第 1 章　特征值，特征向量和相似性

1.0　引言

在每章的开始一节，我们将用一些例子对这一章里要讨论的某些关键内容的动机加以阐述，这些例子说明了它们作为概念或者在应用中是如何产生的.

在整本书中，我们都使用在第 0 章里引进的记号与术语. 读者可以查询索引以寻求不熟悉的名词的定义，不熟悉的记号通常可以通过查阅文献后面有关记号的那一节了解辨认.

1.0.1　基变换与相似性

每一个可逆矩阵都是一个可以变换基的矩阵，而每一个可以变换基的矩阵也都是可逆的(0.10). 于是，如果$\mathcal{B}$是向量空间 V 的一组给定的基，T 是 V 上一个给定的线性变换，又如果 $A={}_{\mathcal{B}}[T]_{\mathcal{B}}$是 T 的$\mathcal{B}$基表示，则 T 的所有可能的基表示的集合是

$$\{{}_{\mathcal{B}_1}[I]_{\mathcal{B}}\,{}_{\mathcal{B}}[T]_{\mathcal{B}}\,{}_{\mathcal{B}}[I]_{\mathcal{B}_1}:\mathcal{B}_1\text{ 是 }V\text{ 的基}\}=\{S^{-1}AS:S\in M_n(\mathbf{F})\text{ 是可逆的}\}$$

这恰好是所有与给定的矩阵 A **相似的**(similar)矩阵之集合. 这样一来，相似但不相等的矩阵正好就是单个线性变换的不同的基表示.

我们可以期待相似矩阵共享许多重要的性质(至少是那些反映基础线性变换本质的性质)，而这是线性代数的一个重要课题. 从关于一个给定矩阵的问题退回到线性变换(这个线性变换的矩阵仅仅是它的众多可能的表达方式之一)的某个本质性质的问题常常是有用的.

相似是本章中一个关键的概念.

1.0.2　限制极值与特征值

本章中第二个关键的概念是**特征向量**(eigenvector)和**特征值**(eigenvalue). 使得 Ax 是
43 x 的一个纯量倍数的非零向量 x 在分析矩阵或者线性变换的构造中起着主导的作用，但是这样的向量出现在使从属于几何限制的实对称二次型最大化(或者最小化)这种更为初等的内容之中：对给定的实对称矩阵 $A\in M_n(\mathbf{R})$，

$$\text{极大化 } x^{\mathrm{T}}Ax,\quad \text{服从条件 } x\in\mathbf{R}^n,\quad x^{\mathrm{T}}x=1 \tag{1.0.3}$$

解决这样一个有限制条件的最优化问题的常规方法是引进 Lagrange 算子 $L=x^{\mathrm{T}}Ax-\lambda x^{\mathrm{T}}x$. 取极值的必要条件是

$$0=\nabla L=2(Ax-\lambda x)=0$$

于是，如果一个满足 $x^{\mathrm{T}}x=1$ 的向量 $x\in\mathbf{R}^n$(从而 $x\neq 0$)是 $x^{\mathrm{T}}Ax$ 的一个极值，它就必定满足方程 $Ax=\lambda x$. 对某个非零向量 x，满足 $Ax=\lambda x$ 的纯量 λ 就是 A 的一个**特征值**(eigenvalue).

问题

1.0.P1　利用 Weierstrass 定理(见附录 E)解释为什么限制极值问题(1.0.3)有解，并导出结论：每一个实对称矩阵至少有一个实的特征值.

1.0.P2　假设 $A\in M_n(\mathbf{R})$是对称的. 证明：$\max\{x^{\mathrm{T}}Ax:\ x\in\mathbf{R}^n,\ x^{\mathrm{T}}x=1\}$是 A 的**最大的**实特征值.

1.1 特征值-特征向量方程

矩阵 $A\in M_n$ 可以视为从 $\mathbf{C}^n$ 到 $\mathbf{C}^n$ 的一个线性变换，也即

$$A: x \to Ax \tag{1.1.1}$$

不过把它视为数的阵列也是有用的．根据数的阵列告诉我们的有关线性变换的知识，A 的这两种概念之间的交互作用是矩阵分析的中心内容以及通向应用之关键．矩阵分析中的一个基本概念就是复方阵的**特征值**之集合．

定义 1.1.2 设 $A\in M_n$．如果纯量 λ 和非零向量 x 满足方程

$$Ax = \lambda x, \quad x \in \mathbf{C}^n, \quad x \neq 0, \quad \lambda \in \mathbf{C} \tag{1.1.3}$$

那么 λ 就称为是 A 的一个特征值，而 x 则称为 A 的一个与 λ 相伴的特征向量．元素对 (λ, x) 称为 A 的一个特征对．

上一个定义中的纯量 λ 和向量 x 天生注定成对出现．定义中的关键点是：**特征向量永远不会是零向量．**

习题 考虑对角矩阵 $D=\mathrm{diag}(d_1, d_2, \cdots, d_n)$．说明为什么标准基向量 $e_i(i=1, \cdots, n)$ 是 D 的特征向量．每一个特征向量 e_i 与哪一个特征值相伴？ ◀

方程 1.1.3 可以改写成 $\lambda x-Ax=(\lambda I-A)x=0$，这是一个正方的齐次线性方程组．如果这个方程组有非平凡的解，那么 λ 就是 A 的一个特征值，而矩阵 $\lambda I-A$ 就是奇异的．反过来，如果 $\lambda\in\mathbf{C}$ 且 $\lambda I-A$ 是奇异的，那么就存在一个非零向量 x，使得 $(\lambda I-A)x=0$，所以 $Ax=\lambda x$，也就是说，(λ, x) 是 A 的一个特征值-特征向量对． 44

定义 1.1.4 $A\in M_n$ 的谱是 A 的所有特征值 $\lambda\in\mathbf{C}$ 组成的集合，我们将这个集合记为 $\sigma(A)$．

对给定的 $A\in M_n$，此时我们还不知道 $\sigma(A)$ 是否是空集，或者如果它非空，我们也不知道它包含有限多个还是无限多个复数．

习题 如果 x 是与 A 的特征值 λ 相伴的特征向量，证明：x 的任何非零纯量倍数也是 A 的与 λ 相伴的特征向量． ◀

如果 x 是 $A\in M_n$ 的与 λ 相伴的特征向量，将它标准化通常更为方便，所谓标准化就是构造一个单位向量 $\xi=x/\|x\|_2$，它依然是 A 的与 λ 相伴的特征向量．标准化并没有选择与 λ 相伴的唯一的特征向量，无论如何，$(\lambda, e^{i\theta}\xi)$ 都是 A 的特征值-特征向量对(对所有 $\theta\in\mathbf{R}$)．

习题 如果 $Ax=\lambda x$，注意有 $\overline{A}\,\overline{x}=\overline{\lambda}\,\overline{x}$．说明为什么 $\sigma(\overline{A})=\overline{\sigma(A)}$．如果 $A\in M_n(\mathbf{R})$ 且 $\lambda\in\sigma(A)$，说明为什么也有 $\overline{\lambda}\in\sigma(A)$． ◀

特征值与特征向量即便没有其他的重要性，它们在代数上也是很有意义的：根据 (1.1.3)，特征向量恰好是那样的非零向量：用 A 来乘与用纯量 λ 来乘有相同的结果．

习题 考虑矩阵

$$A = \begin{bmatrix} 7 & -2 \\ 4 & 1 \end{bmatrix} \in M_2 \tag{1.1.4a}$$

那么 $3\in\sigma(A)$ 且 $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ 是一个与之相伴的特征向量，这是因为

$$A\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 6 \end{bmatrix} = 3\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

又有 $5\in\sigma(A)$. 求一个与特征值 5 相伴的特征向量. ◀

有时一个矩阵的构造使人容易想象出它的一个特征向量，从而其相伴的特征值也容易计算出来.

习题 设 J_n 是 $n\times n$ 矩阵，其元素全都等于 1. 考虑元素全都等于 1 的 n 元向量 e，并设 $x_k=e-ne_k$，其中 $\{e_1, \cdots, e_n\}$ 是 $\mathbf{C}^n$ 的一组标准基. 对 $n=2$，证明 e 和 x_1 是 J_2 的线性无关的特征向量，而 2 和 0 分别是相伴的特征值. 对 $n=3$，证明 e、x_1 以及 x_2 是 J_3 的线性无关的特征向量，而 2、0 和 0 分别是相伴的特征值. 一般来说，证明 $e, x_1, \cdots, x_{n-1}$
45 是 J_n 的线性无关的特征向量，而 $n, 0, \cdots, 0$ 分别是相伴的特征值. ◀

习题 证明 1 和 4 是矩阵

$$A=\begin{bmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{bmatrix}$$

的特征值. 提示：利用特征向量. 记 $A=4I-J_3$，并利用上一习题. ◀

实系数或者复系数 k 次多项式

$$p(t)=a_kt^k+a_{k-1}t^{k-1}+\cdots+a_1t+a_0, \quad a_k\neq 0 \tag{1.1.5a}$$

在矩阵 $A\in M_n$ 处的值有良好的定义，因为我们可以作出给定方阵的整数幂的线性组合. 定义

$$p(A)=a_kA^k+a_{k-1}A^{k-1}+\cdots+a_1A+a_0I \tag{1.1.5b}$$

其中注意通用的约定 $A^0=I$. k 次多项式(1.1.5a)被说成是**首一的**(monic)，如果 $a_k=1$；由于 $a_k\neq 0$，故而 $a_k^{-1}p(t)$ 总是首一的. 当然，首一多项式不可能是零多项式.

有另外一种方法来表示 $p(A)$，它有很重要的推论. **代数基本定理**(fundamental theorem of algebra)(附录 C)确保次数为 $k\geqslant 1$ 的任何首一多项式(1.1.5a)可以表示成恰好 k 个复的或者实的线性因子的乘积：

$$p(t)=(t-\alpha_1)\cdots(t-\alpha_k) \tag{1.1.5c}$$

$p(t)$的这个表达式除了因子的排列顺序外是唯一的. 它告诉我们：对每个 $j=1, \cdots, k$ 有 $p(\alpha_j)=0$，所以每个 α_j 都是方程 $p(t)=0$ 的一个**根**(root)；我们也说成每个 α_j 是 $p(t)$的一个**零点**(zero). 反过来，如果 β 是一个复数，它使得 $p(\beta)=0$，那么 $\beta\in\{\alpha_1, \cdots, \alpha_k\}$，所以一个次数为 $k\geqslant 1$ 的多项式至多有 k 个不同的零点. 在乘积(1.1.5c)中，有些因子可能会重复，例如，$p(t)=t^2+2t+1=(t+1)(t+1)$. 因子$(t-\alpha_j)$重复的次数就是 α_j 作为 $p(t)$零点的**重数**(multiplicity). 分解式(1.1.5c)给出 $p(A)$的分解式：

$$p(A)=(A-\alpha_1I)\cdots(A-\alpha_kI) \tag{1.1.5d}$$

$p(A)$的特征值与 A 的特征值以一种简单的方式联系在一起.

定理 1.1.6 设 $p(t)$是给定的 k 次多项式. 如果 λ, x 是 $A\in M_n$ 的一个特征值-特征向量对，那么 $p(\lambda), x$ 就是 $p(A)$的一个特征值-特征向量对. 反过来，如果 $k\geqslant 1$ 且 μ 是 $p(A)$的一个特征值，那么就存在 A 的某个特征值 λ 使得 $\mu=p(\lambda)$.

证明 我们有

$$p(A)x=a_kA^kx+a_{k-1}A^{k-1}x+\cdots+a_1Ax+a_0x, \quad a_k\neq 0$$

重复应用特征值-特征向量方程又有 $A^jx=A^{j-1}Ax=A^{j-1}\lambda x=\lambda A^{j-1}x=\cdots=\lambda^jx$. 从而

46 $$p(A)x=a_k\lambda^kx+\cdots+a_0x=(a_k\lambda^k+\cdots+a_0)x=p(\lambda)x$$

反过来，如果 μ 是 $p(A)$ 的一个特征值，那么 $p(A)-\mu I$ 是奇异的．由于 $p(t)$ 的次数 $k\geqslant 1$，故而多项式 $q(t)=p(t)-\mu$ 的次数 $k\geqslant 1$，我们就可以将它分解成 $q(t)=(t-\beta_1)\cdots(t-\beta_k)$（对某些复数或者实数 β_1，…，β_k）．由于 $p(A)-\mu I=q(A)=(A-\beta_1 I)\cdots(A-\beta_k I)$ 是奇异的，故而它的某个因子 $A-\beta_j I$ 是奇异的，这就意味着 β_j 是 A 的特征值．但是 $0=q(\beta_j)=p(\beta_j)-\mu$，所以有 $\mu=p(\beta_j)$，如所断言． □

习题 假设 $A\in M_n$．如果 $\sigma(A)=\{-1, 1\}$，那么 $\sigma(A^2)$ 是什么？**小心**：定理 1.1.6 中的第一个结论可以使人找出 $\sigma(A^2)$ 中的一个点，但是你必须求助第二个结论来确定它是否是 $\sigma(A^2)$ 中**仅有的**点． ◀

习题 考虑 $A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$．$A^2$ 等于什么？证明 e_1 是 A 和 A^2 的特征向量，且它们两者都是与特征值 $\lambda=0$ 相伴的．证明 e_2 是 A^2 的特征向量但不是 A 的特征向量．说明为什么定理 1.1.6 的“逆命题”部分只谈到了 $p(A)$ 的特征值，而未谈及其特征向量．证明：除了 e_1 的纯量倍数之外，A 没有其他的特征向量了，并说明为什么有 $\sigma(A)=\{0\}$． ◀

结论 1.1.7 矩阵 $A\in M_n$ 是奇异的，当且仅当 $0\in\sigma(A)$．

证明 矩阵 A 是奇异的，当且仅当对某个 $x\neq 0$ 有 $Ax=0$．而这当且仅当对某个 $x\neq 0$ 有 $Ax=0x$，也就是当且仅当 $\lambda=0$ 是 A 的特征值时才会发生． □

结论 1.1.8 设给定 $A\in M_n$ 以及 λ，$\mu\in\mathbf{C}$．那么，$\lambda\in\sigma(A)$ 当且仅当 $\lambda+\mu\in\sigma(A+\mu I)$．

证明 如果 $\lambda\in\sigma(A)$，则存在一个非零向量 x，使得 $Ax=\lambda x$，从而 $(A+\mu I)x=Ax+\mu x=\lambda x+\mu x=(\lambda+\mu)x$．于是，$\lambda+\mu\in\sigma(A+\mu I)$．反过来，如果 $\lambda+\mu\in\sigma(A+\mu I)$，则存在一个非零向量 y，使得 $Ay+\mu y=(A+\mu I)y=(\lambda+\mu)y=\lambda y+\mu y$．于是 $Ay=\lambda y$，从而 $\lambda\in\sigma(A)$． □

我们现在准备给出一个很重要的结论：每个复矩阵都有非空的谱，也就是说，对每个 $A\in M_n$，存在某个纯量 $\lambda\in\mathbf{C}$ 以及非零的向量 $x\in\mathbf{C}^n$，使得 $Ax=\lambda x$．

定理 1.1.9 设给定 $A\in M_n$．那么 A 有特征值．事实上，对每个给定的非零的 $y\in C^n$，存在一个至多 $n-1$ 次多项式 $g(t)$，使得 $g(A)y$ 是 A 的特征向量．

证明 设 m 是使得向量 y，Ay，A^2y，…，A^ky 线性相关的**最小的**整数 k．那么有 $m\geqslant 1$（由于 $y\neq 0$），且有 $m\leqslant n$（由于 $\mathbf{C}^n$ 中任意 $n+1$ 个向量都是线性相关的）．设 a_0，a_1，…，a_m 是不全为零的纯量，它们使得

$$a_mA^my+a_{m-1}A^{m-1}y+\cdots+a_1Ay+a_0y=0 \tag{1.1.10}$$

如果 $a_m=0$，那么 (1.1.10) 蕴含向量 y，Ay，A^2y，…，$A^{m-1}y$ 线性相关，这与 m 的最小性矛盾．于是 $a_m\neq 0$，我们可以考虑多项式 $p(t)=t^m+(a_{m-1}/a_m)t^{m-1}+\cdots+(a_1/a_m)t+(a_0/a_m)$．恒等式 (1.1.10) 确保 $p(A)y=0$，所以 0，y 是 $p(A)$ 的一个特征值-特征向量对．定理 1.1.6 就确保 $p(t)$ 的 m 个零点中有一个是 A 的特征值．

假设 λ 是 $p(t)$ 的一个零点，它是 A 的一个特征值，分解 $p(t)=(t-\lambda)g(t)$，其中 $g(t)$ 是一个 $m-1$ 次多项式．如果 $g(A)y=0$，则 m 的最小性再次出现矛盾，所以 $g(A)y\neq 0$．但是 $0=p(A)y=(A-\lambda I)(g(A)y)$，所以非零向量 $g(A)y$ 是 A 的一个与特征值 λ 相伴的特征向量． □

上面的讨论表明，对给定的 $A\in M_n$ 可以求得一个次数最多为 n 的多项式，**它至少有一个**零点是 A 的特征值．在下一节里，我们要介绍一个次数恰好为 n 的多项式 $p_A(t)$，它的每个零点都是 A 的特征值，且 A 的每一个特征值也都是 $p_A(t)$ 的零点，也就是说，$p_A(\lambda)=0$ 当且仅当 $\lambda\in\sigma(A)$．

问题

1.1.P1 假设 $A\in M_n$ 非奇异．根据(1.1.7)，这就等价于假设 $0\notin\sigma(A)$．对每一个 $\lambda\in\sigma(A)$，证明 $\lambda^{-1}\in\sigma(A^{-1})$．如果 $Ax=\lambda x$ 且 $x\neq0$，证明 $A^{-1}x=\lambda^{-1}x$．

1.1.P2 设给定 $A\in M_n$．(a)证明 A 的每一行的元素之和为 1 当且仅当 $1\in\sigma(A)$ 且向量 $e=[1, 1, \cdots, 1]^T$ 是与之相伴的特征向量，也就是说，$Ae=e$．(b)假设 A 的每一行的元素之和是 1．如果 A 非奇异，证明 A^{-1} 的每一行元素之和也是 1．此外，对任意给定的多项式 $p(t)$，证明 $p(A)$ 的每一行元素之和都相等．等于什么？

1.1.P3 设 $A\in M_n(\mathbf{R})$．假设 λ 是 A 的一个实的特征值，且 $Ax=\lambda x$，$x\in\mathbf{C}^n$，$x\neq0$．设 $x=u+\mathrm{i}v$，其中 u，$v\in\mathbf{R}^n$ 分别是 x 的实部和虚部，见(0.2.5)．证明 $Au=\lambda u$ 以及 $Av=\lambda v$．解释为什么 u，v 中至少有一个必定不是零，并推出结论：A 有与 λ 相伴的实的特征向量．u 与 v 两者都必定是 A 的特征向量吗？A 能有与一个非实数的特征值相伴的实的特征向量吗？

1.1.P4 考虑分块对角矩阵

$$A=\begin{bmatrix}A_{11} & 0\\ 0 & A_{22}\end{bmatrix},\quad A_{ii}\in M_{n_i}$$

证明 $\sigma(A)=\sigma(A_{11})\cup\sigma(A_{22})$．你必须证明三件事：(a)如果 λ 是 A 的特征值，那么它要么是 A_{11} 的特征值要么是 A_{22} 的特征值；(b)如果 λ 是 A_{11} 的特征值，那么它也是 A 的特征值；(c)如果 λ 是 A_{22} 的特征值，那么它也是 A 的特征值．

1.1.P5 设 $A\in M_n$ 是幂等的，即 $A^2=A$．证明 A 的每个特征值或者是 0 或者是 1．说明为什么 I 是唯一的非奇异的幂等矩阵．

1.1.P6 证明：幂零矩阵的所有特征值都是 0．给出一个非零的幂零矩阵的例子．说明为什么 0 矩阵是仅有的幂零幂等矩阵．

1.1.P7 如果 $A\in M_n$ 是 Hermite 矩阵，证明 A 的所有特征值都是**实数**．

48 1.1.P8 说明，如果我们试图利用(1.1.9)中的推理方法来证明每个实方阵都有实特征值时是如何失效的．

1.1.P9 利用定义(1.1.3)证明实矩阵 $A=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$没有实的特征值．然而，(1.1.9)却显示 A 有复的特征值．实际上，它有两个复的特征值，这两个特征值是什么？

1.1.P10 对如下的例子给出细节，这个例子表明无限维复向量空间上的线性算子可能**没有**特征值．设 $V=\{(a_1, a_2, \cdots): a_i\in\mathbf{C}, i=1, 2, \cdots\}$是由所有无限的形式复数序列组成的向量空间，并用 $S(a_1, a_2, \cdots)=(0, a_1, a_2, \cdots)$来定义 V 上的**右平移算子**(right-shift operator)．验证 S 是一个线性变换．如果 $Sx=\lambda x$，证明有 $x=0$．

1.1.P11 设给定 $A\in M_n$ 以及 $\lambda\in\sigma(A)$．那么 $A-\lambda I$ 是奇异的，所以$(A-\lambda I)\operatorname{adj}(A-\lambda I)=[\det(A-\lambda I)]I=0$，见(0.8.2)．说明为什么存在某个 $y\in\mathbf{C}^n$(有可能 $y=0$)，使得 $\operatorname{adj}(A-\lambda I)=xy^*$．得出结论：$\operatorname{adj}(A-\lambda I)$的每个非零的列都是 A 的与特征值 λ 相伴的特征向量．这个结论为什么仅当 $\operatorname{rank}(A-\lambda I)=n-1$ 时才是有用的？

1.1.P12 假设 λ 是 $A=\begin{bmatrix}a & b\\ c & d\end{bmatrix}\in M_2$ 的一个特征值．利用(1.1.P11)证明：如果$\begin{bmatrix}d-\lambda & -b\\ -c & a-\lambda\end{bmatrix}$有一列不为零，那么它就是 A 的与 λ 相伴的特征向量．为什么这些列中必定有一列是另一列的纯量倍数？利用这个方法求(1.1.4a)中的矩阵与特征值 3 以及 5 相伴的特征向量．

1.1.P13 设 $A\in M_n$，又 λ，x 是 A 的一个特征值-特征向量对．证明：x 是 $\operatorname{adj}A$ 的一个特征向量．

1.2 特征多项式与代数重数

一个复方阵会有多少个特征值？可以怎样用系统的方式来刻画它们的特征？

将特征值-特征向量方程(1.1.3)改写成

$$(\lambda I - A)x = 0, \quad x \neq 0 \tag{1.2.1}$$

从而，$\lambda \in \sigma(A)$当且仅当$\lambda I - A$是奇异的，即当且仅当

$$\det(\lambda I - A) = 0 \tag{1.2.2}$$

定义 1.2.3 视为 t 的形式多项式，则 $A \in M_n$ 的**特征多项式**(characterictic polynomial)就是

$$p_A(t) = \det(tI - A)$$

我们把方程 $p_A(A)=0$ 称为 A 的**特征方程**(characteristic equation).

