论 (3) 也是显然的. 结论 (2) 的证明较为复杂, 这里略去了. $\square$

习 题 13–2

1. 设 $A=\begin{pmatrix}0 & 1 & 1\\ 1 & 0 & 1\\ 1 & -1 & 1\end{pmatrix}$, 求 A^{100}, $\exp A$, $\sqrt[3]{A}$.

2. 设 $A=\begin{pmatrix}1 & -3 & 3\\ -2 & -6 & 13\\ -1 & -4 & 8\end{pmatrix}$, 试求矩阵 B, 使 $B^2=A$.

3. 设 n 阶实矩阵 A 的特征值全是正实数. 证明: 存在实矩阵 B, 使 $B^2=A$.

4. 设 A 的特征值全为 ± 1, 证明: A 与 A^{-1} 相似.

5. 利用矩阵函数, 求出递归数列

$$D_1, D_2, \cdots, D_n, \cdots, \quad D_n=3D_{n-1}-3D_{n-2}+D_{n-3} \quad (n>3)$$

的通项公式 $D_n=f(D_1,D_2,D_3)$.

(提示: 考察矩阵 $A=\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & -3 & 3\end{pmatrix}$)

§3 简单的矩阵方程

本节及以后会经常用到若尔当典范形. 为方便起见, 把 k 阶幂零矩阵记为

$$H_k \overset{\text{def}}{=} \begin{pmatrix}0 & 1 & 0 & \cdots & 0\\ 0 & 0 & 1 & \cdots & 0\\ \vdots & \vdots & \ddots & \ddots & \vdots\\ 0 & 0 & 0 & \ddots & 1\\ 0 & 0 & 0 & \cdots & 0\end{pmatrix} \in M_k(\mathbb{C}).$$

于是若尔当块

$$J_k(c)=cE_k+H_k.$$

方程 $AX=XB$

设 $A\in M_m(\mathbb{C})$, $B\in M_n(\mathbb{C})$. 使 A, B 相似于若尔当典范形:

$$J_A=U^{-1}AU, \quad J_B=V^{-1}BV.$$

设所求的解矩阵是 $X \in M_{m,n}(\mathbb{C})$, 代入原方程后可得

$$UJ_AU^{-1}X = XVJ_BV^{-1},$$

即

$$J_A(U^{-1}XV) = (U^{-1}XV)J_B. \tag{3.1}$$

因此方程 (3.1) 与原方程用相同的形式. 如果方程 (3.1) 的解矩阵是 $\tilde{X}$, 那么 $U\tilde{X}V^{-1}$ 就是原方程的解.

现在我们设 A 和 B 都是若尔当典范形. 并且设 A 的初等因子是 $(\lambda-\lambda_1)^{e_1},\cdots,(\lambda-\lambda_s)^{e_s}$, B 的初等因子是 $(\lambda-\mu_1)^{f_1},\cdots,(\lambda-\mu_t)^{f_t}$. 于是

$$A = \mathrm{diag}(\lambda_1E_{e_1}+H_{e_1},\cdots,\lambda_sE_{e_s}+H_{e_s}),$$
$$B = \mathrm{diag}(\mu_1E_{f_1}+H_{f_1},\cdots,\mu_tE_{f_t}+H_{f_t}).$$

把 X 写成相应的分块矩阵:

$$X = (X_{ij}),\quad X_{ij} \in M_{e_i,f_j}(\mathbb{C}),\ i = 1,\cdots,s,\ j = 1,\cdots,t.$$

代入原方程后, 得到等价的 st 个方程:

$$(\lambda_iE_{e_i}+H_{e_i})X_{ij} = X_{ij}(\mu_jE_{f_j}+H_{f_j}).$$

上述方程又可变形为

$$(\mu_j-\lambda_i)X_{ij} = H_{e_i}X_{ij} - X_{ij}H_{f_j}. \tag{3.2}$$

对任意取定的 i, j, 可分两种情况讨论方程 (3.2).

(1) $\lambda_i \neq \mu_j$. 重复应用公式 (3.2) $m-1$ 次, 可得

$$(\mu_j-\lambda_i)^mX_{ij} = \sum_{p=0}^{m}(-1)^{m-p}\mathrm{C}_m^pH_{e_i}^pX_{ij}H_{f_j}^{m-p}. \tag{3.3}$$

由于

$$H_{e_i}^{e_i} = H_{f_j}^{f_j} = 0,$$

当 $m = e_i + f_j$ 时, 在 (3.3) 式中 $p \geqslant e_i$ 或 $m-p \geqslant f_j$ 两者必居其一. 所以 (3.3) 式右端等于零. 从而 $X_{ij} = 0$.

