第 2 章 酉相似与酉等价

2.0 引言

在第 1 章里，我们通过一般性的非奇异矩阵 S 对于 $A\in M_n$ 的相似性，即对变换 $A\to S^{-1}AS$ 作了初步的研究. 对某种很特别的非奇异矩阵，称为**酉矩阵**(unitary matrix)，S 的逆有简单的形式：$S^{-1}=S^*$. 通过酉矩阵 U 的相似 $A\to U^*AU$ 不仅在概念上比一般的相似性更加简单(共轭转置要比计算逆矩阵简单得多)，而且在数值计算中也有较高的稳定性. 酉相似的一个基本性质是：每一个 $A\in M_n$ 都与一个上三角矩阵酉相似，这个上三角矩阵的对角元素是 A 的特征值. 这个三角形式可以在一般相似性之下进一步加以改进，我们将在第 3 章里研究后者.

变换 $A\to S^*AS$(其中 S 是非奇异的，但不一定是酉矩阵)称为 *** 相合**(* congruence)；我们将在第 4 章里研究它. 注意，通过酉矩阵的相似性既是相似，也是一个 * 相合.

对 $A\in M_{n,m}$，变换 $A\to UAV$(其中 $U\in M_m$ 与 $V\in M_n$ 两者都是酉矩阵)称为**酉等价**(unitary equivalence). 在酉相似之下可以得到的上三角形式在酉等价之下可以大为改进并推广到长方形矩阵的情形：每一个 $A\in M_{n,m}$ 都与一个非负的对角矩阵酉等价，这个对角矩阵的对角元素(A 的奇异值)有极大的重要性.

2.1 酉矩阵与 *QR* 分解

定义 2.1.1 一列向量 x_1，…，x_k 是正交的，如果对所有 $i\neq j$ 以及 i，$j\in\{1,\ \cdots,\ k\}$ 都有 $x_i^*x_j=0$. 此外，如果对所有 $i=1$，…，k 都有 $x_i^*x_i=1$(即这些向量是标准化的)，那么这组向量就是标准正交的. 我们常常说"x_1，…，x_k 是正交的(标准正交的)"，而不说

83 成更正式的"向量组 x_1，…，x_k 是正交的(标准正交的)"，这样更为方便.

习题 如果 y_1，…，$y_k\in\mathbf{C}^n$ 是正交的非零向量，证明：由 $x_i=(y_i^*y_i)^{-1/2}y_i$，$i=1$，…，k 定义的向量 x_1，…，x_k 是标准正交的. ◀

定理 2.1.2 $\mathbf{C}^n$ 中每个标准正交的向量组都是线性无关的.

证明 假设$\{x_1,\ \cdots,\ x_k\}$是一个标准正交组，又设 $0=\alpha_1x_1+\cdots+\alpha_kx_k$. 那么 $0=(\alpha_1x_1+\cdots+\alpha_kx_k)^*(\alpha_1x_1+\cdots+\alpha_kx_k)=\sum_{i,j}\bar{\alpha}_i\alpha_jx_i^*x_j=\sum_{i=1}^k|\alpha_i|^2x_i^*x_i=\sum_{i=1}^k|\alpha_i|^2$，因为诸向量 x_i 是正交的且标准化的. 于是，所有 $\alpha_i=0$，从而$\{x_1,\ \cdots,\ x_k\}$是线性无关的向量组. □

习题 证明 $\mathbf{C}^n$ 中每个非零的正交的向量组都是线性无关的. ◀

习题 如果 x_1，…，$x_k\in\mathbf{C}^n$ 是正交的，证明要么 $k\leqslant n$，要么诸向量 x_i 中至少有 $k-n$ 个是零向量. ◀

当然，线性无关组不一定是标准正交的，不过我们可以对它应用 Gram-Schmidt 标准正交化方法(0.6.4)，从而得到一组具有相同生成子空间的标准正交基.

习题 证明 $\mathbf{R}^n$ 或者 $\mathbf{C}^n$ 的任何非零子空间都有标准正交基(0.6.5). ◀

定义 2.1.3 矩阵 $U\in M_n$ 是酉矩阵，如果 $U^*U=I$. 矩阵 $U\in M_n(\mathbf{R})$是实正交矩阵，

如果 $U^{\mathrm{T}}U=I$.

习题 证明：$U\in M_n$ 与 $V\in M_m$ 是酉矩阵，当且仅当 $U\oplus V\in M_{n+m}$ 是酉矩阵.

习题 验证：(1.3)节的(P19，P20 以及 P21)中的矩阵 Q，U 以及 V 是酉矩阵.

M_n 中的酉矩阵构成一个不寻常的重要集合. 我们在(2.1.4)中列出 U 是酉矩阵的基本等价条件.

定理 2.1.4 如果 $U\in M_n$，则下列诸命题等价.

(a) U 是酉矩阵.

(b) U 是非奇异的，且 $U^*=U^{-1}$.

(c) $U^*U=I$.

(d) U^* 是酉矩阵.

(e) U 的列是标准正交的.

(f) U 的行是标准正交的.

(g) 对所有 $x\in\mathbf{C}^n$，$\|x\|_2=\|Ux\|_2$，即 x 与 Ux 有相同的 **Euclid 范数**(Euclidean norm).

证明 (a)蕴含(b)是因为 U^{-1}(当它存在时)是唯一的矩阵，用它左乘得到 I(0.5)，酉矩阵的定义就是说 U^* 正是这样一个矩阵. 由于 $BA=I$ 当且仅当 $AB=I$(对 A，$B\in M_n$(0.5))，所以(b)蕴含(c). 由于$(U^*)^*=U$，所以(c)蕴含 U^* 是酉矩阵，也就是说，(c)蕴含(d). 这些蕴含关系中每一个的逆命题都可以类似地处理，故而(a)～(d)是等价的.

按照列作分划 $U=[u_1\cdots u_n]$. 那么 $U^*U=I$ 就意味着对所有 $i=1,\cdots,n$ 有 $u_i^*u_i=1$ 以及对所有 $i\neq j$ 有 $u_i^*u_j=0$. 于是，$U^*U=I$ 是 U 的列标准正交的另一种表述方法，从而(a)等价于(e). 类似地，(d)与(f)等价. 84

如果 U 是酉矩阵，且 $y=Ux$，那么 $y^*y=x^*U^*Ux=x^*Ix=x^*x$，所以(a)蕴含(g). 为证明其逆，设 $U^*U=A=[a_{ij}]$，且给定 z，$w\in\mathbf{C}$，又在(g)中取 $x=z+w$. 那么 $x^*x=z^*z+w^*w+2\mathrm{Re}\,z^*w$，且 $y^*y=x^*Ax=z^*Az+w^*Aw+2\mathrm{Re}\,z^*Aw$. (g)确保 $z^*z=z^*Az$ 以及 $w^*w=w^*Aw$，从而对任意的 z 与 w 有 $\mathrm{Re}\,z^*w=\mathrm{Re}\,z^*Aw$. 取 $z=e_p$ 以及 $w=\mathrm{i}e_q$，并计算 $\mathrm{Re}\,\mathrm{i}e_p^{\mathrm{T}}e_q=0=\mathrm{Re}\,\mathrm{i}e_p^{\mathrm{T}}Ae_q=\mathrm{Re}\,\mathrm{i}a_{pq}=-\mathrm{im}\,a_{pq}$，所以 A 的每个元素都是实的. 最后，取 $z=e_p$ 以及 $w=e_q$，并计算 $e_p^{\mathrm{T}}e_q=\mathrm{Re}\,e_p^{\mathrm{T}}e_q=\mathrm{Re}\,e_p^{\mathrm{T}}Ae_q=a_{pq}$，这就告诉我们有 $A=I$，且 U 是酉矩阵. □

定义 2.1.5 线性变换称为 **Euclid 等距**(Euclidean isometry)，如果对所有 $x\in\mathbf{C}^n$ 都有 $\|x\|_2=\|Tx\|_2$. 定理 2.1.4 是说：复方阵 $U\in M_n$ 是 Euclid 等距的(通过 U：$x\to Ux$)，当且仅当它是酉矩阵. 其他种类的等距，请见(5.4.P11～13).

习题 设 $U_\theta=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$，其中 θ 是实参数. (a)证明：给定的 $U\in M_2(\mathbf{R})$是实的正交矩阵，当且仅当要么 $U=U_\theta$，要么对某个 $\theta\in\mathbf{R}$ 有 $U=\begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix}U_\theta$. (b)证明：给定的 $U\in M_2(\mathbf{R})$是实正交的，当且仅当对某个 $\theta\in\mathbf{R}$ 有 $U=U_\theta$ 或者 $U=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}U_\theta$. 这些结论是2×2实正交矩阵的两种不同的表达方式(涉及一个参数 θ). 用几何方式对它们予以解释.

结论 2.1.6 如果 U，$V\in M_n$ 是酉矩阵(实正交矩阵)，那么 UV 也是酉矩阵(实正交矩阵).

习题 利用(2.1.4)的(b)证明(2.1.6). ◀

结论 2.1.7 M_n 中酉矩阵(实正交矩阵)的集合作成一个群. 这个群通常称为 $n\times n$ 酉(实正交)群，它是 $GL(n, \mathbf{C})$ 的一个子群(0.5).

习题 群是对单独一个满足结合律的二元运算("乘法")封闭的集合，且在此集合中含有该运算的恒等元以及逆元. 验证(2.1.7). 提示：封闭性利用(2.1.6)，矩阵乘法是可结合的，$I\in M_n$ 是酉矩阵，而 $U^*=U^{-1}$ 仍然是酉矩阵. ◀

M_n 中酉矩阵的集合(群)有其他非常重要的性质. 矩阵序列的"收敛性"以及"极限"的概念确定将在第 5 章里给出，但在这里可以被理解为其元素的"收敛性"以及"极限". 定义等式 $U^*U=I$ 表明 U 的每一列的 Euclid 范数均为 1，因而 $U=[u_{ij}]$中没有任何元素的绝对值大于 1. 如果我们把酉矩阵的集合视为 $\mathbf{C}^{n^2}$ 的一个子集，这就是说是它的一个**有界**子集. 如果 $U_k=[u_{ij}^{(k)}]$是酉矩阵组成的一个无限序列($k=1, 2, \cdots$)，使得对所有 i，$j=1, 2, \cdots, n$ 都有 $\lim\limits_{k\to\infty} u_{ij}^k=u_{ij}$，那么由恒等式 $U_k^*U_k=I$(对所有 $k=1, 2, \cdots$)我们就看出 $\lim\limits_{k\to\infty} U_k^*U_k=U^*U=I$，
85 其中 $U=[u_{ij}]$. 于是，极限矩阵 U 也是酉矩阵. 这就是说，酉矩阵的集合是 $\mathbf{C}^{n^2}$ 的封闭子集.

由于有限维 Euclid 空间的封闭且有界的子集是一个**紧集**(compact set)(见附录 E)，我们断言：M_n 中酉矩阵的集合(群)是紧的. 对我们的目的来说，这一结论的最重要的推论是如下关于酉矩阵的**选择原理**(selection principle).

引理 2.1.8 设 U_1，U_2，$\cdots\in M_n$ 是一个给定的由酉矩阵组成的无穷序列. 则存在一个无穷子序列 U_{k_1}，U_{k_2}，$\cdots(1\leqslant k_1<k_2<\cdots)$，使得当 $i\to\infty$时，U_{k_i} 的所有的元素都收敛于一个酉矩阵的元素.

证明 这里所需要的全部事实都来自紧集的任何无限子序列，有人或许总是会选取一个收敛的子序列. 我们已经注意到，如果酉矩阵的序列收敛于某个矩阵，那么极限矩阵必定是酉矩阵.

引理确保存在的酉极限未必是唯一的；它有可能与子序列的选择有关. □

习题 考虑酉矩阵序列 $U_k=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}^k$ $(k=1, 2, \cdots)$. 证明其子序列有两个可能的极限. ◀

习题 说明为什么选择原理(2.1.8)也适用于(实的)正交群，也就是说，由实正交矩阵组成的无穷序列有一个无限子序列收敛于一个实正交矩阵. ◀

酉矩阵 U 有这样的性质：U^{-1}等于 U^*. 推广酉矩阵的一种方式是要求 U^{-1}与 U^* 相似. 这样的矩阵组成之集合容易刻画成映射 $A\to A^{-1}A^*$ 的值域(对所有非奇异的 $A\in M_n$).

定理 2.1.9 设 $A\in M_n$ 非奇异. 那么 A^{-1} 相似于 A^*，当且仅当存在一个非奇异的 $B\in M_n$，使得 $A=B^{-1}B^*$.