结论 1.2.4 每一个 $A=[a_{ij}]\in M_n$ 的特征多项式的次数都是 n，且 $p_A(t)=t^n-(\mathrm{tr}A)t^{n-1}+\cdots+(-1)^n\det A$. 此外，$p_A(\lambda)=0$ 当且仅当 $\lambda\in\sigma(A)$，故而 $\sigma(A)$至多包含 n 个复数.

证明 $tI-A$ 的行列式的表达式(0.3.2.1)中的每一个求和项都是 $tI-A$ 中恰好 n 个元素的乘积，每个元素都取自不同的行和列，所以每一个求和项也都是关于 t 的次数至多为 n 的多项式. 一个求和项的次数能达到 n，当且仅当乘积中的每个因子都包含 t，这仅对是对角元素乘积的那个求和项 49

$$(t-a_{11})\cdots(t-a_{nn}) = t^n - (a_{11}+\cdots+a_{nn})t^{n-1} + \cdots \tag{1.2.4a}$$

才会发生. 任何其他的求和项都必定包含一个因子 $-a_{ij}$(对 $i\neq j$)，所以对角元素($t-a_{ii}$)(在与 a_{ij} 同一行)以及($t-a_{ii}$)(在与 a_{ij} 同一列)不可能也是因子. 这样一来，这个求和项的次数不可能大于 $n-2$. 于是，多项式 $p_A(t)$中 t^n 与 t^{n-1} 的系数仅仅在求和项(1.2.4a)中出现. $p_A(t)$中的常数项正好是 $p_A(0)=\det(0I-A)=\det(-A)=(-1)^n\det A$. 剩下的结论是与(1.2.1)、(1.2.2)以及一个次数为 $n\geq 1$ 的多项式至多有 n 个不同的零点这一事实等价的结果. □

习题 证明 $\det(A-tI)=0$ 的根与 $\det(tI-A)=0$ 的根相同，又有 $\det(A-tI)=(-1)^n\det(tI-A)=(-1)^n(t^n+\cdots)$. ◀

特征多项式也可以定义为 $\det(A-tI)=(-1)^n t^n+\cdots$. 我们所取的习惯定义确保特征多项式中 t^n 的系数恒为$+1$.

习题 设 $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}\in M_2$. 证明 A 的特征多项式是

$$p_A(t) = t^2-(a+d)t+(ad-bc) = t^2-(\mathrm{tr}A)t+\det A$$

设 $r=(-\mathrm{tr}A)^2-4\det A=(a-d)^2+4bc$，这是 $p_A(t)$的**判别式**(discriminant)，又令$\sqrt{r}$是 r 的一个固定的平方根. 证明

$$\lambda_1 = \frac{1}{2}(a+d+\sqrt{r}) \quad 与 \quad \lambda_2 = \frac{1}{2}(a+d-\sqrt{r}) \tag{1.2.4b}$$ ◀

这两者中每一个都是 A 的特征值. 验证 $\mathrm{tr}A=\lambda_1+\lambda_2$ 以及 $\det A=\lambda_1\lambda_2$. 说明为什么 $\lambda_1\neq\lambda_2$ 当且仅当 $r\neq 0$. 如果 $A\in M_2(\mathbf{R})$，证明(a)A 的特征值是实的，当且仅当 $r\geq 0$；(b)A 的特征值是实的，当且仅当 $bc\geq 0$；(c)如果 $r<0$，那么 $\lambda_1=\overline{\lambda_2}$，也即 λ_1 是 λ_2 的复共轭.

上一习题用例子说明了，矩阵 $A\in M_n(n>1)$的特征值 λ 可能是 $p_A(t)$的重零点(等价地说，是它的特征方程的一个重根). 的确，$I\in M_n$ 的特征多项式是

$$p_I(t) = \det(tI-I) = \det((t-1)I) = (t-1)^n\det I = (t-1)^n$$

故而其特征值 $\lambda=1$ 是 $p_I(t)$的一个 n 重零点. 应该怎样来说明特征值的这样一种重复计

数呢？

对一个给定的 $A\in M_n(n>1)$，将它的特征多项式分解成 $p_A(t)=(t-\alpha_1)\cdots(t-\alpha_n)$．我们知道，$p_A(t)$的每一个零点 α_i(不记其重数)都是 A 的一个特征值．计算给出

$$p_A(t)=t^n-(\alpha_1+\cdots+\alpha_n)t^{n-1}+\cdots+(-1)^n\alpha_1\cdots\alpha_n \tag{1.2.4c}$$

所以将(1.2.4)与(1.2.4c)作比较即告诉我们：$p_A(t)$的零点之和是 A 的迹，而 $p_A(t)$的零点之积则是 A 的行列式．如果 $p_A(t)$的每个零点的重数都是 1，也就是说，如果只要 $i\neq j$，

50 就有 $\alpha_i\neq\alpha_j$，那么 $\sigma(A)=\{\alpha_1,\cdots,\alpha_n\}$，所以 $\mathrm{tr}A$ **是** A **的特征值之和，而** $\det A$ **则是** A **的特征值之积**．如果即使在 $p_A(t)$的某些零点有大于 1 的重数时这两个命题依然为真，我们就必须按照它们作为特征方程的根的重数来对 A 的特征值加以计数．

定义 1.2.5 设 $A\in M_n$．A 的特征值 λ 的重数是指它作为特征多项式 $p_A(t)$的零点的重数．为清晰起见，我们有时把特征值的重数称为它的代数重数．

从现在开始，$A\in M_n$ **的特征值总是指这个特征值与其相应的(代数)重数的合并称谓**．于是，A 的特征多项式的零点(包含其重数在内)与 A 的特征值(包含其重数在内)是相同的：

$$p_A(t)=(t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n) \tag{1.2.6}$$

其中 $\lambda_1,\cdots,\lambda_n$ 是 A 的 n 个特征值，按照任意次序排列．当我们提及 A 的**不同的特征值**时，指的就是集合 $\sigma(A)$的元素．

现在无需限制就能说：**每个矩阵** $A\in M_n$ **在复数中恰好有** n **个特征值**；A 的迹和行列式分别是它的特征值之和以及乘积．如果 A 是实的，它的特征值中可能有某些不是实的或者全部不是实的．

习题 考虑一个实矩阵 $A\in M_n(\mathbf{R})$．(a)说明为什么 $p_A(t)$的所有系数都是实数．(b)假设 A 有一个特征值 λ 不是实的．利用(a)说明为什么 $\bar\lambda$ 也是 A 的特征值，又为什么 λ 与 $\bar\lambda$ 的代数重数相同．如果 x,λ 是 A 的一个特征对，我们知道 $\bar x,\bar\lambda$ 也是它的一个特征对(为什么?)．注意，x 以及 $\bar x$ 是 A 的与**不同的**特征值 λ 以及 $\bar\lambda$ 相伴的特征向量． ◀

例 1.2.7 设 $x,y\in\mathbf{C}^n$．$I+xy^*$ 的特征值与行列式是什么？利用(0.8.5.11)以及 $\mathrm{adj}(\alpha I)=\alpha^{n-1}I$这一事实，我们来计算

$$\begin{aligned}p_{I+xy^*}(t)&=\det(tI-(I+xy^*))=\det((t-1)I-xy^*)\\&=\det((t-1)I)-y^*\mathrm{adj}((t-1)I)x=(t-1)^n-(t-1)^{n-1}y^*x\\&=(t-1)^{n-1}(t-(1+y^*x))\end{aligned}$$

于是，$I+xy^*$ 的特征值是 $1+y^*x$ 和 1(重数为 $n-1$)，所以 $\det(I+xy^*)=(1+y^*x)\times(1)^{n-1}=1+y^*x$．

例 1.2.8(Brauer 定理) 设 $x,y\in\mathbf{C}^n$，$x\neq0$，且 $A\in M_n$．假设 $Ax=\lambda x$，又设 A 的特征值是 $\lambda,\lambda_2,\cdots,\lambda_n$．那么 $A+xy^*$ 的特征值是什么？首先注意$(t-\lambda)x=(tI-A)x$ 蕴含 $(t-\lambda)\mathrm{adj}(tI-A)x=\mathrm{adj}(tI-A)(tI-A)x=\det(tI-A)x$，这就是

$$(t-\lambda)\mathrm{adj}(tI-A)x=p_A(t)x \tag{1.2.8a}$$

利用(0.8.5.11)来计算

51

$$\begin{aligned}p_{A+xy^*}(t)&=\det(tI-(A+xy^*))=\det((tI-A)-xy^*)\\&=\det(tI-A)-y^*\mathrm{adj}(tI-A)x\end{aligned}$$

用$(t-\lambda)$来乘两边，利用(1.2.8a)就得到

$$\begin{aligned}(t-\lambda)p_{A+xy^*}(t) &= (t-\lambda)\det(tI-A) - y^*(t-\lambda)\operatorname{adj}(tI-A)x \\ &= (t-\lambda)p_A(t) - p_A(t)y^*x\end{aligned}$$

这是一个多项式恒等式

$$(t-\lambda)p_{A+xy^*}(t) = (t-(\lambda+y^*x))p_A(t)$$

左边多项式的零点是 λ 以及 $A+xy^*$ 的 n 个特征值．右边多项式的零点则是 $\lambda+y^*x$，λ，λ_2，…，λ_n．由此推出，$A+xy^*$ 的特征值是 $\lambda+y^*x$，λ_2，…，λ_n．

由于我们现在知道每一个 $n\times n$ 复矩阵都有有限多个特征值，故可以做出如下的定义．

定义 1.2.9 设 $A\in M_n$．则 A 的**谱半径**(spectral radius)是 $\rho(A)=\max\{|\lambda|:\lambda\in\sigma(A)\}$．

习题 说明为什么 $A\in M_n$ 的每个特征值都位于复平面中封闭的有界圆盘 $\{z:z\in\mathbf{C}$ 且 $|z|\leqslant\rho(A)\}$ 内． ◀

习题 假设 $A\in M_n$ 至少有一个非零特征值．说明为什么 $\min\{|\lambda|:\lambda\in\sigma(A)$且$\lambda\neq0\}>0$． ◀

习题 上面两个习题的基础在于如下的事实：$\sigma(A)$是一个非空有限集．解释其原因． ◀

有时一个矩阵的构造使得它的特征多项式容易计算．这就是对角矩阵或者三角矩阵的情形．

习题 考虑上三角矩阵

$$T=\begin{bmatrix} t_{11} & \cdots & t_{1n} \\ & \ddots & \vdots \\ 0 & & t_{nn}\end{bmatrix}\in M_n$$ ◀

证明 $p_T(t)=(t-t_{11})\cdots(t-t_{nn})$，所以 T 的特征值是它的对角元素 t_{11}，t_{22}，…，t_{nn}．如果 T 是下三角矩阵呢？如果 T 是对角矩阵呢？

习题 假设 $A\in M_n$ 是分块上三角的

$$A=\begin{bmatrix} A_{11} & & \bigstar \\ & \ddots & \\ 0 & & A_{kk}\end{bmatrix},\quad A_{ii}\in M_{n_i},\quad i=1,\cdots,n_k$$ ◀

解释为什么 $p_A(t)=p_{A_{11}}(t)\cdots p_{A_{kk}}(t)$，又为什么 A 的特征值是由 A_{11} 的特征值、A_{22} 的特征值，…，以及 A_{kk} 的特征值连同它们的重数一起所构成．这一结论是计算特征值的许多算法之基础．说明为什么上一题是这一题的特殊情形．

定义 1.2.10 设 $A\in M_n$．它的 k 阶主子式（这样的主子式有 $\binom{n}{k}$ 个）之和记为 $E_k(A)$． 52

我们已经在特征多项式

$$p_A(t)=t^n+a_{n-1}t^{n-1}+\cdots+a_2t^2+a_1t+a_0 \qquad (1.2.10a)$$

的两个系数中遇到过主子式和了．如果 $k=1$，那么 $\binom{n}{k}=n$ 且 $E_1(A)=a_{11}+\cdots+a_{nn}=\operatorname{tr}A=-a_{n-1}$；如果 $k=n$，那么 $\binom{n}{k}=1$ 且 $E_n(A)=\det A=(-1)^na_0$．系数与主子式之和之间的更广泛的联系是如下事实的推论：这些系数是 $p_A(t)$在 $t=0$ 的某种导数的显式函数：

$$a_k = \frac{1}{k!}p_A^{(k)}(0), \quad k = 0,1,\cdots,n-1 \tag{1.2.11}$$

利用(0.8.10.2)计算导数

$$p'_A(t) = \mathrm{tr}\ \mathrm{adj}(tI - A)$$

注意到 tr adj$A$ 是 A 的 $n-1$ 阶主子式之和，所以 tr adj$A=E_{n-1}(A)$. 这样就有

$$\begin{aligned}a_1 &= p'_A(t)\,|_{t=0} = \mathrm{tr}\ \mathrm{adj}(tI - A)\,|_{t=0} = \mathrm{tr}\ \mathrm{adj}(-A)\\ &= (-1)^{n-1}\mathrm{tr}\ \mathrm{adj}(A) = (-1)^{n-1}E_{n-1}(A)\end{aligned}$$

现在注意 tr adj$(tI-A)=\sum_{i=1}^{n} p_{A_{(i)}}(t)$是 A 的 n 个 $n-1$ 阶主子矩阵的特征多项式之和，这些主子矩阵记为 $A_{(1)}$，…，$A_{(n)}$. 再次利用(0.8.10.2)计算

$$p''_A(t) = \frac{\mathrm{d}}{\mathrm{d}t}\mathrm{tr}\ \mathrm{adj}(tI - A) = \sum_{i=1}^{n}\frac{\mathrm{d}}{\mathrm{d}t}p_{A_{(i)}}(t) = \sum_{i=1}^{n}\mathrm{tr}\ \mathrm{adj}(tI - A_{(i)}) \tag{1.2.12}$$

每一个求和项 tr adj$(tI-A_{(i)})$都是 $tI-A$ 的一个主子式的 $n-1$ 个 $n-2$ 阶主子式之和，所以每一个求和项都是 $tI-A$ 的某些 $n-2$ 阶主子式之和. $tI-A$ 的$\binom{n}{n-2}$个 $n-2$ 阶主子式中的每一个都在(1.2.12)中出现两次：删去标号分别是 k 以及 l 的行与列得到的主子式既在 $i=k$ 也在 $i=l$ 时出现. 这样就有

$$\begin{aligned}a_2 &= \frac{1}{2}p''_A(t)\,|_{t=0} = \frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\ \mathrm{adj}(tI - A_{(i)})\,|_{t=0} = \frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\ \mathrm{adj}(-A_{(i)})\\ &= \frac{1}{2}(-1)^{n-2}\sum_{i=1}^{n}\mathrm{tr}\ \mathrm{adj}(A_{(i)}) = \frac{1}{2}(-1)^{n-2}(2E_{n-2}(A)) = (-1)^{n-2}E_{n-2}(A)\end{aligned}$$

重复此法显示有 $p_A^{(k)}(0)=k!\ (-1)^{n-k}E_{n-k}(A)$，$k=0$，1，…，$n-1$，故而特征多项式(1.2.11)的系数是

$$a_k = \frac{1}{k!}p_A^{(k)}(0) = (-1)^{n-k}E_{n-k}(A), \quad k = 0,1,\cdots,n-1$$

从而

53

$$p_A(t) = t^n - E_1(A)t^{n-1} + \cdots + (-1)^{n-1}E_{n-1}(A)t + (-1)^nE_n(A) \tag{1.2.13}$$

记住有恒等式(1.2.6)，我们给出下面的定义.

定义 1.2.14 n 个复数λ_1，…，λ_n 的第 k 个初等对称函数($k\leqslant n$)是

$$S_k(\lambda_1,\cdots,\lambda_n) = \sum_{1\leqslant i_1<\cdots<i_k\leqslant n}\prod_{j=1}^{k}\lambda_{i_j}$$

注意，这个和式有$\binom{n}{k}$个求和项. 如果 $A\in M_n$ 且λ_1，…，λ_n 为它的特征值，我们就定义 $S_k(A)=S_k(\lambda_1，\cdots，\lambda_n)$.

习题 $S_1(\lambda_1，\cdots，\lambda_n)$和 $S_n(\lambda_1，\cdots，\lambda_n)$等于什么？说明：如果对数组$\lambda_1$，…，$\lambda_n$ 重新编号或者排序，为什么每一个函数 $S_k(\lambda_1，\cdots，\lambda_n)$都不会改变. ◀

对(1.2.6)作计算给出

$$p_A(t) = t^n - S_1(A)t^{n-1} + \cdots + (-1)^{n-1}S_{n-1}(A)t + (-1)^nS_n(A) \tag{1.2.15}$$

(1.2.13)与(1.2.15)比较就给出矩阵的特征值的初等对称函数与其主子式的和之间的如下

恒等式.

定理 1.2.16 设 $A\in M_n$. 那么 $S_k(A)=E_k(A)$(对每个 $k=1,\cdots,n$).

下一个定理表明，一个奇异的复矩阵总可以稍加平移使之成为非奇异的. 这个重要的事实常常使得我们可以利用连续性方法从非奇异矩阵的性质推导出有关奇异矩阵的结果.

定理 1.2.17 设 $A\in M_n$. 则存在某个 $\delta>0$，使得只要 $\varepsilon\in\mathbf{C}$ 且 $0<|\varepsilon|<\delta$，$A+\varepsilon I$ 就是非奇异的.

证明 结论 1.1.8 确保 $\lambda\in\sigma(A)$ 的充分必要条件是 $\lambda+\varepsilon\in\sigma(A+\varepsilon I)$. 这样一来，$0\in\sigma(A+\varepsilon I)$ 当且仅当对某个 $\lambda\in\sigma(A)$ 有 $\lambda+\varepsilon=0$，即当且仅当对某个 $\lambda\in\sigma(A)$ 有 $\varepsilon=-\lambda$. 如果 A 的所有特征值都为零，就取 $\delta=1$. 如果 A 的某个特征值不为零，则令 $\delta=\min\{|\lambda|:\lambda\in\sigma(A)\text{ 且 }\lambda\neq0\}$. 如果我们选取任何一个满足 $0<|\varepsilon|<\delta$ 的 ε，我们确信必有 $-\varepsilon\notin\sigma(A)$，所以 $0\notin\sigma(A+\varepsilon I)$ 且 $A+\varepsilon I$ 是非奇异的. □

在多项式 $p(t)$ 的导数与其零点的重数之间有一个有用的联系：α 是 $p(t)$ 的一个重数为 $k\geqslant1$ 的零点，当且仅当可以将 $p(t)$ 写成形式

$$p(t)=(t-\alpha)^k q(t)$$

其中 $q(t)$ 是一个满足 $q(\alpha)\neq0$ 的多项式. 对此恒等式微分给出 $p'(t)=k(t-\alpha)^{k-1}q(t)+(t-\alpha)^k q'(t)$，它表明 $p'(\alpha)=0$ 当且仅当 $k>1$. 如果 $k\geqslant2$，那么 $p''(t)=k(k-1)(t-\alpha)^{k-2}\times q(t)+$若干个多项式项，其中每一项都含有一个因子 $(t-\alpha)^m$，$m\geqslant k-1$，所以，$p''(\alpha)=0$ 当且仅当 $k>2$. 重复这一计算表明，α 是 $p(t)$ 的 k 重零点，当且仅当 $p(\alpha)=p'(\alpha)=\cdots=p^{(k-1)}(\alpha)=0$ 以及 $p^{(k)}(\alpha)\neq0$. 54

定理 1.2.18 设 $A\in M_n$ 并假设 $\lambda\in\sigma(A)$ 的代数重数为 k. 那么 $\operatorname{rank}(A-\lambda I)\geqslant n-k$，且其中的等式当 $k=1$ 时成立.