(2) $\lambda_i = \mu_j$. 这时 (3.2) 式变成

$$H_{e_i}X_{ij} = X_{ij}H_{f_j}. \tag{3.4}$$

设 $X_{ij}=(x_{pq})$, (3.4) 式可展开为

$$\begin{pmatrix} x_{21} & x_{22} & x_{23} & \cdots & x_{2,f_j} \\ x_{31} & x_{32} & x_{33} & \cdots & x_{3,f_j} \\ x_{41} & x_{42} & x_{43} & \cdots & x_{4,f_j} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix} = \begin{pmatrix} 0 & x_{11} & x_{12} & \cdots & x_{1,f_j-1} \\ 0 & x_{21} & x_{22} & \cdots & x_{2,f_j-1} \\ 0 & x_{31} & x_{32} & \cdots & x_{3,f_j-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_{e_i,1} & x_{e_i,2} & \cdots & x_{e_i,f_j-1} \end{pmatrix}.$$

不难得到此矩阵方程的解.

当 $e_i=f_j$ 时:

$$X_{ij}=\begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_{e_i} \\ 0 & x_1 & x_2 & \cdots & x_{e_i-1} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & x_2 \\ 0 & 0 & 0 & \cdots & x_1 \end{pmatrix},$$

当 $e_i<f_j$ 时:

$$X_{ij}=\begin{pmatrix} 0 & \cdots & 0 & x_1 & x_2 & \cdots & x_{e_i} \\ 0 & \cdots & 0 & 0 & x_1 & \cdots & x_{e_i-1} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots & x_1 \end{pmatrix},$$

当 $e_i>f_j$ 时:

$$X_{ij}=\begin{pmatrix} x_1 & x_2 & \cdots & x_{f_j} \\ 0 & x_1 & \cdots & x_{f_j-1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

综合这两种情况, 可以知道解 X_{ij} 所含有的参数的个数等于:

$$\begin{cases} 0, & \lambda_i\neq\mu_j, \\ \min\{e_i,f_j\}, & \lambda_i=\mu_j. \end{cases} \tag{3.5}$$

把这些 X_{ij} 合在一起, 就得到解 $X=(X_{ij})$.

不难验证, 解的集合

$$C(A,B) \overset{\text{def}}{=} \{X \in M_{m,n}(\mathbb{C}) \mid AX = XB\}$$

是 $M_{m,n}(\mathbb{C})$ 中的一个线性子空间. 它的维数就是一般解 X 中所含参数的个数. 由 (3.5) 式可以得到以下结论.

命题 3.1 设 $A \in M_m(\mathbb{C})$, $B \in M_n(\mathbb{C})$, A, B 的初等因子组分别是 $(\lambda-\lambda_1)^{e_1}, \cdots, (\lambda-\lambda_s)^{e_s}$ 与 $(\lambda-\mu_1)^{f_1}, \cdots, (\lambda-\mu_t)^{f_t}$. 则矩阵方程 $AX = XB$ 的解空间 $C(A,B)$ 的维数是

$$\dim C(A,B) = \sum_{i=1}^{s} \sum_{\substack{j=1 \\ \mu_j=\lambda_i}}^{t} \min\{e_i, f_j\}. \quad \square$$

推论 3.2 矩阵方程 $AX = XB$ 有非零解的充分必要条件是 A 与 B 有公共的特征值. $\square$

例 3.1 解矩阵方程

$$\left(\begin{array}{cc|c|c} 1 & 1 & & \\ & 1 & & \\ \hline & & 1 & \\ \hline & & & 2 \end{array}\right) X = X \left(\begin{array}{c|ccc} 2 & & & \\ \hline & 1 & 1 & \\ & & 1 & 1 \\ & & & 1 \end{array}\right).$$

解:

$$X = \left(\begin{array}{c|ccc} 0 & 0 & a & b \\ 0 & 0 & 0 & a \\ \hline 0 & 0 & 0 & c \\ \hline d & 0 & 0 & 0 \end{array}\right), \quad a, b, c, d \in \mathbb{C}. \quad \square$$

矩阵的中心化子

与一个矩阵 $A \in M_n(\mathbb{C})$ 可交换的矩阵的集合称为 A 的**中心化子** (*centralizer*), 记为 $C(A)$. 即

$$C(A) = \{X \in M_n(\mathbb{C}) \mid AX = XA\}.$$

$C(A)$ 是 $M_n(\mathbb{C})$ 中的一个线性子空间. 要确定 $C(A)$ 中的矩阵相当于解矩阵方程 $AX = XA$. 关于矩阵的中心化子的维数有以下结论.