证明 如果对某个非奇异的 $B\in M_n$ 有 $A=B^{-1}B^*$，那么 $A^{-1}=(B^*)^{-1}B$ 且 $B^*A^{-1}\times(B^*)^{-1}=B(B^*)^{-1}=(B^{-1}B^*)^*=A^*$，所以 A^{-1}通过相似矩阵 B^* 与 A^* 相似. 反过来，如果 A^{-1} 与 A^* 相似，那么就存在一个非奇异的 $S\in M_n$，使得 $SA^{-1}S^{-1}=A^*$，从而 $S=A^*SA$. 置 $S_\theta=e^{i\theta}S$(对 $\theta\in\mathbf{R}$)，所以 $S_\theta=A^*S_\theta A$，且 $S_\theta^*=A^*S_\theta^*A$. 将这两个恒等式相加给出 $H_\theta=A^*H_\theta A$，其中 $H_\theta=S_\theta+S_\theta^*$ 是 Hermite 的. 如果 H_θ 是奇异的，那么就会存在一个非零的 $x\in\mathbf{C}^n$，使得 $0=H_\theta x=S_\theta x+S_\theta^*x$，所以 $-x=S_\theta^{-1}S_\theta^*x=e^{-2i\theta}S^{-1}S^*x$，且

$S^{-1}S^{*}x=-\mathrm{e}^{2\mathrm{i}\theta}x$. 选取一个值 $\theta=\theta_0\in[0,2\pi)$，使得 $-\mathrm{e}^{2\mathrm{i}\theta_0}$ 不是 $S^{-1}S^{*}$ 的特征值；所产生的 Hermite 矩阵 $H=H_{\theta_0}$ 就是非奇异的，且有性质 $H=A^{*}HA$.

现在选取任意一个复的 α，使得 $|\alpha|=1$，且 α 不是 A^{*} 的特征值. 置 $B=\beta(\alpha I-A^{*})H$，其中复参数 $\beta\neq0$ 有待选取，注意 B 是非奇异的. 我们希望有 $A=B^{-1}B^{*}$，即 $BA=B^{*}$. 计算 $B^{*}=H(\bar{\beta}\bar{\alpha}I-\bar{\beta}A)$ 以及 $BA=\beta(\alpha I-A^{*})HA=\beta(\alpha HA-A^{*}HA)=\beta(\alpha HA-H)=H(\alpha\beta A-\beta I)$. 如果我们能选取一个非零的 β，使得 $\beta=-\bar{\beta}\bar{\alpha}$，我们就完成了，但是如果 $\alpha=\mathrm{e}^{\mathrm{i}\varphi}$，那么就有 $\beta=\mathrm{e}^{\mathrm{i}(\pi-\varphi)/2}$，这也完成了证明. □ 86

如果酉矩阵作为 2×2 分块矩阵出现，那么它落在对角线之外的那些块的秩相等，它的对角线块的秩通过一个简单的公式相联系.

引理 2.1.10 设酉矩阵 $U\in M_n$ 被分划成 $U=\begin{bmatrix}U_{11} & U_{12}\\ U_{21} & U_{22}\end{bmatrix}$，其中 $U_{11}\in M_k$. 这样就有 $\operatorname{rank}U_{12}=\operatorname{rank}U_{21}$ 以及 $\operatorname{rank}U_{22}=\operatorname{rank}U_{11}+n-2k$. 特别地，$U_{12}=0$ 当且仅当 $U_{21}=0$，在此情形 U_{11} 与 U_{22} 为酉矩阵.

证明 关于秩的那两个论断立即由零性互补法则(0.7.5)推出，并用到 $U^{-1}=\begin{bmatrix}U_{11}^{*} & U_{21}^{*}\\ U_{12}^{*} & U_{22}^{*}\end{bmatrix}$这一事实.

习题 利用上一引理来证明：酉矩阵是上三角的，当且仅当它是对角矩阵. ◀

平面旋转与 Householder 矩阵是特殊的(也是非常简单的)酉矩阵，它们在建立某些基本的矩阵分解过程中起着重要的作用.

例 2.1.11(平面旋转) 设 $1\leqslant i<j\leqslant n$，并令

$$U(\theta;i,j)=\begin{bmatrix}1 &&&&&&&&&\\ &\ddots&&&&&&&&\\ &&1&&&&&&&\\ &&&\cos\theta&&&&-\sin\theta&&\\ &&&&1&&&&&\\ &&&&&\ddots&&&&\\ &&&&&&1&&&\\ &&&\sin\theta&&&&\cos\theta&&\\ &&&&&&&&1&\\ &&&&&&&&&\ddots\\ &&&&&&&&&&1\end{bmatrix}$$

表示用 $\cos\theta$ 代替 $n\times n$ 单位矩阵的第 i 行第 i 列和第 j 行第 j 列的元素，用 $-\sin\theta$ 代替它的第 i 行第 j 列的元素，而用 $\sin\theta$ 代替它的第 j 行第 i 列的元素所得到的结果. 矩阵 $U(\theta;i,j)$ 称为**平面旋转**(plane rotation)或者 **Givens 旋转**(Givens rotation).

习题 验证：对任何一对指数 i，$j(1\leqslant i<j\leqslant n)$ 以及任何参数 $\theta\in[0,2\pi)$，$U(\theta;i,j)\in M_n(\mathbf{R})$都是实正交的. 矩阵 $U(\theta;i,j)$在 $\mathbf{R}^n$ 的 i，j 坐标平面上执行一个旋转(旋转任意角度 θ). 用 $U(\theta;i,j)$左乘只影响被乘的矩阵的第 i 行和第 j 行，而用 $U(\theta;i,j)$右乘只影响被乘的矩阵的第 i 列和第 j 列. ◀

习题 验证 $U(\theta; i, j)^{-1}=U(-\theta; i, j)$. ◀

例 2.1.12(Householder 矩阵) 设 $w\in\mathbf{C}^n$ 是一个非零向量. Householder 矩阵 $U_w\in M_n$
87 定义为 $U_w=I-2(w^*w)^{-1}ww^*$. 如果 w 是单位向量, 则有 $U_w=I-2ww^*$.

习题 证明: Householder 矩阵 U_w 既是酉矩阵, 也是 Hermite 矩阵, 所以 $U_w^{-1}=U_w$. ◀

习题 设 $w\in\mathbf{R}^n$ 是一个非零向量. 证明 Householder 矩阵 U_w 是实正交的且是对称的. 为什么 U_w 的每一个特征值或者是 $+1$, 或者是 -1? ◀

习题 证明 Householder 矩阵 U_w 在子空间 $w^\perp$ 上的作用是恒等元, 而它在由 w 生成的一维子空间上的作用是一个反射. 也就是说: $U_wx=x$ 成立, 如果 $x\perp w$ 且 $U_ww=-w$. ◀

习题 利用(0.8.5.11)证明: 对所有 n 有 $\det U_w=-1$. 于是, 对所有 n 以及每个非零的 $w\in\mathbf{R}^n$, Householder 矩阵 $U_w\in M_n(\mathbf{R})$ 是实正交矩阵, 它从来就不是**真旋转矩阵**(proper rotation matrix)(真旋转矩阵是行列式为 $+1$ 的实正交矩阵). ◀

习题 利用(1.2.8)证明 Householder 矩阵的特征值永远是 $-1, 1, \cdots, 1$, 并说明为什么它的行列式总是 -1. ◀

习题 设 $n\geqslant 2$, 并设 $x, y\in\mathbf{R}^n$ 是单位向量. 如果 $x=y$, 令 w 是任意一个与 x 正交的实单位向量. 如果 $x\neq y$, 令 $w=x-y$. 证明 $U_wx=y$. 导出结论: 任意的 $x\in\mathbf{R}^n$ 可以由实的 Householder 矩阵变换成任何一个满足 $\|x\|_2=\|y\|_2$ 的向量 $y\in\mathbf{R}^n$. ◀

习题 在 $\mathbf{C}^n$ 中情形不同. 证明在那里不存在 $w\in\mathbf{C}^n$ 使得 $U_we_1=\mathrm{i}e_1$. ◀

Householder 矩阵以及纯量酉矩阵可以用来构造一个酉矩阵, 它将 $\mathbf{C}^n$ 中任意给定的向量变换成 $\mathbf{C}^n$ 中有同样 Euclid 范数的另外任意一个向量.

定理 2.1.13 设给定 $x, y\in\mathbf{C}^n$, 并假设 $\|x\|_2=\|y\|_2>0$. 如果 $y=\mathrm{e}^{\mathrm{i}\theta}x$(对某个实的 θ), 令 $U(y,x)=\mathrm{e}^{\mathrm{i}\theta}I_n$; 反之, 设 $\phi\in[0,2\pi)$ 使得 $x^*y=\mathrm{e}^{\mathrm{i}\phi}|x^*y|$(如果 $x^*y=0$, 就取 $\phi=0$); 设 $w=\mathrm{e}^{\mathrm{i}\phi}x-y$; 又令 $U(y,x)=\mathrm{e}^{\mathrm{i}\phi}U_w$, 其中 $U_w=I-2(w^*w)^{-1}ww^*$ 是一个 Householder 矩阵. 这样 $U(y,x)$ 就是一个酉矩阵, 且是本性 Hermite 的, $U(y,x)x=y$, 又只要 $z\perp x$, 就有 $U(y,x)z\perp y$. 如果 x 与 y 是实的, 那么 $U(y,x)$ 是实正交矩阵: 如果 $y=x$, 则有 $U(y,x)=I$, 反之则 $U(y,x)$ 是实的 Householder 矩阵 U_{x-y}.

证明 如果 x 与 y 线性相关(也就是说, 如果对某个实的 θ 有 $y=\mathrm{e}^{\mathrm{i}\theta}x$), 那么这些结论容易验证. 如果 x 与 y 线性无关, 则 Cauchy-Schwartz 不等式(0.6.3)就确保有 $x^*x\neq|x^*y|$. 计算

$$\begin{aligned} w^*w &= (\mathrm{e}^{\mathrm{i}\phi}x-y)^*(\mathrm{e}^{\mathrm{i}\phi}x-y)=x^*x-\mathrm{e}^{-\mathrm{i}\phi}x^*y-\mathrm{e}^{\mathrm{i}\phi}y^*x+y^*y \\ &= 2(x^*x-\mathrm{Re}(\mathrm{e}^{-\mathrm{i}\phi}x^*y))=2(x^*x-|x^*y|) \end{aligned}$$

和

88 $$w^*x=\mathrm{e}^{-\mathrm{i}\phi}x^*x-y^*x=\mathrm{e}^{-\mathrm{i}\phi}x^*x-\mathrm{e}^{-\mathrm{i}\phi}|y^*x|=\mathrm{e}^{-\mathrm{i}\phi}(x^*x-|x^*y|)$$

最后计算

$$\mathrm{e}^{\mathrm{i}\phi}U_wx=\mathrm{e}^{\mathrm{i}\phi}(x-2(w^*w)^{-1}ww^*x)=\mathrm{e}^{\mathrm{i}\phi}(x-(\mathrm{e}^{\mathrm{i}\phi}x-y)\mathrm{e}^{-\mathrm{i}\phi})=y$$

如果 z 与 x 正交, 那么 $w^*z=-y^*z$, 且

$$y^*U(y,x)z=\mathrm{e}^{\mathrm{i}\phi}\left(y^*z-\frac{1}{\|x\|_2^2-|x^*y|}(\mathrm{e}^{\mathrm{i}\phi}y^*x-\|y\|_2^2)(-y^*z)\right)$$

$$= e^{i\phi}(y^* z + (-y^* x)) = 0$$

由于 U_w 是酉矩阵，且是 Hermite 矩阵，故而 $U(y, x)=(e^{i\phi}I)U_w$ 是酉矩阵(它是两个酉矩阵的乘积)，且是本性 Hermite 的，见(0.2.5). □

习题 设 $y\in\mathbf{C}^n$ 是一个给定的单位向量，又设 e_1 是 $n\times n$ 单位矩阵的第一列. 利用上一个定理中的方法构造 $U(y, e_1)$，并验证它的第一列是 y(它应该如此，因为 $y=U(y, e_1)e_1$). ◀

习题 设 $x\in\mathbf{C}^n$ 是一个给定的非零向量. 说明为什么上面定理中构造出来的矩阵 $U(\|x\|_2 e_1, x)$是本性 Hermite 的酉矩阵，它把 x 变成$\|x\|_2 e_1$. ◀

下面的复矩阵或者实矩阵的 QR 分解在理论上与计算上都有相当的重要性.

定理 2.1.14(QR 分解) 设给定 $A\in M_{n,m}$.

(a) 如果 $n\geqslant m$，则存在一个具有标准正交列向量的 $Q\in M_{n,m}$ 以及一个具有非负主对角元素的上三角矩阵 $R\in M_m$，使得 $A=QR$.

(b) 如果 $\operatorname{rank}A=m$，那么(a)中的因子 Q 和 R 是唯一确定的，且 R 的主对角元素全为正数.

(c) 如果 $m=n$，那么(a)中的因子 Q 是酉矩阵.

(d) 存在一个酉矩阵 $Q\in M_n$ 以及一个有非负对角元素的上三角矩阵 $R\in M_{n,m}$，使得$A=QR$.

(e) 如果 A 是实的，那么(a)、(b)、(c)以及(d)中的因子 Q 与 R 也可以取成实的.

