得 $\boldsymbol{A}$ 的特征值 $\lambda_1=\lambda_2=1$，为二重特征值。令特征向量为 $\boldsymbol{u}=[u_1,u_2]^{\mathrm{T}}$，则由 $\boldsymbol{A}\boldsymbol{u}=\lambda\boldsymbol{u}$ 及 $\lambda=1$，可列出方程组

$$u_1+u_2=u_1$$

$$u_2=u_2$$

解之，得 $u_2=0$，u_1 任意。换言之，若 $\alpha\neq 0$，则 $\alpha\begin{bmatrix}1\\0\end{bmatrix}$ 都是矩阵 $\boldsymbol{A}$ 的特征向量，即 $\boldsymbol{A}$ 的任何两个特征向量都是线性相关的。这意味着，矩阵 $\boldsymbol{T}$ 必定是奇异矩阵。因此，不可能通过 $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}$ 将例中的矩阵 $\boldsymbol{A}$ 对角化。事实上，对于特征值 $\lambda=1$ 而言，其多重度 $m_1=2$。因此，根据定理 8.2.5 知，只有当 $\mathrm{rank}(\boldsymbol{A}-\lambda\boldsymbol{I})=n-m_1=2-2=0$ 时，矩阵 $\boldsymbol{A}$ 是可对角化的。然而，现在由于

$$\boldsymbol{A}-\lambda\boldsymbol{I}=\begin{bmatrix}1&1\\0&1\end{bmatrix}-\begin{bmatrix}1&0\\0&1\end{bmatrix}=\begin{bmatrix}0&1\\0&0\end{bmatrix}$$

其秩等于 1，与定理 8.2.5 的条件直接相矛盾，所以检验定理 8.2.5 的条件直接知，本例的矩阵不可对角化。

8.3 Cayley-Hamilton 定理及其应用

如上节所述，一个 $n\times n$ 一般矩阵 $\boldsymbol{A}$ 的特征值由特征多项式 $\det(\boldsymbol{A}-\lambda\boldsymbol{I})$ 决定。不仅如此，特征多项式还与矩阵的求逆、矩阵幂和矩阵指数函数的计算密切相关，所以有必要对特征多项式作更深入的讨论与分析。

8.3.1 Cayley-Hamilton 定理

Cayley-Hamilton 定理是关于一般矩阵的特征多项式的重要结果。从这一定理出发，很容易解决矩阵的求逆、矩阵幂和矩阵指数函数的计算等问题。为了引出 Cayley-Hamilton 定理，先介绍几个与多项式有关的重要概念。

当 $p_n\neq 0$ 时，n 称为多项式 $p(x)=p_nx^n+p_{n-1}x^{n-1}+\cdots+p_1x+p_0$ 的阶数。一个 n 阶多项式称为首一多项式 (monic polynomial)，若 x^n 的系数等于 1。

若 $p(\boldsymbol{A})=p_n\boldsymbol{A}^n+p_{n-1}\boldsymbol{A}^{n-1}+\cdots+p_1\boldsymbol{A}+p_0\boldsymbol{I}=\boldsymbol{O}$，则称 $p(x)=p_nx^n+p_{n-1}x^{n-1}+\cdots+p_1x+p_0$ 是使矩阵 $\boldsymbol{A}$ 零化的多项式，简称零化多项式 (annihilating polynomial)。

对于一个 $n\times n$ 矩阵 $\boldsymbol{A}$，令 m 是使得幂 $\boldsymbol{I},\boldsymbol{A},\cdots,\boldsymbol{A}^m$ 线性相关的最小整数。于是，有方程式

$$p_m\boldsymbol{A}^m+p_{m-1}\boldsymbol{A}^{m-1}+\cdots+p_1\boldsymbol{A}+p_0\boldsymbol{I}=\boldsymbol{O}_{n\times n} \tag{8.3.1}$$

式中，$\boldsymbol{A}^m$ 的系数不为零。多项式 $p(x)=p_mx^m+p_{m-1}x^{m-1}+\cdots+p_1x+p_0$ 称为矩阵 $\boldsymbol{A}$ 的最小多项式。

下面的定理表明，特征多项式 $p(x)=\det(\boldsymbol{A}-x\boldsymbol{I})$ 是使矩阵 $\boldsymbol{A}_{n\times n}$ 零化的多项式。

定理 8.3.1 (Cayley-Hamilton 定理)　每一个正方矩阵 $\boldsymbol{A}_{n\times n}$ 都满足其特征方程，即若特征多项式具有式 (8.1.7) 的形式，则

$$p_n\boldsymbol{A}^n + p_{n-1}\boldsymbol{A}^{n-1} + \cdots + p_1\boldsymbol{A} + p_0\boldsymbol{I} = \boldsymbol{O} \tag{8.3.2}$$

式中，$\boldsymbol{I}$ 和 $\boldsymbol{O}$ 分别为 $n\times n$ 单位矩阵和零矩阵。

证明　逆矩阵的定义公式 $\boldsymbol{B}^{-1} = \dfrac{1}{\det(\boldsymbol{B})}\mathrm{adj}(\boldsymbol{B})$ 可以等价写作 $\boldsymbol{B}\mathrm{adj}(\boldsymbol{B}) = \det(\boldsymbol{B})\boldsymbol{I}_n$，故有

$$(\boldsymbol{A} - x\boldsymbol{I})\mathrm{adj}(\boldsymbol{A} - x\boldsymbol{I}) = \det(\boldsymbol{A} - x\boldsymbol{I})\boldsymbol{I}$$