命题 3.3 设 $A \in M_n(\mathbb{C})$, A 的初等因子组是 $(\lambda-\lambda_1)^{e_1}, \cdots, (\lambda-\lambda_s)^{e_s}$. 则矩阵 A 的中心化子 $C(A)$ 的维数是

$$\dim C(A) = \sum_{i=1}^{s} \sum_{\substack{j=1 \\ \lambda_j=\lambda_i}}^{s} \min\{e_i, e_j\}. \quad \Box \tag{3.6}$$

从公式 (3.6) 可以得出 $n \leqslant \dim C(A) \leqslant n^2$. 我们观察两个极端的情形.

推论 3.4 (1) $\dim C(A) = n^2$, 即 $C(A) = M_n(\mathbb{C})$ 的充分必要条件是 A 为标量矩阵 aE;

(2) $\dim C(A) = n$ 当且仅当 A 的初等因子两两互素, 当且仅当 $\chi_A(\lambda) = m_A(\lambda)$.

证明: (1) 为使 $\dim C(A)$ 达到极大, 首先 A 的所有特征值都要相等, 这时 $\dim C(A) \leqslant s(e_1 + \cdots + e_s) = sn$. 因而 $s = n$, 即每个 $e_i = 1$. 使得 A 成为标量矩阵. 本结论的充分性是显然的.

(2) 根据公式 (3.6), $\dim C(A) = n = e_1 + \cdots + e_s$ 当且仅当 A 的各初等因子具有不同的特征值, 也就是两两互素. 这等价于 A 的极小多项式等于所有初等因子的乘积, 后者就是 A 的特征多项式. $\Box$

矩阵多项式方程

我们这里只讨论最简单的情形. 设

$$f(\lambda) = a_0\lambda^m + a_1\lambda^{m-1} + \cdots + a_{m-1}\lambda + a_m \in \mathbb{C}[\lambda],$$

求解矩阵方程

$$f(X) = 0, \quad X \in M_n(\mathbb{C}).$$

显然 $f(\lambda)$ 是 X 的化零多项式, 因此 $f(X) = 0$ 当且仅当 $m_X(\lambda) \mid f(\lambda)$. $f(\lambda)$ 的任何因式都可以作为 X 的极小多项式. 对于每个取定的极小多项式又可以确定 X 的所有可能的初等因子组. 对于每个选定的初等因子组可以写出它的若尔当典范形 J. 方程 $f(X) = 0$ 的所有可能的解都具有形式 TJT^{-1}, $T \in GL(n, \mathbb{C})$. 上述讨论可以归纳成以下命题.

命题 3.5 设

$$f(\lambda) = a_0(\lambda-\lambda_1)^{k_1}(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_t)^{k_t},$$

其中 λ_i 互不相同. 则 $f(X)=0$ 当且仅当 X 的每个初等因子均形如 $(\lambda-\lambda_i)^{e_i}$ $(e_i \leqslant k_i,\ i=1,\cdots,t)$. □

例 3.2 求解矩阵方程

$$(X-E)^2(X-2E)=0, \quad X\in M_3(\mathbb{C}).$$

解: $f(\lambda)=(\lambda-1)^2(\lambda-2)$, 因此解集是一些相似类的并集, 这些相似类的初等因子组是: (1) $\lambda-1,(\lambda-1)^2$; (2) $\lambda-1,\lambda-1,\lambda-1$; (3) $\lambda-2,\lambda-2,\lambda-2$; (4) $(\lambda-1)^2,\lambda-2$; (5) $\lambda-1,\lambda-1,\lambda-2$; (6) $\lambda-1,\lambda-2,\lambda-2$. □

WIMS **网上游戏** WIMS 的对谈式练习中有一个名为“交叉相乘” (Cross multiplication), 给出矩阵 A 和 C, 要你求矩阵方程 $AB-BA=C$ 的解 B. 试试你解矩阵方程的本领吧.

习 题 13–3

1. 设

$$A=U\left(\begin{array}{cc|cc|ccc} -1 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & -1 & 0 & 0 & 0 & 0 & 0\\ \hline 0 & 0 & 2 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 2 & 0 & 0 & 0\\ \hline 0 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)U^{-1},\ B=V\left(\begin{array}{c|cc|ccc} -1 & 0 & 0 & 0 & 0 & 0\\ \hline 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0\\ \hline 0 & 0 & 0 & 2 & 1 & 0\\ 0 & 0 & 0 & 0 & 2 & 1\\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right)V^{-1}.$$

求解 $AX=XB$.