证明 设 $a_1\in\mathbf{C}^n$ 是 A 的第一列，$r_1=\|a_1\|_2$，又设 U_1 是一个酉矩阵，它使得 $U_1 a_1 = r_1 e_1$. 定理 2.1.13 对这样的矩阵给出了一个明显的构造，它或者是一个纯量的酉矩阵，或者是一个纯量的酉矩阵与一个 Householder 矩阵的乘积. 分划

$$U_1 A = \begin{bmatrix} r_1 & \bigstar \\ 0 & A_2 \end{bmatrix}$$

其中 $A_2\in M_{n-1,m-1}$. 设 $a_2\in\mathbf{C}^{n-1}$是 A_2 的第一列，并令 $r_2=\|a_2\|_2$. 再次利用(2.1.13)来构造一个酉矩阵 $V_2\in M_{n-1}$，使得 $V_2 a_2 = r_2 e_1$，再令 $U_2 = I_1\oplus V_2$. 那么

$$U_2 U_1 A = \begin{bmatrix} r_1 & & \bigstar \\ 0 & r_2 & \\ 0 & 0 & A_3 \end{bmatrix}$$

重复这一结构 m 次就得到 89

$$U_m U_{m-1}\cdots U_2 U_1 A = \begin{bmatrix} R \\ 0 \end{bmatrix}$$

其中 $R\in M_m$ 是上三角的，其主对角线元素是 r_1，…，r_m，它们全都是非负的. 设 $U=U_m U_{m-1}\cdots U_2 U_1$. 分划 $U^* = U_1^* U_2^* \cdots U_{m-1}^* U_m^* = [Q \quad Q_2]$，其中 $Q\in M_{n,m}$ 的列是标准正交的(它包含了一个酉矩阵的前 m 个列). 这样就有 $A=QR$，如所希望的那样. 如果 A 是列满秩的，则 R 是非奇异的，所以它的主对角线元素全是正的.

假设 $\operatorname{rank}A=m$，且 $A=QR=\tilde{Q}\tilde{R}$，其中 R 与$\tilde{R}$是上三角的且有正的主对角线元素，而 Q 与$\tilde{Q}$都有标准正交的列向量. 那么 $A^*A=R^*(Q^*Q)R=R^*IR=R^*R$，且还有 $A^*A=\tilde{R}^*\tilde{R}$，所以 $R^*R=\tilde{R}^*\tilde{R}$且$\tilde{R}^{-*}R^*=\tilde{R}R^{-1}$. 这就是说下三角矩阵等于一个上三角矩阵，所以它们两者必定都是对角矩阵：$\tilde{R}R^{-1}=D$ 是对角的，且它必定有正的主对角线元素，这

是因为$\widetilde{R}$与R^{-1}这两者的主对角线元素都是正的．但是$\widetilde{R}=DR$蕴含$D=\widetilde{R}R^{-1}=\widetilde{R}^{-*}R^{*}=(DR)^{-*}R^{*}=D^{-1}R^{-*}R^{*}=D^{-1}$，所以$D^2=I$，从而$D=I$．我们断言有$\widetilde{R}=R$以及$\widetilde{Q}=Q$.

(c)中的结论由列向量标准正交的方阵是酉矩阵这一事实推出．

如果在(d)中有$n\geqslant m$，我们可以从(a)中的分解开始，设$\widetilde{Q}=[Q\quad Q_2]\in M_n$是酉矩阵，令$\widetilde{R}=\begin{bmatrix}R\\0\end{bmatrix}\in M_{n,m}$，并注意到$A=QR=\widetilde{Q}\widetilde{R}$．如果$n<m$，我们可以采用(a)中的构造(用Householder变换的一列纯量倍数左乘)并在n步后停止，这时就得到分解式$U_n\cdots U_1A=[R\quad \bigstar]$，而$R$是上三角的．★这个块中的元素不一定为零.

最后的结论(e)从(2.1.13)中的如下结论推出：(a)与(d)的结构中所包含的酉矩阵U_i可以全部取为实矩阵．□

习题 证明：任何形如$B=A^*A$的$B\in M_n(A\in M_n)$可以写成$B=LL^*$，其中$L\in M_n$是下三角的，且有非负的对角元素．说明为什么这个分解是唯一的，如果A是非奇异的．这是B的 **Cholesky 分解**，每一个正定的或半正定的矩阵都可以用此种方式进行分解，见(7.2.9).

$A\in M_{n,m}$的QR分解的某些简单的变量可能是有用的．首先假设$n\leqslant m$，并令$A^*=QR$，其中$Q\in M_{n,m}$有标准正交的列，而$R\in M_m$是上三角的．这样，$A=R^*Q^*$就是形如

$$A=LQ \tag{2.1.15a}$$

的一个分解，其中$Q\in M_{n,m}$有标准正交的行，且$L\in M_n$是下三角的．如果$\widetilde{Q}=\begin{bmatrix}Q\\\widetilde{Q}_2\end{bmatrix}$是酉矩阵，我们就有形如

$$A=[L\quad 0]\,\widetilde{Q} \tag{2.1.15b}$$

的分解.

现在设K_p是(实正交以及对称的)$p\times p$反序矩阵(0.9.5.1)，它有令人愉悦的性质$K_p^2=I_p$．对方阵$R\in M_p$，如果R是上三角的，那么矩阵$L=K_pRK_p$是下三角的；L的主

90 对角元素就是R的主对角元素，只不过次序相反而已.

如果像在(2.1.14a)中那样有$n\geqslant m$以及$AK_m=QR$，那么$A=(QK_m)(K_mRK_m)$，这是这种形式的一个分解(其中$Q\in M_n$是酉矩阵，而$R\in M_n$是上三角矩阵)，那么

$$A=QL \tag{2.1.17a}$$

其中$Q\in M_{n,m}$有标准正交的列，而$L\in M_m$是下三角的．如果$\widetilde{Q}=[Q\quad Q_2]$是酉矩阵，我们就有形如

$$A=\widetilde{Q}\begin{bmatrix}L\\0\end{bmatrix} \tag{2.1.17b}$$

的分解.

如果$n\leqslant m$，我们可以将(2.1.17a)以及(2.1.17b)应用于A^*，就得到形如

$$A=RQ=[R\quad 0]\,\widetilde{Q} \tag{2.1.17c}$$

的分解，其中$R\in M_n$是上三角的，$Q\in M_{n,m}$有标准正交的行，且$\widetilde{Q}\in M_m$是酉矩阵．如果$n\leqslant m$，且将(2.1.14d)应用于AK_m，我们就得到$A=(QK_n)(K_n[R\quad \bigstar]K_m)$，这是形如

$$A = \tilde{Q} L \tag{2.1.17d}$$

的分解，其中$\tilde{Q}\in M_n$ 是酉矩阵，而 $L\in M_{n,m}$则是下三角的.

一个重要的几何事实是：任何两个有相同个数的标准正交向量组都通过酉变换联系在一起.

定理 2.1.18 如果 $X=[x_1\ \cdots\ x_k]\in M_{n,k}$ 与 $Y=[y_1\ \cdots\ y_k]\in M_{n,k}$ 有标准正交的列，那么存在一个酉矩阵 $U\in M_n$，使得 $Y=UX$. 如果 X 与 Y 是实的，那么 U 可以取为实的.

证明 将标准正交向量组 x_1，…，x_k 与 y_1，…，y_k 中的每一个都扩充成为 $\mathbf{C}^n$ 的一组标准正交基，见(0.6.4)和(0.6.5). 这也就是构造酉矩阵 $V=[X\quad X_2]$以及 $W=[Y\quad Y_2]\in M_n$. 那么 $U=WV^*$ 是酉矩阵，且$[Y\quad Y_2]=W=UV=[UX\quad UX_2]$，所以 $Y=UX$. 如果 X 与 Y 是实的，则矩阵$[X\quad X_2]$与$[Y\quad Y_2]$可以选为实的正交矩阵(它们的列是 $\mathbf{R}^n$ 的标准正交基). □

问题

2.1.P1 如果 $U\in M_n$ 是酉矩阵，证明 $\det|U|=1$.

2.1.P2 如果 $U\in M_n$ 是酉矩阵，又令 λ 是 U 的一个给定的特征值. 证明：(a) $|\lambda|=1$ 以及(b) x 是 U 的与 λ 相伴的(右)特征向量，当且仅当 x 是 U 的与 λ 相伴的左特征向量.

2.1.P3 给定实参数 θ_1，θ_2，…，θ_n，证明 $U=\mathrm{diag}(e^{i\theta_1}, e^{i\theta_2}, \cdots, e^{i\theta_n})$是酉矩阵. 证明每一个对角酉矩阵都有此形状. 91

2.1.P4 刻画实对角正交矩阵的特征.

2.1.P5 证明：M_n 中的置换矩阵(0.9.5)是实正交矩阵群的一个子群(子群指的是本身也作成群的子集合). M_n 中有多少不同的置换矩阵?

2.1.P6 给出 3×3 正交群的一个参数表示. 2×2 正交群的两个表示给出在接在(2.1.5)后面的习题中.

2.1.P7 假设 A，$B\in M_n$ 且 $AB=I$. 对下面关于 $BA=I$ 的论证提供细节：每个 $y\in\mathbf{C}^n$ 都可以表示成 $y=A(By)$，所以 $\mathrm{rank}A=n$，因此 $\dim(\mathrm{nullspace}(A))=0$(0.2.3.1). 计算$A(AB-BA)=A(I-BA)=A-(AB)A=A-A=0$，所以 $AB-BA=0$.

2.1.P8 矩阵 $A\in M_n$ 是**复正交的**(complex orthogonal)，如果 $A^{\mathrm{T}}A=I$. (a)证明：复正交矩阵是酉矩阵，当且仅当它是实的. (b)设 $S=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}\in M_2(\mathbf{R})$. 证明 $A(t)=(\cosh t)I+(i\sinh t)S\in M_2$ 对所有 $t\in\mathbf{R}$ 都是复正交矩阵，但是 $A(t)$仅当 $t=0$ 时是酉矩阵. 双曲函数由 $\cosh t=(e^t+e^{-t})/2$，$\sinh t=(e^t-e^{-t})/2$ 来定义. (c)证明，与酉矩阵不同，复正交矩阵的集合不是一个有界的集合，因此它也就不是紧集. (d)证明：给定阶的复正交矩阵的集合作成一个群. 给定阶的实正交矩阵组成的更小的(以及紧的)群常称为**正交群**(orthogonal group). (e)如果 $A\in M_n$ 是复正交的，证明$|\det A|=1$，考虑(b)中的 $A(t)$来证明 A 可以有$|\lambda|\neq 1$ 的特征值 λ. (f)如果 $A\in M_n$ 是复正交矩阵，证明$\bar{A}$，A^{T} 以及 A^* 全都是复正交的，且都是非奇异的. A 的行(列)是正交的吗? (g)刻画复对角正交矩阵的特征. 与 2.1.P4 作比较. (h)证明 $A\in M_n$ 既是复正交矩阵，也是酉矩阵，当且仅当它是实正交矩阵.

2.1.P9 如果 $U\in M_n$ 是酉矩阵，证明$\bar{U}$，U^{T} 以及 U^* 全都是酉矩阵.

2.1.P10 如果 $U\in M_n$ 是酉矩阵，证明 x，$y\in\mathbf{C}^n$ 是正交的，当且仅当 Ux 与 Uy 是正交的.

2.1.P11 非奇异的矩阵 $A\in M_n$ 是**斜正交的**(skew orthogonal)，如果 $A^{-1}=-A^{\mathrm{T}}$. 证明：A 是斜正交的，当且仅当$\pm iA$ 是正交的. 更一般地，如果 $\theta\in\mathbf{R}$，证明：$A^{-1}=e^{i\theta}A^{\mathrm{T}}$ 当且仅当 $e^{i\theta/2}A$ 是正交的. 对 $\theta=0$ 以及 π 结论如何?

2.1.P12 证明：如果 $A\in M_n$ 与一个酉矩阵相似，那么 A^{-1} 与 A^* 相似.

2.1.P13 考虑 $\mathrm{diag}\left(2,\ \frac{1}{2}\right)\in M_2$ 并证明：与酉矩阵相似的矩阵集合是使得 A^{-1} 与 A^* 相似的那种矩阵 A 组成的集合的一个真子集.

2.1.P14 证明 M_n 中的酉矩阵群与 M_n 中的复正交矩阵群的交是 M_n 中的实正交矩阵群.

2.1.P15 如果 $U\in M_n$ 是酉矩阵，$\alpha\subset\{1,\ \cdots,\ n\}$，且 $U[\alpha\quad\alpha^c]=0$，(0.7.1)表明 $U[\alpha^c\quad\alpha]=0$，且 $U[\alpha]$与 $U[\alpha^c]$是酉矩阵.

2.1.P16 设 $x,\ y\in\mathbf{R}^n$ 是给定的线性无关的单位向量，令 $w=x+y$. 考虑 Palais 矩阵 $P_{x,y}=I-2(w^{\mathrm{T}}w)^{-1}ww^{\mathrm{T}}+2yx^{\mathrm{T}}$. 证明：(a)$P_{x,y}=(I-2(w^{\mathrm{T}}w)^{-1}ww^{\mathrm{T}})(I-2xx^{\mathrm{T}})=U_wU_x$ 是两个实 Householder 矩阵的乘积，所以它是实正交矩阵；(b)$\det P_{x,y}=+1$，所以 $P_{x,y}$ 总是真旋转矩阵；(c)$P_{x,y}x=y$ 以及 $P_{x,y}y=-x+2(x^{\mathrm{T}}y)y$；(d)$P_{x,y}z=z$，如果 $z\in\mathbf{R}^n$，$z\perp x$ 以及 $z\perp y$；(e)$P_{x,y}$ 在$(n-2)$维子空间$(\mathrm{span}\{x,\ y\})^{\perp}$上的作用像是恒等元，且它是在二维子空间 $\mathrm{span}\{x,\ y\}$上将 x 变成 y 的真旋转；(f)如果$n=3$，说明为什么 $P_{x,y}$是将 x 变成 y 且保持向量的叉积 $x\times y$ 不变的唯一的真旋转；(g)$P_{x,y}$的特征值是 $x^{\mathrm{T}}y\pm i(1-(x^{\mathrm{T}}y)^2)^{1/2}=\mathrm{e}^{\pm i\theta}$，1，…，1，其中 $\cos\theta=x^{\mathrm{T}}y$.