将 $p(x) = \det(\boldsymbol{A} - x\boldsymbol{I}) = p_nx^n + p_{n-1}x^{n-1} + \cdots + p_1x + p_0$ 代入上式，便有

$$(\boldsymbol{A} - x\boldsymbol{I})\mathrm{adj}(\boldsymbol{A} - x\boldsymbol{I}) = (p_nx^n + p_{n-1}x^{n-1} + \cdots + p_1x + p_0)\boldsymbol{I} \tag{1}$$

上式表明，伴随矩阵 $\mathrm{adj}(\boldsymbol{A} - x\boldsymbol{I})$ 必然是一个关于 x 的 $n-1$ 次矩阵多项式，不妨令其为

$$\mathrm{adj}(\boldsymbol{A} - x\boldsymbol{I}) = x^{n-1}\boldsymbol{B}_{n-1} + x^{n-2}\boldsymbol{B}_{n-2} + \cdots + x\boldsymbol{B}_1 + \boldsymbol{B}_0 \tag{2}$$

式中，$\boldsymbol{B}_{n-1}, \boldsymbol{B}_{n-2}, \cdots, \boldsymbol{B}_1$ 均为 $n\times n$ 常数矩阵。将式 (2) 代入式 (1)，得

$$\begin{aligned}&(\boldsymbol{A} - x\boldsymbol{I})(x^{n-1}\boldsymbol{B}_{n-1} + x^{n-2}\boldsymbol{B}_{n-2} + \cdots + x\boldsymbol{B}_1 + \boldsymbol{B}_0)\\ =&(p_nx^n + p_{n-1}x^{n-1} + \cdots + p_1x + p_0)\boldsymbol{I}\end{aligned}$$

比较上式两边同幂次项 x^k 的系数，即可得到 $n+1$ 个方程：

$$\begin{aligned}-\boldsymbol{B}_{n-1} &= p_n\boldsymbol{I}\\ -B_{n-2} + \boldsymbol{A}\boldsymbol{B}_{n-1} &= p_{n-1}\boldsymbol{I}\\ &\cdots\\ -\boldsymbol{B}_0 + \boldsymbol{A}\boldsymbol{B}_1 &= p_1\boldsymbol{I}\\ -\boldsymbol{A}\boldsymbol{B}_0 &= p_0\boldsymbol{I}\end{aligned}$$

在上述前 n 个方程中，两边分别左乘矩阵 $\boldsymbol{A}^n, \boldsymbol{A}^{n-1}, \cdots, \boldsymbol{A}$，然后将所有 $n+1$ 个方程相加，立即有 $\boldsymbol{O} = p_n\boldsymbol{A}^n + p_{n-1}\boldsymbol{A}^{n-2} + \cdots + p_1\boldsymbol{A} + p_0\boldsymbol{I}$，这正是所期望的结果。 ■

以上证明是大多数文献采用的传统证明方法。最近，Jain 与 Gunawardena 从矩阵分解的角度，给出了另外一种证明。对此证明感兴趣的读者可参考文献 [236, p.180]。

Cayley-Hamilton 定理有很多非常有趣和重要的应用。例如，利用 Cayley-Hamilton 定理，也能够直接证明两个相似矩阵具有相同的特征值。

考查两个相似矩阵的特征多项式。令 $\boldsymbol{B} = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$ 是 $\boldsymbol{A}$ 的相似矩阵，并且已知矩阵 $\boldsymbol{A}$ 的特征多项式 $p(\boldsymbol{x}) = \det(\boldsymbol{A} - x\boldsymbol{I}) = p_nx^n + p_{n-1}x^{n-1} + \cdots + p_1x + p_0$。根据 Cayley-Hamilton 定理知 $p(\boldsymbol{A}) = p_n\boldsymbol{A}^n + p_{n-1}\boldsymbol{A}^{n-1} + \cdots + p_1\boldsymbol{A} + p_0\boldsymbol{I} = \boldsymbol{O}$。

对于相似矩阵 $\boldsymbol{B}$，由于

$$\boldsymbol{B}^k = (\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S})(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S})\cdots(\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}) = \boldsymbol{S}^{-1}\boldsymbol{A}^k\boldsymbol{S}$$

故有

$$
\begin{aligned}
p(\boldsymbol{B}) &= p_n\boldsymbol{B}^n + p_{n-1}\boldsymbol{B}^{n-1} + \cdots + p_1\boldsymbol{B} + p_0\boldsymbol{I} \\
&= p_n\boldsymbol{S}^{-1}\boldsymbol{A}^n\boldsymbol{S} + p_{n-1}\boldsymbol{S}^{-1}\boldsymbol{A}^{n-1}\boldsymbol{S} + \cdots + p_1\boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S} + p_0\boldsymbol{I} \\
&= \boldsymbol{S}^{-1}\left(p_n\boldsymbol{A}^n + p_{n-1}\boldsymbol{A}^{n-1} + \cdots + p_1\boldsymbol{A} + p_0\boldsymbol{I}\right)\boldsymbol{S} \\
&= \boldsymbol{S}^{-1}p(\boldsymbol{A})\boldsymbol{S} \\
&= \boldsymbol{O}
\end{aligned}
$$

在得到最后一个式子时，代入了 Cayley-Hamilton 定理的结果 $p(\boldsymbol{A}) = \boldsymbol{O}$。换言之，两个相似矩阵具有相同的特征多项式，从而它们具有相同的特征值。