2. 设

$$A=\left(\begin{array}{ccc|cc|c} 1 & 1 & 0 & 0 & 0 & 0\\ 0 & 1 & 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0\\ \hline 0 & 0 & 0 & 1 & 1 & 0\\ 0 & 0 & 0 & 0 & 1 & 0\\ \hline 0 & 0 & 0 & 0 & 0 & 2 \end{array}\right).$$

求 $C(A)$ 以及 $\dim C(A)$.

3. 写出矩阵方程

$$X^2-2X-3E=0, \quad X\in M_3(\mathbb{C}),$$

的解的初等因子组.

4. 不用命题 3.1 的方法证明推论 3.2.

§4 矩阵的广义逆

设有线性方程组

$$AX = B, \quad A \in M_{m,n}(\mathbb{C}),\ B \in M_{m,1}(\mathbb{C}), \tag{4.1}$$

当 $m = n$, A 可逆时. (4.1) 有唯一解 $X = A^{-1}B$. 而当 $m \neq n$, 或虽有 $m = n$, 但 A 退化时, 方程 (4.1) 不一定有解, 即使有解也不一定唯一. 不过我们可以降低要求, 寻找矩阵 $G \in M_{n,m}(\mathbb{C})$, 使得当 (4.1) 有解时 GB 一定是一个解. 这样的矩阵就被称为广义逆.

让我们分析一下. B 使 (4.1) 有解当且仅当存在 $C \in M_{n,1}(\mathbb{C})$ 使得 $B = AC$. 因此对于有解的 B 使得 GB 是解等价于对于任意的 $C \in M_{n,1}(\mathbb{C})$, $G(AC)$ 是解, 这又等价于 $AGAC = AC$ 对任意的 C 成立. 最后一个条件等价于 $AGA = A$. 因此有以下定义.

定义 4.1 设 $A \in M_{m,n}(\mathbb{C})$, 如果矩阵 $G \in M_{n,m}(\mathbb{C})$ 满足

$$AGA = A,$$

则称 G 是 A 的一个 $\{1\}$ **广义逆** (*generalized inverse*), 简称 $\{1\}$ 逆.

当 A 可逆时, 显然 $G = A^{-1}$. 因此广义逆确实是逆矩阵的推广.

下面我们要研究 $\{1\}$ 广义逆是否存在以及存在时如何计算.

先把 A 化成等价正规形

$$PAQ = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}, \quad P \in GL(m, \mathbb{C}),\ Q \in GL(n, \mathbb{C}).$$

则 $AGA = A$ 等价于

$$P^{-1} \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1}GP^{-1} \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1} = P^{-1} \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} Q^{-1},$$

记

$$Q^{-1}GP^{-1} = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix}, \quad X_{11} \in M_r(\mathbb{C}),\ X_{22} \in M_{n-r,m-r}(\mathbb{C}),$$

代入前式可得

$$\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{pmatrix} \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}.$$

比较左右两端后可以知道 $AGA = A$ 等价于 $X_{11} = E_r$, 也就是说

$$G = Q\begin{pmatrix} E_r & U \\ V & W \end{pmatrix} P \in M_{n,m}(\mathbb{C}).$$

这样就得到了以下命题.

命题 4.1 设 $A \in M_{m,n}(\mathbb{C})$ 是秩等于 r 的矩阵, 且有

$$PAQ = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}, \quad P \in GL(m,\mathbb{C}),\ Q \in GL(n,\mathbb{C}).$$

则 A 的 $\{1\}$ 逆是以下形式的矩阵:

$$Q\begin{pmatrix} E_r & U \\ V & W \end{pmatrix} P \in M_{n,m}(\mathbb{C}),$$

其中 $U \in M_{r,m-r}(\mathbb{C})$, $V \in M_{n-r,r}(\mathbb{C})$, $W \in M_{n-r,m-r}(\mathbb{C})$ 是任意的矩阵.

从这个命题可以看出矩阵 A 的 $\{1\}$ 逆一定存在, 而且一般不唯一. 当且仅当 A 是可逆矩阵时, A 的 $\{1\}$ 逆唯一, 就是逆矩阵 A^{-1}.

当 $r = m$, 即 A 是行满秩矩阵时, A 的 $\{1\}$ 逆具有形式

$$G = Q\begin{pmatrix} E_m \\ V \end{pmatrix} P,$$

于是

$$AG = P^{-1}(E_m\ 0)Q^{-1}Q\begin{pmatrix} E_m \\ V \end{pmatrix} P = E_m,$$

即 G 是 A 的右逆.