2.1.P17 假设 $A\in M_{n,m}$，$n\geqslant m$，且 $\mathrm{rank}A=m$. 描述将 Gram-Schmidt 方法应用到 A 的列时的步骤(从左向右进行). 说明为什么逐列往下做时这个程序会产生出一个具有标准正交列的显式矩阵 $Q\in M_{n,m}$以及一个上三角矩阵 $R\in M_m$，使得 $Q=AR$. 这一分解与(2.1.14)中的分解有怎样的联系?

2.1.P18 设 $A\in M_n$ 如同在(2.1.14)中那样被分解成 $A=QR$，按照列来分划 $A=[a_1\quad\cdots\quad a_n]$以及 $Q=[q_1\quad\cdots\quad q_n]$，又设 $R=[r_{ij}]_{i,j=1}^n$. (a)说明为什么对每个 $k=1,\ \cdots,\ n$，$\{q_1,\ \cdots,\ q_k\}$都是 $\mathrm{span}\{a_1,\ \cdots,\ a_k\}$的一组标准正交基. (b)证明：对每个 $k=2,\ \cdots,\ n$，r_{kk} 都是 a_k 到 $\mathrm{span}\{a_1,\ \cdots,\ a_{k-1}\}$的 Euclid 距离.

2.1.P19 设 $X=[x_1\quad\cdots\quad x_m]\in M_{n,m}$，假设 $\mathrm{rank}X=m$，且如在(2.1.14)中那样分解 $X=QR$. 设 $Y=QR^{-*}=[y_1\quad\cdots\quad y_m]$. (a)证明 Y 的列是子空间$\mathcal{S}=\mathrm{span}\{x_1,\ \cdots,\ x_m\}$的一组基，且 $Y^*X=I_m$，所以 $y_i^*x_j=0$(如果 $i\neq j$)且每一个 $y_i^*x_i=1$. 给定$\mathcal{S}$的基 $x_1,\ \cdots,\ x_m$，它的**对偶基**(dual basis)(有时称为 reciprocal basis)就是 $y_1,\ \cdots,\ y_m$. (b)说明为什么 $x_1,\ \cdots,\ x_m$ 的对偶基是唯一的，也就是说，如果 $Z\in M_{n,m}$的列在$\mathcal{S}$中，且 $Z^*X=I$，那么 $Z=Y$. (c)证明向量组 $y_1,\ \cdots,\ y_m$ 的对偶基就是 $x_1,\ \cdots,\ x_m$. (d)如果 $n=m$，证明 X^{-*} 的列是 $\mathbf{C}^n$ 的一组与基 $x_1,\ \cdots,\ x_n$ 对偶的基.

2.1.P20 如果 $U\in M_n$ 是酉矩阵，证明 $\mathrm{adj}U=(\det U)U^*$，并断言 $\mathrm{adj}U$ 是酉矩阵.

2.1.P21 说明：如果用**复正交矩阵**代替**酉矩阵**，为什么(2.1.10)依然为真. 导出结论：复正交矩阵是上三角的，当且仅当它是对角的. 复的对角正交矩阵看起来像什么样子?

2.1.P22 假设 $X,\ Y\in M_{n,m}$有标准正交的列. 证明：X 与 Y 有相同的值域(列空间)，当且仅当存在一个酉矩阵 $U\in M_m$，使得 $X=YU$.

2.1.P23 设 $A\in M_n$，令 $A=QR$ 是 QR 分解，设 $R=[r_{ij}]$，又按照它们的列来分划 A，Q 以及 R：$A=[a_1\quad\cdots\quad a_n]$，$Q=[q_1\quad\cdots\quad q_n]$，$R=[r_1\quad\cdots\quad r_n]$. 说明为什么$|\det A|=\det R=r_{11}\cdots r_{nn}$以及为什么$\|a_i\|_2=\|r_i\|_2\geqslant r_{ii}$(对每个 $i=1,\ \cdots,\ n$)，其中等式对某个 i 成立，当且仅当 $a_i=r_{ii}q_i$. 导出结论 $|\det A|\leqslant\prod_{i=1}^{n}\|a_i\|_2$，其中等式成立，当且仅当要么(a)$a_i=0$，要么(b)$A$ 的列向量正交(即 $A^*A=\mathrm{diag}(\|a_1\|_2^2,\ \cdots,\ \|a_n\|_2^2)$). 这就是 Hadamard 不等式.

2.1.P24 设 $E=[e_{ij}]\in M_3$，其中每个 $e_{ij}=+1$. (a)证明：E 的积和式(0.3.2)是 $\mathrm{per}E=6$. (b)设 $B=[b_{ij}]\in M_3$，其中每个 $b_{ij}=\pm1$. 利用 Hadamard 不等式证明：不存在正负号的选取方式，能使得有 $\mathrm{per}E=\det B$.

2.1.P25 如果 $U\in M_n$ 是酉矩阵，且 $r\in\{1,\ \cdots,\ n\}$，说明为什么复合矩阵 $C_r(U)$是酉矩阵.

2.1.P26 说明为什么(a)每一个 $A\in M_n$ 都可以分解成 $A=H_1\cdots H_{n-1}R$，其中每个 H_i 是 Householder 矩阵，而 R 是上三角矩阵；(b)每一个酉矩阵 $U\in M_n$ 可以分解成 $U=H_1\cdots H_{n-1}D$，其中每个 H_i 是 Householder 矩阵，而 D 是对角酉矩阵；(c)每一个实正交矩阵 $Q\in M_n(\mathbf{R})$ 可以分解成 $Q=H_1\cdots H_{n-1}D$，其中每个 H_i 是实的 Householder 矩阵，而 $D=\mathrm{diag}(1,\cdots,1,\pm 1)=\mathrm{diag}(1,\cdots,1,(-1)^{n-1},\det Q)$.

下面三个问题给出上一问题的类似结果，在这三个结果中用平面旋转代替了 Householder 矩阵.

2.1.P27 设 $n\geqslant 2$ 以及 $x=[x_i]\in\mathbf{R}^n$，如果 $x_n=x_{n-1}=0$，令 $\theta_1=0$；反之则选取 $\theta_1\in[0,2\pi)$，使得 $\cos\theta_1=x_{n-1}/\sqrt{x_n^2+x_{n-1}^2}$ 以及 $\sin\theta_1=-x_n/\sqrt{x_n^2+x_{n-1}^2}$. 设 $x^{(1)}=[x_i^{(1)}]=U(\theta_1;n-1,n)x$. 证明 $x_n^{(1)}=0$ 以及 $x_{n-1}^{(1)}\geqslant 0$. 令 $x^{(2)}=[x_i^{(2)}]=U(\theta_2;n-2,n-1)U(\theta_1;n-1,n)x$. 你如何选取 θ_2，使得 $x_n^{(2)}=x_{n-1}^{(2)}=0$ 以及 $x_{n-2}^{(2)}\geqslant 0$? 如果 $1\leqslant k<n$，说明如何来构造 k 个平面旋转组成的序列 $U_1,\cdots,U_k$，使得向量 $x^{(k)}=[x_i^{(k)}]=U_k,\cdots,U_1x$ 满足 $x_n^{(k)}=\cdots=x_{n-k+1}^{(k)}=0$ 以及 $x_{n-k}^{(k)}\geqslant 0$. 为什么有 $\|x\|_2=\|x^{(k)}\|_2$?

2.1.P28 设 $A\in M_{n,m}(\mathbf{R})$，其中 $n\geqslant m$. (a)说明怎样构造平面旋转的一个有限序列 $U_1,\cdots,U_N$，使得 $U_N\cdots U_1A=\begin{bmatrix}B\\0\end{bmatrix}$，其中 $B=[b_{ij}]\in M_m(\mathbf{R})$ 是上三角的，而 $b_{11},\cdots,b_{m-1,m-1}$ 中的每一个都是非负的. (b)说明为什么向上三角型的这种化简可以通过一列 $N=m\left(n-\frac{m+1}{2}\right)$ 个平面旋转来达到. 这些平面旋转中有一些可能是恒等旋转，所以有可能只需要少于 N 个非平凡的平面旋转. (c)利用(a)来证明：每一个 $A\in M_n(\mathbf{R})$ 都可以分解成 $A=U_1,\cdots,U_NR$，其中 $N=n(n-1)/2$，每一个 U_i 是一个平面旋转，$R=[r_{ij}]$ 是上三角的，而每一个 $r_{11},\cdots,r_{n-1,n-1}$(但 r_{nn} 不一定)都是非负的.

2.1.P29 说明为什么每个实正交矩阵 $Q\in M_n(\mathbf{R})$ 都可以分解成 $Q=U_1\cdots U_ND$，其中 $N=n(n-1)/2$，每个 U_i 是一个平面旋转，而 $D=\mathrm{diag}(1,\cdots,1,\det Q)=\mathrm{diag}(1,\cdots,1,\pm 1)\in M_n(\mathbf{R})$.

进一步的阅读参考 有关满足(2.1.9)条件的矩阵的更多的信息，请见 C. R. DePrima 以及 C. R. Johnson, The range of $A^{-1}A^*$ in $GL(n,\mathbf{C})$, *Linear Algebra Appl.* 9(1974)209-222.

2.2 酉相似

由于对酉矩阵 U 有 $U^*=U^{-1}$，故而 M_n 上由 $A\to U^*AU$ 给出的变换是相似变换，如果 U 是酉矩阵. 这种特殊类型的相似称为**酉相似**(unitary similarity).

定义 2.2.1 设给定 $A,B\in M_n$. 我们称 A 与 B **酉相似**(unitary similar)，如果存在一个酉矩阵 $U\in M_n$，使得 $A=UBU^*$. 如果 U 可以取为实的(从而它是实正交的)，那么就说 A 与 B **实正交相似**(real orthogonally similar). 我们称 A 可以**酉对角化**(unitarily diagonalizable)，如果它与一个对角矩阵酉相似；我们称 A 可以**实正交对角化**(real orthogonally diagonalizable)，如果它与一个对角矩阵实正交相似.

习题 证明酉相似是一个等价关系. ◀

定理 2.2.2 设 $U\in M_n$ 与 $V\in M_m$ 是酉矩阵. 令 $A=[a_{ij}]\in M_{n,m}$ 以及 $B=[b_{ij}]\in M_{n,m}$，又假设 $A=UBV$. 那么 $\sum_{i,j=1}^{n,m}|b_{ij}|^2=\sum_{i,j=1}^{n,m}|a_{ij}|^2$. 特别地，这个恒等式当 $m=n$ 以及 $V=U^*$ 时，即当 A 与 B 酉相似时是满足的.

证明 只要验证 $\mathrm{tr}B^*B=\mathrm{tr}A^*A$ 即可，见(0.2.5). 计算给出 $\mathrm{tr}A^*A=\mathrm{tr}((UBV)^*$

94 $(UBV))=\mathrm{tr}(V^*B^*U^*UBV))=\mathrm{tr}V^*B^*BV=\mathrm{tr}B^*BVV^*=\mathrm{tr}B^*B$. □

习题 证明：矩阵$\begin{bmatrix}3 & 1\\ -2 & 0\end{bmatrix}$与$\begin{bmatrix}1 & 1\\ 0 & 2\end{bmatrix}$相似，但并非是酉相似.

酉相似蕴含相似，但反之不然. 酉相似这个等价关系将 M_n 分划成比相似这个等价关系**更精细的**等价类. 与相似类同的是，酉相似对应于基的改变，不过是特殊类型的——酉相似对应的是从一组**标准正交**基到另一组标准正交基的改变. ◀

习题 利用(2.1.11)的记号，说明为什么在实正交相似下，通过平面旋转 $U(\theta;i,j)$ 改变的仅仅是标号为 i 与 j 的行与列. ◀

习题 利用(2.1.13)的记号，说明为什么对任何 $A\in M_n$ 有 $U(y,x)^*AU(y,x)=U_w^*AU_w$，这就是说，通过形如 $U(y,x)$ 的本性 Hermite 酉矩阵给出的酉相似就是通过 Householder 矩阵给出的酉相似. 通过 Householder 矩阵给出的酉(或者实正交)相似常称为 **Householder 变换**. ◀

鉴于计算或者理论的原因，通过酉相似将一个给定的矩阵变换成另一个特殊形式的矩阵，常会带来方便. 这里给出两个例子.

例 2.2.3(酉相似于对角元素相等的矩阵) 设给定 $A=[a_{ij}]\in M_n$. 我们断言存在一个酉矩阵 $U\in M_n$，使得 $UAU^*=B=[b_{ij}]$的所有主对角元素都相等，如果 A 是实的，则 U 可以取为实正交的. 如果这一断言为真，那么 $\mathrm{tr}A=\mathrm{tr}B=nb_{11}$，所以 B 的每个主对角元素都等于 A 的主对角元素之平均值.