证明 对矩阵 $A\in M_n$ 的特征多项式 $p_A(t)$ 应用上一结论，该矩阵有一个重数为 $k\geqslant1$ 的特征值 λ. 如果令 $B=A-\lambda I$，那么 0 就是 B 的一个重数为 k 的特征值，从而有 $p_B^{(k)}(0)\neq0$. 但是 $p_B^{(k)}(0)=k!\,(-1)^{n-k}E_{n-k}(B)$，故有 $E_{n-k}(B)\neq0$. 特别地，$B=A-\lambda I$ 的某个 $n-k$ 阶主子式不为零，所以 $\operatorname{rank}(A-\lambda I)\geqslant n-k$. 如果 $k=1$，我们就可以说得更多一些：$A-\lambda I$ 是奇异的，故而 $n>\operatorname{rank}(A-\lambda I)\geqslant n-1$，这就意味着：如果特征值 λ 的代数重数为 1，那么 $\operatorname{rank}(A-\lambda I)=n-1$. □

问题

1.2.P1 设 $A\in M_n$. 利用恒等式 $S_n(A)=E_n(A)$ 验证(1.1.7).

1.2.P2 对矩阵 $A\in M_{m,n}$ 以及 $B\in M_{n,m}$，用直接计算证明 $\operatorname{tr}(AB)=\operatorname{tr}(BA)$. 对任何 $A\in M_n$ 以及任何非奇异的 $S\in M_n$，推出有 $\operatorname{tr}(S^{-1}AS)=\operatorname{tr}A$. 对任何 A，$B\in M_n$，利用行列式函数的积性证明 $\det(S^{-1}AS)=\det A$. 导出结论：M_n 上的行列式函数是相似不变量.

1.2.P3 设 $D\in M_n$ 是对角矩阵. 计算其特征多项式 $p_D(t)$ 并证明 $p_D(D)=0$.

1.2.P4 假设 $A\in M_n$ 是幂等的. 利用(1.2.15)以及(1.1.P5)证明：$p_A(t)$ 的每个系数都是整数(正的、负的或者零).

1.2.P5 利用(1.1.P6)证明：幂零矩阵的迹为零. 幂零矩阵的特征多项式是什么？

1.2.P6 如果 $A\in M_n$，且 $\lambda\in\sigma(A)$ 的重数为 1，我们知道 $\operatorname{rank}(A-\lambda I)=n-1$. 考虑其逆命题：如果 $\operatorname{rank}(A-\lambda I)=n-1$，$\lambda$ 必定是 A 的特征值吗？它必定有重数 1 吗？

1.2.P7 利用(1.2.13)来确定三对角矩阵

$$\begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

的特征多项式．考虑可以怎样来利用此方法计算一般的 $n\times n$ 三对角矩阵的特征多项式．

1.2.P8　设给定 $A\in M_n$ 与 $\lambda\in\mathbf{C}$，假设 A 的特征值是 $\lambda_1,\cdots,\lambda_n$．说明为什么 $p_{A+\lambda I}(t)=p_A(t-\lambda)$，并由此恒等式得出结论：$A+\lambda I$ 的特征值是 $\lambda_1+\lambda,\cdots,\lambda_n+\lambda$．

1.2.P9　用显式计算 $S_2(\lambda_1,\cdots,\lambda_6)$，$S_3(\lambda_1,\cdots,\lambda_6)$，$S_4(\lambda_1,\cdots,\lambda_6)$以及 $S_5(\lambda_1,\cdots,\lambda_6)$．

55 1.2.P10　如果 $A\in M_n(\mathbf{R})$且 n 是**奇数**，证明 A 至少有一个实的特征值．

1.2.P11　设 V 是域 $\mathbf{F}$ 上的一个向量空间．线性变换 $T:V\rightarrow V$ 的特征值是一个纯量 $\lambda\in\mathbf{F}$，它使得存在一个非零向量 $v\in V$ 满足 $Tv=\lambda v$．如果 $\mathbf{F}=\mathbf{C}$，且 V 是有限维的，证明：每个线性变换 $T:V\rightarrow V$ 都有特征值．给出例子来说明，如果其中有一个假设被减弱(V 不是有限维的或者 $\mathbf{F}\neq\mathbf{C}$)，那么 T 有可能没有特征值．

1.2.P12　设给定 $x=[x_i]$，$y=[y_i]\in\mathbf{C}^n$ 以及 $a\in\mathbf{C}$，又令 $A=\begin{bmatrix} 0_n & x \\ y^* & a \end{bmatrix}\in M_{n+1}$．用两种方法证明 $p_A(t)=t^{n-1}(t^2-at-y^*x)$：(a)利用 Cauchy 展开式(0.8.5.10)计算 $p_A(t)$．(b)说明为什么 $\operatorname{rank}A\leqslant 2$，并利用(1.2.13)计算 $p_A(t)$．为什么只需要计算 $E_1(A)$和 $E_2(A)$，又为什么仅需要考虑形如 $\begin{bmatrix} 0 & x_i \\ \bar{y}_i & a \end{bmatrix}$的主子矩阵？证明 A 的特征值是$(a\pm\sqrt{a^2+4y^*x})/2$ 加上 $n-1$ 个零特征值．

1.2.P13　设 x，$y\in\mathbf{C}^n$，$a\in\mathbf{C}$，而 $B\in M_n$．考虑加边矩阵 $A=\begin{bmatrix} B & x \\ y^* & a \end{bmatrix}\in M_{n+1}$．(a)利用(0.8.5.10)证明

$$p_A(t)=(t-a)p_B(t)-y^*(\operatorname{adj}(tI-B))x \tag{1.2.19}$$

(b) 如果 $B=\lambda I_n$，导出

$$p_A(t)=(t-\lambda)^{n-1}(t^2-(a+\lambda)t+a\lambda-y^*x) \tag{1.2.20}$$

并得出结论：$\begin{bmatrix} \lambda I_n & x \\ y^* & a \end{bmatrix}$的特征值是 λ(重数为 $n-1$)以及$\frac{1}{2}(a+\lambda\pm((a-\lambda)^2+4y^*x)^{1/2})$．

1.2.P14　设 $n\geqslant 3$，$B\in M_{n-2}$且 λ，$\mu\in\mathbf{C}$．考虑分块矩阵

$$A=\begin{bmatrix} \lambda & \bigstar & \bigstar \\ 0 & \mu & 0 \\ 0 & \bigstar & B \end{bmatrix}$$

其中的记号★所代表的元素不一定是零．证明 $p_A(t)=(t-\lambda)(t-\mu)p_B(t)$．

1.2.P15　假设 $A(t)\in M_n$ 是一个给定的连续的矩阵值的函数，又每一个向量值函数 $x_1(t),\cdots,x_n(t)\in\mathbf{C}^n$ 都满足常微分方程组$x_j'(t)=A(t)x_j(t)$．设 $X(t)=[x_1(t)\cdots x_n(t)]$以及 $W(t)=\det X(t)$．利用(0.8.10)以及(0.8.2.11)，并对下面的论证

$$\begin{aligned} W'(t) &= \sum_{j=1}^{n}\det(X(t)\underset{j}{\leftarrow}x_j'(t)) = \operatorname{tr}[\det(X(t)\underset{i}{\leftarrow}x_j'(t))]_{i,j=1}^{n} \\ &= \operatorname{tr}((\operatorname{adj}X(t))X'(t)) = \operatorname{tr}((\operatorname{adj}X(t))A(t)X(t)) = W(t)\operatorname{tr}A(t) \end{aligned}$$

给出细节．这样一来，$W(t)$就满足纯量微分方程 $W'(t)=\operatorname{tr}A(t)W(t)$，它的解是关于 Wronski 行列式的 **Abel 公式**

$$W(t)=W(t_0)\mathrm{e}^{\int_{t_0}^{t}\operatorname{tr}A(s)\mathrm{d}s}$$

由此得出结论：如果对 $t=t_0$，向量 $x_1(t)$，…，$x_n(t)$线性无关，那么对所有的 t 它们都线性无关．你如何利用(1.2.P2)中的恒等式 $\mathrm{tr}BC=\mathrm{tr}CB$?

1.2.P16 设给定 $A\in M_n$ 和 x，$y\in\mathbf{C}^n$．设 $f(t)=\det(A+txy^{\mathrm{T}})$．利用(0.8.5.11)证明 $f(t)=\det A+\beta t$，这是 t 的一个线性函数．β 等于什么呢？对任何 $t_1\neq t_2$，证明 $\det A=(t_2f(t_1)-t_1f(t_2))/(t_2-$ 56
$t_1)$．现在考虑

$$A=\begin{bmatrix} d_1 & b & \cdots & b \\ c & d_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ c & \cdots & c & d_n \end{bmatrix}\in M_n$$

$x=y=e$(元素全为 1 的向量)，$t_1=b$ 且 $t_2=c$．设 $q(t)=(d_1-t)\cdots(d_n-t)$．证明：如果 $b\neq c$，则有 $\det A=(bq(c)-cq(b))/(b-c)$；而如果 $b=c$，则有 $\det A=q(b)-bq'(b)$．如果 $d_1=\cdots=d_n=0$，证明：如果 $b\neq c$，则有 $p_A(t)=(b(t+c)^n-c(t+b)^n)/(b-c)$；而如果 $b=c$，则有 $p_A(t)=(t+b)^{n-1}(t-(n-1)b)$．

1.2.P17 设 A，$B\in M_n$ 以及 $C=\begin{bmatrix} 0_n & A \\ B & 0_n \end{bmatrix}$．利用(0.8.5.13)和(0.8.5.14)证明 $p_C(t)=p_{AB}(t^2)=p_{BA}(t^2)$，并详细说明为什么这就蕴含 AB 和 BA 有相同的特征值．说明为什么这就确证有 $\mathrm{tr}AB=\mathrm{tr}BA$ 以及 $\det AB=\det BA$．还要说明为什么有 $\det(I+AB)=\det(I+BA)$．

1.2.P18 设 $A\in M_3$．说明为什么 $p_A(t)=t^3-(\mathrm{tr}A)t^2+(\mathrm{tr}\ \mathrm{adj}A)t-\det A$．

1.2.P19 假设 $A=[a_{ij}]\in M_n$ 的所有元素或者是零，或者是 1，并假设 A 的所有特征值 λ_1，…，λ_n 都是正实数．说明为什么 $\det A$ 是一个正整数，并对如下结论提供证明细节：

$$n\geqslant \mathrm{tr}A=\frac{1}{n}(\lambda_1+\cdots+\lambda_n)n\geqslant n(\lambda_1\cdots\lambda_n)^{1/n}=n(\det A)^{1/n}\geqslant n$$

导出结论：所有 $\lambda_i=1$，所有 $a_{ii}=1$ 且 $\det A=1$．

1.2.P20 对任何 $A\in M_n$，证明 $\det(I+A)=I+E_1(A)+\cdots+E_n(A)$．

1.2.P21 设给定 $A\in M_n$ 以及非零向量 x，$v\in\mathbf{C}^n$．假设 $c\in\mathbf{C}$，$v^*x=1$，$Ax=\lambda x$，且 A 的特征值是 λ，λ_2，…，λ_n．证明 Google 矩阵 $A(c)=cA+(1-c)\lambda xv^*$ 的特征值是 λ，$c\lambda_2$，…，$c\lambda_n$．

1.2.P22 考虑(0.9.6.2)中的 $n\times n$ 循环矩阵 C_n．对给定的 $\varepsilon>0$，令 $C_n(\varepsilon)$是用 ε 代替 C_n 的位于$(n,1)$处的元素所得到的矩阵．证明：$C_n(\varepsilon)$的特征多项式是 $p_{C_n(\varepsilon)}(t)=t^n-\varepsilon$，它的谱是 $\sigma(C_n(\varepsilon))=\{\varepsilon^{1/n}e^{2\pi ik/n}:k=0,1,\cdots,n-1\}$，且 $I+C_n(\varepsilon)$的谱半径是 $\rho(I+C_n(\varepsilon))=1+\varepsilon^{1/n}$．

1.2.P23 如果 $A\in M_n$ 是奇异的，且有不同的特征值，证明它有一个 $n-1$ 阶非奇异的主子式．

注记 主子式之**和**出现在关于特征多项式系数的讨论之中．主子式之**积**也以自然的方式出现，见(7.8.11)．

1.3 相似性

我们知道，M_n 中矩阵的相似变换与 $\mathbf{C}^n$ 上在另一组基下表示它的基本线性变换相对应．于是，研究相似性可以看成是研究线性变换固有的性质或者是它们所有的基表示都共 57
有的性质．

定义 1.3.1 设给定 A，$B\in M_n$．我们说 B 相似于 A，如果存在一个非奇异的矩阵 $S\in M_n$，使得

$$B=S^{-1}AS$$

变换 $A\to S^{-1}AS$ 称为由相似矩阵 S 给出的相似变换．我们说 B 置换相似于 A，即如果存在

一个置换矩阵 P，使得 $B=P^{\mathrm{T}}AP$. 关系"B 相似于 A"有时简记为 $B\sim A$.

结论 1.3.2 相似性是 M_n 上的一个等价关系，也就是说，相似性是自反的、对称的以及传递的，见(0.11).

与任何等价关系类似，相似性将集合 M_n 分划成不相交的等价类. 每一个等价类是 M_n 中与一个给定矩阵(这个类的一个代表元)相似的所有矩阵组成之集合. 一个等价类中所有的矩阵都是相似的. 不同等价类中的矩阵是不相似的. 关键的结论是处于一个相似类中的矩阵共同享有许多重要的性质. 其中有一些性质在这里提到，而对于**相似不变量**(similarity invariant)(例如 Jordan 标准型)的完整描述放在第 3 章讲述.

定理 1.3.3 设 A，$B\in M_n$. 如果 B 相似于 A，那么 A 与 B 有相同的特征多项式.

证明 计算

$$\begin{aligned}p_B(t)&=\det(tI-B)=\det(tS^{-1}S-S^{-1}AS)=\det(S^{-1}(tI-A)S)\\&=\det S^{-1}\det(tI-A)\det S=(\det S)^{-1}(\det S)\det(tI-A)=\det(tI-A)=p_A(t)\end{aligned}$$ □

推论 1.3.4 设 A，$B\in M_n$，并假设 A 与 B 相似. 那么

(a) A 与 B 有同样的特征值.

(b) 如果 B 是对角矩阵，那么它的主对角线上的元素就是 A 的特征值.

(c) $B=0$(这是一个对角矩阵)当且仅当 $A=0$.

(d) $B=I$(这是一个对角矩阵)当且仅当 $A=I$.

习题 验证上面推论中的结论. ◀

例 1.3.5 对相似性来说，有相同的特征值是一个必要但非充分的条件. 考虑 $\begin{bmatrix}0&1\\0&0\end{bmatrix}$与$\begin{bmatrix}0&0\\0&0\end{bmatrix}$，它们有同样的特征值，但并不是相似的(为什么不相似?).

习题 假设 A，$B\in M_n$ 是相似的且 $q(t)$是一个给定的多项式. 证明：$q(A)$与 $q(B)$相似. 特别地，证明：对任何 $\alpha\in\mathbf{C}$，$A+\alpha I$ 与 $B+\alpha I$ 相似. ◀

习题 设 A，B，C，$D\in M_n$. 假设 $A\sim B$ 以及 $C\sim D$，两者都通过同一个相似矩阵 S

相似. 证明 $A+C\sim B+D$ 以及 $AC\sim BD$. ◀

习题 设 A，$S\in M_n$ 并假设 S 是非奇异的. 证明：对所有 $k=1$，…，n 都有 $S_k(S^{-1}AS)=S_k(A)$，并说明为什么对所有 $k=1$，…，n 有 $E_k(S^{-1}AS)=E_k(A)$. 这样一来，所有主子式之和(1.2.10)是相似不变量，而并非只有行列式与迹是相似不变量. ◀

习题 说明为什么秩是一个相似不变量：如果 $B\in M_n$ 与 $A\in M_n$ 相似，那么 $\operatorname{rank}B=\operatorname{rank}A$. 提示：见(0.4.6). ◀

由于对角矩阵特别简单且有很好的性质，我们乐于知道何种矩阵与对角矩阵相似.

定义 1.3.6 如果 $A\in M_n$ 与一个对角矩阵相似，那么就说 A 是可以对角化的.

定理 1.3.7 设给定 $A\in M_n$. 那么 A 与一个形如

$$\begin{bmatrix}\Lambda&C\\0&D\end{bmatrix},\quad\Lambda=\operatorname{diag}(\lambda_1,\cdots,\lambda_k),\quad D\in M_{n-k},\quad 1\leqslant k<n\tag{1.3.7.1}$$

的分块矩阵相似，当且仅当在 $\mathbf{C}^n$ 中存在 k 个线性无关的向量，它们每一个都是 A 的特征向量. 矩阵 A 是可以对角化的，当且仅当存在 n 个线性无关的向量，它们每一个都是 A 的特征向量. 如果 $x^{(1)}$，…，$x^{(n)}$ 是 A 的线性无关的特征向量且 $S=[x^{(1)}\ \cdots\ x^{(n)}]$，那么

$S^{-1}AS$是对角矩阵．如果A与一个形如(1.3.7.1)的矩阵相似，那么Λ的对角元素是A的特征值；如果A与对角矩阵Λ相似，那么Λ的对角元素是A所有的特征值．

证明 假设$k<n$，且n元向量$x^{(1)}$，…，$x^{(k)}$是线性无关的，又对每个$i=1$，…，k有$Ax^{(i)}=\lambda_i x^{(i)}$．设$\Lambda=\text{diag}(\lambda_1,\cdots,\lambda_k)$，$S_1=[x^{(1)} \ \cdots \ x^{(k)}]$，并选取任意一个$S_2\in M_n$，使得$S=[S_1 \ \ S_2]$是非奇异的．计算

$$\begin{aligned}S^{-1}AS &= S^{-1}[Ax^{(1)}\cdots Ax^{(k)}\,AS_2] = S^{-1}[\lambda_1 x^{(1)}\cdots\lambda_k x^{(k)}\,AS_2]\\ &=[\lambda_1 S^{-1}x^{(1)}\cdots\lambda_k S^{-1}x^{(k)}\,S^{-1}AS_2] = [\lambda_1 e_1\cdots\lambda_k e_k S^{-1}AS_2]\\ &=\begin{bmatrix}\Lambda & C\\ 0 & D\end{bmatrix},\quad \Lambda=\text{diag}(\lambda_1,\cdots,\lambda_k),\begin{bmatrix}C\\ D\end{bmatrix}=S^{-1}AS_2\end{aligned}$$

反过来，如果S是非奇异的，$S^{-1}AS=\begin{bmatrix}\Lambda & C\\ 0 & D\end{bmatrix}$，且我们给出分划$S=[S_1 \ \ S_2]$，其中$S_1\in M_{n,k}$，那么$S_1$的列就是线性无关的，且$[AS_1 \ \ AS_2]=AS=S\begin{bmatrix}\Lambda & C\\ 0 & D\end{bmatrix}=[S_1\Lambda \ \ S_1C+S_2D]$．于是，$AS_1=S_1\Lambda$，所以$S_1$的每一列都是$A$的特征向量．

如果$k=n$且有$\mathbf{C}^n$的一组基$\{x^{(1)},\cdots,x^{(n)}\}$，使得对每个$i=1$，…，$n$有$Ax^{(i)}=\lambda_i x^{(i)}$，令$\Lambda=\text{diag}(\lambda_1,\cdots,\lambda_n)$以及$S=[x^{(1)} \ \cdots \ x^{(n)}]$，后者是非奇异的．我们上面的计算表明$S^{-1}AS=\Lambda$．反过来，如果$S$是非奇异的，且$S^{-1}AS=\Lambda$，那么$AS=S\Lambda$，故而$S$的每一列都是$A$的特征向量．

有关特征值的最后面那些结论可以从检查特征多项式得出来：如果$k<n$，则有$p_A(t)=p_\Lambda(t)p_D(t)$以及，如果$k=n$，则有$p_A(t)=p_\Lambda(t)$． □ 59

原则上讲，定理1.3.7的证明是对一个可以对角化的矩阵$A\in M_n$进行对角化的一种算法：求出A的所有n个特征值，求出n个与之相伴的(而且是线性无关的!)特征向量，并作出矩阵S．然而，除了很小的例子之外，这**不是**一种有实用价值的计算方法．

习题 证明$\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$是**不可以**对角化的．提示：如果它可以对角化，它就会与零矩阵相似．换一种问法，有多少个线性无关的特征向量与特征值0相伴呢？ ◀

习题 设$q(t)$是给定的多项式．如果A是可以对角化的，证明：$q(A)$也是可以对角化的．如果$q(A)$是可以对角化的，那么A必定是可以对角化的吗？为什么？ ◀

习题 如果λ是$A\in M_n$的一个特征值，其重数为$m\geq 1$，证明：A不可对角化，如果$\text{rank}(A-\lambda I)>n-m$． ◀

习题 如果在$\mathbf{C}^n$中存在k个线性无关的向量，它们中的每一个都是$A\in M_n$的与一个给定的特征值λ相伴的特征向量，详细说明为什么λ的(代数)重数至少为k． ◀

如果所有特征值都不相同，则可以确保对角化．这一事实的基础在于如下的与某些特征值有关的重要引理．

引理 1.3.8 设λ_1，…，λ_k是$A\in M_n$的$k\geq 2$个不同的特征值(即如果$i\neq j$且$1\leq i$，$j\leq k$，就有$\lambda_i\neq\lambda_j$)，又假设对每个$i=1$，…，k，$x^{(i)}$是与λ_i相伴的特征向量．那么诸向量$x^{(1)}$，…，$x^{(k)}$是线性无关的．

证明 假设存在复纯量α_1，…，α_k，使得$\alpha_1 x^{(1)}+\alpha_2 x^{(2)}+\cdots+\alpha_k x^{(k)}=0$．设$B_1=(A-$

$\lambda_2 I)(A-\lambda_3 I)\cdots(A-\lambda_k I)$(乘积中略去了 $A-\lambda_1 I$). 由于对每个 $i=1$，…，n，$x^{(i)}$是与特征值 λ_i 相伴的特征向量，我们就有 $B_1 x^{(i)}=(\lambda_i-\lambda_2)(\lambda_i-\lambda_3)\cdots(\lambda_i-\lambda_k)x^{(i)}$，它当 $2\leqslant i\leqslant k$ 时为零(这些因子中有一个为零)，而当 $i=1$ 时不为零(没有因子为零且 $x^{(1)}\neq 0$). 从而

$$\begin{aligned}0&=B_1(\alpha_1 x^{(1)}+\alpha_2 x^{(2)}+\cdots+\alpha_k x^{(k)})=\alpha_1 B_1 x^{(1)}+\alpha_2 B_1 x^{(2)}+\cdots+\alpha_k B_1 x^{(k)}\\&=\alpha_1 B_1 x^{(1)}+0+\cdots+0=\alpha_1 B_1 x^{(1)}\end{aligned}$$

这就确保有 $\alpha_1=0$，这是由于 $B_1 x^{(1)}\neq 0$. 对每个 $j=2$，…，k 重复这种论证方法，用类似于定义 B_1 的乘积来定义 B_j，不过在其中要略去因子 $A-\lambda_j I$. 对每个 j，我们求得 $\alpha_j=0$，故而 $\alpha_1=\cdots=\alpha_k=0$，从而 $x^{(1)}$，…，$x^{(k)}$是线性无关的. □

定理 1.3.9 如果 $A\in M_n$ 有 n 个不同的特征根，那么 A 是可以对角化的.