当 $r = n$, 即 A 是列满秩矩阵时, A 的 $\{1\}$ 逆具有形式

$$G = Q(E_n\ U)P,$$

于是

$$GA = Q(E_n\ U)PP^{-1}\begin{pmatrix} E_n \\ 0 \end{pmatrix} Q^{-1} = E_n,$$

即 G 是 A 的左逆.

上面的讨论有一个前提, 即方程 $AX = B$ 有解. 那么在方程不一定有解时又怎么办呢? 我们自然会想起用最小二乘解 (参见第五章 §6) 作为方程的最佳近似解. 因此有必要扩大广义逆的应用范围, 即找出矩阵 G, 使得 GB 成为方程 $AX = B$ 的最小二乘解. 我们先在实数域内讨论.

命题 4.2 设 $A \in M_{m,n}(\mathbb{R})$, $G \in M_{n,m}(\mathbb{R})$. 则以下两个性质是等价的:

(1) 对任意的列矩阵 $B \in M_{m,1}(\mathbb{R})$, $X = GB$ 一定是 $AX = B$ 的最小二乘解;

(2) $AGA = A$, $(AG)^{\mathrm{T}} = AG$.

证明: (1) $\Rightarrow$ (2) 因为当 $AX = B$ 有解时, 最小二乘解就是方程的解, 因此 G 应该是 A 的 $\{1\}$ 逆, 即满足 $AGA = A$. 再设 W 是 A 的列向量张成的 $\mathbb{R}^m$ 中的线性子空间(参见第四章 §6), $W^{\perp}$ 是 W 的正交补空间. 则最小二乘解 $X = GB$ 满足 $AX = AGB \in W$, $B - AX = B - AGB \in W^{\perp}$. 记 $P = AG$. 因此对于任意的列矩阵 $Y, Z \in M_{m,1}(\mathbb{R})$, 有 $PY \in W$, $(E_m - P)Z \in W^{\perp}$. 于是

$$(PY)^{\mathrm{T}}(E_m - P)Z = 0 \Longrightarrow P^{\mathrm{T}} - P^{\mathrm{T}}P = 0,$$

$$[(E_m - P)Z]^{\mathrm{T}}PY = 0 \Longrightarrow P - P^{\mathrm{T}}P = 0.$$

从而 $P^{\mathrm{T}} = P^{\mathrm{T}}P = P$.

(2) $\Rightarrow$ (1) 设 $X = GB$. 则

$$A^{\mathrm{T}}AX = A^{\mathrm{T}}AGB = [(AG)^{\mathrm{T}}A]^{\mathrm{T}}B = [AGA]^{\mathrm{T}}B = A^{\mathrm{T}}B.$$

根据第五章 §6 的讨论, X 是 $AX = B$ 的最小二乘解当且仅当 $A^{\mathrm{T}}AX = A^{\mathrm{T}}B$, 因此 GB 确是一个最小二乘解. □

如果读者学习过酉空间 (第八章 §6) 的话, 就知道欧几里得空间的内积可以推广到复数域上, 从而在复线性空间中也可以定义向量的长度和角度. 这样的复线性空间称为酉空间. 酉空间中两个列矩阵 $X, Y \in M_{m,1}(\mathbb{C})$ 间的内积是 $\overline{X}^{\mathrm{T}}Y$, 与欧几里得空间的内积相比, 多了一个共轭. 因此对于复矩阵的线性方程组 $AX = B$, 同样可以定义它的最小二乘解. 所以命题 4.2 对于复矩阵也适用, 唯一的改变是把条件 (2) 中的 $(AG)^{\mathrm{T}} = AG$ 改为 $\overline{(AG)}^{\mathrm{T}} = AG$. 以后我们就在更广泛的复矩阵范围内讨论广义逆.

根据以上讨论, 再增加一点条件, 就能得到以下更严格的广义逆的定义.

定义 4.2 设 $A \in M_{m,n}(\mathbb{C})$. 如果矩阵 $G \in M_{n,m}(\mathbb{C})$ 满足以下条件:

(1) $AGA = A$;

(2) $GAG = G$;

(3) $\overline{(AG)}^{\mathrm{T}} = AG$;

(4) $\overline{(GA)}^{\mathrm{T}} = GA$;

就称 G 是 A 的一个 **Moore-Penrose 广义逆**, 简称 M-P 逆.

为了证明 M-P 逆的存在性, 我们先对矩阵作满秩分解.

引理 4.3 设矩阵 $A \in M_{m,n}(\mathbb{C})$ 的秩等于 r, 则一定存在矩阵 $U \in M_{m,r}(\mathbb{C})$, $V \in M_{r,n}(\mathbb{C})$, 使得

$$A = UV.$$

这个分解称为 A 的**满秩分解**.