首先考虑复的情形以及 $n=2$. 由于我们可以用 $A-\left(\frac{1}{2}\mathrm{tr}A\right)I$ 代替 $A\in M_2$，故而不失一般性，我们假设 $\mathrm{tr}A=0$，在此情形 A 的两个特征值是$\pm\lambda$(对某个 $\lambda\in\mathbf{C}$). 我们希望确定一个单位向量 u 使得 $u^*Au=0$. 如果 $\lambda=0$，设 u 是任意一个使得 $Au=0$ 的单位向量. 如果 $\lambda\neq0$，令 w 与 z 是与不同的特征值$\pm\lambda$ 相伴的任意单位特征向量. 设 $x(\theta)=e^{i\theta}w+z$，它对所有 $\theta\in\mathbf{R}$ 都不等于零，这是因为 w 与 z 是线性无关的. 计算 $x(\theta)^*Ax(\theta)=\lambda(e^{i\theta}w+z)^*(e^{i\theta}w-z)=2i\lambda\mathrm{Im}(e^{i\theta}z^*w)$. 如果 $z^*w=e^{i\phi}|z^*w|$，那么 $x(-\phi)^*Ax(-\phi)=0$. 设 $u=x(-\phi)/\|x(-\phi)\|_2$. 现在设 $v\in\mathbf{C}^2$ 是任意一个与 u 正交的单位向量，并令 $U=[u\quad v]$. 那么 U 是酉矩阵，且$(U^*AU)_{11}=u^*Au=0$. 但是 $\mathrm{tr}(U^*AU)=0$，所以也有$(U^*AU)_{22}=0$.

现在假设 $n=2$，且 A 是实的. 如果 $A=[a_{ij}]$的对角元素不相等，考虑平面旋转矩阵 $U_\theta=\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix}$. 计算发现 $U_\theta AU_\theta^{\mathrm{T}}$ 的对角元素相等，如果有$(\cos^2\theta-\sin^2\theta)(a_{11}-a_{22})=2\sin\theta\cos\theta(a_{12}+a_{21})$，所以，如果可以选取到 $\theta\in(0,\pi/2)$使得 $\cot2\theta=(a_{12}+a_{21})/(a_{11}-a_{22})$，那么就可以做到使其对角元素均相等.

我们现在已经指出了：任何 2×2 复矩阵 A 都酉相似于一个两个对角线元素都等于 A
95 的对角元素之平均值的矩阵，如果 A 是实的，则相似可以取为实正交的.

现在假设 $n>2$，并定义 $f(A)=\max\{|a_{ii}-a_{jj}|: i,j=1,2,\cdots,n\}$. 如果 $f(A)>0$，就对满足 $f(A)=|a_{ii}-a_{jj}|$的一对指标 i，j，令 $A_2=\begin{bmatrix}a_{ii} & a_{ij}\\ a_{ji} & a_{jj}\end{bmatrix}$(可能存在若干对指标都能取到这个最大的正的差值，这时就任取其中的一对). 设 $U_2\in M_2$ 是酉矩阵，当 A 是实

的时它是实的，且使得 $U_2^* A_2 U_2$ 的两个主对角线元素都等于 $\frac{1}{2}(a_{ii}+a_{jj})$. 以在(2.1.11)中从 2×2 平面旋转构造出 $U(\theta;\ i,\ j)$的同样的方式从 U_2 构造出 $U(i,\ j)\in M_n$. 酉相似 $U(i,\ j)^* AU(i,\ j)$只影响到行与列在 i 与 j 的元素，所以它保持 A 的每一个主对角线元素不变，除非该元素的位置在 i 与 j 处，这样的元素以平均值 $\frac{1}{2}(a_{ii}+a_{jj})$取代之. 对任何 $k\neq i,\ j$，三角不等式确保有

$$\begin{aligned}\left|a_{kk}-\frac{1}{2}(a_{ii}+a_{jj})\right| &= \left|\frac{1}{2}(a_{kk}-a_{ii})+\frac{1}{2}(a_{kk}-a_{jj})\right| \\ &\leqslant \frac{1}{2}\left|a_{kk}-a_{ii}\right|+\frac{1}{2}\left|a_{kk}-a_{jj}\right| \leqslant \frac{1}{2}f(A)+\frac{1}{2}f(A)=f(A)\end{aligned}$$

其中的等式仅当纯量 $a_{kk}-a_{ii}$ 与 $a_{kk}-a_{jj}$ 两者都位于复平面的同一条射线上且 $|a_{kk}-a_{ii}|=|a_{kk}-a_{jj}|$时才成立. 这两个条件蕴含 $a_{ii}=a_{jj}$，故而由此推出：对所有 $k\neq i,\ j$ 有 $|a_{kk}-\frac{1}{2}(a_{ii}+a_{jj})|<f(A)$. 这样一来，我们刚刚构造出来的酉相似矩阵中满足 $f(A)=|a_{kk}-a_{\ell\ell}|$的有限多对指标对 $k,\ \ell$ 就会减少一对. 如果需要，就重复这一做法，以此来处理剩下来的任何指标对并得到一个酉矩阵 U(如果 A 是实的，它还是实的)，使得 $f(U^* AU)<f(A)$.

最后，考虑紧集 $R(A)=\{U^* AU:\ U\in M_n$ 是酉矩阵$\}$. 由于 f 是 $R(A)$上一个非负值的连续函数，它在其中取得最小值，也就是说，存在某个 $B\in R(A)$，使得对所有 $A\in R(A)$都有 $f(A)\geqslant f(B)\geqslant 0$. 如果 $f(B)>0$，我们就恰好看到存在一个酉矩阵 U(如果 A 是实的，它也是实的)，使得 $f(B)>f(U^* AU)$. 这个矛盾表明 $f(B)=0$，故而 B 的所有对角元素都相等.

例 2.2.4(与上 Hessenberg 矩阵酉相似) 设给定 $A=[a_{ij}]\in M_n$. 下面的构造表明 A 与一个第一条次对角线元素非负的上 Hessenberg 矩阵是酉相似的. 设 a_1 是 A 的第一列，它被分划成 $a_1^{\mathrm{T}}=[a_{11}\quad \xi^{\mathrm{T}}]$，其中 $\xi\in\mathbf{C}^{n-1}$. 如果 $\xi=0$，就令 $U_1=I_{n-1}$；反之，就利用(2.1.13)来构造 $U_1=U(\|\xi\|_2 e_1,\ \xi)\in M_{n-1}$，它是将 ξ 变成 e_1 的正倍数的酉矩阵. 构造酉矩阵 $V_1=I_1\oplus U_1$ 并注意到 V_1A 的第一列是向量$[a_{11}\quad \|\xi\|_2\quad 0]^{\mathrm{T}}$. 此外，$A_1=(V_1A)V_1^*$与 V_1A 的第一列相同，且与 A 酉相似. 将它分划成

$$\mathcal{A}_1=\begin{bmatrix} a_{11} & \bigstar \\ \begin{bmatrix}\|\xi\|_2 \\ 0\end{bmatrix} & A_2\end{bmatrix},\qquad A_2\in M_{n-1}$$ 96

按照同样的方式，再次利用(2.1.13)作出一个酉矩阵 U_2，它将 A_2 的第一列变成第二个元素之下所有元素皆为零而第二个元素为非负数的向量. 设 $V_2=I_2\oplus U_2$，并令 $\mathcal{A}_2=V_2AV_2^*$. 这个相似并不影响 $\mathcal{A}_1$ 的第一列. 经过至多 $n-1$ 步，这个构造就产生出一个上 Hessenberg 矩阵 $\mathcal{A}_{n-1}$，它与 A 酉相似，且次对角线元素非负.

习题 如果 A 是 Hermite 矩阵或者斜 Hermite 矩阵，说明为什么上面例子中的构造产生出一个与 A 酉相似的三对角的 Hermite 矩阵或者三对角的斜 Hermite 矩阵. ◀

定理 2.2.2 对于两个给定的矩阵是否为酉相似提供了一个必要但非充分的条件. 它可以增补一些附加的恒等式以共同给出必要且充分的条件. 如下的简单概念起着关键的作

用. 设 s, t 是两个给定的非交换变量. s 与 t 的非负幂组成的任何有限的形式乘积

$$W(s,t) = s^{m_1}t^{n_1}s^{m_2}t^{n_2}\cdots s^{m_k}t^{n_k}, \quad m_1,n_1,\cdots,m_k,n_k \geqslant 0 \tag{2.2.5}$$

称为一个**关于 s 与 t 的字**(word in s and t). 字 $W(s, t)$的**长度**(length)是非负整数 $m_1+n_1+m_2+n_2+\cdots+m_k+n_k$, 即这个字中的所有指数之和. 如果给定 $A\in M_n$, 我们将**关于 A 与 A^* 的字**(word in A and A^*)定义为

$$W(A,A^*) = A^{m_1}(A^*)^{n_1}A^{m_2}(A^*)^{n_2}\cdots A^{m_k}(A^*)^{n_k}$$

由于 A 与 A^* 的幂不一定可交换, 因而有可能无法通过在乘积中重新排列各个项来简化 $W(A, A^*)$的表达式.

假设 A 与 $B\in M_n$ 酉相似, 也就是说, 对某个酉矩阵 $U\in M_n$ 有 $A=UBU^*$. 对任何字 $W(s, t)$我们有

$$\begin{aligned}W(A,A^*) &= (UBU^*)^{m_1}(UB^*U^*)^{n_1}\cdots(UBU^*)^{m_k}(UB^*U^*)^{n_k}\\ &= UB^{m_1}U^*U(B^*)^{n_1}U^*\cdots UB^{m_k}U^*U(B^*)^{n_k}U^*\\ &= UB^{m_1}(B^*)^{n_1}\cdots B^{m_k}(B^*)^{n_k}U^* = UW(B,B^*)U^*\end{aligned}$$

故而 $W(A, A^*)$与 $W(B, B^*)$酉相似. 从而有 $\mathrm{tr}W(A, A^*)=\mathrm{tr}W(B, B^*)$. 如果取字 $W(s, t)=ts$, 我们就得到(2.2.2)中的恒等式.

如果考虑所有可能的字 $W(s, t)$, 这种观察到的结果就会对两个矩阵的酉相似给出无穷多个必要条件. 我们将不加证明地陈述 W. Specht 的一个定理, 此定理保证了这些必要条件也是充分的.

定理 2.2.6 两个矩阵 A, $B\in M_n$ 是酉相似的, 当且仅当对关于两个非交换变量的每一个字 $W(s, t)$都有

$$\mathrm{tr}W(A,A^*) = \mathrm{tr}W(B,B^*) \tag{2.2.7}$$

通过给出一个违反(2.2.7)的特殊的字, Specht 定理可以用来证明两个矩阵不是酉相似的. 然而, 除了在特殊情形(见 2.2.P6), 它在证明两个给定的矩阵是酉相似的这一点上
97 可能没什么用, 这是因为必须验证的条件有无穷多个. 幸运的是, Specht 定理的一个改进的形式确保只需对有限多个字查验迹恒等式(2.2.7)就够了, 这就提供了一个实际可操作的判别法则来对阶比较小的矩阵的酉相似性进行判别.

定理 2.2.8 设给定 A, $B\in M_n$.

(a) A 与 B 是酉相似的, 当且仅当对关于两个非交换变量的长度至多为

$$n\sqrt{\frac{2n^2}{n-1}+\frac{1}{4}}+\frac{n}{2}-2$$

的每一个字 $W(s, t)$, (2.2.7)都是满足的.

(b) 如果 $n=2$, 则 A 与 B 是酉相似的, 当且仅当(2.2.7)对三个字 $W(s, t)=s$; s^2 以及 st 是满足的.

(c) 如果 $n=3$, 则 A 与 B 是酉相似的, 当且仅当(2.2.7)对七个字 $W(s, t)=s$; s^2, st; s^3, s^2t; s^2t^2; 以及 s^2t^2st 是满足的.