下面再介绍 Cayley-Hamilton 定理的几个重要应用。

8.3.2 逆矩阵和广义逆矩阵的计算

若矩阵 $\boldsymbol{A}_{n\times n}$ 非奇异，则用 $\boldsymbol{A}^{-1}$ 右乘 (或左乘) 式 (8.3.2) 两边，立即有

$$
p_n\boldsymbol{A}^{n-1} + p_{n-1}\boldsymbol{A}^{n-2} + \cdots + p_2\boldsymbol{A} + p_1\boldsymbol{I} + p_0\boldsymbol{A}^{-1} = \boldsymbol{O}
$$

由此即可得到逆矩阵的计算公式

$$
\boldsymbol{A}^{-1} = -\frac{1}{p_0}(p_n\boldsymbol{A}^{n-1} + p_{n-1}\boldsymbol{A}^{n-2} + \cdots + p_2\boldsymbol{A} + p_1\boldsymbol{I}) \tag{8.3.3}
$$

例 8.3.1 已知矩阵

$$
\boldsymbol{A} = \begin{bmatrix} 1 & 5 \\ 4 & 6 \end{bmatrix}
$$

其特征多项式为

$$
\det(\boldsymbol{A} - x\boldsymbol{I}) = \begin{vmatrix} 1-x & 5 \\ 4 & 6-x \end{vmatrix} = (1-x)(6-x) - 5\times 4 = x^2 - 7x - 14
$$

即 $p_0 = -14$, $p_1 = -7$，$p_2 = 1$。将这些值代入式 (8.3.3)，立即得

$$
\boldsymbol{A}^{-1} = \frac{1}{14}(\boldsymbol{A} - 7\boldsymbol{I}) = \frac{1}{14}\left(\begin{bmatrix} 1 & 5 \\ 4 & 6 \end{bmatrix} - 7\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} -\frac{3}{7} & \frac{5}{14} \\ \frac{2}{7} & -\frac{1}{14} \end{bmatrix}
$$

例 8.3.2 由矩阵

$$
\boldsymbol{A} = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}
$$

得矩阵的二次幂

$$
\boldsymbol{A}^2 = \begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} 2 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 5 \\ 0 & 4 & 0 \\ 0 & 0 & 9 \end{bmatrix}
$$

和特征多项式

$$\det(\boldsymbol{A}-x\boldsymbol{I})=\begin{vmatrix}2-x & 0 & 1\\ 0 & 2-x & 0\\ 0 & 0 & 3-x\end{vmatrix}=(2-x)^2(3-x)$$
$$=-x^3+7x^2-16x+12$$

即 $p_0=12,\ p_1=-16,\ p_2=7,\ p_3=-1$。将这些值连同矩阵 $\boldsymbol{A},\boldsymbol{A}^2$ 一起代入矩阵求逆公式 (8.3.3)，则有

$$\begin{aligned}\boldsymbol{A}^{-1}&=-\frac{1}{12}\left(-\boldsymbol{A}^2+7\boldsymbol{A}-16\boldsymbol{I}\right)\\&=-\frac{1}{12}\left(-\begin{bmatrix}4&0&5\\0&4&0\\0&0&9\end{bmatrix}+7\begin{bmatrix}2&0&1\\0&2&0\\0&0&3\end{bmatrix}-16\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}\right)\\&=\frac{1}{6}\begin{bmatrix}3&0&-1\\0&3&0\\0&0&2\end{bmatrix}\end{aligned}$$

Cayley-Hamilton 定理还可以用于求任意一个复矩阵的广义逆矩阵。这一结果是 Decell 于 1965 年得到的 [116]。

定理 8.3.2 令矩阵 $\boldsymbol{A}$ 是任意一个 $m\times n$ 矩阵，并且令

$$f(\lambda)=(-1)^m(a_0\lambda^m+a_1\lambda^{m-1}+\cdots+a_{m-1}\lambda+a_m),\quad a_0=1 \tag{8.3.4}$$

是矩阵乘积 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的特征多项式 $(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}-\lambda\boldsymbol{I})$。若 k 是满足 $a_k\neq 0$ 的最大整数，则 $\boldsymbol{A}$ 的广义逆矩阵由

$$\boldsymbol{A}^{\dagger}=-a_k^{-1}\boldsymbol{A}^{\mathrm{H}}\left[(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{k-1}+a_1(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{k-2}+\cdots+a_{k-2}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})+a_{k-1}\boldsymbol{I}\right] \tag{8.3.5}$$

确定。当 $k=0$ 是使 $a_k\neq 0$ 的最大整数时，广义逆矩阵 $\boldsymbol{A}^{\dagger}=\boldsymbol{O}$。

证明 参见文献 [116]。

在有些文献 (例如文献 [425]) 中，称上述定理为 Decell 定理。注意，Decell 定理中的整数 k 就是矩阵 $\boldsymbol{A}$ 的秩，即 $k=\mathrm{rang}(\boldsymbol{A})$。