证明: 由于存在 $P \in GL(m,\mathbb{C})$, $Q \in GL(n,\mathbb{C})$, 使得

$$PAQ = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} E_r \\ 0 \end{pmatrix} \begin{pmatrix} E_r & 0 \end{pmatrix},$$

所以

$$A = \left[P^{-1}\begin{pmatrix} E_r \\ 0 \end{pmatrix}\right] \left[\begin{pmatrix} E_r & 0 \end{pmatrix} Q^{-1}\right]. \quad \square$$

当然实际作分解时不一定要象引理证明那样先化成对角形再分解. 只要取 A 的列向量的极大无关组作为矩阵 U, 再把 A 的列向量用这个极大无关组线性表示, 以表示系数作为列向量构成的矩阵就是 V. 当然也可以反过来, 把 A 的行向量的极大无关组作为矩阵 V, 再算出 U.

例 4.1 作出以下矩阵

$$A = \begin{pmatrix} 1 & -1 & -3 & 1 \\ -2 & 3 & 8 & -1 \\ 3 & 1 & -1 & 7 \\ -1 & 0 & 1 & -2 \end{pmatrix}$$

的满秩分解.

解: 把 A 的列向量组记为 $\alpha_1, \alpha_2, \alpha_3, \alpha_4$, α_1, α_2 是这个向量组的极大无关组, 且

$$\alpha_3 = -\alpha_1 + 2\alpha_2, \quad \alpha_4 = 2\alpha_1 + \alpha_2,$$

因此 $\operatorname{rank} A = 2$. 以 α_1, α_2 作为列向量可得矩阵 U, 以 A 的相应列向量关于这个极大无关组的表示系数作为列向量可得矩阵 V:

$$A = UV = \begin{pmatrix} 1 & -1 \\ -2 & 3 \\ 3 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & 2 & 1 \end{pmatrix}. \quad \square$$

定理 4.4 对于任意的矩阵 $A \in M_{m,n}(\mathbb{C})$, A 的 M-P 逆存在且唯一, 记为 A^+. 对于 A 的任意满秩分解 $A = UV$, 有

$$A^+ = \overline{V}^{\mathrm{T}}(V\overline{V}^{\mathrm{T}})^{-1}(\overline{U}^{\mathrm{T}}U)^{-1}\overline{U}^{\mathrm{T}}. \tag{4.2}$$

证明: 先证唯一性. 设 G, H 都是 A 的 M-P 逆. 则

$$\begin{aligned} G &= (GA)G = \overline{A}^{\mathrm{T}}\overline{G}^{\mathrm{T}}G = (\overline{A}^{\mathrm{T}}\overline{H}^{\mathrm{T}}\overline{A}^{\mathrm{T}})\overline{G}^{\mathrm{T}}G \\ &= HA\overline{A}^{\mathrm{T}}\overline{G}^{\mathrm{T}}G = HAGAG = HAG, \\ H &= H(AH) = H\overline{H}^{\mathrm{T}}\overline{A}^{\mathrm{T}} = H\overline{H}^{\mathrm{T}}(\overline{A}^{\mathrm{T}}\overline{G}^{\mathrm{T}}\overline{A}^{\mathrm{T}}) \\ &= H\overline{H}^{\mathrm{T}}\overline{A}^{\mathrm{T}}AG = HAHAG = HAG, \end{aligned}$$

从而 $G = H$.

矩阵 U 的列向量线性无关, 矩阵 $\overline{U}^{\mathrm{T}}U$ 可以被看成酉空间中线性无关向量组的格拉姆矩阵 (参见第五章习题 5–3 的第 10 题), 因而行列式不等于 0. 同理矩阵 $V\overline{V}^{\mathrm{T}}$ 的行列式也不等于 0. 所以这两个矩阵都可逆, 公式 (4.2) 是有意义的. 验证 A^+ 满足定义 4.2 的 4 条性质留给读者作为练习. $\square$

从命题 4.2 可以知道, 用 M-P 逆导出的 A^+B 一定是方程 $AX = B$ 的一个最小二乘解. 这个结论实际上只用到了定义 4.2 中的条件 (1) 与 (3). 当最小二乘解不唯一时, 用 M-P 逆导出的解只是其中的一个解. 实际上条件 (4) 保证了 A^+B 是最小二乘解中长度最小的解.