证明 设对每个 $i=1$，…，n，$x^{(i)}$是与特征值 λ_i 相伴的特征向量. 由于所有的特征值都不相同，引理 1.3.8 确保向量 $x^{(1)}$，…，$x^{(n)}$是线性无关的. 这样定理 1.3.7 就保证了 A 可以对角化. □

60 有不同的特征值是可以对角化的**充分**条件，不过当然，它不是**必要**条件.

习题 给出一个可以对角化但不是有不同特征值的矩阵的例子. ◀

习题 设 A，$P\in M_n$，并假设 P 是一个置换矩阵，所以 P 的每一个元素或者为 0，或者为 1，且有 $P^{\mathrm{T}}=P^{-1}$，见(0.9.5). 证明：置换相似 PAP^{-1}对 A 的对角元素改动了次序. 对任意一个给定的对角矩阵 $D\in M_n$，说明为什么存在一个置换相似 PDP^{-1}，它把 D 的对角元素安排成任意给定的次序. 特别地，说明为什么 P 可以这样来选取，使得任何重复的对角元素都邻接在一起出现. ◀

一般而言，矩阵 A，$B\in M_n$ **不可**交换，但是如果 A 和 B 两者皆为对角矩阵，那么它们总是可以交换的. 后一结论可以稍作推广. 在这方面，下一引理是有助益的.

引理 1.3.10 设给定 $B_1\in M_{n_1}$，…，$B_d\in M_{n_d}$，并设 B 是直和

$$B=\begin{bmatrix}B_1 & & 0\\ & \ddots & \\ 0 & & B_d\end{bmatrix}=B_1\oplus\cdots\oplus B_d$$

那么 B 是可以对角化的，当且仅当 B_1，…，B_d 是可以对角化的.

证明 如果对每个 $i=1$，…，d，存在一个非奇异的 $S_i\in M_{n_i}$，使得 $S_i^{-1}B_i S_i$ 是对角矩阵，又如果我们定义 $S=S_1\oplus\cdots\oplus S_d$，然后我们来检验 $S^{-1}BS$ 是对角的.

关于其逆命题，我们用归纳法. 对 $d=1$ 没什么要证明的. 假设 $d\geqslant 2$，且假设此结论已经对不多于 $d-1$ 个直和项的直和得到证明. 设 $C=B_1\oplus\cdots\oplus B_{d-1}$，$n=n_1+\cdots+n_{d-1}$，并设 $m=n_d$. 设 $S\in M_{n+m}$是非奇异的，且

$$S^{-1}BS=S^{-1}(C\oplus B_d)S=\Lambda=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_{n+m})$$

将此恒等式改写成 $BS=S\Lambda$. 分划 $S=[s_1\quad s_2\quad\cdots\quad s_{n+m}]$，其中

$$s_i=\begin{bmatrix}\xi_i\\ \eta_i\end{bmatrix}\in\mathbf{C}^{n+m},\quad \xi_i\in\mathbf{C}^n,\eta_i\in\mathbf{C}^m,i=1,2,\cdots,n+m$$

那么 $Bs_i=\lambda_i s_i$ 蕴含 $C\xi_i=\lambda_i\xi_i$ 以及 $B_d\eta_i=\lambda_i\eta_i$(对 $i=1$，2，…，$n+m$). $[\xi_1\quad\cdots\quad\xi_{n+m}]\in M_{n,n+m}$的行秩是 n，因为这个矩阵由非奇异矩阵 S 的前 n 行组成. 从而它的列秩也是 n，所以向量组 ξ_1，…，ξ_{n+m}包含由 n 个向量组成的线性无关组，其中每一个都是 C 的特征向量.

定理 1.3.7 确保 C 可以对角化，而归纳假设就保证了它的直和项 B_1，…，B_d 全都是可以对角化的. $[\eta_1 \quad \cdots \quad \eta_{n+m}] \in M_{n,n+m}$ 的行秩是 m，故而向量组 η_1，…，η_{n+m} 包含 m 个向量的线性无关组. 由此推出 B_d 也是可以对角化的. □

定义 1.3.11 两个矩阵 A，$B \in M_n$ 说成是可同时对角化的，如果存在单独一个非奇异的矩阵 $S \in M_n$，使得 $S^{-1}AS$ 和 $S^{-1}BS$ 这两者都是对角矩阵.

习题 设 A，B，$S \in M_n$，并假设 S 非奇异. 证明 A 与 B 可交换，当且仅当 $S^{-1}AS$ 与 $S^{-1}BS$ 可交换. ◀

习题 如果 A，$B \in M_n$ 是可同时对角化的，证明它们可交换. 提示：对角矩阵可交换. ◀

习题 证明：如果 $A \in M_n$ 是可以对角化的，且 $\lambda \in \mathbf{C}$，那么 A 与 λI 是可同时对角化的. ◀

定理 1.3.12 设 A，$B \in M_n$ 是可对角化的. 那么 A 与 B 可交换，当且仅当它们是可同时对角化的.

证明 假设 A 与 B 可交换，对 A 与 B 两者作相似变换使 A 对角化(但并不一定使 B 对角化)，并将 A 的重特征值组合在一起. 如果 μ_1，…，μ_d 是 A 的不同的特征值，而 n_1，…，n_d 分别是它们的重数，那么我们就可以假设

$$A = \begin{bmatrix} \mu_1 I_{n_1} & & & 0 \\ & \mu_2 I_{n_2} & & \\ & & \ddots & \\ 0 & & & \mu_d I_{n_d} \end{bmatrix}, \quad \mu_i \neq \mu_j, \quad i \neq j \tag{1.3.13}$$

由于 $AB=BA$，(0.7.7)就保证

$$B = \begin{bmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_d \end{bmatrix}, \quad 每个\ B_i \in M_{n_i} \tag{1.3.14}$$

是与 A 共形的分块对角矩阵. 由于 B 是可对角化的，(1.3.10)确保每个 B_i 都是可对角化的. 设 $T_i \in M_{n_i}$ 是非奇异的且使得 $T_i^{-1}B_iT_i$ 为对角矩阵(对每个 $i=1$，…，d)，令

$$T = \begin{bmatrix} T_1 & & & 0 \\ & T_2 & & \\ & & \ddots & \\ 0 & & & T_d \end{bmatrix} \tag{1.3.15}$$

那么 $T_i^{-1}\mu_i I_{n_i} T_i = \mu_i I_{n_i}$，所以 $T^{-1}AT=A$ 与 $T^{-1}BT$ 两者同为对角矩阵. 相反的结论包含在稍早的一个习题中. □

我们希望对定理 1.3.12 有一个包容任意多个可交换的可对角化矩阵的形式. 我们的研究中心是不变子空间的概念以及与之相伴的分块三角矩阵的概念.

定义 1.3.16 矩阵族 $\mathcal{F} \subseteq M_n$ 是矩阵的一个非空有限或者无限的集合；交换族是这样一族矩阵，其中每一对矩阵都是可交换的. 对给定的 $A \in M_n$，我们称子空间 $W \subseteq \mathbf{C}^n$ 是 A-不变的，如果对每个 $w \in W$ 都有 $Aw \in W$. 子空间 $W \subseteq \mathbf{C}^n$ 称为是平凡的，如果或者 $W=$

$\{0\}$，或者 $W=\mathbf{C}^n$；反之则称它是非平凡的. 对一个给定的族 $\mathcal{F}\subseteq M_n$，我们称子空间 $W\subseteq \mathbf{C}^n$ 是 $\mathcal{F}$-不变的，如果对每个 $A\in\mathcal{F}$，W 都是 A-不变的. 给定的族 $\mathcal{F}\subseteq M_n$ 称为是可约的，
62 如果 $\mathbf{C}^n$ 中某个非平凡的子空间是 $\mathcal{F}$-不变的，反之则称 $\mathcal{F}$ 是不可约的.

习题 对 $A\in M_n$，证明：$\mathbf{C}^n$ 的一维 A-不变子空间中的每一个非零元素都是 A 的一个特征向量. ◀

习题 假设 $n\geqslant 2$，且 $S\in M_n$ 是非奇异的. 分划 $S=[S_1\quad S_2]$，其中 $S_1\in M_{n,k}$，而 $S_2\in M_{n,n-k}(1<k<n)$. 说明为什么

$$S^{-1}S_1=[e_1\ \cdots\ e_k]=\begin{bmatrix}I_k\\0\end{bmatrix},\quad S^{-1}S_2=[e_{k+1}\ \cdots\ e_n]=\begin{bmatrix}0\\I_{n-k}\end{bmatrix}$$ ◀

不变子空间与分块三角矩阵是同一个有价值的事物的两个面：前者是关于线性代数的，而后者则是关于矩阵分析的. 设 $A\in M_n(n\geqslant 2)$，并假设 $W\subseteq\mathbf{C}^n$ 是一个 k 维子空间$(1<k<n)$. 选取 W 的一组基 s_1，…，s_k，并令 $S_1=[s_1\cdots s_k]\in M_{n,k}$. 选取任意的 s_{k+1}，…，s_n，使得 s_1，…，s_n 是 $\mathbf{C}^n$ 的一组基，设 $S_2=[s_{k+1}\cdots s_n]\in M_{n,n-k}$，并令 $S=[S_1\quad S_2]$. S 的列是线性无关的，所以它是非奇异的. 如果 W 是 A-不变的，那么对每个 $j=1$，…，k 都有 $As_j\in W$，所以每个 As_j 都是 s_1，…，s_k 的线性组合，这就是说，对某个 $B\in M_k$ 有 $AS_1=S_1B$. 如果 $AS_1=S_1B$，那么就有 $AS=[AS_1\quad AS_2]=[S_1B\quad AS_2]$，从而

$$\begin{aligned}S^{-1}AS&=[S^{-1}S_1B\quad S^{-1}AS_2]=\left[\begin{bmatrix}I_k\\0\end{bmatrix}B\quad S^{-1}AS_2\right]\\&=\begin{bmatrix}B&C\\0&D\end{bmatrix},\quad B\in M_k,1\leqslant k\leqslant n-1\end{aligned}\tag{1.3.17}$$

结论是：A 与一个分块三角矩阵(1.3.17)相似，如果它有一个 k 维不变子空间. 不过我们可以稍微再多说一点：我们知道 $B\in M_k$ 有一个特征值，故可假设对某个纯量 λ 以及一个非零的 $\xi\in\mathbf{C}^k$ 有 $B\xi=\lambda\xi$. 这样就有 $0\neq S_1\xi\in W$ 以及 $A(S_1\xi)=(AS_1)\xi=S_1B\xi=\lambda(S_1\xi)$，这就意味着 A 在 W 中有一个特征向量.

反过来，如果 $S=[S_1\quad S_2]\in M_n$ 是非奇异的，$S_1\in M_{n,k}$，且 $S^{-1}AS$ 有分块三角型(1.3.17)，这样就有

$$AS_1=AS\begin{bmatrix}I_k\\0\end{bmatrix}=S\begin{bmatrix}B&C\\0&D\end{bmatrix}\begin{bmatrix}I_k\\0\end{bmatrix}=[S_1\quad S_2]\begin{bmatrix}B_1\\0\end{bmatrix}=S_1B$$

所以 S_1 的列生成的(k 维)子空间是 A-不变的. 我们把上面的讨论总结成如下的结论.

结论 1.3.18 假设 $n\geqslant 2$. 给定的 $A\in M_n$ 与一个形如(1.3.17)的分块三角矩阵相似，当且仅当 $\mathbf{C}^n$ 的某个非平凡的子空间是 A-不变的. 此外，如果 $W\subseteq\mathbf{C}^n$ 是一个非零的 A-不变子空间，那么 W 中某个向量就是 A 的一个特征向量. 一个给定的族 $\mathcal{F}\subseteq M_n$ 是可约的，当且仅当存在某个 $k\in\{2,\ \cdots,\ n-1\}$ 以及一个非奇异的 $S\in M_n$，使得对每个 $A\in\mathcal{F}$，$S^{-1}AS$ 有(1.3.17)的形式.

下面的引理是许多后续结果的核心要素.

引理 1.3.19 设 $\mathcal{F}\subset M_n$ 是一个交换族. 那么 $\mathbf{C}^n$ 中某个非零向量是每个 $A\in\mathcal{F}$ 的特征
63 向量.

证明 总有一个非零的 $\mathcal{F}$-不变子空间，即 $\mathbf{C}^n$. 设 $m=\min\{\dim V：V$ 是 $\mathbf{C}^n$ 的非零的

$\mathcal{F}$-不变子空间},并令 W 是任意一个给定的$\mathcal{F}$-不变子空间,它使得 $\dim W=m$. 设给定任意的 $A\in\mathcal{F}$. 由于 W 是$\mathcal{F}$-不变的,故而它是 A-不变的,所以(1.3.18)确保存在某个非零的 $x_0\in W$ 以及某个 $\lambda\in\mathbf{C}$,使得 $Ax_0=\lambda x_0$. 考虑子空间 $W_{A,\lambda}=\{x\in W: Ax=\lambda x\}$. 这样就有 $x_0\in W_{A,\lambda}$,所以 $W_{A,\lambda}$ 是 W 的一个非零的子空间. 对任何 $B\in\mathcal{F}$以及任何 $x\in W_{A,\lambda}$,W 的$\mathcal{F}$-不变量确保 $Bx\in W$. 利用$\mathcal{F}$的交换性,我们来计算

$$A(Bx)=(AB)x=(BA)x=B(Ax)=B(\lambda x)=\lambda(Bx)$$

它表明 $Bx\in W_{A,\lambda}$. 于是,$W_{A,\lambda}$ 是$\mathcal{F}$-不变且非零的,所以 $\dim W_{A,\lambda}\geqslant m$. 但是 $W_{A,\lambda}\subseteq W$,所以 $\dim W_{A,\lambda}\leqslant m$,从而 $W=W_{A,\lambda}$. 现在我们就已经指出了:对每个 $A\in\mathcal{F}$,存在某个纯量 λ_A,使得对所有 $x\in W$ 都有 $Ax=\lambda_A x$,所以 W 中每个非零向量都是$\mathcal{F}$中每个矩阵的一个特征向量. □

习题 考虑上面证明中的非零的$\mathcal{F}$-不变子空间 W. 说明为什么 $m=\dim W=1$. ◀

习题 假设$\mathcal{F}\subset M_n$ 是一个交换族. 证明:存在一个非奇异的 $S\in M_n$,使得对每个 $A\in\mathcal{F}$,$S^{-1}AS$ 都有分块三角型(1.3.17)(对 $k=1$). ◀

引理 1.3.19 关注的是任意非零基数的交换族. 我们的下一个结果表明,定理 1.3.12 可以延拓到由可对角化矩阵组成的任意的交换族上去.

定义 1.3.20 一个族$\mathcal{F}\subset M_n$ 说成是**可同时对角化的**(simultaneously diagonalizable),如果存在单独一个非奇异的 $S\in M_n$,使得对每个 $A\in\mathcal{F}$,$S^{-1}AS$ 都是对角的.

定理 1.3.21 *设$\mathcal{F}\subset M_n$ 是一个可对角化矩阵族. 那么$\mathcal{F}$是一个交换族,当且仅当它是一个可同时对角化的族. 此外,对任意给定的 $A_0\in\mathcal{F}$以及 A_0 的特征值的任意给定的排序 $\lambda_1,\cdots,\lambda_n$,都存在一个非奇异的 $S\in M_n$,使得对每个 $B\in\mathcal{F}$,$S^{-1}A_0S=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$ 以及 $S^{-1}BS$ 都是对角的.*

证明 如果$\mathcal{F}$是可同时对角化的,那么根据上一习题,它就是一个交换族. 我们用对 n 的归纳法证明其逆. 如果 $n=1$,就没什么需要证明的,因为每一个族既是交换的,又是对角的. 假设 $n\geqslant 2$,又设对每个 $k=1,2,\cdots,n-1$,任何由 $k\times k$ 可对角化矩阵组成的交换族都是可同时对角化的. 如果$\mathcal{F}$中每个矩阵都是纯量矩阵,就没什么要证明的,故而我们可以假设 $A\in\mathcal{F}$是一个 $n\times n$ 可对角化矩阵,它有不同的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_k$,且 $k\geqslant 2$,又设对每个 $B\in\mathcal{F}$有 $AB=BA$,且每个 $B\in\mathcal{F}$都是可对角化的. 利用(1.3.12)中的论证方法,我们将其转化成 A 形如(1.3.13)中的情形. 由于每个 $B\in\mathcal{F}$都与 A 可交换,(0.7.7)就确保每一个 $B\in\mathcal{F}$都有(1.3.14)的形式. 设 $B,\hat{B}\in\mathcal{F}$,所以 $B=B_1\oplus\cdots\oplus B_k$,且$\hat{B}=\hat{B}_1\oplus\cdots\oplus\hat{B}_k$,其中 B_i 与$\hat{B}_i$ 中每一个都有同样的大小,且它们的阶不超过 $n-1$. B 与$\hat{B}$的交换性以及可对角化性质蕴含 B_i 与$\hat{B}_i$ 的交换性以及可对角化性质(对每个 $i=1,\cdots,d$). 根据归纳假设,存在 k 个有适当大小的相似矩阵 $T_1,T_2,\cdots,T_k$,它们中每一个都可以使$\mathcal{F}$中每一个矩阵的对应的分块对角化. 这样一来,直和(1.3.15)就使得$\mathcal{F}$中每一个矩阵对角化了. 64

我们证明了,存在一个非奇异的 $T\in M_n$,使得对每个 $B\in\mathcal{F}$,$T^{-1}BT$ 都是对角矩阵. 这样就对某个置换矩阵 P 有 $T^{-1}A_0T=P\mathrm{diag}(\lambda_1,\cdots,\lambda_n)P^{\mathrm{T}}$,$P^{\mathrm{T}}(T^{-1}A_0T)P=(TP)^{-1}A_0\times(TP)=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$与$(TP)^{-1}B(TP)=P^{\mathrm{T}}(T^{-1}BT)P$ 都是对角的(对每个 $B\in\mathcal{F}$)(0.9.5). □

注 我们将两件重要的事项推迟到第 3 章:(1)给定 $A,B\in M_n$,怎样来确定 A 是否与 B 相似?(2)在不知道其

特征向量的情况下，如何才能得知给定的矩阵是否是可以对角化的？

尽管 AB 与 BA 不一定相同(即便当两个乘积都有定义时，这两个乘积的大小也未必相同)，它们的特征值有可能有相当多是相同的．的确，如果 A 和 B 两者均为方阵，那么 AB 与 BA 恰好有完全相同的特征值．这些重要的事实可以从一个简单而十分有用的结论中推导出来．

习题 设给定 $X\in M_{m,n}$．说明为什么 $\begin{bmatrix} I_m & X \\ 0 & I_n \end{bmatrix}\in M_{m+n}$ 是非奇异的，并验证它的逆是 $\begin{bmatrix} I_m & -X \\ 0 & I_n \end{bmatrix}$． ◀

定理 1.3.22 假设 $A\in M_{m,n}$ 以及 $B\in M_{n,m}(m\leqslant n)$．那么 BA 的 n 个特征值是 AB 的 m 个特征值加上 $n-m$ 个零，这就是说，$p_{BA}(t)=t^{n-m}p_{AB}(t)$．如果 $m=n$ 且 A 与 B 中至少有一个非奇异，那么 AB 与 BA 相似．