根据上述定理，Decell 提出了计算 Moore-Penrose 逆矩阵的下列方法 [116]。

(1) 构造矩阵序列 $\boldsymbol{A}_0,\boldsymbol{A}_1,\cdots,\boldsymbol{A}_k$

$$\begin{array}{lll}\boldsymbol{A}_0=\boldsymbol{O}, & -1=q_0, & \boldsymbol{B}_0=\boldsymbol{I};\\ \boldsymbol{A}_1=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}, & \mathrm{tr}(\boldsymbol{A}_1)=q_1, & \boldsymbol{B}_1=\boldsymbol{A}_1-q_1\boldsymbol{I};\\ \boldsymbol{A}_2=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}_1, & \dfrac{\mathrm{tr}(\boldsymbol{A}_2)}{2}=q_2, & \boldsymbol{B}_2=\boldsymbol{A}_2-q_2\boldsymbol{I};\\ \cdots & \cdots & \cdots\\ \boldsymbol{A}_{k-1}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}_{k-2}, & \dfrac{\mathrm{tr}(\boldsymbol{A}_{k-1})}{k-1}=q_{k-1}, & \boldsymbol{B}_{k-1}=\boldsymbol{A}_{k-1}-q_{k-1}\boldsymbol{I};\\ \boldsymbol{A}_k=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}_{k-1}, & \dfrac{\mathrm{tr}(\boldsymbol{A}_k)}{k}=q_k, & \boldsymbol{B}_k=\boldsymbol{A}_k-q_k\boldsymbol{I}\end{array}$$

Faddeev 证明了 [149, pp.260~265]，按照以上方法构造的系数 $q_i = -a_i, i = 1, 2, \cdots, k$。

(2) 计算 Moore-Penrose 逆矩阵

$$\begin{aligned}\boldsymbol{A}^{\dagger} &= -a_k^{-1}\boldsymbol{A}^{\mathrm{H}}\left[(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{k-1} + a_1(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{k-2} + \cdots + a_{k-2}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{k} + a_{k-1}\boldsymbol{I}\right] \\ &= -a_k^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{B}_{k-1}\end{aligned} \tag{8.3.6}$$

8.3.3 矩阵幂的计算

给定任意一个矩阵 $\boldsymbol{A}_{n\times n}$ 和一个整数 k，称 $\boldsymbol{A}^k$ 是矩阵 $\boldsymbol{A}$ 的 k 次幂。如果 k 比较大，显然矩阵幂 $\boldsymbol{A}^k$ 的计算是一件很繁琐的事。幸运的是，Cayley-Hamilton 定理为这个问题提供了一种简单的解决方法。

考查多项式除法 $f(x)/g(x)$，其中，$g(x) \neq 0$。根据 Euclidean 除法知，存在两个多项式 $q(x)$ 和 $r(x)$，使得

$$f(x) = g(x)q(x) + r(x)$$

式中，$q(x)$ 和 $r(x)$ 分别称为商和余项，并且余项 $r(x)$ 的阶数小于 $g(x)$ 的阶数或 $r(x) = 0$。

令矩阵 $\boldsymbol{A}$ 的特征多项式为

$$p(x) = p_n x^n + p_{n-1}x^{n-1} + \cdots + p_1 x + p_0$$

则对于任何一个 x，其 K 次幂

$$x^K = p(x)q(x) + r(x) \tag{8.3.7}$$

当 x 是特征方程 $p(x) = 0$ 的一个根时，上式变为

$$x^K = r(x) = r_0 + r_1 x + \cdots + r_{n-1}x^{n-1} \tag{8.3.8}$$

因为 $r(x)$ 的阶数小于 $p(x)$ 的阶数 n。

用 $\boldsymbol{A}$ 代替标量 x，则式 (8.3.7) 变为

$$\boldsymbol{A}^K = p(\boldsymbol{A})q(\boldsymbol{A}) + r(\boldsymbol{A}) \tag{8.3.9}$$

根据 Cayley-Hamilton 定理知，若 $p(x)$ 是矩阵 $\boldsymbol{A}$ 的特征多项式，则 $p(\boldsymbol{A}) = \boldsymbol{O}$。因此，式 (8.3.9) 简化为

$$\boldsymbol{A}^K = r(\boldsymbol{A}) = r_0\boldsymbol{I} + r_1\boldsymbol{A} + \cdots + r_{n-1}\boldsymbol{A}^{n-1} \tag{8.3.10}$$

式 (8.3.10) 给出了计算矩阵幂 $\boldsymbol{A}^K$ 的方法。

算法 8.3.1 (矩阵幂的计算)[236, p.172]

步骤 1　构造特征多项式

$$p(x) = \det(\boldsymbol{A} - x\boldsymbol{I}) = p_n x^n + p_{n-1}x^{n-1} + \cdots + p_1 x + p_0 \tag{8.3.11}$$

步骤 2 计算特征方程 $p(x)=0$ 的 n 个特征根即特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$。

步骤 3 将特征值代入式 (8.3.8)，得到一组线性方程：

$$\left.\begin{array}{c} r_0+\lambda_1 r_1+\cdots+\lambda_1^{n-1}r_{n-1}=\lambda_1^K \\ r_0+\lambda_2 r_1+\cdots+\lambda_2^{n-1}r_{n-1}=\lambda_2^K \\ \cdots \\ r_0+\lambda_n r_1+\cdots+\lambda_n^{n-1}r_{n-1}=\lambda_n^K \end{array}\right\} \tag{8.3.12}$$

解之，得 $r_0,r_1,\cdots,r_{n-1}$。

步骤 4 计算矩阵幂

$$\boldsymbol{A}^K=r_0\boldsymbol{I}+r_1\boldsymbol{A}+\cdots+r_{n-1}\boldsymbol{A}^{n-1}$$