例 4.2 用 $I_{m,n}$ 表示元素全为 1 的 $m \times n$ 矩阵. 则 $\operatorname{rank} I_{m,n} = 1$, 其满秩分解为

$$I_{m,n} = I_{m,1}I_{1,n}.$$

所以

$$I_{m,n}^+ = I_{n,1}(I_{1,n}I_{n,1})^{-1}(I_{1,m}I_{m,1})^{-1}I_{1,m} = \frac{1}{mn}I_{n,m}. \quad \square$$

例 4.3 设有分块矩阵

$$A=\begin{pmatrix} A_1 & 0 \\ 0 & 0 \end{pmatrix}\in M_{m,n}(\mathbb{C}),\ A_1\in GL(r,\mathbb{C}).$$

则 $\operatorname{rank} A=r$, 其满秩分解 $A=UV$ 为

$$U=\begin{pmatrix} A_1 \\ 0 \end{pmatrix},\quad V=(\,E_r\quad 0\,).$$

由

$$\overline{U}^{\mathrm{T}}U=\overline{A_1}^{\mathrm{T}}A_1,\quad V\overline{V}^{\mathrm{T}}=E_r,$$

可得

$$A^+=\overline{V}^{\mathrm{T}}(V\overline{V}^{\mathrm{T}})^{-1}(\overline{U}^{\mathrm{T}}U)^{-1}\overline{U}^{\mathrm{T}}=\begin{pmatrix} A_1^{-1} & 0 \\ 0 & 0 \end{pmatrix}.\quad\square$$

例 4.4 设 $A=\operatorname{diag}(a_1,a_2,\cdots,a_n)$, 则

$$A^+=\operatorname{diag}(a_1^+,a_2^+,\cdots,a_n^+),$$

这里对于一个数 a, 记

$$a^+=\begin{cases} 0, & \text{当 } a=0; \\ a^{-1}, & \text{当 } a\neq 0. \end{cases}$$

这和把数 a 当成 1×1 矩阵时的 M-P 逆完全一样. $\square$

网上游戏 WIMS 的对谈式练习中有一个"满秩分解" (Rankmult), 要你求出一个矩阵的满秩分解式.

习 题 13–4

1. 设 $A\in M_{m,n}(\mathbb{C})$, G 是 A 的 $\{1\}$ 逆. 如果对于 $B\in M_{m,1}(\mathbb{C})$, 方程 $AX=B$ 有解. 令

$$X=GB+(E_n-GA)Z$$

其中 $Z\in M_{n,1}(\mathbb{C})$ 是任意的列矩阵. 验证 X 是 $AX=B$ 的解. 你能证明方程的所有解都能表示成这种形式吗?

2. 验证 (4.2) 式定义的 A^+ 确实是 A 的 M-P 逆.

3. 证明 M-P 逆的以下性质:

(1) $(A^+)^+=A$;

(2) $(AA^+)^2=AA^+$;

(3) $(A^+A)^2 = A^+A$;

4. 计算

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

的 M-P 逆.

5. 利用 M-P 逆求以下方程组的最小二乘解:

$$\begin{cases} x_1 - x_2 + 2x_3 = 3 \\ -x_1 + 2x_2 - 3x_3 + x_4 = 6 \\ x_2 - x_3 + x_4 = 0. \end{cases}$$

6. 设

$$A = U\begin{pmatrix} \mathrm{diag}(a_1, \cdots, a_r) & 0 \\ 0 & 0 \end{pmatrix} V \in M_{m,n}(\mathbb{R}), \quad U \in O(m, \mathbb{R}),\ V \in O(n, \mathbb{R}).$$

证明

$$A^+ = V^{\mathrm{T}}\begin{pmatrix} \mathrm{diag}(a_1^+, \cdots, a_r^+) & 0 \\ 0 & 0 \end{pmatrix} U^{\mathrm{T}}.$$

7. 举例说明 $(AB)^+ = B^+A^+$ 不一定正确.

§ 5　矩阵特征值的范围

我们先证明一个判别矩阵可逆性的阿达马 (Jacques-Salomon Hadamard, 1865–1963, 法国人) 判则.

定理 5.1 如果矩阵 $A = (a_{ij}) \in M_n(\mathbb{C})$ 满足以下的阿达马条件:

$$H_i \overset{\text{def}}{=} |a_{ii}| - \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| > 0, \quad i = 1, \cdots, n,$$

则矩阵 A 非退化.