证明 计算给出

$$\begin{bmatrix} I_m & -A \\ 0 & I_n \end{bmatrix}\begin{bmatrix} AB & 0 \\ B & 0_n \end{bmatrix}\begin{bmatrix} I_m & A \\ 0 & I_n \end{bmatrix}=\begin{bmatrix} 0_m & 0 \\ B & BA \end{bmatrix}$$

而上一习题确保 $C_1=\begin{bmatrix} AB & 0 \\ B & 0_n \end{bmatrix}$ 与 $C_2=\begin{bmatrix} 0_m & 0 \\ B & BA \end{bmatrix}$ 相似．C_1 的特征值是 AB 的特征值加上 n 个零．C_2 的特征值是 BA 的特征值加上 m 个零．由于 C_1 与 C_2 的特征值是相同的，这就得到定理的第一个论断．定理的最后论断得自如下结论：如果 A 非奇异且 $m=n$，那么就有 $AB=A(BA)A^{-1}$． □

定理 1.3.22 有许多应用，其中的一些出现在下面几章中．这里只是其中的四个应用．

例 1.3.23(下秩矩阵的特征值) 假设 $A\in M_n$ 分解成 $A=XY^{\mathrm{T}}$，其中 $X,Y\in M_{n,r}$，且 $r<n$．这样 A 的特征值就与 $r\times r$ 矩阵 $Y^{\mathrm{T}}X$ 的特征值相同，再加上 $n-r$ 个零．例如，考虑 $n\times n$ 全 1 矩阵 $J_n=ee^{\mathrm{T}}$(0.2.8)．其特征值是 1×1 矩阵 $e^{\mathrm{T}}e=[n]$ 的特征值，即 n 加上 $n-1$ 个零．任何形如 $A=xy^{\mathrm{T}}$ 的矩阵的特征值(其中 $x,y\in\mathbf{C}^n$)(A 的秩至多为 1)是 $y^{\mathrm{T}}x$ 再加上 $n-1$ 个零．任意一个形如 $A=xy^{\mathrm{T}}+zw^{\mathrm{T}}=[x\quad z][y\quad w]^{\mathrm{T}}$ 的矩阵(其中 $x,y,z,w\in\mathbf{C}^n$)(A 的秩至多为 2)是 $[y\quad w]^{\mathrm{T}}[x\quad z]=\begin{bmatrix} y^{\mathrm{T}}x & y^{\mathrm{T}}z \\ w^{\mathrm{T}}x & w^{\mathrm{T}}z \end{bmatrix}$ 的两个特征值(1.2.4b)加上 $n-2$ 个零．

65

例 1.3.24(Cauchy 行列式恒等式) 设给定非奇异的 $A\in M_n$ 以及 $x,y\in\mathbf{C}^n$．那么

$$\begin{aligned}\det(A+xy^{\mathrm{T}}) &= (\det A)(\det(I+A^{-1}xy^{\mathrm{T}})) = (\det A)\prod_{i=1}^{n}\lambda_i(I+A^{-1}xy^{\mathrm{T}}) \\ &= (\det A)\prod_{i=1}^{n}(1+\lambda_i(A^{-1}xy^{\mathrm{T}})) = (\det A)(1+y^{\mathrm{T}}A^{-1}x) \quad (\text{应用}(1.3.23)) \\ &= \det A+y^{\mathrm{T}}((\det A)A^{-1})x = \det A+y^{\mathrm{T}}(\operatorname{adj}A)x\end{aligned}$$

对任何 $A\in M_n$，Cauchy 恒等式 $\det(A+xy^{\mathrm{T}})=\det A+y^{\mathrm{T}}(\operatorname{adj}A)x$ 都成立，现在它可以由连续性得出．有关不同的方法，见(0.8.5)．

例 1.3.25 对任何 $n\geqslant 2$，考虑 $n\times n$ 实对称 Hankel 矩阵

$$A=[i+j]_{i,j=1}^{n}=\begin{bmatrix}2&3&4&\cdots\\3&4&5&\cdots\\4&5&6&\cdots\\\vdots&&&\ddots\end{bmatrix}=ve^{\mathrm{T}}+ev^{\mathrm{T}}=[v\quad e][e\quad v]^{\mathrm{T}}$$

其中 $e\in\mathbf{R}^n$ 的每个元素都是 1，且 $v=[1\quad 2\quad\cdots\quad n]^{\mathrm{T}}$. A 的特征值与

$$B=[e\quad v]^{\mathrm{T}}[v\quad e]=\begin{bmatrix}e^{\mathrm{T}}v&e^{\mathrm{T}}e\\v^{\mathrm{T}}v&v^{\mathrm{T}}e\end{bmatrix}=\begin{bmatrix}\dfrac{n(n+1)}{2}&n\\\dfrac{n(n+1)(2n+1)}{6}&\dfrac{n(n+1)}{2}\end{bmatrix}$$

的特征值相同，再加上 $n-2$ 个零. 按照(1.2.4b)，B 的特征值(一正一负)是

$$n(n+1)\left[\frac{1}{2}\pm\sqrt{\frac{2n+1}{6(n+1)}}\right]$$

例 1.3.26 对任何 $n\geqslant 2$，考虑 $n\times n$ 实斜对称的 Toeplitz 矩阵

$$A=[i-j]_{i,j=1}^{n}=\begin{bmatrix}0&-1&-2&\cdots\\1&0&-1&\cdots\\2&1&0&\cdots\\\vdots&&&\ddots\end{bmatrix}=ve^{\mathrm{T}}-ev^{\mathrm{T}}=[v\quad -e][e\quad v]^{\mathrm{T}}$$

其中 $e\in\mathbf{R}^n$ 的每个元素都是 1，且 $v=[1\quad 2\quad\cdots\quad n]^{\mathrm{T}}$. 除了 $n-2$ 个零之外，A 的特征值与

$$B=[e\quad v]^{\mathrm{T}}[v\quad -e]=\begin{bmatrix}e^{\mathrm{T}}v&-e^{\mathrm{T}}e\\v^{\mathrm{T}}v&-v^{\mathrm{T}}e\end{bmatrix}$$

的特征值相同，再次利用(1.2.4b)得出后者的特征值是 $\pm\dfrac{ni}{2}\sqrt{\dfrac{n^2-1}{3}}$. 66

关于 AB 相对于 BA 的特征值的定理 1.3.22 仅仅是这方面内容的一部分，我们将在(3.2.11)中再次回到这个论题.

如果 $A\in M_n$ 可以对角化，且 $A=S\Lambda S^{-1}$，那么对任何 $a\neq 0$，aS 也都可以使 A 对角化. 于是，对角化相似从来就不是唯一的. 尽管如此，A 与一个特殊的对角矩阵的**每一个**相似都可以从一个给定的相似得出.

定理 1.3.27 假设 $A\in M_n$ 可以对角化，设 $\mu_1,\cdots,\mu_d$ 是它的不同的特征值，相应的重数分别为 $n_1,\cdots,n_d$. 设 $S,T\in M_n$ 是非奇异的，又假设 $A=S\Lambda S^{-1}$，其中 Λ 是形如(1.3.13)的对角矩阵.

(a) $A=T\Lambda T^{-1}$ 当且仅当 $T=S(R_1\oplus\cdots\oplus R_d)$，其中每一个 $R_i\in M_{n_i}$ 都是非奇异的.

(b) 如果将 $S=[S_1\quad\cdots\quad S_d]$ 以及 $T=[T_1\quad\cdots\quad T_d]$ 与 Λ 共形地分划，那么 $A=S\Lambda S^{-1}=T\Lambda T^{-1}$，当且仅当对每个 $i=1,\cdots,d$，S_i 的列空间与 T_i 的列空间相同.

(c) 如果 A 有不同的特征值，且 $S=[s_1\quad\cdots\quad s_n]$ 与 $T=[t_1\quad\cdots\quad t_n]$ 按照它们的列来进行分划，那么 $A=S\Lambda S^{-1}=T\Lambda T^{-1}$，当且仅当存在一个非奇异的对角矩阵 $R=\mathrm{diag}(r_1,\cdots,r_n)$，使得 $T=SR$，当且仅当对每个 $i=1,\cdots,n$，列 s_i 是对应的列 t_i 的一个非零的纯量倍数.

证明 我们有 $S\Lambda S^{-1}=T\Lambda T^{-1}$，当且仅当 $(S^{-1}T)\Lambda=\Lambda(S^{-1}T)$，当且仅当 $S^{-1}T$ 是与

Λ 共形的分块对角矩阵(0.7.7)，即当且仅当 $S^{-1}T=R_1\oplus\cdots\oplus R_d$ 且每一个 $R_i\in M_{n_i}$ 都是非奇异的. 对于(b)，注意到如果 $1\leqslant k\leqslant n$，那么 $X\in M_{n,k}$ 的列空间包含在 $Y\in M_{n,k}$ 的列空间中，当且仅当存在某个 $C\in M_k$，使得 $X=YC$. 进一步，如果 $\operatorname{rank}X=\operatorname{rank}Y=k$，那么 C 必定是非奇异的. 结论(c)是(a)与(b)的特例. □

如果实矩阵是通过一个复矩阵相似的，那么它们也能通过一个实矩阵相似吗？对于可交换的实矩阵，(1.3.21)有实的形式吗？下面的引理是回答这类问题的关键.

引理 1.3.28 设 $S\in M_n$ 非奇异，而 $S=C+\mathrm{i}D$，其中 C，$D\in M_n(\mathbf{R})$. 那么存在一个实数 τ，使得 $T=C+\tau D$ 是非奇异的.

证明 如果 C 是非奇异的，取 $\tau=0$. 如果 C 是奇异的，考虑多项式 $p(t)=\det(C+tD)$，它不是常数(零次)多项式，这是因为 $p(0)=\det C=0\neq\det S=p(\mathrm{i})$. 由于 $p(t)$在复平面上仅有有限多个零点，故而存在一个实数 τ，使得 $p(\tau)\neq 0$，所以 $C+\tau D$ 是非奇异的. □

定理 1.3.29 设$\mathcal{F}=\{A_\alpha:\alpha\in\mathcal{I}\}\subset M_n(\mathbf{R})$以及$\mathcal{G}=\{B_\alpha:\alpha\in\mathcal{I}\}\subset M_n(\mathbf{R})$是给定的实矩阵族. 如果存在一个非奇异的 $S\in M_n$，使得对每个 $\alpha\in\mathcal{I}$ 都有 $A_\alpha=SB_\alpha S^{-1}$，那么就存在一个非奇异的 $T\in M_n(\mathbf{R})$，使得对每个 $\alpha\in\mathcal{I}$ 都有 $A_\alpha=TB_\alpha T^{-1}$. 特别地，两个在 $\mathbf{C}$ 上相似的实矩阵在 $\mathbf{R}$ 上也是相似的.

证明 设 $S=C+\mathrm{i}D$ 是非奇异的，其中 C，$D\in M_n(\mathbf{R})$. 上面的引理确保存在一个实数 τ，使得 $T=C+\tau D$ 是非奇异的. 相似性 $A_\alpha=SB_\alpha S^{-1}$ 等价于恒等式 $A_\alpha(C+\mathrm{i}D)=A_\alpha S=SB_\alpha=(C+\mathrm{i}D)B_\alpha$. 令这个恒等式的实部和虚部分别相等表明 $A_\alpha C=CB_\alpha$ 以及 $A_\alpha D=DB_\alpha$. 由此有 $A_\alpha C=CB_\alpha$ 以及 $A_\alpha(\tau D)=(\tau D)B_\alpha$，所以 $A_\alpha T=TB_\alpha$ 且 $A_\alpha=TB_\alpha T^{-1}$. □

上面定理的一个直接推论是(1.3.21)的实的形式.

推论 1.3.30 设$\mathcal{F}=\{A_\alpha:\alpha\in\mathcal{I}\}\subset M_n(\mathbf{R})$是具有实特征值的可对角化的实矩阵族. 那么$\mathcal{F}$是一个交换族，当且仅当存在一个非奇异的实矩阵 T，使得 $T^{-1}A_\alpha T=\Lambda_\alpha$ 对每个 $A\in\mathcal{F}$ 都是对角矩阵. 此外，对任何给定的 $\alpha_0\in\mathcal{I}$ 以及 A_{α_0} 的特征值的任意一种给定的排序 λ_1，…，λ_n，存在一个非奇异的 $T\in M_n(\mathbf{R})$，使得对每个 $\alpha\in\mathcal{I}$，$T^{-1}A_{\alpha_0}T=\operatorname{diag}(\lambda_1,\cdots,\lambda_n)$与 $T^{-1}A_\alpha T$ 都是对角矩阵.

证明 对"必要性"这部分结论，将上面的定理应用到族$\mathcal{F}=\{A_\alpha:\alpha\in\mathcal{I}\}$以及$\mathcal{G}=\{\Lambda_\alpha:\alpha\in\mathcal{I}\}$上. 而"充分性"这部分的结论可以如同(1.3.21)中那样得出. □

关于相似性的最后一个定理表明，复矩阵的特征值与其主对角元素之间**仅有的**联系是：它们各自的和相等.

定理 1.3.31(Mirsky) 设给定一个整数 $n\geqslant 2$ 与复的纯量 λ_1，…，λ_n 以及 d_1，…，d_n. 则存在一个 $A\in M_n$，其特征值为 λ_1，…，λ_n，而主对角元素为 d_1，…，d_n 的充分必要条件是 $\sum_{i=1}^{n}\lambda_i=\sum_{i=1}^{n}d_i$. 如果 λ_1，…，λ_n 与 d_1，…，d_n 全是实数，且有同样的和，那么就存在一个 $A\in M_n(\mathbf{R})$，使得 λ_1，…，λ_n 为其特征值，而 d_1，…，d_n 为其主对角元素.

证明 我们知道，对任何 $A\in M_n$ 有 $\operatorname{tr}A=E_1(A)=S_1(A)$(1.2.16)，这就确立了所述条件的必要性. 我们尚需证明它是充分的.

如果 $k\geqslant 2$，且 λ_1，…，λ_k 以及 d_1，…，d_k 是任意给定的复的纯量，它们使得 $\sum_{i=1}^{k}\lambda_i=$

$\sum_{i=1}^{k} d_i$，我们断言：上双对角矩阵

$$T(\lambda_1,\cdots,\lambda_k)=\begin{bmatrix}\lambda_1 & 1 & & \\ & \lambda_2 & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_k\end{bmatrix}\in M_k$$

与一个以 d_1，…，d_k 为对角元素的矩阵相似，该矩阵有所断言之性质. 设 $L(s,t)=\begin{bmatrix}1 & 0\\ s-t & 1\end{bmatrix}$，所以 $L(s,t)^{-1}=\begin{bmatrix}1 & 0\\ t-s & 1\end{bmatrix}$.

首先考虑 $k=2$ 的情形，此时 $\lambda_1+\lambda_2=d_1+d_2$. 计算相似性给出

$$\begin{aligned}L(\lambda_1,d_1)T(\lambda_1,\lambda_2)L(\lambda_1,d_1)^{-1}&=\begin{bmatrix}1 & 0\\ \lambda_1-d_1 & 1\end{bmatrix}\begin{bmatrix}\lambda_1 & 1\\ 0 & \lambda_2\end{bmatrix}\begin{bmatrix}1 & 0\\ d_1-\lambda_1 & 1\end{bmatrix}\\ &=\begin{bmatrix}d_1 & \bigstar\\ \bigstar & \lambda_1+\lambda_2-d_1\end{bmatrix}=\begin{bmatrix}d_1 & \bigstar\\ \bigstar & d_2\end{bmatrix}\end{aligned}$$

这里我们利用了假设 $\lambda_1+\lambda_2-d_1=d_1+d_2-d_1=d_2$. 这就验证了 $k=2$ 时的结论.

我们来用归纳法. 假设结论已经对某个 $k\geqslant 2$ 得到了证明，且有 $\sum_{i=1}^{k+1}\lambda_i=\sum_{i=1}^{k+1}d_i$. 分划 $T(\lambda_1,\cdots,\lambda_{k+1})=[T_{ij}]_{i,j=1}^{2}$，其中 $T_{11}=T(\lambda_1,\lambda_2)$，$T_{12}=E_2$，$T_{21}=0$，而 $T_{22}=T(\lambda_3,\cdots,\lambda_{k+1})$（其中 $E_2=[e_2\quad 0\quad\cdots\quad 0]\in M_{2,k-1}$，而 $e_2=[0\quad 1]^{\mathrm{T}}\in\mathbf{C}^2$）. 令 $\mathcal{L}=L(\lambda_1,d_1)\oplus I_{k-1}$ 并计算 $\mathcal{L}\,T(\lambda_1,\cdots,\lambda_{k+1})\mathcal{L}^{-1}$

$$=\begin{bmatrix}L(\lambda_1,d_1) & 0\\ 0 & I_{k-1}\end{bmatrix}\begin{bmatrix}T(\lambda_1,\lambda_2) & E_2\\ 0 & T(\lambda_3,\cdots,\lambda_{k+1})\end{bmatrix}\begin{bmatrix}L(d_1,\lambda_1) & 0\\ 0 & I_{k-1}\end{bmatrix}$$

$$=\begin{bmatrix}\begin{bmatrix}d_1 & \bigstar\\ \bigstar & \lambda_1+\lambda_2-d_1\end{bmatrix} & E_2\\ 0 & T(\lambda_3,\cdots,\lambda_{k+1})\end{bmatrix}=\begin{bmatrix}d_1 & \bigstar\\ \bigstar & T(\lambda_1+\lambda_2-d_1,\lambda_3,\cdots,\lambda_{k+1})\end{bmatrix}=\begin{bmatrix}d_1 & \bigstar\\ \bigstar & D\end{bmatrix}$$

$D=T(\lambda_1+\lambda_2-d_1,\lambda_3,\cdots,\lambda_{k+1})\in M_k$ 的特征值之和是 $\sum_{i=1}^{k+1}\lambda_i-d_1=\sum_{i=1}^{k+1}d_i-d_1=\sum_{i=2}^{k+1}d_i$，所以归纳假设就确保存在一个非奇异的 $S\in M_k$，使得 SDS^{-1} 的对角元素是 d_2，…，d_{k+1}. 这样一来，$\begin{bmatrix}1 & 0\\ 0 & S\end{bmatrix}\begin{bmatrix}d_1 & \bigstar\\ \bigstar & D\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & S\end{bmatrix}^{-1}=\begin{bmatrix}d_1 & \bigstar\\ \bigstar & SDS^{-1}\end{bmatrix}$ 就有对角元素 d_1，d_2，…，d_{k+1}.