下面举例加以说明。

例 8.3.3 已知

$$\boldsymbol{A}=\begin{bmatrix}1 & 1/2\\ 2 & 1\end{bmatrix}$$

计算 $\boldsymbol{A}^{731}$。

解 利用式 (8.3.11) 构造特征多项式

$$p(x)=\det(\boldsymbol{A}-x\boldsymbol{I})=\begin{vmatrix}1-x & 1/2\\ 2 & 1-x\end{vmatrix}=x^2-2x$$

令 $p(x)=0$，求出 $\boldsymbol{A}$ 的特征值 $\lambda_1=0$ 和 $\lambda_2=2$。

将特征值代入式 (8.3.12) 得线性方程组

$$r_0+0r_1=0^{731}$$

$$r_0+2r_1=2^{731}$$

解之，得 $r_0=0$ 和 $r_1=2^{730}$。

计算矩阵幂，得

$$\boldsymbol{A}^{731}=2^{730}\boldsymbol{A}=2^{730}\begin{bmatrix}1 & 1/2\\ 2 & 1\end{bmatrix}=\begin{bmatrix}2^{730} & 2^{729}\\ 2^{731} & 2^{730}\end{bmatrix}$$

8.3.4 矩阵指数函数的计算

类似于标量指数函数

$$\mathrm{e}^{at}=1+at+\frac{1}{2!}a^2t^2+\cdots+\frac{1}{k!}a^kt^k+\cdots$$

对一个已知矩阵 $\boldsymbol{A}$，可以定义矩阵指数函数 (matrix exponential function)

$$\mathrm{e}^{\boldsymbol{A}t}=\boldsymbol{I}+\boldsymbol{A}t+\frac{1}{2!}\boldsymbol{A}^2t^2+\cdots+\frac{1}{k!}\boldsymbol{A}^kt^k+\cdots \tag{8.3.13}$$

由于

$$e^{(\boldsymbol{A}+\boldsymbol{B})t} = \boldsymbol{I} + (\boldsymbol{A}+\boldsymbol{B})t + \frac{(\boldsymbol{A}+\boldsymbol{B})^2}{2}t^2 + \cdots \tag{8.3.14}$$

$$\begin{aligned} e^{\boldsymbol{A}t}e^{\boldsymbol{B}t} &= \left(\boldsymbol{I} + \boldsymbol{A}t + \frac{\boldsymbol{A}^2t^2}{2} + \cdots\right)\left(\boldsymbol{I} + \boldsymbol{B}t + \frac{\boldsymbol{B}^2t^2}{2} + \cdots\right) \\ &= \boldsymbol{I} + (\boldsymbol{A}+\boldsymbol{B})t + \frac{\boldsymbol{A}^2t^2}{2} + \boldsymbol{A}\boldsymbol{B}t^2 + \frac{\boldsymbol{B}^2t^2}{2} + \cdots \end{aligned} \tag{8.3.15}$$

两式相减，得

$$e^{(\boldsymbol{A}+\boldsymbol{B})t} - e^{\boldsymbol{A}t}e^{\boldsymbol{B}t} = (\boldsymbol{B}\boldsymbol{A} - \boldsymbol{A}\boldsymbol{B})\frac{t^2}{2} + \cdots \tag{8.3.16}$$

因此，$e^{(\boldsymbol{A}+\boldsymbol{B})t} = e^{\boldsymbol{A}t}e^{\boldsymbol{B}t}$ 只对 $\boldsymbol{A}\boldsymbol{B} = \boldsymbol{B}\boldsymbol{A}$ 成立。

容易看出

$$e^{\boldsymbol{O}} = e^{\boldsymbol{O}0} = \boldsymbol{I}, \qquad \frac{d}{dt}e^{\boldsymbol{A}t} = \boldsymbol{A}e^{\boldsymbol{A}t}$$

矩阵指数函数可以用来表示一阶微分方程的解。在工程应用中，经常会遇到线性一阶微分方程组

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{A}\boldsymbol{x}(t), \qquad \boldsymbol{x}(0) = \boldsymbol{x}_0$$

其中，$\boldsymbol{A}$ 为常数矩阵。上述一阶微分方程组的解可以写作 $\boldsymbol{x}(t) = e^{\boldsymbol{A}t}\boldsymbol{x}_0$。因此，线性一阶微分方程组的求解等价于计算矩阵指数函数 $e^{\boldsymbol{A}t}$。

利用 Cayley-Hamilton 定理，可以证明 n 阶线性矩阵微分方程的解的唯一性。

定理 8.3.3 令 $\boldsymbol{A}$ 是一个 $n \times n$ 常数矩阵，其特征多项式为

$$p(\lambda) = \det(\lambda\boldsymbol{I} - \boldsymbol{A}) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0 \tag{8.3.17}$$

并且 n 阶矩阵微分方程

$$\boldsymbol{\Phi}^{(n)}(t) + c_{n-1}\boldsymbol{\Phi}^{(n-1)}(t) + \cdots + c_1\boldsymbol{\Phi}'(t) + c_0\boldsymbol{\Phi}(t) = \boldsymbol{O} \tag{8.3.18}$$

满足初始条件

$$\boldsymbol{\Phi}(0) = \boldsymbol{I}, \quad \boldsymbol{\Phi}'(0) = \boldsymbol{A}, \quad \boldsymbol{\Phi}''(0) = \boldsymbol{A}^2, \quad \cdots, \boldsymbol{\Phi}^{(n-1)} = \boldsymbol{A}^{n-1} \tag{8.3.19}$$