(d) 如果 $n=4$, A 与 B 是酉相似的, 当且仅当(2.2.7)对如下表中的 20 个字 $W(s, t)$是满足的:

s	s^2, st
s^3, s^2t	$s^4, s^3t, s^2t^2, stst$
s^3t^2	$s^2ts^2t, s^2t^2st, t^2s^2ts$
s^3t^2st	$s^3t^2s^2t, s^3t^3st, t^3s^3ts$
$s^3ts^2tst, s^2t^2sts^2t$	$s^3t^3s^2t^2$

两个实矩阵是酉相似的，当且仅当它们是实正交相似的，见(2.5.21)．于是，(2.2.8)中的判别法对两个实矩阵 A 与 B 实正交相似是必要且充分的．

问题

2.2.P1 设 $A=[a_{ij}]\in M_n(\mathbf{R})$ 是对称的，但不是对角的，又选取指标 i，j 满足 $i<j$，使得 $|a_{ij}|=\max\{|a_{pq}|: p<q\}$．由 $\cot 2\theta=(a_{ii}-a_{jj})/2a_{ij}$ 来定义 θ，令 $U(\theta; i, j)$ 是平面旋转(2.1.11)，又令 $B=U(\theta;i,j)^{\mathrm{T}}AU(\theta;i,j)=[b_{pq}]$．证明 $b_{ij}=0$，$\sum_{p,q=1}^{n}|b_{pq}|^2=\sum_{p,q=1}^{n}|a_{pq}|^2$ 以及

$$\sum_{p\neq q}|b_{pq}|^2=\sum_{p\neq q}|a_{pq}|^2-2|a_{ij}|^2\leqslant\left(1-\frac{2}{n^2-n}\right)\sum_{p\neq q}|a_{pq}|^2$$

说明为什么通过这种方式选取的平面旋转所给出的一列实正交相似矩阵(在每一步做一个平面旋转，它使得对角线之外大小最大的元素变为零)收敛于一个对角矩阵，它的对角元素是 A 的特征值．作为这一过程的一个副产品，怎样才能得到对应的特征向量呢？这就是计算实对称矩阵特征值的 **Jacobi 方法**．在实践中是用一种规避计算任意三角函数或者它们的反函数的算法来实施 Jacobi 方法，见 Golub 与 VanLoan(1996)． 98

2.2.P2 Givens 的计算实矩阵特征值的方法也用到平面旋转，不过是以不同的方式．对 $n\geqslant 3$，对下面的论证提供细节：每个 $A=[a_{ij}]\in M_n(\mathbf{R})$ 都与一个实的下 Hessenberg 矩阵实正交相似，这个实的下 Hessenberg 矩阵必定是三对角的，如果 A 是对称的，见(0.9.9)以及(0.9.10)．如同在上一个问题中那样，选取一个形如 $U(\theta; 1, 3)$ 的平面旋转 $U_{1,3}$，使得 $U_{1,3}^*AU_{1,3}$ 的处于位置(1，3)的元素为零．选取形如 $U_{1,4}=U(\theta; 1, 4)$ 的另一个平面旋转，使得 $U_{1,4}^*(U_{1,3}^*AU_{1,3})U_{1,4}$ 的处于位置(1，4)的元素为零，继续此法直到用一列实正交相似将其第一行的其余元素全部变为零．然后再从第二行的处于位置(2，4)的元素开始，并将处于位置(2，4)，(2，5)，…，(2，n)的元素全部变为零．说明为什么这个过程不会扰乱此前已经得到的元素零，又为什么如果 A 是对称的，它还能保持对称性．继续此程序直至 $n-3$ 行，在经过由平面旋转给出的有限多个实正交相似之后，就产生一个下 Hessenberg 矩阵．此矩阵是三对角的，如果 A 是对称的．然而，A 的特征值不能像在 Jacobi 方法中那样显现出来，它们必须要用进一步的计算才能得到．

2.2.P3 设 $A\in M_2$．(a)证明：对(2.2.8b)中那三个字中的每一个字都有 $\mathrm{tr}W(A, A^*)=\mathrm{tr}W(A^{\mathrm{T}}, \bar{A})$．(b)说明为什么每一个 2×2 复矩阵都与它的转置是酉相似的．

2.2.P4 设 $A\in M_3$．(a)证明：对(2.2.8c)中所列出的前六个字中的每一个字都有 $\mathrm{tr}W(A, A^*)=\mathrm{tr}W(A^{\mathrm{T}}, \bar{A})$，并得出结论：$A$ 与 A^{T} 酉相似，当且仅当 $\mathrm{tr}(A^2(A^{*2})AA^*)=\mathrm{tr}((A^{\mathrm{T}})^2\bar{A}^2A^{\mathrm{T}}\bar{A})$．(b)说明为什么 A 与 A^{T} 酉相似，当且仅当 $\mathrm{tr}(AA^*(A^*A-AA^*)A^*A)=0$．(c)利用(b)或者(c)中的判别法则证明：矩阵

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

与它的转置不是酉相似的．然而注意，每一个复方阵都与它的转置相似(3.2.3)．

2.2.P5 如果 $A\in M_n$，且存在一个酉矩阵 $U\in M_n$ 使得 $A^*=UAU^*$，证明 U 与 $A+A^*$ 可交换．将此结论应用到上一问题中的 3×3 矩阵并得出结论：如果它与自己的转置酉相似，则任何这样的酉相似

矩阵必定都是对角的．证明不存在对角的酉相似矩阵，可以把这个矩阵变成自己的转置，所以它不与自己的转置酉相似．

2.2.P6 设给定 $A\in M_n$ 以及 B，$C\in M_m$．利用(2.2.6)或者(2.2.8)证明：B 与 C 是酉相似的，当且仅当下面诸条件中有任何一个满足．

(a) $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$与$\begin{bmatrix} A & 0 \\ 0 & C \end{bmatrix}$酉相似．

(b) $B\oplus\cdots\oplus B$ 与 $C\oplus\cdots\oplus C$ 酉相似，如果两个直和包含同样多的直和项．

(c) $A\oplus B\oplus\cdots\oplus B$ 与 $A\oplus C\oplus\cdots\oplus C$ 酉相似，如果两个直和包含同样多的直和项．

2.2.P7 给出两个 2×2 矩阵的例子，它们满足恒等式(2.2.2)，但它们并不是酉相似的．说明原因．

99 2.2.P8 设 A，$B\in M_2$ 并令 $C=AB-BA$．利用例 2.2.3 证明：对某个纯量 λ 有 $C^2=\lambda I$．

2.2.P9 设 $A\in M_n$ 并假设 $\mathrm{tr}A=0$．利用(2.2.3)证明：A 可以写成两个幂零矩阵之和．反之，如果 A 可以写成幂零矩阵之和，说明为什么 $\mathrm{tr}A=0$．

2.2.P10 设 $n\geqslant2$ 是给定的整数，并定义 $\omega=e^{2\pi i/n}$．(a)说明为什么 $\sum_{k=0}^{n}\omega^{k\ell}=0$，除非对某个 $m=0$，±1，±2，…有 $\ell=mn$(在此情形，这个和等于 n)．(b)设 $F_n=n^{-1/2}[\omega^{(i-1)(j-1)}]_{i,j=1}^{n}$ 表示 $n\times n$**Fourier 矩阵**．证明 F_n 是对称的酉矩阵，且是共轭对合的：$F_nF_n^*=F_n\overline{F_n}=I$．(c)设 C_n 表示基本的循环置换矩阵(0.9.6.2)．说明为什么 C_n 是酉矩阵(实正交矩阵)．(d)设 $D=\mathrm{diag}(1,\omega,\omega^2,\cdots,\omega^{n-1})$，并证明 $C_nF_n=F_nD$，所以 $C_n=F_nDF_n^*$ 且对所有 $k=1$，2，…都有 $C_n^k=F_nD^kF_n^*$．(e)设 A 表示循环矩阵(0.9.6.1)，它的第一行是$[a_1\ \cdots\ a_n]$，此矩阵表示成(0.9.6.3)的和式．说明为什么 $A=F_n\Lambda F_n^*$，其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，$A$ 的特征值是

$$\lambda_\ell=\sum_{k=0}^{n-1}a_{k+1}\omega^{k(\ell-1)},\quad \ell=1,\cdots,n \tag{2.2.9}$$

而 λ_1，…，λ_n 是向量 $n^{1/2}F_n^*Ae_1$ 的元素．于是，Fourier 矩阵就对每个循环矩阵提供了显明的酉对角化．(f)如果存在某个 $i\in\{1,\cdots,n\}$使得 $|a_i|>\sum_{j\neq i}|a_j|$，从(2.2.9)推导出 A 是非奇异的．我们可以将此判别法则重新表述如下：如果一个循环矩阵是奇异的且第一行是$[a_1\cdots a_n]$，那么那个行向量是**平衡的**(balanced)，见(7.2.P28)．(g)记 $F_n=\mathcal{C}_n+i\mathcal{S}_n$，其中$\mathcal{C}_n$ 与$\mathcal{S}_n$ 都是实的．$\mathcal{C}_n$ 与$\mathcal{S}_n$ 的元素是什么？矩阵 $H_n=\mathcal{C}_n+\mathcal{S}_n$ 是 **Hartley 矩阵**．(h)证明$\mathcal{C}_n^2+\mathcal{S}_n^2=I$，$\mathcal{C}_n\mathcal{S}_n=\mathcal{S}_n\mathcal{C}_n=0$，$H_n$ 是对称的，而且 H_n 是实正交的．(i)设 K_n 表示反序矩阵(0.9.5.1)．证明$\mathcal{C}_nK_n=K_n\mathcal{C}_n=\mathcal{C}_n$，$\mathcal{S}_nK_n=K_n\mathcal{S}_n=-\mathcal{S}_n$，且 $H_nK_n=K_nH_n$，所以$\mathcal{C}_n$，$\mathcal{S}_n$ 以及 H_n 都是中心对称的．已知对任何形如 $A=E+K_nF$ 的矩阵(其中 E 与 F 是实循环矩阵，$E=E^T$ 以及 $F=-F^T$)，$H_nAH_n=\Lambda$ 都是对角的．Λ 的对角元素(这样一个矩阵 A 的特征值)是向量 $n^{1/2}H_nAe_1$ 的元素．特别地，Hartley 矩阵对每个实对称循环矩阵提供了显式实正交对角化．

注记以及进一步的阅读参考 有关(2.2.6)的原始证明，见 W. Specht，Zur Theorie der Matrizen Ⅱ，*Jahresber. Deutsch. Math.-Verein.* 50(1940)19-23；在[Kap]中有一个现代的证明．有关(2.2.8)中所谈及的事项之综述，见 D. Đjoković以及 C. R. Johnson，Unitarily achievable zero patterns and traces of words in A and A^*，*Linear Algebra Appl.* 421(2007)63-68.(2.2.8d)中所列出的字出现在 D. Đjoković，Poincaré series of some pure and mixed trace algebras of two generic matrices，*J. Algebra* 309(2007)654-671 的定理 4.4 中．一个 4×4 复矩阵与它的转置酉相似，当且仅当(2.2.P4(b))中那种类型的七个零迹恒等式满足，见 S. R. Garcia，D. E. Poore 以及 J. E. Tener，Unitary equivalence to a complex symmetric matrix：low dimensions，*Linear Algebra Appl.* 437(2012)271-284 的定理 1．对两个非奇异的矩阵 A，$B\in M_n$(2.2.6)有一个近似的形式：A 与 B 是酉相似的，当且仅当对关于两个非交换变量的每一个字 $W(s,t)$都有 $|\mathrm{tr}W$

100 $(A,A^*)-\mathrm{tr}W(B,B^*)|\leqslant1$ 以及 $|\mathrm{tr}W((A,A^*)^{-1})-\mathrm{tr}W((B,B^*)^{-1})|\leqslant1$，见 L. W. Marcoux，

M. Mastnak 以及 H. Radjavi, An appxoximate, multivariable version of Specht's theorem, *Linear Multilinear Algebra* 55(2007)159-173.

2.3 酉三角化以及实正交三角化

初等矩阵论中基本上最有用的事实可能是 I. Schur 给出的一个定理：任何复方阵 A 与以 A 的特征值作为对角元素的一个三角矩阵酉相似，此三角矩阵的对角元素以任意指定的次序排列．我们的证明与一系列由酉相似给出的定义有关．

定理 2.3.1(Schur 型；Schur 三角化) 设 $A\in M_n$ 有特征值 $\lambda_1,\cdots,\lambda_n$，它们以任意指定的次序排列，又设 $x\in\mathbf{C}^n$ 是满足 $Ax=\lambda_1 x$ 的单位向量．

(a) 存在一个酉矩阵 $U=[x\quad u_2\quad\cdots\quad u_n]\in M_n$，使得 $U^*AU=T=[t_{ij}]$ 是以 $t_{ii}=\lambda_i$，$i=1,\cdots,n$ 为对角元素的上三角矩阵．

(b) 如果 $A\in M_n(\mathbf{R})$ 仅有实的特征值，那么可选取 x 为实的，且存在一个实正交矩阵 $Q=[x\quad q_2\quad\cdots\quad q_n]\in M_n(\mathbf{R})$，使得 $Q^TAQ=T=[t_{ij}]$ 是以 $t_{ii}=\lambda_i(i=1,\cdots,n)$ 为对角元素的上三角矩阵．

证明 设 x 是 A 的与特征值 λ_1 相伴的标准化特征向量，即 $x^*x=1$，且 $Ax=\lambda_1 x$．设 $U_1=[x\quad u_2\quad\cdots\quad u_n]$ 是任意一个第一列为 x 的酉矩阵．例如，可以如同在(2.1.13)中那样取 $U_1=U(x,e_1)$，或者如同在 2.3.P1 中那样去做．这样就有

$$U_1^*AU_1=U_1^*[Ax\quad Au_2\quad\cdots\quad Au_n]=U_1^*[\lambda_1 x\quad Au_2\quad\cdots\quad Au_n]$$

$$=\begin{bmatrix}x^*\\u_2^*\\\vdots\\u_n^*\end{bmatrix}[\lambda_1 x\quad Au_2\quad\cdots\quad Au_n]=\begin{bmatrix}\lambda_1 x^*x & x^*Au_2 & \cdots & x^*Au_n\\ \lambda_1 u_2^*x & & & \\ \vdots & & A_1 & \\ \lambda_1 u_n^*x & & & \end{bmatrix}\begin{bmatrix}\lambda_1 & \bigstar\\ 0 & A_1\end{bmatrix}$$