证明: 如果行列式 $\det A = 0$, 则存在不全为 0 的 $x_1, \cdots, x_n \in \mathbb{C}$ 成为以下齐次线性方程组的解:

$$\sum_{j=1}^{n} a_{ij}x_j = 0, \quad i = 1, \cdots, n.$$

设 $|x_i| = \max\{|x_1|, \cdots, |x_n|\} > 0$, 则

$$|a_{ii}||x_i| \leqslant \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}||x_j| \leqslant |x_i| \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|.$$

因此

$$|a_{ii}| \leqslant \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|,$$

这与阿达马条件矛盾. □

条件 $H_i > 0$ 意味着对角元 a_{ii} 的模严格大于第 i 行其余元素的模之和. 这样的元素 a_{ii} 称为**支配的** (*dominant*). 阿达马条件要求矩阵的所有的对角元都是支配的. 这种矩阵也称为对角占优的.

推论 5.2 如果矩阵 A 满足阿达马条件, 则

$$|\det A| \geqslant H_1 H_2 \cdots H_n > 0.$$

证明: 引入辅助矩阵 $B = (b_{ij})$, 其中

$$b_{ij} = \frac{a_{ij}}{H_i}, \quad i, j = 1, \cdots, n.$$

显然有

$$|b_{ii}| - \sum_{\substack{j=1\\ j\neq i}}^{n} |b_{ij}| = 1, \quad i = 1, \cdots, n.$$

用 λ_0 表示矩阵 B 的任意特征值, λ_0 对应的特征向量设为 $(x_1, \cdots, x_n)$, 其中模最大的设为 $|x_k| > 0$, 则

$$\lambda_0 x_k = \sum_{j=1}^{n} b_{kj} x_j.$$

于是

$$|\lambda_0||x_k| \geqslant |b_{kk}||x_k| - \sum_{\substack{j=1\\ j\neq i}}^{n} |b_{kj}||x_j| \geqslant |x_k| \left(|b_{kk}| - \sum_{\substack{j=1\\ j\neq i}}^{n} |b_{kj}| \right) = |x_k|.$$

除以 $|x_k|$ 后可得

$$|\lambda_0| \geqslant 1.$$

而行列式 $\det B$ 的值等于 B 的所有特征值的乘积, 这些特征值的模都大于等于 1, 因此

$$|\det B| \geqslant 1.$$

另一方面, 由定义有

$$\det B = \frac{\det A}{H_1 H_2 \cdots H_n},$$

所以

$$|\det A| \geqslant H_1 H_2 \cdots H_n. \quad \square$$

在满足阿达马条件、且有相同的 $H_1, \cdots, H_n$ 的值的矩阵类中, 上述估计是最好的. 只需取 $A = \operatorname{diag}(H_1, \cdots, H_n)$ 就可看出.

另一方面由于 $\det A = \det A^{\mathrm{T}}$, 我们可以得到关于列的阿达马条件:

$$G_i \stackrel{\text{def}}{=} |a_{ii}| - \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ji}| > 0, \quad i = 1, \cdots, n,$$

$$|\det A| \geqslant G_1 G_2 \cdots G_n.$$

利用阿达马判则, 可以得到对矩阵特征值的估计.

定理 5.3 (盖施戈林 (Gershgorin) 圆盘定理) *设 λ_0 是矩阵 $A = (a_{ij}) \in M_n(\mathbb{C})$ 的特征值, 则存在某个 $1 \leqslant i \leqslant n$ 使得*

$$|a_{ii} - \lambda_0| \leqslant \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|.$$

*即 λ_0 落在某个以 a_{ii} 为圆心的圆内, 这样的圆称为***盖施戈林圆**.

证明: 因为 $|A - \lambda_0 E| = 0$, 所以对这个矩阵, 阿达马条件至少有一个不成立. 这就是定理的结论. $\square$

利用关于列的阿达马不等式, 可得另外 n 个圆:

$$|a_{ii} - \lambda_0| \leqslant \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ji}|, \quad i = 1, \cdots, n.$$

特别当矩阵 A 是实对称矩阵时, 它的特征值都是实数, 这时盖施戈林圆成为实数轴上的区间.

习 题 13–5

1. 证明: 如果对称矩阵 $A=(a_{ij})\in M_n(\mathbb{R})$ 满足条件

$$a_{ii}>\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|>0,\quad i=1,\cdots,n,$$

则 A 是正定矩阵.

2. 设

$$A=\begin{pmatrix} \dfrac{n^2}{2} & 1 & 2 & \cdots & n-2 & n-1\\ n-1 & \dfrac{n^2}{2} & 1 & \cdots & n-3 & n-2\\ n-2 & n-1 & \dfrac{n^2}{2} & \cdots & n-4 & n-3\\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 1 & 2 & 3 & \cdots & n-1 & \dfrac{n^2}{2}\end{pmatrix},$$

证明: $\det A\geqslant\left(\dfrac{n}{2}\right)^n$.