则 $\boldsymbol{\Phi}(t) = e^{\boldsymbol{A}t}$ 是 n 阶矩阵微分方程式 (8.3.18) 的唯一解。

证明 [278] 令 $\boldsymbol{\Phi}_1(t)$ 和 $\boldsymbol{\Phi}_2(t)$ 是 n 阶矩阵微分方程的两个解，并令 $\boldsymbol{\Phi}(t) = \boldsymbol{\Phi}_1(t) - \boldsymbol{\Phi}_2(t)$。由于 $\boldsymbol{\Phi}(t)$ 满足矩阵微分方程式 (8.3.18)，故 $\boldsymbol{\Phi}(t)$ 的每一个元素 $\phi_{ij}(t)$ 都满足标量微分方程

$$\phi_{ij}^{(n)}(t) + c_{n-1}\phi_{ij}^{(n-1)}(t) + \cdots + c_1\phi_{ij}'(t) + c_0\phi_{ij}(t) = 0 \tag{1}$$

又由于解 $\boldsymbol{\Phi}_1(t)$ 和 $\boldsymbol{\Phi}_2(t)$ 都满足初始条件式 (8.3.19)，故有 $\boldsymbol{\Phi}(0) = \boldsymbol{\Phi}'(0) = \cdots = \boldsymbol{\Phi}^{(n-1)}(0) = \boldsymbol{O}$，即 $\boldsymbol{\Phi}(0)$ 的每一个元素都等于零，即有

$$\phi_{ij}(0) = \phi_{ij}'(0) = \cdots = \phi_{ij}^{(n-1)}(0) = 0, \quad \forall\, i, j \tag{2}$$

求解具有初始条件式 (2) 的标量微分方程式 (1)，得 $\phi_{ij}(t)=0,\forall\, t\in R$，$i,j=1,2,\cdots,n$。因此，$\boldsymbol{\Phi}(t)\equiv\boldsymbol{O},\forall\, t\in R$，即 $\boldsymbol{\Phi}_1(t)=\boldsymbol{\Phi}_2(t)$。

现在令 $\boldsymbol{A}$ 为一 $n\times n$ 常数矩阵，具有特征多项式

$$p(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\lambda^n+c_{n-1}\lambda^{n-1}+\cdots+c_1\lambda+c_0$$

若 $\boldsymbol{\Phi}(t)=\mathrm{e}^{\boldsymbol{A}t}$，则

$$\boldsymbol{\Phi}'(t)=\boldsymbol{A}\mathrm{e}^{\boldsymbol{A}t},\quad \boldsymbol{\Phi}''(t)=\boldsymbol{A}^2\mathrm{e}^{\boldsymbol{A}t},\cdots,\boldsymbol{\Phi}^{(n)}(t)=\boldsymbol{A}^n\mathrm{e}^{\boldsymbol{A}t} \tag{3}$$

于是，由 Cayley-Hamilton 定理知

$$\begin{aligned}&\boldsymbol{\Phi}^{(n)}(t)+c_{n-1}\boldsymbol{\Phi}^{(n-1)}(t)+\cdots+c_1\boldsymbol{\Phi}'(t)+c_0\boldsymbol{\Phi}(t)\\&=(\boldsymbol{A}^n+c_{n-1}\boldsymbol{A}^{n-1}+\cdots+c_1\boldsymbol{A}+c_0\boldsymbol{I})\mathrm{e}^{\boldsymbol{A}t}=p(\boldsymbol{A})\mathrm{e}^{\boldsymbol{A}t}\\&=\boldsymbol{O}\end{aligned}$$

这恰好是矩阵微分方程式 (8.3.18)。又由式 (3) 有

$$\boldsymbol{\Phi}(0)=\boldsymbol{I},\ \boldsymbol{\Phi}'(0)=\boldsymbol{A},\ \boldsymbol{\Phi}''(0)=\boldsymbol{A}^2,\ \cdots,\ \boldsymbol{\Phi}^{(n-1)}(0)=\boldsymbol{A}^{n-1}$$

此式即是初始条件式 (8.3.19)。因此，$\boldsymbol{\Phi}(t)=\mathrm{e}^{\boldsymbol{A}t}$ 是满足具有初始条件式 (8.3.19) 的矩阵微分方程式 (8.3.18) 的解。 ∎

定理 8.3.3 保证了矩阵微分方程的唯一解的存在性。下面的定理给出了求这一唯一解的方法。

定理 8.3.4 令 $\boldsymbol{A}$ 是一个 $n\times n$ 常数矩阵，其特征多项式为

$$p(\lambda)=\det(\lambda\boldsymbol{I}-\boldsymbol{A})=\lambda^n+c_{n-1}\lambda^{n-1}+\cdots+c_1\lambda+c_0$$

则满足初始条件式 (8.3.19) 的矩阵微分方程式 (8.3.18) 的解由

$$\boldsymbol{\Phi}(t)=\mathrm{e}^{\boldsymbol{A}t}=x_1(t)\boldsymbol{I}+x_2(t)\boldsymbol{A}+x_3(t)\boldsymbol{A}^2+\cdots+x_n(t)\boldsymbol{A}^{n-1} \tag{8.3.20}$$