如果 λ_1，…，λ_n 与 d_1，…，d_n 全是实数，则上面的推理中所有的矩阵以及相似变换矩阵也全都是实矩阵. □

习题 写出上一证明中从 $k=2$ 到 $k=3$ 的归纳证明细节. ◀

问题

1.3.P1 令 A，$B\in M_n$. 假设 A 与 B 是可对角化且可交换的. 设 λ_1，…，λ_n 是 A 的特征值，而 μ_1，…，μ_n 是 B 的特征值. (a)证明 $A+B$ 的特征值是 $\lambda_1+\mu_{i_1}$，$\lambda_2+\mu_{i_2}$，…，$\lambda_n+\mu_{i_n}$，这里 i_1，…，i_n 是 1，…，n 的某个排列. (b)如果 B 是幂零的，说明为什么 A 与 $A+B$ 有相同的特征值. (c) AB 的特征值是什么？

1.3.P2 如果 A，$B\in M_n$，且 A 与 B 可交换，证明关于 A 的任何多项式与关于 B 的任何多项式都是可交换的．

1.3.P3 如果 $A\in M_n$，$SAS^{-1}=\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，且 $p(t)$是一个多项式，证明 $p(A)=S^{-1}p(\Lambda)S$ 以及 $p(\Lambda)=\mathrm{diag}(p(\lambda_1),\cdots,p(\lambda_n))$．如果可以将 A 对角化，这就对计算 $p(A)$提供了一个简单的方法．

1.3.P4 如果 $A\in M_n$ 有不同的特征值 $\alpha_1,\cdots,\alpha_n$ 且与一个给定的矩阵 $B\in M_n$ 可交换，证明 B 是可以对角化的，且存在一个次数至多为 $n-1$ 的多项式 $p(t)$，使得 $B=p(A)$．

1.3.P5 给出两个不可同时对角化的交换矩阵的例子．这与(1.3.12)矛盾吗？为什么？

1.3.P6 (a) 如果 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，证明 $p_\Lambda(\Lambda)$是零矩阵．(b)假设 $A\in M_n$ 可对角化．说明为什么 $p_A(t)=p_\Lambda(t)$以及 $p_A(A)=Sp_\Lambda(\Lambda)S^{-1}$．导出 $p_A(A)$是零矩阵这一结论．

1.3.P7 矩阵 $A\in M_n$ 是 $B\in M_n$ 的**平方根**(square root)，如果 $A^2=B$．证明：每一个可对角化的矩阵 $B\in M_n$ 都有平方根．$B=\begin{bmatrix}0&1\\0&0\end{bmatrix}$有平方根吗？为什么？

1.3.P8 如果 A，$B\in M_n$ 且其中至少有一个矩阵有不同的特征值(对另外一个矩阵，甚至没有假设它可以对角化)，对下述的几何推理提供细节，A 与 B 可交换，当且仅当它们可同时对角化：一个方向的证明容易；对另一方向的证明，假设 B 有不同的特征值且 $Bx=\lambda x(x\neq 0)$．那么 $B(Ax)=A(Bx)=A\lambda x=\lambda Ax$，所以对某个 $\mu\in\mathbf{C}$ 有 $Ax=\mu x$(为什么？见(1.2.18))．于是，我们可以用使得 B 对角化的特征向量组成的同一个矩阵来将 A 对角化．自然，A 的特征值不一定都是不相同的．

1.3.P9 考虑奇异矩阵 $A=\begin{bmatrix}1&0\\0&0\end{bmatrix}$以及 $B=\begin{bmatrix}0&0\\1&0\end{bmatrix}$．证明：$AB$ 与 BA 不相似，但是它们的确有同样的特征值．

1.3.P10 设给定 $A\in M_n$，又设 $\lambda_1,\cdots,\lambda_k$ 是 A 的不同的特征值．对每个 $i=1,2,\cdots,k$，假设 $x_1^{(i)}$，$x_2^{(i)},\cdots,x_{n_i}^{(i)}$是 A 的与特征值λ_i 相伴的一列线性无关的特征向量．证明：所有这样的向量组成的向量组 $x_1^{(1)},x_2^{(1)},\cdots,x_{n_1}^{(1)},\cdots,x_1^{(k)},x_2^{(k)},\cdots,x_{n_k}^{(k)}$ 是线性无关的．

1.3.P11 对如下的关于(1.3.19)的另一种证明给出证明细节：(a)假设 A，$B\in M_n$ 可交换，$x\neq 0$，且 $Ax=\lambda x$．考虑一列向量 $x,Bx,B^2x,B^3x,\cdots$．假设 k 是使得 B^kx 是它前面元素的线性组合这一结论成立的最小正整数；$S=\mathrm{span}\{x,Bx,B^2x,\cdots,B^{k-1}x\}$是 B-不变的，从而它包含 B 的一个特征向量．但是 $AB^jx=B^jAx=B^j\lambda x=\lambda B^jx$，所以$\mathcal{S}$中每个非零向量都是 A 的特征向量．由此推出结论：A 与 B 有一个共同的特征向量．(b)如果$\mathcal{F}=\{A_1,A_2,\cdots,A_m\}\subset M_n$ 是一个有限的交换族，利用归纳法证明它有一个共同的特征向量．如果 $y\neq 0$ 是 $A_1,A_2,\cdots,A_{m-1}$的共同的特征向量，考虑 $y,A_my,A_m^2y,A_m^3y,\cdots$．(c)如果$\mathcal{F}\subset M_n$ 是一个无限的交换族，那么在$\mathcal{F}$中不存在多于 n^2 个矩阵，它们可以是线性无关的．选取一个极大线性无关组，并说明为什么这一组有限多个矩阵集合的共同特征向量也是$\mathcal{F}$的共同特征向量．

1.3.P12 设 A，$B\in M_n$，并假设 A 或者 B 是非奇异的．如果 AB 可对角化，证明 BA 也可对角化．考虑 $A=\begin{bmatrix}0&1\\0&0\end{bmatrix}$以及 $B=\begin{bmatrix}1&1\\0&0\end{bmatrix}$，以此来证明：如果 A 与 B 两者都是奇异的，则此结论不一定为真．

1.3.P13 证明：两个可对角化的矩阵是相似的，当且仅当它们的特征多项式相同．对于两个并非都能对角化的矩阵，此结论是否成立？

1.3.P14 假设 $A\in M_n$ 可以对角化．(a)证明 A 的秩等于它的非零特征值的个数．(b)证明 $\mathrm{rank}A=\mathrm{rank}A^k$(对所有 $k=1,2,\cdots$)．(c)证明：A 是幂零的，当且仅当 $A=0$．(d)如果 $\mathrm{tr}A=0$，证明 $\mathrm{rank}A\neq 1$．

(e)利用以上四个结果中的每一个来证明 $B=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$ 是不可对角化的.

1.3.P15 设给定 $A\in M_n$ 以及一个多项式 $p(t)$. 如果 A 可以对角化，证明 $p(A)$ 可以对角化. 其逆命题呢？

1.3.P16 设 $A\in M_n$，并假设 $n>\operatorname{rank}A=r\geqslant 1$. 如果 A 与 $B\oplus 0_{n-r}$ 相似(故而 $B\in M_r$ 是非奇异的)，证明 A 有一个非奇异的 $r\times r$ 主子矩阵(这就是说，A 是主秩的(0.7.6)). 如果 A 是主秩的，它必定与 $B\oplus 0_{n-r}$ 相似吗？

1.3.P17 设给定 A，$B\in M_n$. 证明存在一个非奇异的矩阵 $T\in M_n(\mathbf{R})$ 使得 $A=TBT^{-1}$ 成立的充分必要条件是：存在一个非奇异的 $S\in M_n$，使得 $A=SBS^{-1}$ 与 $\overline{A}=S\overline{B}S^{-1}$ 两者都成立.

1.3.P18 假设 A，$B\in M_n$ 是共轭对合的，也即 $A\overline{A}=B\overline{B}=I$. 证明 A 和 B 在 $\mathbf{C}$ 上是相似的，当且仅当它们在 $\mathbf{R}$ 上是相似的. 70

1.3.P19 设 B，$C\in M_n$，并定义 $\mathcal{A}=\begin{bmatrix}B & C\\ C & B\end{bmatrix}\in M_{2n}$. 设 $Q=\frac{1}{\sqrt{2}}\begin{bmatrix}I_n & I_n\\ I_n & -I_n\end{bmatrix}$，验证 $Q^{-1}=Q=Q^{\mathrm{T}}$. 令 $\mathcal{K}_{2n}=\begin{bmatrix}0_n & I_n\\ I_n & 0_n\end{bmatrix}$. (a) M_{2n} 中具有 $\mathcal{A}$ 的块状结构的矩阵称为 2×2 **分块中心对称的**(block centrosymmetric). 证明：$A\in M_{2n}$ 是 2×2 分块中心对称的，当且仅当 $\mathcal{K}_{2n}A=A\mathcal{K}_{2n}$. 由此恒等式导出结论：非奇异的 2×2 分块中心对称矩阵的逆也是 2×2 分块中心对称的，而且 2×2 分块中心对称矩阵的乘积也是 2×2 分块中心对称的. (b)证明 $Q^{-1}\mathcal{A}Q=(B+C)\oplus(B-C)$. (c)说明为什么有 $\det\mathcal{A}=\det(B^2+CB-BC-C^2)$ 以及 $\operatorname{rank}\mathcal{A}=\operatorname{rank}(B+C)+\operatorname{rank}(B-C)$. (d)说明为什么 $\begin{bmatrix}0 & C\\ C & 0\end{bmatrix}$ 与 $C\oplus(-C)$ 相似，又为什么它的特征值按照“±”成对出现. 如果 C 是实的，那么它的特征值有什么性质？一个更加确切的命题，请见(4.6.P20).

1.3.P20 将任何 A，$B\in M_n$ 表示成 $A=A_1+\mathrm{i}A_2$ 以及 $B=B_1+\mathrm{i}B_2$，其中 A_1，A_2，B_1，$B_2\in M_n(\mathbf{R})$. 定义 $R_1(A)=\begin{bmatrix}A_1 & A_2\\ -A_2 & A_1\end{bmatrix}\in M_{2n}(\mathbf{R})$. 证明如下结论.

(a) $R_1(A+B)=R_1(A)+R_1(B)$，$R_1(AB)=R_1(A)R_1(B)$，以及 $R(I_n)=I_{2n}$.

(b) 如果 A 是非奇异的，那么 $R_1(A)$ 是非奇异的，$R_1(A)^{-1}=R_1(A^{-1})$，且 $R_1(A)^{-1}=\begin{bmatrix}X & Y\\ -Y & X\end{bmatrix}$ 与 $R_1(A)$ 有同样的分块结构.

(c) 如果 S 是非奇异的，那么 $R_1(SAS^{-1})=R_1(S)R_1(A)R_1(S)^{-1}$.

(d) 如果 A 与 B 相似，那么 $R_1(A)$ 与 $R_1(B)$ 也相似.

设 A 的特征值为 λ_1，…，λ_n，令 $S=\begin{bmatrix}I_n & \mathrm{i}I_n\\ 0 & I_n\end{bmatrix}$ 以及 $U=\frac{1}{\sqrt{2}}\begin{bmatrix}I_n & \mathrm{i}I_n\\ \mathrm{i}I_n & I_n\end{bmatrix}$. 证明如下之结论.

(e) $S^{-1}=\overline{S}$ 以及 $U^{-1}=\overline{U}=U^*$.

(f) $S^{-1}R_1(A)S=\begin{bmatrix}A & 0\\ -A_2 & \overline{A}\end{bmatrix}$ 以及 $U^{-1}R_1(A)U=\begin{bmatrix}A & 0\\ 0 & \overline{A}\end{bmatrix}$.

(g) $R_1(A)$ 的特征值与 $A\oplus\overline{A}$ 的特征值相同，它们是 λ_1，…，λ_n，$\overline{\lambda_1}$，…，$\overline{\lambda_n}$(更确切的结论见(1.3.P30)).

(h) $\det R_1(A)=|\det A|^2\geqslant 0$ 且 $\operatorname{rank}R_1(A)=2\operatorname{rank}A$.

(i) 如果 $R_1(A)$ 非奇异，那么 A 也非奇异.

(j) $\mathrm{i}I_n$ 不与 $-\mathrm{i}I_n$ 相似，但是 $R_1(\mathrm{i}I_n)$ 与 $R_1(-\mathrm{i}I_n)$ 相似，故而(d)中的结论不是可逆的.

(k) $p_{R_1(A)}(t)=p_A(t)p_{\bar{A}}(t)$.

(l) $R_1(A^*)=R_1(A)^{\mathrm{T}}$，所以 A 是 Hermite 矩阵，当且仅当 $R_1(A)$是(实的)对称矩阵，又 A 是酉矩阵，当且仅当 $R_1(A)$是实正交矩阵.

(m) A 与 A^* 可交换，当且仅当 $R_1(A)$与 $R_1(A)^{\mathrm{T}}$ 可交换，也就是说，复矩阵 A 是正规的，当且仅当实矩阵 $R_1(A)$是正规的，见(2.5).

(n) $M_{2n}(\mathbf{R})$中具有 $R_1(A)$的分块结构的矩阵称为一个**复型矩阵**(matrix of complex type). 设 $S_{2n}=\begin{bmatrix}0_n & I_n\\ -I_n & 0_n\end{bmatrix}$. 证明：$A\in M_{2n}(\mathbf{R})$是一个复型矩阵，当且仅当 $S_{2n}A=AS_{2n}$. 由此恒等式推出：一个实的复型矩阵的逆是复型矩阵，且实的复型矩阵之积是复型矩阵.

分块矩阵 $R_1(A)$是 A 的**实表示**(real representation)的一个例子；它推广到**复表示**(complex representation)见(4.4.P29)，也称为**四元数型矩阵**(matrix of quaternion type).

1.3.P21 利用与上一问题中同样的记号，定义 $R_2(A)=\begin{bmatrix}A_1 & A_2\\ A_2 & -A_1\end{bmatrix}\in M_{2n}(\mathbf{R})$. 设 $V=\frac{1}{\sqrt{2}}\begin{bmatrix}-\mathrm{i}I_n & -\mathrm{i}I_n\\ I_n & -I_n\end{bmatrix}$，并考虑 $R_2(\mathrm{i}I_n)=\begin{bmatrix}0 & I_n\\ I_n & 0\end{bmatrix}$以及 $R_2(I_n)=\begin{bmatrix}I_n & 0\\ 0 & -I_n\end{bmatrix}$. 证明如下结论.

(a) $V^{-1}=V^*$，$R_2(I_n)^{-1}=R_2(I_n)=R_2(I_n)^*$，$R_2(\mathrm{i}I_n)^{-1}=R_2(\mathrm{i}I_n)=R_2(\mathrm{i}I_n)^*$，以及 $R_2(\mathrm{i}I_n)=V^{-1}R_2(I_n)V$.

(b) $A=B$ 当且仅当 $R_2(A)=R_2(B)$以及 $R_2(A+B)=R_2(A)+R_2(B)$.

(c) $R_2(A)=V\begin{bmatrix}0 & \bar{A}\\ A & 0\end{bmatrix}V^{-1}$.

(d) $\det R_2(A)=(-1)^n|\det A|^2$，见(0.8.5.13).

(e) $R_2(A)$是非奇异的，当且仅当 A 是非奇异的.

(f) 特征多项式与特征值：$p_{R_2(A)}(t)=\det(t^2I-A\bar{A})=p_{A\bar{A}}(t^2)$(0.8.5.13)，所以，如果$\mu_1$，…，$\mu_n$ 是 $A\bar{A}$的特征值，那么$\pm\mu_1$，…，$\pm\mu_n$ 就是$R_2(A)$的特征值. 此外，$p_{R_2(A)}(t)$有实系数，所以 $A\bar{A}$的非实的特征值成对共轭出现.

(g) $R_2(AB)=R_2(A\cdot I_n\cdot B)=R_2(A)R_2(I_n)R_2(B)$.

(h) $R_2(\bar{A})=R_2(I_n)R_2(A)R_2(I_n)$，所以 $R_2(\bar{A})$与 $R_2(A)$相似，且 $R_2(A\bar{B}C)=R_2(A)R_2(B)R_2(C)$.

(i) $-R_2(A)=R_2(-A)=R_2(\mathrm{i}I_n\cdot A\cdot\mathrm{i}I_n)=(R_2(\mathrm{i}I_n)R_2(I_n))\cdot R_2(A)\cdot(R_2(\mathrm{i}I_n)R_2(I_n))^{-1}$，所以 $R_2(-A)$与 $R_2(A)$相似.

(j) $R_2(A)R_2(B)=V(\bar{A}B\oplus A\bar{B})V^{-1}$.

(k) 如果 A 非奇异，那么 $R_2(A)^{-1}=R_2(\bar{A}^{-1})$.

(l) $R_2(A)^2=R_1(\bar{A}A)=R_2(A\bar{A})R_2(I_n)$.

(m) 如果 S 非奇异，那么 $R_2(SA\bar{S}^{-1})=(R_2(S)R_2(I_n))\cdot R_2(A)\cdot(R_2(S)R_2(I_n))^{-1}$，所以 $R_2(SA\bar{S}^{-1})$与 $R_2(A)$相似. 其逆命题见(4.6.P19)：如果 $R_2(A)$与 $R_2(B)$相似，那么就存在一个非奇异的 S，使得 $B=SA\bar{S}^{-1}$.

(n) $R_2(A^{\mathrm{T}})=R_2(A)^{\mathrm{T}}$，所以 A 是(复)对称的，当且仅当 $R_2(A)$是(实)对称的.

(o) A 是酉矩阵，当且仅当 $R_2(A)$是实正交矩阵.

分块矩阵 $R_2(A)$是 A 的**实表示**的第二个例子.

1.3.P22 设 A，$B\in M_n$. 证明：A 与 B 相似，当且仅当存在 X，$Y\in M_n$，它们中至少有一个是非奇异的，且使得 $A=XY$ 以及 $B=YX$.

1.3.P23　设 $B\in M_n$ 且 $C\in M_{n,m}$，又定义 $\mathcal{A}=\begin{bmatrix} B & C \\ 0 & 0_m \end{bmatrix}\in M_{n+m}$. 证明：$\mathcal{A}$ 与 $B\oplus 0_m$ 相似，当且仅当 $\operatorname{rank}[B\quad C]=\operatorname{rank}B$，也就是说，当且仅当存在某个 $X\in M_{n,m}$，使得 $C=BX$.

1.3.P24　对给定的整数 $n\geqslant 3$，令 $\theta=2\pi/n$ 以及 $A=[\cos(j\theta+k\theta)]_{j,k=1}^{n}\in M_n(\mathbf{R})$. 证明 $A=[x\quad y][x\quad y]^{\mathrm{T}}$，其中 $x=[\alpha\ \alpha^2\ \cdots\ \alpha^n]^{\mathrm{T}}$，$y=[\alpha^{-1}\ \alpha^{-2}\ \cdots\ \alpha^{-n}]^{\mathrm{T}}$，且 $\alpha=\mathrm{e}^{2\pi\mathrm{i}/n}$. 证明 A 的特征值是 $n/2$ 与 $-n/2$，外加 $n-2$ 个零.

1.3.P25　设给定 $x, y\in\mathbf{C}^n$，并假设 $y^*x\neq -1$. (a)验证 $(I+xy^*)^{-1}=I-cxy^*$，其中 $c=(1+y^*x)^{-1}$. (b)设 $\Lambda=\operatorname{diag}(\lambda_1,\cdots,\lambda_n)$并假设 $y^*x=0$. 说明为什么

$$A=(I+xy^*)\Lambda(I-xy^*)=\Lambda+xy^*\Lambda-\Lambda xy^*-(y^*\Lambda x)xy^*$$

的特征值是 $\lambda_1,\cdots,\lambda_n$. 注意：$A$ 的元素是整数，如果 x，y 以及 Λ 的元素都是整数. 利用这一结论来构造一个元素为整数且特征值为 1，2 以及 7 的有趣的 3×3 矩阵，验证你构造的矩阵有所述的特征值.

1.3.P26　设 $e_1,\cdots,e_n$ 与 $\varepsilon_1,\cdots,\varepsilon_m$ 分别表示 $\mathbf{C}^n$ 与 $\mathbf{C}^m$ 的规范的标准正交基. 考虑 $n\times m$ 分块矩阵 $P=[P_{ij}]\in M_{mn}$，其中的每一个块 $P_{ij}\in M_{m,n}$ 由 $P_{ij}=\varepsilon_j e_i^{\mathrm{T}}$ 给出. (a)证明 P 是置换矩阵. (b)任何矩阵 $A\in M_{mn}$ 通过 P 给出的相似性都给出一个矩阵 $\tilde{A}=PAP^{\mathrm{T}}$，其元素是 A 的元素的一个重新排列. 将 A 与 $\tilde{A}$ 作适当的分划，使得我们可以用简单的方式来描述这个重新排列. 将 $A=[A_{ij}]\in$ 72 M_{mn} 写成一个 $m\times m$ 分块矩阵，其中每一个块 $A_{kl}=[a_{ij}^{(k,l)}]\in M_n$，又将 $\tilde{A}=[\tilde{A}_{ij}]$写成一个 $n\times n$ 分块矩阵，其中每一个块 $\tilde{A}_{ij}\in M_m$. 说明为什么 $\tilde{A}_{pq}$ 的位于(i,j)处的元素是 A_{ij} 的位于(p,q)处的元素(对所有 $i,j=1,\cdots,m$ 以及 $p,q=1,\cdots,n$)，即 $\tilde{A}_{pq}=[a_{pq}^{(i,j)}]$. 由于 A 与 $\tilde{A}$ 是置换相似的，故而它们有相同的特征值和有相同的行列式等. (c) A 的元素中各种特殊的模式在 $\tilde{A}$ 的元素中也产生特殊的模式(反之亦然). 例如，说明为什么(i)所有的分块 A_{ij} 都是上三角的，当且仅当 $\tilde{A}$ 是分块上三角的；(ii)所有的分块 A_{ij} 都是上 Hessenberg 的，当且仅当 $\tilde{A}$ 是分块上 Hessenberg 的；(iii)所有的块 A_{ij} 都是对角的，当且仅当 $\tilde{A}$ 是分块对角的；(iv) A 是分块上三角的且所有的块 A_{ij} 都是上三角的，当且仅当 $\tilde{A}$ 是分块对角的，且它所有的主对角块都是上三角的.

1.3.P27　(1.3.P26 续)设 $A=[A_{kl}]\in M_{mn}$ 是一个给定的 $m\times m$ 分块矩阵，其中每一个 $A_{kl}=[a_{ij}^{(k,l)}]\in M_n$. 又假设每一个块 A_{kl} 都是上三角的. 说明为什么 A 的特征值与 $\tilde{A}_{11}\oplus\cdots\oplus\tilde{A}_{nn}$ 的特征值相同，其中 $\tilde{A}_{pp}=[a_{pp}^{(i,j)}]$(对 $p=1,\cdots,n$). 于是，A 的特征值只与块 A_{ij} 的主对角元素有关. 特别地，$\det A=(\det\tilde{A}_{11})\cdots(\det\tilde{A}_{nn})$. 如果每一个块 A_{ij} 的对角元素都是常数(故而存在纯量 α_{kl} 使得对所有 $i=1,\cdots,n$ 以及所有 $k,l=1,\cdots,m$ 都有 $a_{ii}^{(k,l)}=\alpha_{kl}$)，那么你能对 A 的特征值以及行列式说些什么？

1.3.P28　设给定 $A\in M_{m,n}$ 以及 $B\in M_{n,m}$. 证明 $\det(I_m+AB)=\det(I_n+BA)$.

1.3.P29　设 $A=[a_{ij}]\in M_n$. 假设每一个 $a_{ii}=0$(对 $i=1,\cdots,n$)以及 $a_{ij}\in\{-1,1\}$(对所有 $i\neq j$). 说明为什么 $\det A$ 是整数. 利用 Cauchy 恒等式(1.3.24)证明：如果 A 有任何一个元素从 -1 变成 $+1$，则 $\det A$ 的奇偶性不发生改变，也就是说，如果它原来是偶数，它仍保持是偶数，如果它原来是奇数，则仍保持为奇数. 证明：$\det A$ 的奇偶性与 $\det(J_n-I)$的奇偶性相同，而后者的奇偶性与 n 的奇偶性相反. 导出结论：如果 n 是偶数，那么 A 是非奇异的.