因为 U_1 的列是标准正交的．子矩阵 $A_1=[u_i^*Au_j]_{i,j=2}^n\in M_{n-1}$ 的特征值是 $\lambda_2,\cdots,\lambda_n$．如果 $n=2$，我们就完成了所希望的酉三角化．如若不然，设 $\xi\in\mathbf{C}^{n-1}$ 是 A_1 的一个与 λ_2 相伴的单位特征向量，并对 A_1 执行上面的化简．如果 $U_2\in M_{n-1}$ 是任意一个以 ξ 为第一列的酉矩阵，那么我们就看到了

$$U_2^*A_1U_2=\begin{bmatrix}\lambda_2 & \bigstar\\ 0 & A_2\end{bmatrix}$$

设 $V_2=[1]\oplus U_2$，并计算酉相似

$$(U_1V_2)^*AU_1V_2=V_2^*U_1^*AU_1V_2=\begin{bmatrix}\lambda_1 & \bigstar & \bigstar\\ 0 & \lambda_2 & \bigstar\\ 0 & 0 & A_2\end{bmatrix}$$

继续这一化简以产生出酉矩阵 $U_i\in M_{n-i+1}$，$i=1,\cdots,n-1$ 以及酉矩阵 $V_i\in M_n$，$i=2,\cdots,n-2$．矩阵 $U=U_1V_2V_3\cdots V_{n-2}$ 是酉矩阵，而 U^*AU 是上三角的．

如果 $A\in M_n(\mathbf{R})$ 所有的特征值都是实的，那么上面的算法中所有的特征向量以及酉矩阵都可以取为实的(1.1.P3 以及(2.1.13))． □

习题 利用(2.3.1)中的记号，设 U^*A^TU 是上三角的．设 $V=\overline{U}$，并说明为什么 V^*AV 是下三角的．

例 2.3.2 如果 A 的特征值重新排序，并执行相应的上三角化(2.3.1)，T 的位于主对角线上方的元素可能不同. 考虑

$$T_1=\begin{bmatrix}1&1&4\\0&2&2\\0&0&3\end{bmatrix},\quad T_2=\begin{bmatrix}2&-1&3\sqrt{2}\\0&1&\sqrt{2}\\0&0&3\end{bmatrix},\quad U=\frac{1}{\sqrt{2}}\begin{bmatrix}1&1&0\\1&-1&0\\0&0&\sqrt{2}\end{bmatrix}$$

验证 U 是酉矩阵，且 $T_2=UT_1U^*$.

习题 (**Schur 不等式；标准化的缺陷**) 如果 $A=[a_{ij}]\in M_n$ 有特征值 $\lambda_1,\cdots,\lambda_n$，且它与一个上三角矩阵 $T=[t_{ij}]\in M_n$ 酉相似，T 的对角元素是 A 的特征值按照某种次序的排列. 将(2.2.2)应用于 A 以及 T 以证明 ◀

$$\sum_{i=1}^{n}|\lambda_i|^2=\sum_{i,j=1}^{n}|a_{ij}|^2-\sum_{i<j}|t_{ij}|^2\leqslant\sum_{i,j=1}^{n}|a_{ij}|^2=\operatorname{tr}(AA^*)\qquad(2.3.2\text{a})$$

其中等式当且仅当 T 是对角矩阵时成立.

习题 如果 $A=[a_{ij}]$ 与 $B=[b_{ij}]\in M_2$ 有相同的特征值，又如果 $\sum_{i,j=1}^{2}|a_{ij}|^2=\sum_{i,j=1}^{2}|b_{ij}|^2$，利用(2.2.8)中的判别法证明 A 与 B 是酉相似的. 然而，考虑

$$A=\begin{bmatrix}1&3&0\\0&2&4\\0&0&3\end{bmatrix}\quad\text{以及}\quad B=\begin{bmatrix}1&0&0\\0&2&5\\0&0&3\end{bmatrix}\qquad(2.3.2\text{b})$$

它们有同样的特征值以及同样的元素平方之和. 利用(2.2.8)中的判别法或者利用(2.4.5.1)后面的习题证明 A 与 B 不是酉相似的. 然而无论如何，A 与 B 是相似的. 为什么? ◀

(2.3.1)有一个有用的推广：由复矩阵组成的一个交换族可以通过单独一个酉相似同

时化简为上三角型.

定理 2.3.3 设 $\mathcal{F}\subseteq M_n$ 是非空的交换族. 则存在一个酉矩阵 $U\in M_n$，使得对每个 $A\in\mathcal{F}$，U^*AU 都是上三角的.

证明 回到(2.3.1)的证明. 在该证明的每一步选取特征向量(以及酉矩阵)时，利用(1.3.19)选取对每一个 $A\in\mathcal{F}$ 所共有的一个单位特征向量，并构作一个以这个公共特征向量作为第一列的酉矩阵，它用同样的方式(通过酉相似)缩减 $\mathcal{F}$ 中的每一个矩阵. 相似则将交换性保留下来，分划矩阵的乘法计算显示，如果两个形如 $\begin{bmatrix}A_{11}&A_{12}\\0&A_{22}\end{bmatrix}$ 以及 $\begin{bmatrix}B_{11}&B_{12}\\0&B_{22}\end{bmatrix}$ 的矩阵可交换，那么 A_{22} 与 B_{22} 也可交换. 我们断言：(2.3.1)中关于 U 的所有组成成分都可以对交换族的所有成员用同样的方式选取. □

在(2.3.1)中我们可以指定 T 的主对角线(即可以预先指定 A 的特征值出现在缩减过程中的排列次序)，但是(2.3.3)没有声明这一点. 在缩减的每一步，所用到的公共特征向量是与 $\mathcal{F}$ 中**每一个**矩阵的**某个**特征值相伴的，但是我们或许不能指出它是哪一个. 根据由(1.3.19)所保证的公共特征向量，我们必须随意选取特征值.

下一个习题说明了：在实相似之下寻求实矩阵可能达致的三角型时，为什么会出现拟

三角矩阵以及拟对角矩阵，见(0.9.4).

习题 证明：$a\pm ib$ 是实 2×2 矩阵 $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ 的特征值.

如果一个实矩阵 A 有任何非实的特征值，就没有希望通过一个实的相似将它化简为上三角型 T，因为 T 的某个主对角线元素(A 的特征值)就会不是实的. 然而，我们总可以通过实正交相似将 A 化为一个实的拟三角型，成对共轭的非实特征值与 2×2 分块相伴.

定理 2.3.4(实 Schur 型) 设给定 $A\in M_n(\mathbf{R})$.

(a) 存在一个实的非奇异的 $S\in M_n(\mathbf{R})$，使得 $S^{-1}AS$ 是实的上拟三角矩阵

$$\begin{bmatrix} A_1 & & & \bigstar \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_m \end{bmatrix},\quad 每个\ A_i\ 都是\ 1\times1\ 或\ 2\times2\ 的 \tag{2.3.5}$$

它具有下述性质：(i)它的 1×1 对角块给出 A 的实特征值；(ii)它的每一个 2×2 对角块有特殊的形式，它给出 A 的一对共轭的非实的特征值：

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix},\quad a,b\in\mathbf{R},b>0,且\ a\pm ib\ 是\ A\ 的特征值 \tag{2.3.5a}$$

(iii) 它的对角块由 A 的特征值完全确定；它们可以按照任意预先指定的次序出现. 103

(b) 存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$，使得 $Q^{\mathrm{T}}AQ$ 是具有如下性质的实的上拟三角矩阵：(i)它的 1×1 对角块给出 A 的实特征值；(ii)它的每一个 2×2 对角块给出一对共轭的非实的特征值(不过没有特定的形式)；(iii)对角块的排序可以按照如下的意义预先加以指定：如果 A 的实特征值以及成对共轭的非实的特征值按照预先指定的次序列出，那么 $Q^{\mathrm{T}}AQ$ 的各个对角块 A_1，…，A_m 代表的实特征值与成对共轭的非实的特征值也按照同样的次序排列.

证明 (a)(2.3.1)的证明指出怎样用与任意给定的实特征对相对应的实正交相似来缩减 A，那样的缩减产生出实的 1×1 对角块以及一个形如 $\begin{bmatrix} \lambda & * \\ 0 & \mathcal{A} \end{bmatrix}$ 的缩减的矩阵. 问题 1.3.P33 描述了如何通过与特征对 λ，x 对应的实相似来缩减 A，其中 λ 不是实的。这种缩减产生出有特殊形式(2.3.5a)的 2×2 对角块 B 以及一个形如 $\begin{bmatrix} B & * \\ 0 & \mathcal{A} \end{bmatrix}$ 的缩减的矩阵. 仅需有限多次缩减就可以构造出一个非奇异的 S，使得 $S^{-1}AS$ 有所结论中的上拟三角型. 通过在每次缩减时选择一个特殊的特征值以及与之对应的特征向量，我们可以控制对角块出现的次序.

(b) 假设给定 A 的实特征值以及成对共轭的非实特征值的一个排序，又令 S 是一个非奇异的实矩阵，它使得 $S^{-1}AS$ 有(2.3.5)的形状，其对角块按照指定的次序排列. 利用(2.1.14)将 S 分解成 $S=QR$，其中 Q 是实正交矩阵，而 R 是实的上三角矩阵. 将 $R=[r_{ij}]$ 按照与(2.3.5)共形地予以分划并计算 $S^{-1}AS=R^{-1}Q^{\mathrm{T}}AQR$，所以

$$Q^{\mathrm{T}}AQ = R\begin{bmatrix} A_1 & & & \bigstar \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_m \end{bmatrix}R^{-1} = \begin{bmatrix} R_{11}A_1R_{11}^{-1} & & & \bigstar \\ & R_{22}A_2R_{22}^{-1} & & \\ & & \ddots & \\ 0 & & & R_{mm}A_mR_{mm}^{-1} \end{bmatrix}$$

是上拟三角矩阵，它的 1×1 对角块与(2.3.5)的那些是相同的，而它的 2×2 对角块与(2.3.5)的对应的分块是相似的. □

上面的定理有一个用交换族来表达的形式：一个由实矩阵组成的交换族可以通过单独一个实相似或者实正交相似被同时简化成共同的上拟三角型. 用如下的说法来描述(2.3.5)的分划的结构是方便的：它按照与给定的拟对角矩阵 $D=J_{n_1}\oplus\cdots\oplus J_{n_m}\in M_n$ 共形地加以分划，其中 J_k 记 $k\times k$ 全 1 矩阵(0.2.8)，而每个 n_j 取值或者为 1，或者为 2.

定理 2.3.6 设 $\mathcal{F}\subseteq M_n(\mathbf{R})$ 是一个非空的交换族.

(a) 存在一个非奇异的 $S\in M_n(\mathbf{R})$ 以及一个拟对角矩阵 $D=J_{n_1}\oplus\cdots\oplus J_{n_m}\in M_n$，使得：(i)对每个 $A\in\mathcal{F}$，$S^{-1}AS$ 都是形如

$$\begin{bmatrix} A_1(A) & & & \bigstar \\ & A_2(A) & & \\ & & \ddots & \\ 0 & & & A_m(A) \end{bmatrix} \tag{2.3.6.1}$$

的实的上拟三角矩阵，它与 D 共形地分划；(ii)如果 $n_i=2$，则对每个 $A\in\mathcal{F}$ 我们有

$$A_j(A)=\begin{bmatrix} a_j(A) & b_j(A) \\ -b_j(A) & a_j(A)\end{bmatrix}\in M_2(\mathbf{R}) \tag{2.3.6.2}$$

且 $a_j(A)\pm \mathrm{i}b_j(A)$ 是 A 的特征值；(iii)对每个满足 $n_j=2$ 的 $j\in\{1,\cdots,m\}$，存在某个 $A\in\mathcal{F}$，使得 $b_j(A)\neq0$. 如果 $\mathcal{F}$ 中每个矩阵都只有实特征值，那么对每个 $A\in\mathcal{F}$，$S^{-1}AS$ 都是上三角的.

(b) 存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$ 以及一个拟对角矩阵 $D=J_{n_1}\oplus\cdots\oplus J_{n_m}\in M_n$，使得：(i)对每个 $A\in\mathcal{F}$，$Q^{\mathrm{T}}AQ$ 都是形如(2.3.6.1)的与 D 共形分划的上拟三角矩阵；(ii)对每个满足 $n_j=2$ 的 $j\in\{1,\cdots,m\}$，存在某个 $A\in\mathcal{F}$，使得 $A_j(A)$ 有一对共轭的非实特征值. 如果 $\mathcal{F}$ 中每个矩阵都只有实特征值，那么对每个 $A\in\mathcal{F}$，$Q^{\mathrm{T}}AQ$ 都是上三角的.