给出，式中，$x_k(t)$ 是 n 阶标量微分方程

$$x^{(n)}(t)+c_{n-1}x^{(n-1)}(t)+\cdots+c_1x'(t)+c_0x(t)=0 \tag{8.3.21}$$

满足初始条件

$$\left.\begin{aligned}x_1(0)&=1\\x_1'(0)&=0\\&\cdots\\x_1^{(n-1)}(0)&=0\end{aligned}\right\},\qquad\left.\begin{aligned}x_2(0)&=0\\x_2'(0)&=1\\&\cdots\\x_2^{(n-1)}(0)&=0\end{aligned}\right\},\quad\cdots,\quad\left.\begin{aligned}x_n(0)&=0\\x_n'(0)&=0\\&\cdots\\x_n^{(n-1)}(0)&=1\end{aligned}\right\}$$

的解。

证明 [278] 令 $x_1(t), x_2(t), \cdots, x_n(t)$ 是 n 阶标量微分方程 (8.3.21) 满足定理初始条件的唯一一组解，并令

$$\boldsymbol{\Phi}(t) = x_1(t)\boldsymbol{I} + x_2(t)\boldsymbol{A} + x_3\boldsymbol{A}^2 + \cdots + x_n(t)\boldsymbol{A}^{n-1} \tag{1}$$

则

$$\begin{aligned}
&\boldsymbol{\Phi}^{(n)}(t) + c_{n-1}\boldsymbol{\Phi}^{(n-1)}(t) + \cdots + c_1\boldsymbol{\Phi}'(t) + c_0\boldsymbol{\Phi}(t) \\
=&[x_1^{(n)}(t) + c_{n-1}x_1^{(n-1)}(t) + \cdots + c_1x_1'(t) + c_0x_1(t)]\boldsymbol{I} + \\
&[x_2^{(n)}(t) + c_{n-1}x_2^{(n-1)}(t) + \cdots + c_1x_2'(t) + c_0x_2(t)]\boldsymbol{A} + \cdots + \\
&[x_n^{(n)}(t) + c_{n-1}x_n^{(n-1)}(t) + \cdots + c_1x_n'(t) + c_0x_n(t)]\boldsymbol{A}^{n-1} \\
=&0\boldsymbol{I} + 0\boldsymbol{A} + \cdots + 0\boldsymbol{A}^{n-1}
\end{aligned}$$

即有

$$\boldsymbol{\Phi}^{(n)}(t) + c_{n-1}\boldsymbol{\Phi}^{(n-1)}(t) + \cdots + c_1\boldsymbol{\Phi}'(t) + c_0\boldsymbol{\Phi}(t) = \boldsymbol{O}, \quad \forall\, t \in R \tag{2}$$

将定理初始条件代入式 (1) 及其 $1, 2, \cdots, n-1$ 阶导数，又有

$$\left.\begin{aligned}
\boldsymbol{\Phi}(0) &= x_1(0)\boldsymbol{I} + x_2(0)\boldsymbol{A} + \cdots + x_n(0)\boldsymbol{A}^{n-1} = \boldsymbol{I} \\
\boldsymbol{\Phi}'(0) &= x_1'(0)\boldsymbol{I} + x_2'(0)\boldsymbol{A} + \cdots + x_n'(0)\boldsymbol{A}^{n-1} = \boldsymbol{A} \\
&\cdots \\
\boldsymbol{\Phi}^{(n-1)}(0) &= x_1^{(n-1)}(0)\boldsymbol{I} + x_2^{(n-1)}(0)\boldsymbol{A} + \cdots + x_n^{(n-1)}(0)\boldsymbol{A}^{n-1} = \boldsymbol{A}^{n-1}
\end{aligned}\right\} \tag{3}$$

因此，由式 (1) 定义的矩阵 $\boldsymbol{\Phi}$ 是满足具有初始条件式 (3) 的矩阵微分方程式 (2) 的解。根据定理 8.3.3 知，这个解是唯一的。又因为式 (1) 的右边等于 $\mathrm{e}^{\boldsymbol{A}t}$，故式 (1) 即是式 (8.3.20)，定理得证。 ■

式 (8.3.20) 给出了计算矩阵指数函数 $\mathrm{e}^{\boldsymbol{A}t}$ 的一种有效方法。

例 8.3.4 已知矩阵

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

利用定理 8.3.4 求 $\mathrm{e}^{\boldsymbol{A}t}$。

解 根据矩阵 $\boldsymbol{A}$，可以写出相对应的标量微分方程组

$$\begin{aligned}
\frac{\mathrm{d}x(t)}{\mathrm{d}t} &= x + 0y + z \\
\frac{\mathrm{d}y(t)}{\mathrm{d}t} &= 0x + y + 0z \\
\frac{\mathrm{d}z(t)}{\mathrm{d}t} &= 0x + 0y + 2z
\end{aligned}$$

由特征方程 $\det(\lambda\boldsymbol{I} - \boldsymbol{A}) = 0$ 求得矩阵 $\boldsymbol{A}$ 的特征值 $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 2$，即特征值 1 的多重度为 2。

由定理 8.3.4 知，矩阵指数函数

$$e^{\boldsymbol{A}t} = x_1(t)\boldsymbol{I} + x_2(t)\boldsymbol{A} + x_3(t)\boldsymbol{A}^2$$