1.3.P30　假设 $A\in M_n$ 是可以对角化的，且 $A=S\Lambda S^{-1}$，其中 Λ 有(1.3.13)的形式. 如果 f 是一个复值函数，其定义域包含 $\sigma(A)$，我们就定义 $f(A)=Sf(\Lambda)S^{-1}$，其中 $f(\Lambda)=f(\mu_1)I_{n_1}\oplus\cdots\oplus f(\mu_d)I_{n_d}$. $f(A)$与对角化所用的相似矩阵的选取有关吗(这个矩阵从来都不是唯一的)？利用定理 1.3.27 证明它与相似矩阵的选取无关；也即，如果 $A=S\Lambda S^{-1}=T\Lambda T^{-1}$，证明 $Sf(\Lambda)S^{-1}=Tf(\Lambda)T^{-1}$. 如果 A 有实的特征值，证明 $\cos^2 A+\sin^2 A=I$.

1.3.P31 设 a，$b\in\mathbf{C}$. 证明 $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ 的特征值是 $a\pm \mathrm{i}b$.

1.3.P32 设 $x\in\mathbf{C}^n$ 是一个给定的非零向量，记 $x=u+\mathrm{i}v$，其中 u，$v\in\mathbf{R}^n$. 证明：向量 x，$\bar{x}\in\mathbf{C}^n$ 是线性无关的，当且仅当向量 u，$v\in\mathbf{R}^n$ 是线性无关的.

1.3.P33 假设 $A\in M_n(\mathbf{R})$ 有一个非实的特征值 λ，记 $\lambda=a+\mathrm{i}b$，其中 a，$b\in\mathbf{R}$ 且 $b>0$. 令 x 是 A 的一个与 λ 相伴的特征向量，记 $x=u+\mathrm{i}v$，其中 u，$v\in\mathbf{R}^n$. (a)说明为什么 $\bar{\lambda}$，$\bar{x}$ 是 A 的一个特征对. (b)说明为什么 x 与 $\bar{x}$ 是线性无关的，并导出 u 与 v 是线性无关的. (c)证明 $Au=au-bv$ 以及 $Av=bu+av$，所以 $A[u \quad v]=[u \quad v]B$，其中 $B=\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$. (d)设 $S=[u \quad v \quad S_1]\in M_n(\mathbf{R})$ 是非奇异的. 说明为什么 $S^{-1}[u \quad v]=\begin{bmatrix} I_2 \\ 0 \end{bmatrix}$，并检验 $S^{-1}AS=S^{-1}[A[u \quad v] \quad AS_1]=S^{-1}[[u \quad v]B \quad AS_1]=\begin{bmatrix} B & \bigstar \\ 0 & A_1 \end{bmatrix}$，其中 $A_1\in M_{n-2}$. 于是，有非实的特征值 λ 的实方阵与一个 2×2 分块上三角矩阵(其左上块揭示出 λ 的实部和虚部)实相似. (e)说明为什么 λ 与 $\bar{\lambda}$ 中每一个作为 A_1 的特征值的重数都比它们作为 A 的特征值的重数小 1.

1.3.P34 如果 A，$B\in M_n$ 是相似的，证明 $\mathrm{adj}A$ 与 $\mathrm{adj}B$ 是相似的.

1.3.P35 集合 $\mathcal{A}\subseteq M_n$ 是一个**代数**(algebra)，如果(ⅰ)$\mathcal{A}$ 是一个子空间，且(ⅱ)只要 A，$B\in\mathcal{A}$，就有 $AB\in\mathcal{A}$. 对下面的结论提供细节并组织成**关于矩阵代数的 Burnside 定理**(Burnside's theorem on matrix algebras)：设 $n\geqslant 2$，且 $\mathcal{A}\subseteq M_n$ 是一个给定的代数. 那么 $\mathcal{A}=M_n$ 当且仅当 $\mathcal{A}$ 是不可约的.

(a) 如果 $n\geqslant 2$，且代数 $\mathcal{A}\subseteq M_n$ 是可约的，那么 $\mathcal{A}\neq M_n$. 这就是 Burnside 定理的直接含义，需要做点工作来证明：如果 $\mathcal{A}$ 是不可约的，那么 $\mathcal{A}=M_n$. 在下面，$\mathcal{A}\subseteq M_n$ 是一个给定的代数，且 $\mathcal{A}^*=\{A^*: A\in\mathcal{A}\}$.

(b) 如果 $n\geqslant 2$，且 $\mathcal{A}$ 是不可约的，那么 $\mathcal{A}\neq\{0\}$.

(c) 如果 $x\in\mathbf{C}^n$ 不是零，那么 $\mathcal{A}x=\{Ax: A\in\mathcal{A}\}$ 是 $\mathbf{C}^n$ 的一个 $\mathcal{A}$-不变子空间.

(d) 如果 $n\geqslant 2$，$x\in\mathbf{C}^n$ 不是零，且 $\mathcal{A}$ 是不可约的，那么 $\mathcal{A}x=\mathbf{C}^n$.

(e) 对任何给定的 $x\in\mathbf{C}^n$，$\mathcal{A}^*x=\{A^*x: A\in\mathcal{A}\}$ 是 $\mathbf{C}^n$ 的一个子空间.

(f) 如果 $n\geqslant 2$，$x\in\mathbf{C}^n$ 不是零，且 $\mathcal{A}$ 是不可约的，那么 $\mathcal{A}^*x=\mathbf{C}^n$.

(g) 如果 $n\geqslant 2$，且 $\mathcal{A}$ 是不可约的，那么存在一个 $A\in\mathcal{A}$，使得 $\mathrm{rank}A=1$.

(h) 如果 $n\geqslant 2$，$\mathcal{A}$ 是不可约的，且存在非零的 y，$z\in\mathbf{C}^n$，使得 $yz^*\in\mathcal{A}$，那么 $\mathcal{A}$ 包含每个秩 1 矩阵.

(i) 如果 $\mathcal{A}$ 包含每个秩 1 矩阵，那么 $\mathcal{A}=M_n$，见(0.4.4i).

1.3.P36 设 A，$B\in M_n$，并假设 $n\geqslant 2$. **由 A 与 B 生成的代数**(记为 $\mathcal{A}(A, B)$)是关于 A 以及 B 的所有字(word)的集合所生成的子空间(2.2.5). (a)如果 A 与 B 没有共同的特征向量，说明为什么 $\mathcal{A}(A, B)=M_n$. (b)设 $A=\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ 以及 $B=A^{\mathrm{T}}$，证明 A 与 B 没有共同的特征向量，所以 $\mathcal{A}(A, B)=M_2$. 通过给出由关于 A 与 B 的字所组成的 M_2 的一组基来给出一个直接的证明.

1.3.P37 设 $A\in M_n$ 是中心对称的. 如果 $n=2m$ 且 A 表示成分块的形式(0.9.10.2)，证明：A 通过实正交矩阵 $Q=\frac{1}{\sqrt{2}}\begin{bmatrix} I_m & I_m \\ -K_m & K_m \end{bmatrix}$ 而与 $(B-K_mC)\oplus(B+K_mC)$ 相似. 如果 $n=2m+1$，且 A 表示成分块的形式(0.9.10.3)，证明 A 相似于

$$(B-K_mC)\oplus\begin{bmatrix} \alpha & \sqrt{2}x^{\mathrm{T}} \\ \sqrt{2}K_my & B+K_mC \end{bmatrix}, \text{通过} \quad Q=\frac{1}{\sqrt{2}}\begin{bmatrix} I_m & 0 & I_m \\ 0 & \sqrt{2} & 0 \\ -K_m & 0 & K_m \end{bmatrix}$$

且 Q 是实正交矩阵.

1.3.P38 设 J_n 是全 1 矩阵(0.2.8)，且 $B(t)=(1-t)I_n+tJ_n$，其中 $n\geqslant 2$. (a)描述 $B(t)$的元素. 说明为什么它的特征值是 $1+(n-1)t$ 以及 $1-t$，重数是 $n-1$. (b)验证：如果 $1\neq t\neq -(n-1)^{-1}$，那么 $B(t)$是非奇异的，且有 $B(t)^{-1}=(1-t)^{-1}(I_n-t(1+(n-1)t)^{-1}J_n)$.

1.3.P39 设给定 $A\in M_n$，并假设 $\mathrm{tr}A=0$. 如果 A 可对角化，说明为什么 $\mathrm{rank}A\neq 1$. (1.3.5)中有一个矩阵的秩为 1，而迹为零. 对于它你能得出何种结论？ 74

1.3.P40 A，$B\in M_n$ 的 Jordan 积是$]A$，$B[=AB+BA$. 矩阵 A 与 B 称为是**反交换的**(anticommute)，如果$]A$，$B[=0$，见(0.7.7). (a)给出一个包含无穷多个不同矩阵的矩阵交换族的例子. (b)设 $\mathcal{F}=\{A_1, A_2, \cdots\}$是一个矩阵族，它使得当 $i\neq j$ 时有$]A_i$，$A_j[=0$，但是对所有 $i=1, 2, \cdots$ 有 $A_i^2\neq 0$；这就是说，$\mathcal{F}$中没有矩阵与其自身是反交换的. 证明：$I\notin\mathcal{F}$，且$\mathcal{F}$中矩阵的任何有限集合都是线性无关的. 推出结论：$\mathcal{F}$至多包含 n^2-1 个矩阵. (c)如果$\mathcal{F}=\{A_1, A_2, \cdots\}$是由两两反交换、可对角化的不同的矩阵所组成的族，证明它是有限族，且$\{A_1^2, A_2^2, \cdots\}$是可对角化矩阵组成的一个(有限)交换族.

1.3.P41 如果 $A\in M_n$ 没有不同的特征值，那么不存在与 A 相似的矩阵能有不同的特征值，但是似乎有某个与 A 对角等价的矩阵有不同的特征值. (a)如果 D_1，$D_2\in M_n$ 是非奇异的对角矩阵，又如果 D_1AD_2 有不同的特征值，说明为什么存在一个非奇异的 $D\in M_n$，使得 DA 有不同的特征值. (b)如果 $A\in M_n$ 是严格三角的，且 $n\geqslant 2$，说明为什么不存在与 A 对角等价的矩阵有不同的特征值. (c)设 $n=2$，$A=\begin{bmatrix} a & b \\ c & d\end{bmatrix}$，并假设不存在与 A 对角等价的矩阵能有不同的特征值. 那么 $A_z=\begin{bmatrix} 1 & 0 \\ 0 & z\end{bmatrix}A=\begin{bmatrix} a & b \\ zc & zd\end{bmatrix}$对所有非零的 $z\in C$ 都有一个二重特征值. 证明：$p_{A_z}(t)$的判别式是 $(a+dz)^2-4(ad-bc)z=d^2z^2+(2ad-4(ad-bc))z+a^2$. 为什么这个判别式对所有非零的 $z\in\mathbf{C}$ 都等于零？说明为什么 $d=0$，$a=0$，以及 $bc=0$，并得出结论：A 是严格三角的. (d)在 $n=2$ 的情形，说明为什么 $A\in M_n$ 不与有不同特征值的矩阵对角等价，当且仅当 A 是奇异的，且 A 的每个 $n-1$ 阶主子式都为零. (e)(d)中的结论已知对所有 $n\geqslant 2$ 都是正确的.

注记以及进一步的阅读参考 定理 1.3.31 属于 L. Mirsky(1958)，我们的证明取材自 E. Carlen 与 E. Lieb. Mirsky 与 Horn，Short proofs of theorems of Mirsky and Horn on diagonals and eigenvalues of matrices，*Electron. J. Linear Algebra* 18(2009)438-441. 已知 Mirsky 定理的一个补充结果：如果给定 n^2 个复数 $\lambda_1, \cdots, \lambda_n$ 以及 $a_{ij}(i, j=1, \cdots, n, i\neq j)$，那么就存在 n 个复数 $a_{11}, \cdots, a_{nn}$，使得 $\lambda_1, \cdots, \lambda_n$ 是 $A=[a_{ij}]_{i,j=1}^n$ 的特征值，见 S. Frieland，Matrices with prescribed off-diagonal elements，*Israel J. Math.* 11(1975)184-189. (1.3. P35)中 Burnside 定理的证明取自 I. Halperin 与 P. Rosenthal，Burnside's theorem on algebra of matrices，*Amer. Math. Monthly* 87(1980)810. 有关其他可供选择的方法，见 Radjavi 与 Rosenthal(2000)以及 V. Lomonosov 与 P. Rosenthal，The simplest proof of Burnside's theorem on matrix algebras，*Linear Algebra Appl.* 383(2004)45-47. 作为(1.3. P41(e))中论断的证明，见 M. D. Choi，Z. Huang，C. K. Li 以及 N. S. Sze，Every invertible matrix is diagonally equivalent to a matrix with distinct eigenvalues，*Linear Algebra Appl.* 436(2012)3773-3776.

1.4 左右特征向量与几何重数

矩阵的特征向量之重要性不仅在于它在对角化中所起的作用，而且也在于它在各种应用中的成效. 我们先来给出有关特征值的一个重要结果. 75

结论 1.4.1 设 $A\in M_n$. (a)A 与 A^{T} 的特征值相同. (b)A^* 的特征值是 A 的特征值的复共轭.

证明 由于 $\det(tI-A^{\mathrm{T}})=\det(tI-A)^{\mathrm{T}}=\det(tI-A)$，我们有 $p_{A^{\mathrm{T}}}(t)=p_A(t)$，所以 $p_{A^{\mathrm{T}}}(\lambda)=0$ 当且仅当 $p_A(\lambda)=0$. 类似地，$\det(\bar{t}I-A^*)=\det[(tI-A)^*]=\overline{\det(tI-A)}$，所以 $p_{A^*}(\bar{t})=\overline{p_A(t)}$，又 $p_{A^*}(\bar{\lambda})=0$ 当且仅当 $p_A(\lambda)=0$. □

习题 如果 x，$y\in\mathbf{C}^n$ 两者都是 $A\in M_n$ 的与特征值 λ 相伴的特征向量，证明 x 与 y 的任何非零的线性组合也是它的与 λ 相伴的特征向量. 导出结论：与一个特别的 $\lambda\in\sigma(A)$ 相伴的所有特征向量组成的集合与零向量合起来作成 $\mathbf{C}^n$ 的一个**子空间**(subspace). ◀

习题 上一习题中所描述的子空间是 $A-\lambda I$ 的**零空间**(null space)，它就是齐次线性方程组 $(A-\lambda I)x=0$ 的解集. 说明为什么这个子空间的维数是 $n-\operatorname{rank}(A-\lambda I)$. ◀

定义 1.4.2 设 $A\in M_n$. 对给定的 $\lambda\in\sigma(A)$，满足 $Ax=\lambda x$ 的所有向量 $x\in\mathbf{C}^n$ 组成的集合称为 A 的与特征值 λ 相伴的**特征空间**(eigenspace). 这个特征空间中的每一个非零的元素都是 A 的与 λ 相伴的特征向量.

习题 证明：A 的与特征值 λ 相伴的特征空间是一个 A-不变子空间，但是一个 A-不变子空间不一定就是 A 的特征空间. 说明为什么**最小的** A-不变子空间(它是这样一个 A-不变子空间：它不包含有严格来说更低维度的非零的 A-不变子空间)W 是 A 的**单独一个**特征向量所生成的子空间，也就是说 $\dim W=1$. ◀

定义 1.4.3 设 $A\in M_n$ 且 λ 是 A 的一个特征值. A 的与特征值 λ 相伴的特征空间的维数是 λ 的**几何重数**(geometric multiplicity). λ 作为 A 的特征多项式的零点的重数是 λ 的**代数重数**(algebraic multiplicity). 如果用到重数这一术语而未加申明与 λ 之联系，那么就都是指其代数重数. 我们称 λ 是**单重的**(simple)，如果它的代数重数为 1；称它是**半单的**(semisimple)，如果它的代数重数与几何重数相等.

用多于一种方式来考虑 $A\in M_n$ 的特征值 λ 的几何重数常常是有用的：由于几何重数是 $A-\lambda I$ 的零空间的维数，它等于 $n-\operatorname{rank}(A-\lambda I)$. 它也是与 λ 相伴的线性无关特征向量的最大个数. 定理 1.2.18 与定理 1.3.7 两者都包含特征值的几何重数与代数重数之间的不等式，不过是从两种不同的观点着眼.

习题 利用(1.2.18)说明为什么特征值的代数重数大于或者等于它的几何重数. 如果代数重数为 1，为什么几何重数也等于 1? ◀

习题 利用(1.3.7)说明为什么特征值的几何重数小于或者等于它的代数重数. 如果代数重数为 1，为什么几何重数必定也为 1? ◀

76 **习题** 对如下各个矩阵验证下面的命题且它们的特征值 $\lambda=1$.

(a) $A_1=\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}$：几何重数=代数重数=1；单重的.

(b) $A_2=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$：几何重数=代数重数=2；半单的.

(c) $A_3=\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix}$：几何重数=1；代数重数=2. ◀

定义 1.4.4 设 $A\in M_n$. 我们说 A 是**有亏的**(defective)，如果 A 的某个特征值的几何重数严格小于它的代数重数. 如果 A 的每个特征值的几何重数都与它的代数重数相同，我们就称 A 是**无亏的**(nondefective). 如果 A 的每个特征值的几何重数都是 1，我们就称 A

是**无损的**(nonderogatory)；否则就称为**有损的**(derogatory).

一个矩阵可对角化，当且仅当它是无亏的；它有不同的特征值，当且仅当它是无损的且是无亏的.

习题 在上一习题中说明，为什么 A_1 是无亏的；A_2 是无亏的且是有损的；而 A_3 则是有亏的且是无损的. ◀

例 1.4.5 尽管 A 与 A^{T} 有相同的特征值，它们与给定特征值相伴的特征空间有可能是不同的. 例如，设 $A=\begin{bmatrix}2 & 3\\ 0 & 4\end{bmatrix}$. 那么 A 的与特征值 2 相伴的(一维)特征空间是由 $\begin{bmatrix}1\\ 0\end{bmatrix}$ 生成的，而 A^{T} 的与特征值 2 相伴的特征空间是由 $\begin{bmatrix}1\\ -3/2\end{bmatrix}$ 生成的.

定义 1.4.6 非零向量 $y\in\mathbf{C}^n$ 是 $A\in M_n$ 的一个与 A 的特征值 λ 相伴的**左特征向量**(left eigenvector)，如果 $y^*A=\lambda y^*$. 如果需要表述清晰，我们就把(1.1.3)中的向量 x 称为**右特征向量**(right eigenvector)；当上下文中不要求区分时，我们继续把 x 称为特征向量.

结论 1.4.6a 设 $x\in\mathbf{C}^n$ 是非零向量，而 $A\in M_n$，又假设 $Ax=\lambda x$. 如果 $x^*A=\mu x^*$，那么 $\lambda=\mu$.

证明 我们可以假设 x 是一个单位向量. 计算给出 $\mu=\mu x^*x=(x^*A)x=x^*Ax=x^*(Ax)=x^*(\lambda x)=\lambda x^*x=\lambda$. □

习题 证明：与 $A\in M_n$ 的特征值 λ 相伴的左特征向量 y 是 A^* 的与 $\bar{\lambda}$ 相伴的右特征向量；还要证明 $\bar{y}$ 是 A^{T} 的与 λ 相伴的右特征向量. ◀

习题 假设 $A\in M_n$ 可以对角化，S 非奇异，且 $S^{-1}AS=\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$. 按照列来分划 $S=[x_1\ \ \cdots\ \ x_n]$ 与 $S^{-*}=[y_1\ \ \cdots\ \ y_n]$(0.2.5). 恒等式 $AS=S\Lambda$ 告诉我们，S 的每一列 x_j 都是 A 的与特征值 λ_j 相伴的右特征向量. 说明：为什么 $(S^{-*})^*A=\Lambda(S^{-*})^*$；$S^{-*}$ 的每一列 y_j 都是 A 的与特征值 λ_j 相伴的左特征向量；对每个 $j=1,\cdots,n$ 有 $y_j^*x_j=1$，而只要 $i\neq j$ 就有 $y_i^*x_j=0$. ◀

请不要将左特征向量仅仅贬低为是在理论上与右特征向量平行的另一种概念. 每一种类型的特征向量都传递出有关矩阵的不同信息，而了解这两种类型的特征向量是怎样相互影响的，可能是非常有用的. 接下来我们来检查上一个习题中关于不一定能对角化的矩阵的那些结果的一种形式. 77

定理 1.4.7 设给定 $A\in M_n$，非零向量 $x,y\in\mathbf{C}^n$ 以及纯量 $\lambda,\mu\in\mathbf{C}$. 假设 $Ax=\lambda x$ 以及 $y^*A=\mu y^*$.

(a) 如果 $\lambda\neq\mu$，那么 $y^*x=0$.

(b) 如果 $\lambda=\mu$ 且 $y^*x\neq 0$，那么存在一个形如 $S=[x\ \ S_1]$ 的非奇异的 $S\in M_n$，使得 $S^{-*}=[y/(x^*y)\ Z_1]$ 以及

$$A=S\begin{bmatrix}\lambda & 0\\ 0 & B\end{bmatrix}S^{-1},\quad B\in M_{n-1}\tag{1.4.8}$$

反过来，如果 A 与一个形如(1.4.8)的分块矩阵相似，那么它就有一对与特征值 λ 相伴的非正交的左右特征向量.