式中，$x_1(t), x_2(t), x_3(t)$ 是三阶标量微分方程

$$x'''(t) + c_2x''(t) + c_1x'(t) + c_0x(t) = 0$$

满足初始条件

$$\left.\begin{aligned} x_1(0) &= 1 \\ x_1'(0) &= 0 \\ x_1''(0) &= 0 \end{aligned}\right\}, \quad \left.\begin{aligned} x_2(0) &= 0 \\ x_2'(0) &= 1 \\ x_2''(0) &= 0 \end{aligned}\right\}, \quad \left.\begin{aligned} x_3(0) &= 0 \\ x_3'(0) &= 0 \\ x_3''(0) &= 1 \end{aligned}\right\}$$

的解。

由于矩阵 $\boldsymbol{A}$ 的特征值 $\lambda_1 = 1, \lambda_2 = 1, \lambda_3 = 2$ 知，标量微分方程 $x'''(t) + c_2x''(t) + c_1x'(t) + c_0x(t) = 0$ 的通解由

$$x(t) = a_1te^t + a_2e^t + a_3e^{2t}$$

给出。

由上述通解公式易得以下结果：

(1) 将初始条件 $x_1(0) = 1, x_1'(0) = 0, x_1''(0) = 0$ 代入通解公式，得

$$\begin{cases} a_2 + a_3 = 1 \\ a_1 + a_2 + 2a_3 = 0 \\ 2a_1 + a_2 + 4a_3 = 0 \end{cases} \Rightarrow \begin{cases} a_1 = -2 \\ a_2 = 0 \\ a_3 = 1 \end{cases}$$

即满足初始条件 $x_1(0) = 1, x_1'(0) = 0, x_1''(0) = 0$ 的特解为

$$x_1(t) = -2te^t + e^{2t}$$

(2) 由初始条件 $x_2(0) = 0, x_2'(0) = 1, x_2''(0) = 0$ 得

$$\begin{cases} a_2 + a_3 = 0 \\ a_1 + a_2 + 2a_3 = 1 \\ 2a_1 + a_2 + 4a_3 = 0 \end{cases} \Rightarrow \begin{cases} a_1 = 1 \\ a_2 = 2 \\ a_3 = -1 \end{cases}$$

从而得满足初始条件 $x_2(0) = 0, x_2'(0) = 1, x_2''(0) = 0$ 的特解为

$$x_2(t) = te^t + 2e^t - e^{2t}$$

(3) 由初始条件 $x_3(0) = 0, x_3'(0) = 0, x_3''(0) = 1$ 得

$$\begin{cases} a_2 + a_3 = 0 \\ a_1 + a_2 + 2a_3 = 0 \\ 2a_1 + a_2 + 4a_3 = 1 \end{cases} \Rightarrow \begin{cases} a_1 = -1 \\ a_2 = -1 \\ a_3 = 1 \end{cases}$$

换言之，满足初始条件 $x_3(0)=0, x_3'(0)=0, x_3''(0)=1$ 的特解为

$$x_3(t) = -te^t - e^t + e^{2t}$$

计算知

$$\boldsymbol{A}^2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

因此，由定理 8.3.4 可求得

$$\begin{aligned} e^{\boldsymbol{A}t} &= x_1(t)\boldsymbol{I} + x_2(t)\boldsymbol{A} + x_3(t)\boldsymbol{A}^2 \\ &= (-2tc^t + e^{2t}) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} + (te^t + 2e^t - e^{2t}) \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} + \\ &\quad (-te^t - e^t + e^{2t}) \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix} \end{aligned}$$

即有

$$e^{\boldsymbol{A}t} = \begin{bmatrix} -2te^t + e^t + e^{2t} & 0 & -2te^t - e^t + 2e^{2t} \\ 0 & -2te^t + e^t + e^{2t} & 0 \\ 0 & 0 & -4te^t + 3e^{2t} \end{bmatrix}$$

除了矩阵指数函数外，利用矩阵幂，还可以定义以下矩阵函数：

矩阵 $\boldsymbol{A}_{n\times n}$ 的正弦

$$\sin(\boldsymbol{A}) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} \boldsymbol{A}^{2k+1} \tag{8.3.22}$$

矩阵 $\boldsymbol{A}_{n\times n}$ 的余弦

$$\cos(\boldsymbol{A}) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} \boldsymbol{A}^{2k} \tag{8.3.23}$$

总结以上讨论，可以得出结论：Cayley-Hamliton 定理为矩阵求逆、矩阵幂的计算、线性微分方程的求解 (或等价于矩阵指数函数的计算) 提供了非常有效的工具。很难想象，如果没有 Cayley-Hamilton 定理，即便只是 3×3 矩阵的幂函数 $\boldsymbol{A}^{798}$ 的计算也将显得异常复杂和耗时，人工计算将是一件难于承受的负担！

8.4 Hermitian 矩阵的特征值分解

前面几节关于特征值和特征向量的所有讨论都是针对一般矩阵的，并没有要求矩阵一定是实对称或者复共轭对称的。然而，在统计与信息科学中，经常遇到实对称矩阵或 Hermitian(复共轭对称) 矩阵。例如，对于实观测数据向量 $\boldsymbol{x}(t)$，其自相关矩阵 $\boldsymbol{R} = \mathrm{E}\{\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{T}}(t)\}$ 是实对称矩阵，而复观测数据向量 $\boldsymbol{x}(t)$ 的自相关矩阵 $\boldsymbol{R} = \mathrm{E}\{\boldsymbol{x}(t)\boldsymbol{x}^{\mathrm{H}}(t)\}$ 为 Hermitian 矩阵。另外一方面，由于实对称矩阵是 Hermitian 矩阵的特例，而 Hermitian