证明 (a) 设 y 是 A 的与 μ 相伴的左特征向量，而 x 则是 A 的与 λ 相伴的右特征向

量．用两种方式处理 y^*Ax：

$$y^*Ax = y^*(\lambda x) = \lambda(y^*x) = (\mu y^*)x = \mu(y^*x)$$

由于 $\lambda\neq\mu$，故而仅当 $y^*x=0$ 时有 $\lambda y^*x=\mu y^*x$．

(b) 假设 $Ax=\lambda x$，$y^*A=\lambda y^*$，且 $y^*x\neq 0$．如果用 $y/(x^*y)$代替 y，我们可以假设 $y^*x=1$．令 $S_1\in M_{n,n-1}$ 的列是 y 的正交补的任意一组基(所以 $y^*S_1=0$)，并考虑 $S=[x\quad S_1]\in M_n$．设 $z=[z_1\quad \zeta^{\mathrm{T}}]^{\mathrm{T}}$(其中 $\zeta\in\mathbf{C}^{n-1}$)，并假设 $Sz=0$．那么

$$0 = y^*Sz = y^*(z_1x+S_1\zeta) = z_1(y^*x)+(y^*S_1)\zeta = z_1$$

所以 $z_1=0$ 且 $0=Sz=S_1\zeta$，这蕴含 $\zeta=0$，这是因为 S_1 是列满秩的．我们断言 S 是非奇异的．用 $\eta\in\mathbf{C}^n$ 分划 $S^{-*}=[\eta\quad Z_1]$，并计算

$$I_n = S^{-1}S = \begin{bmatrix}\eta^*\\ Z_1^*\end{bmatrix}[x\quad S_1] = \begin{bmatrix}\eta^*x & \eta^*S_1\\ Z_1^*x & Z_1^*S_1\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & I_{n-1}\end{bmatrix}$$

其中包含四个恒等式．恒等式 $\eta^*S_1=0$ 蕴含 η 与 y 的正交补正交，所以对某个纯量 α 有 $\eta=\alpha y$．恒等式 $\eta^*x=1$ 告诉我们 $\eta^*x=(\alpha y)^*x=\bar{\alpha}(y^*x)=\bar{\alpha}=1$，所以 $\eta=y$．利用恒等式 $\eta^*S_1=y^*S_1=0$ 与 $Z_1^*x=0$，以及 x 与 y 的特征向量的性质，计算相似矩阵

$$\begin{aligned}S^{-1}AS &= \begin{bmatrix}y^*\\ Z_1^*\end{bmatrix}A[x\quad S_1] = \begin{bmatrix}y^*Ax & y^*AS_1\\ Z_1^*Ax & Z_1^*AS_1\end{bmatrix}\\ &= \begin{bmatrix}(\lambda y^*)x & (\lambda y^*)S_1\\ Z_1^*(\lambda x) & Z_1^*AS_1\end{bmatrix} = \begin{bmatrix}\lambda(y^*x) & \lambda(y^*S_1)\\ \lambda(Z_1^*x) & Z_1^*AS_1\end{bmatrix} = \begin{bmatrix}\lambda & 0\\ 0 & Z_1^*AS_1\end{bmatrix}\end{aligned}$$

这就验证了(1.4.8)．

反过来，假设存在一个非奇异的 S，使得 $A=S([\lambda]\oplus B)S^{-1}$．设 x 是 S 的第一列，y 是 S^{-*} 的第一列，且分划 $S=[x\quad S_1]$以及 $S^{-*}=[y\quad Z_1]$．则恒等式 $S^{-1}S=I$ 的位于(1，
78 1)处的元素告诉我们 $y^*x=1$；恒等式

$$[Ax\quad AS_1] = AS = S([\lambda]\oplus B) = [\lambda x\quad S_1B]$$

的第一列告诉我们 $Ax=\lambda x$；而恒等式

$$\begin{bmatrix}y^*A\\ Z_1^*A\end{bmatrix} = S^{-1}A = ([\lambda]\oplus B)S^{-1} = \begin{bmatrix}\lambda y^*\\ BZ_1^*\end{bmatrix}$$

的第一行告诉我们 $y^*A=\lambda y^*$． □

(1.4.7a)中的结论是**双正交原理**(principle of biorthogonality)．人们可能还会问：如果与**同一个**特征值相伴的左右特征向量或者正交，或者线性相关，那时将会发生什么？这些情形将在(2.4.11.1)中讨论．

相似不改变矩阵的特征值，它的特征向量在相似之下以一种简单的方式进行变换．

定理 1.4.9 设 A，$B\in M_n$，并假设对某个非奇异的 S 有 $B=S^{-1}AS$．如果 $x\in\mathbf{C}^n$ 是 B 的与特征值 λ 相伴的右特征向量，那么 Sx 就是 A 的与特征值 λ 相伴的右特征向量．如果 $y\in\mathbf{C}^n$ 是 B 的与特征值 λ 相伴的左特征向量，那么 $S^{-*}y$ 就是 A 的与 λ 相伴的左特征向量．

证明 如果 $Bx=\lambda x$，那么 $S^{-1}ASx=\lambda x$，或者 $A(Sx)=\lambda(Sx)$．由于 S 是非奇异的，且 $x\neq 0$，$Sx\neq 0$，故而 Sx 是 A 的一个特征向量．如果 $y^*B=\lambda y^*$，那么 $y^*S^{-1}AS=\lambda y^*$，或者 $(S^{-*}y)^*A=\lambda(S^{-*}y)^*$． □

关于主子矩阵的特征值的信息可能对特征值的代数重数不可能小于它的几何重数这一基本结论加以改进.

定理 1.4.10 设给定 $A\in M_n$ 以及 $\lambda\in\mathbf{C}$，又设 $k\geqslant1$ 是给定的正整数. 考虑如下三个命题.

(a) λ 是 A 的几何重数至少为 k 的特征值.

(b) 对每一个 $m=n-k+1,\cdots,n$，λ 是 A 的每一个 $m\times m$ 主子矩阵的特征值.

(c) λ 是 A 的代数重数至少为 k 的特征值.

那么(a)蕴含(b)，且(b)蕴含(c). 特别地，特征值的代数重数至少与其几何重数一样大.

证明 (a)$\Rightarrow$(b)：设 λ 是 A 的几何重数至少为 k 的特征值，这就意味着 $\operatorname{rank}(A-\lambda I)\leqslant n-k$. 假设 $m>n-k$. 那么 $A-\lambda I$ 的每个 $m\times m$ 子式是零. 特别地，$A-\lambda I$ 的每个 $m\times m$ 主子式是零，故而 $A-\lambda I$ 的每个 $m\times m$ 主子矩阵是奇异的. 于是，λ 是 A 的每个 $m\times m$ 主子矩阵的特征值.

(b)$\Rightarrow$(c)：假设对每个 $m\geqslant n-k+1$，λ 是 A 的每个 $m\times m$ 主子矩阵的特征值. 那么，$A-\lambda I$ 的每个阶至少为 $n-k+1$ 的主子式都是零，所以，对所有 $j\geqslant n-k+1$，每个主子式的和 $E_j(A-\lambda I)=0$. 这样(1.2.13)与(1.2.11)确保有 $p_{A-\lambda I}^{(i)}(0)=0$(对 $i=0,1,\cdots,k-1$). 但是 $p_{A-\lambda I}(t)=p_A(t+\lambda)$，所以 $p_A^{(i)}(\lambda)=0$(对 $i=0,1,\cdots,k-1$). 这就是说，λ 是 $p_A(t)$的重数至少为 k 的零点. □

几何重数为 1 的特征值 λ 可能有代数重数 2 或者更高，但这只有在与 λ 相伴的左右特征向量为正交时才可能发生. 然而，如果 λ 的代数重数为 1，那么它的几何重数为 1，与 λ 相伴的左右特征向量不可能是正交的. 我们推导出这些结果的方法依赖于如下的引理.

引理 1.4.11 设给定 $A\in M_n$，$\lambda\in\mathbf{C}$ 以及非零向量 x，$y\in\mathbf{C}^n$. 假设 λ 作为 A 的特征值的几何重数为 1，$Ax=\lambda x$，且 $y^*A=\lambda y^*$. 那么存在一个非零的 $\gamma\in\mathbf{C}$，使得 $\operatorname{adj}(\lambda I-A)=\gamma xy^*$.

证明 我们有 $\operatorname{rank}(\lambda I-A)=n-1$，故而 $\operatorname{rank}\operatorname{adj}(\lambda I-A)=1$，这就是 $\operatorname{adj}(\lambda I-A)=\xi\eta^*$(对某个非零的 ξ，$\eta\in\mathbf{C}^n$)，见(0.8.2). 但是$(\lambda I-A)(\operatorname{adj}(\lambda I-A))=\det(\lambda I-A)I=0$，所以$(\lambda I-A)\xi\eta^*=0$，且$(\lambda I-A)\xi=0$，它蕴含对某个非零的纯量 α 有 $\xi=\alpha x$. 按照类似的方式利用恒等式$(\operatorname{adj}(\lambda I-A))(\lambda I-A)=0$，我们得出结论：对某个非零的纯量 β 有 $\eta=\beta y$. 于是有 $\operatorname{adj}(\lambda I-A)=\alpha\beta xy^*$. □

定理 1.4.12 设给定 $A\in M_n$，$\lambda\in\mathbf{C}$ 以及非零向量 x，$y\in\mathbf{C}^n$. 假设 λ 是 A 的一个特征值，$Ax=\lambda x$，且 $y^*A=\lambda y^*$.

(a) 如果 λ 的代数重数为 1，那么 $y^*x\neq0$.

(b) 如果 λ 的几何重数为 1，那么它的代数重数为 1，当且仅当 $y^*x\neq0$.

证明 对于(a)与(b)这两种情形，λ 的几何重数都为 1. 上一个引理告诉我们，存在一个非零的 $\gamma\in\mathbf{C}$，使得 $\operatorname{adj}(\lambda I-A)=\gamma xy^*$. 这样就有 $p_A(\lambda)=0$ 以及 $p_A'(\lambda)=\operatorname{tr}\operatorname{adj}(\lambda I-A)=\gamma y^*x$，见(0.8.10.2). 在(a)中我们假设了代数重数是 1，所以 $p_A'(\lambda)\neq0$，从而 $y^*x\neq0$. 在(b)中我们假设了 $y^*x\neq0$，所以 $p_A'(\lambda)\neq0$，故而其代数重数为 1. □

问题

1.4.P1 设给定非零向量 x，$y\in\mathbf{C}^n$，$A=xy^*$，又设 $\lambda=y^*x$. 证明：(a)λ 是 A 的特征值；(b)x 是 A 的与 λ 相伴的右特征向量，而 y 则是 A 的与 λ 相伴的左特征向量；(c)如果 $\lambda\neq0$，那么它是 A 的仅

有的非零特征值(代数重数=1). 说明为什么与 y 正交的任何向量都在 A 的零空间中. 特征值 0 的几何重数是什么? 说明为什么 A 可对角化, 当且仅当 $y^*x\neq 0$?

1.4.P2 设 $A\in M_n$ 是斜对称的. 证明 $p_A(t)=(-1)^n p_A(-t)$, 并导出结论: 如果 λ 是 A 的重数为 k 的特征值, 那么 $-\lambda$ 亦然. 如果 n 是奇数, 说明为什么 A 必定是奇异的. 说明为什么 A 的每一个奇数阶的主子式都是奇异的. 利用斜对称矩阵是主秩的这一事实(0.7.6)来证明 $\operatorname{rank}A$ 必定是偶数.

1.4.P3 假设 $n\geqslant 2$, 又令 $T=[t_{ij}]\in M_n$ 是上三角的. (a)设 x 是 T 的与特征值 t_{nn} 相伴的特征向量, 说明为什么 e_n 是与 t_{nn} 相伴的左特征向量. 如果对每个 $i=1,\cdots,n-1$ 都有 $t_{ii}\neq t_{nn}$, 证明 x 最后那个元素**必定**不是零. (b)设 $k\in\{1,\cdots,n-1\}$. 证明存在 T 的与特征值 t_{kk} 相伴的特征向量 x, x 的最后 $n-k$ 个元素都是零, 即 $x^{\mathrm{T}}=[\xi^{\mathrm{T}}\quad 0]^{\mathrm{T}}$, 其中 $\xi\in\mathbf{C}^k$. 如果对所有 $i=1,\cdots,k-1$ 有 $t_{ii}\neq t_{kk}$, 说明为什么 x 的第 k 个元素**必定**不是零.

80

1.4.P4 假设 $A\in M_n$ 是三对角的, 且有为零的主对角线. 设 $S=\operatorname{diag}(-1,1,-1,\cdots,(-1)^n)$, 并证明 $S^{-1}AS=-A$. 如果 λ 是 A 的重数为 k 的特征值, 说明为什么 $-\lambda$ 也是 A 的重数为 k 的特征值. 如果 n 是奇数, 证明 A 是奇异的.

1.4.P5 考虑分块三角矩阵

$$A=\begin{bmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{bmatrix},\quad A_{ii}\in M_{n_i},\quad i=1,2$$

如果 $x\in\mathbf{C}^{n_1}$ 是 A_{11} 的与 $\lambda\in\sigma(A_{11})$ 相伴的右特征向量, 又如果 $y\in\mathbf{C}^{n_2}$ 是 A_{22} 的与 $\mu\in\sigma(A_{22})$ 相伴的左特征向量, 证明 $\begin{bmatrix}x\\0\end{bmatrix}\in\mathbf{C}^{n_1+n_2}$ 和 $\begin{bmatrix}0\\y\end{bmatrix}$ 分别是 A 的与 λ 和 μ 相伴的右特征向量以及左特征向量. 利用这个结论来证明, A 的特征值就是 A_{11} 的特征值与 A_{22} 的特征值之合集.

1.4.P6 假设 $A\in M_n$ 有一个元素皆为正数的左特征向量以及一个元素皆为正数的右特征向量, 两者皆与特征值 λ 相伴. (a)证明: 对于任何一个异于 λ 的特征值, A 没有与之相伴的每个元素均为非负的左特征向量或者右特征向量. (b)如果 λ 的几何重数为 1, 证明它的代数重数也为 1. A 的这一性质(确保存在与 A 的一个特定的特征值相伴的正的左右特征向量的充分条件)见(8.2.2)以及(8.4.4).

1.4.P7 在这个问题中我们来概略描述求 $A\in M_n$ 的最大模特征值以及与之相伴的特征向量的**幂方法**(power method)的一种简单的形式. 假设 $A\in M_n$ 有不同的特征值 $\lambda_1,\cdots,\lambda_n$, 且恰好存在一个特征值 λ_n 有最大模 $\rho(A)$. 如果 $x^{(0)}\in\mathbf{C}^n$ 不正交于与 λ_n 相伴的一个左特征向量, 证明序列

$$x^{(k+1)}=\frac{1}{(x^{(k)*}x^{(k)})^{1/2}}Ax^{(k)},\quad k=0,1,2,\cdots$$

收敛于 A 的一个特征向量, 且向量 $Ax^{(k)}$ 与 $x^{(k)}$ 中一个给定元素的比值收敛于 λ_n.

1.4.P8 继续采用(1.4.P7)中的假设以及记号. A 的更多的特征值(特征向量)可以通过将幂方法与释放出低一阶方阵的**压缩法**(deflation)组合起来进行计算, 这个新的低一阶矩阵的谱(计入重数)包含 A 的除了一个特征值以外所有其他的特征值. 设 $S\in M_n$ 是非奇异的, 且其第一列是与特征值 λ_n 相伴的特征向量 $y^{(n)}$. 证明 $S^{-1}AS=\begin{bmatrix}\lambda_n & *\\ 0 & B\end{bmatrix}$, 且 $B\in M_{n-1}$ 的特征值是 $\lambda_1,\cdots,\lambda_{n-1}$. 另一个特征值可以重复用压缩法由 B 计算出来.

1.4.P9 设 $A\in M_n$ 有特征值 $\lambda_1,\cdots,\lambda_{n-1},0$, 所以 $\operatorname{rank}A\leqslant n-1$. 假设 A 的最后一行是前 $n-1$ 行的线性组合. 分划 $A=\begin{bmatrix}B & x\\ y^{\mathrm{T}} & \alpha\end{bmatrix}$, 其中 $B\in M_{n-1}$. (a)说明为什么存在一个 $z\in\mathbf{C}^{n-1}$, 使得 $y^{\mathrm{T}}=z^{\mathrm{T}}B$

81

以及 $\alpha=z^{\mathrm{T}}x$. 为什么 $\begin{bmatrix}z\\-1\end{bmatrix}$ 是 A 的与特征值 0 相伴的左特征向量? (b)证明 $B+xz^{\mathrm{T}}\in M_{n-1}$ 有特

征值 λ_1，…，λ_{n-1}．这个构造是另一种类型的**压缩法**；有关压缩法的进一步的例子，见(1.3.P33)．(c)如果已知 A 的一个特征值 λ，说明如何对一个合适的置换 P 将这种结构应用到$P(A-\lambda I)P^{-1}$．

1.4.P10 设 $T\in M_n$ 是一个非奇异的矩阵，它的列是 $A\in M_n$ 的左特征向量．证明 T^{-*} 的列是 A 的右特征向量．

1.4.P11 假设 $A\in M_n$ 是一个不可约的上 Hessenberg 矩阵(0.9.9)．说明为什么对每个 $\lambda\in\mathbf{C}$ 有 $\operatorname{rank}(A-\lambda I)\geqslant n-1$，并导出结论：$A$ 的每个特征值的几何重数皆为 1，即 A 是非损的．

1.4.P12 设 λ 是 $A\in M_n$ 的一个特征值．(a)证明 $A-\lambda I$ 的每一组 $n-1$ 列都是线性无关的，当且仅当不存在 A 的与 λ 相伴的特征向量有一个为零的元素．(b)如果不存在 A 的与 λ 相伴的特征向量有一个为零的元素，为什么 λ 必定几何重数为 1?

1.4.P13 设给定 $A\in M_n$ 以及非零向量 x，$y\in\mathbf{C}^n$，又设 λ，λ_2，…，λ_n 是 A 的特征值．假设 $Ax=\lambda x$ 以及 $y^*A=\lambda y^*$，又 λ 的几何重数为 1．那么(1.4.11)是说 $\operatorname{adj}(\lambda I-A)=\gamma xy^*$ 以及 $\gamma\neq 0$．(a)说明为什么 $\gamma y^*x=\operatorname{tr}(\lambda I-A)=E_{n-1}(\lambda I-A)=S_{n-1}(\lambda I-A)=(\lambda-\lambda_2)(\lambda-\lambda_3)\cdots(\lambda-\lambda_n)$．(b)由(a)推出结论：$y^*x\neq 0$ 当且仅当 λ 是单重特征值．(c)参数 γ 不是零，而无论 λ 的重数是多少．如果 λ 是单重的，说明为什么 $\gamma=(\lambda-\lambda_2)\cdots(\lambda-\lambda_n)/y^*x$．作为计算 γ 的不同方式，见(2.6.P12)．(d)说明为什么 x 与 y 的每个元素都不是零$\Leftrightarrow\lambda I-A$ 的每个主子式都不是零$\Leftrightarrow\operatorname{adj}(\lambda I-A)$的每个**主对角**元素都不是零$\Leftrightarrow\operatorname{adj}(\lambda I-A)$的**每一个**元素都不是零．

1.4.P14 设 $A\in M_n$ 并令 $t\in\mathbf{C}$．说明为什么$(A-tI)\operatorname{adj}(A-tI)=\operatorname{adj}(A-tI)(A-tI)=p_A(t)I$．现在假设 λ 是 A 的一个特征值．证明：(a)$\operatorname{adj}(A-\lambda I)$的每个非零的列都是 A 的与 λ 相伴的一个特征向量；(b)$\operatorname{adj}(A-\lambda I)$的每个非零的行都是 A 的与 λ 相伴的一个左特征向量的共轭转置；(c)$\operatorname{adj}(A-\lambda I)\neq 0$ 当且仅当 λ 的几何重数为 1；(d)如果 λ 是 $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}$的一个特征值，那么 $\begin{bmatrix} d-\lambda & -b \\ -c & a-\lambda \end{bmatrix}$的每一个非零的列都是 A 的与 λ 相伴的一个特征向量，每一个非零的行都是 A 的与 λ 相伴的一个左特征向量的共轭转置．

1.4.P15 假设 λ 是 $A\in M_n$ 的一个单重特征值，又假设 x，y，z，$w\in\mathbf{C}^n$，$Ax=\lambda x$，$y^*A=\lambda y^*$，$y^*z\neq 0$ 以及 $w^*x\neq 0$．证明 $A-\lambda I+\kappa zw^*$ 对所有 $\kappa\neq 0$ 都是非奇异的．说明为什么可以取 $z=x$．

1.4.P16 证明复的三对角 Toeplitz 矩阵

$$A=\begin{bmatrix} a & b & & \\ c & a & \ddots & \\ & \ddots & \ddots & b \\ & & c & a \end{bmatrix}\in M_n,\quad bc\neq 0 \tag{1.4.13}$$

是可以对角化的，且有谱 $\sigma(A)=\left\{a+2\sqrt{bc}\cos\left(\frac{\pi\kappa}{n+1}\right)：\kappa=1，\cdots，n\right\}$，其中 $\operatorname{Re}\sqrt{bc}\geqslant 0$ 且 $\operatorname{Im}\sqrt{bc}>0$，如果 bc 是实的，且是负数．

1.4.P17 如果在(1.4.13)中有 $a=2$ 以及 $b=c=-1$，证明 $\sigma(A)=\left\{4\sin^2\left(\frac{\pi\kappa}{2(n+1)}\right)：\kappa=1，\cdots，n\right\}$．