证明 (a)遵循(2.3.3)证明中的归纳模式，只需要构造一个非奇异的实矩阵，它以同样的方式(通过相似性)缩减 $\mathcal{F}$ 中的每一个矩阵就够了. 利用(1.3.19)来选取每一个 $A\in\mathcal{F}$ 所共有的单位特征向量 $x\in\mathbf{C}^n$. 记 $x=u+\mathrm{i}v$，其中 u，$v\in\mathbf{R}^n$. 这里有两种可能性，第一种可能性是(i)$\{u,v\}$是线性相关的. 在此情形，存在一个实的单位向量 $w\in\mathbf{R}^n$ 以及不全为零的实的纯量 α 与 β，使得 $u=\alpha w$ 以及 $v=\beta w$. 这样 $x=(\alpha+\mathrm{i}\beta)w$ 与 $w=(\alpha+\mathrm{i}\beta)^{-1}x$ 就是每一个 $A\in\mathcal{F}$ 的实单位特征向量. 设 Q 是一个实正交矩阵，它的第一列是 w，并注意到对每个 $A\in\mathcal{F}$ 有 $Q^{\mathrm{T}}AQ=\begin{bmatrix}\lambda(A) & * \\ 0 & *\end{bmatrix}$，其中 $\lambda(A)$ 是 A 的一个实特征值. 第二种可能性是(ii)$\{u,v\}$是线性无关的. 在此情形(1.3.P3)指出了怎样构造一个实的非奇异的矩阵 S，使得对每个 $A\in\mathcal{F}$，有 $S^{-1}AS=\begin{bmatrix}A_1(A) & * \\ 0 & *\end{bmatrix}$，其中 $A_1(A)$ 有(2.3.6.2)的形式. 如果 $b_1(A)\neq0$，那么 $a_1(A)\pm\mathrm{i}b_1(A)$ 是 A 的一对共轭的非实特征值. 然而，如果 $b_1(A)=0$，则 $a_1(A)$ 是 A 的二重实特征值. 如果对每个 $A\in\mathcal{F}$ 有 $b_1(A)=0$(例如，如果 $\mathcal{F}$ 中每个矩阵都只有实特征值)，那么就将 2×2 块分成两个 1×1 块.

(b) 设 S 是非奇异的实矩阵，它有(a)中所断言的性质，又设 $S=QR$ 是 QR 分解

(2.1.14). 按照在(2.3.4)的证明中同样的方法，可以证明 Q 有所断言的性质. □

正如在(2.3.3)中那样，我们不能控制与上一个定理中的对角块相对应的特征值出现的次序，我们不得不根据(1.3.19)所保证存在的公共特征向量让特征值随意出现. 105

习题 设 $A\in M_n$. 说明为什么 A 与 $\overline{A}$ 可交换，当且仅当 $A\overline{A}$是实的. ◀

习题 设 $A=\begin{bmatrix}1 & \mathrm{i}\\ -\mathrm{i} & 1\end{bmatrix}$. 证明 $A\overline{A}$是实的，且 $\mathrm{Re}A$ 与 $\mathrm{Im}A$ 可交换. ◀

习题 设 $A\in M_n$ 并记 $A=B+\mathrm{i}C$，其中 B 与 C 是实的. 证明：$A\overline{A}=\overline{A}A$ 成立当且仅当 $BC=CB$. ◀

使得 $A\overline{A}$是实数的矩阵之集合$\mathcal{S}=\{A\in M_n: A\overline{A}=\overline{A}A\}$大于实矩阵的集合 $M_n(\mathbf{R})$，不过它们有一个共同的重要性质：任何实方阵实正交相似于一个实的上拟三角矩阵，而$\mathcal{S}$中的任何矩阵都实正交相似于一个**复的**上拟三角矩阵.

推论 2.3.7 设 $A\in M_n$，并假设 $A\overline{A}=\overline{A}A$. 则存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$以及一个拟对角矩阵 $D=J_{n_1}\oplus\cdots\oplus J_{n_m}\in M_n$，使得 $Q^{\mathrm{T}}AQ\in M_n$ 是一个形如(2.3.6.1)的复的上拟三角矩阵，它与 D 共形地分划且有如下性质：对每个使得 $n_j=2$ 成立的 $j\in\{1,\cdots,m\}$，$\mathrm{Re}A_j$ 或者 $\mathrm{Im}A_j$ 中至少有一个有一对共轭的非实特征值. 如果 $\mathrm{Re}A_j$ 与 $\mathrm{Im}A_j$ 中的每一个都只有实特征值，那么 $Q^{\mathrm{T}}AQ\in M_n$ 是上三角矩阵.

证明 记 $A=B+\mathrm{i}C$，其中 B 与 C 是实数. 这些假设以及上一个习题保证了 B 与 C 可交换. 由(2.3.6b)推出，存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$以及一个拟对角矩阵 $D=J_{n_1}\oplus\cdots\oplus J_{n_m}\in M_n$，使得 $Q^{\mathrm{T}}BQ$ 与 $Q^{\mathrm{T}}CQ$ 中的每一个都是形如(2.3.6.1)的实的上拟三角矩阵，它与 D 共形地分划. 此外，对每个使得 $n_j=2$ 成立的 $j\in\{1,\cdots,m\}$，$A_j(B)$或者 $A_j(C)$中至少有一个有一对共轭的非实特征值. 由此推出，$Q^{\mathrm{T}}AQ=Q^{\mathrm{T}}(B+\mathrm{i}C)Q=Q^{\mathrm{T}}BQ+\mathrm{i}Q^{\mathrm{T}}CQ$ 是一个复的上拟三角矩阵，它与 D 共形地分划. 如果 B 与 C 中的每一个都只有实特征值，那么每个 $n_j=1$，且 $Q^{\mathrm{T}}BQ$ 与 $Q^{\mathrm{T}}CQ$ 中每一个都是上三角的. □

问题

2.3.P1 设 $x\in\mathbf{C}^n$ 是一个给定的单位向量，并记 $x=[x_1\quad y^{\mathrm{T}}]^{\mathrm{T}}$，其中 $x_1\in\mathbf{C}$，而 $y\in\mathbf{C}^{n-1}$. 选取 $\theta\in\mathbf{R}$，使得 $\mathrm{e}^{\mathrm{i}\theta}x_1\geqslant 0$，并定义 $z=\mathrm{e}^{\mathrm{i}\theta}x=[z_1\quad \zeta^{\mathrm{T}}]^{\mathrm{T}}$，其中 $z_1\in\mathbf{R}$ 是非负的，而 $\zeta\in\mathbf{C}^{n-1}$. 考虑 Hermite 矩阵

$$V_x=\left[\begin{array}{c:c} z_1 & \zeta^* \\ \hdashline \zeta & -I+\dfrac{1}{1+z_1}\zeta\zeta^* \end{array}\right]$$

利用分划的乘法来计算 $V_x^*V_x=V_x^2$. 我们推出 $U=\mathrm{e}^{-\mathrm{i}\theta}V_x=[x\quad u_2\quad\cdots\quad u_n]$是酉矩阵，它的第一列是给定的向量 x.

2.3.P2 如果 $x\in\mathbf{R}^n$ 是一个给定的单位向量，指出如何精简(2.3.P1)中所描述的构造，从而产生出一个实正交矩阵 $Q\in M_n(\mathbf{R})$，它的第一列就是 x. 证明你的构造是成功的. 106

2.3.P3 设 $A\in M_n(\mathbf{R})$. 说明为什么 A 的非实的特征值必定共轭成对出现.

2.3.P4 考虑族$\mathcal{F}=\left\{\begin{bmatrix}0 & -1\\ 0 & -1\end{bmatrix},\begin{bmatrix}1 & 1\\ 0 & -1\end{bmatrix}\right\}$，并证明(2.3.3)中的可交换性猜想虽然足以蕴含对$\mathcal{F}$同时酉上三角化，但却并非是必要的.

2.3.P5 设$\mathcal{F}=\{A_1,\cdots,A_k\}\subset M_n$ 是一个给定的族，又设$\mathcal{G}=\{A_iA_j: i, j=1, 2,\cdots,k\}$是$\mathcal{F}$中的矩阵两两成对的乘积组成的族. 如果$\mathcal{G}$是交换的，已知$\mathcal{F}$可以同时酉上三角化，当且仅当每个**换位子**

(commutator)$A_iA_j-A_jA_i$ 的每个特征值都是零. 指出$\mathcal{G}$的交换性假设比$\mathcal{F}$的交换性假设更弱. 证明(2.3.P4)中的族$\mathcal{F}$有对应的$\mathcal{G}$, $\mathcal{G}$是交换的且也满足零特征值条件.

2.3.P6 设给定 A, $B\in M_n$, 并假设 A 与 B 同时相似于上三角矩阵, 也就是说, 对某个非奇异的 $S\in M_n$, $S^{-1}AS$ 与 $S^{-1}BS$ 两者均为上三角矩阵. 证明 $AB-BA$ 的每个特征值都必定是零.

2.3.P7 如果一个给定的 $A\in M_n$ 可以写成 $A=Q\Delta Q^{\mathrm{T}}$, 其中 $Q\in M_n$ 是复的正交矩阵, 而 $\Delta\in M_n$ 是上三角矩阵, 证明 A 至少有一个特征向量 $x\in\mathbf{C}^n$ 使得 $x^{\mathrm{T}}x\neq 0$. 考虑 $A=\begin{bmatrix}1 & \mathrm{i}\\ \mathrm{i} & -1\end{bmatrix}$, 证明: 并非每一个 $A\in M_n$ 都能通过复的正交相似来实现上三角化.

2.3.P8 设 $Q\in M_n$ 是复正交矩阵, 又假设 $x\in\mathbf{C}^n$ 是 Q 的与一个特征值 $\lambda\neq\pm 1$ 相伴的特征向量. 证明 $x^{\mathrm{T}}x=0$. (2.1.P8a)给出一个例子, 这个例子中给出一族 2×2 复正交矩阵, 它们的两个特征值都异于± 1. 证明: 这些矩阵中没有一个能通过复正交相似转化为上三角型.

2.3.P9 设 λ_1, $\cdots$, λ_n 是 $A\in M_n$ 的特征值, 假设 x 是满足 $Ax=\lambda x$ 的非零向量, 又设给定 $y\in\mathbf{C}^n$ 以及 $\alpha\in\mathbf{C}$. 对下面的论证提供细节, 以证明加边矩阵$\mathcal{A}=\begin{bmatrix}\alpha & y^*\\ x & A\end{bmatrix}\in M_{n+1}$的特征值是$\begin{bmatrix}\alpha & y^*x\\ 1 & \lambda\end{bmatrix}$的两个特征值加上$\lambda_2$, $\cdots$, λ_n: 作出一个酉矩阵 U, 它的第一列是 $x/\|x\|_2$. 设 $V=[1]\oplus U$, 并证明 $V^*\mathcal{A}V=\begin{bmatrix}B & \bigstar\\ 0 & C\end{bmatrix}$, 其中 $B=\begin{bmatrix}\alpha & y^*x/\|x\|_2\\ \|x\|_2 & \lambda\end{bmatrix}\in M_2$, 而 $C\in M_{n-2}$有特征值 λ_2, $\cdots$, λ_n. 考虑 B 通过 $\mathrm{diag}(1, \|x\|_2^{-1})$所作的相似. 如果 $y\perp x$, 推出结论: $\mathcal{A}$的特征值是 α, λ, λ_2, $\cdots$, λ_n. 说明为什么$\begin{bmatrix}\alpha & y^*\\ x & A\end{bmatrix}$与$\begin{bmatrix}A & x\\ y^* & \alpha\end{bmatrix}$的特征值相同.

2.3.P10 设 $A=[a_{ij}]\in M_n$ 以及 $c=\max\{|a_{ij}|: 1\leqslant i, j\leqslant n\}$. 用两种方法证明 $|\det A|\leqslant c^n n^{n/2}$: (a)设 λ_1, $\cdots$, λ_n 是 A 的特征值. 利用算术-几何平均不等式以及(2.3.2a)来说明为什么 $|\det A|^2 = |\lambda_1\cdots\lambda_n|^2\leqslant((|\lambda_1|^2+\cdots+|\lambda_n|^2)/n)^n\leqslant\left(\sum_{i,j=1}^n|a_{ij}|^2/n\right)^n\leqslant(nc^2)^n$.

(b) 利用(2.1.P23)中的 Hadamard 不等式

2.3.P11 利用(2.3.1)证明: 如果 $A\in M_n$ 的所有的特征值都是零, 那么 $A^n=0$.

2.3.P12 设 $A\in M_n$, 又设 λ_1, $\cdots$, λ_n 是它的特征值, 再令 $r\in\{1, \cdots, n\}$. (a)利用(2.3.1)证明: 复合矩阵 $C_r(A)$的特征值是$\binom{n}{r}$个可能的乘积$\lambda_{i_1}\cdots\lambda_{i_r}$, 其中 $1\leqslant i_1<i_2<\cdots<i_r\leqslant n$. (b)说明为什么

107 $\mathrm{tr}C_r(A)=S_r(\lambda_1, \cdots, \lambda_n)=E_r(A)$; 见(1.2.14)以及(1.2.16). (c)如果 A 的特征值排列成 $|\lambda_1|\geqslant\cdots\geqslant|\lambda_n|$, 说明为什么 $C_r(A)$的谱半径是 $\rho(C_r(A))=|\lambda_1\cdots\lambda_r|$. (d)说明为什么 $p_A(t)=\sum_{k=0}^n(-1)^k t^{n-k}\mathrm{tr}C_r(A)$,从而有 $\det(I+A)=\sum_{k=0}^n\mathrm{tr}C_r(A)$. (e)对下面的推理提供细节: 如果 A 是非奇异的, 那么

$$\begin{aligned}\det(A+B)&=\det A\det(I+A^{-1}B)=\det A\sum_{k=0}^n\mathrm{tr}C_r(A^{-1}B)\\&=\det A\sum_{k=0}^n\mathrm{tr}(C_r(A^{-1})C_r(B))=\det A\sum_{k=0}^n\mathrm{tr}(C_r(A)^{-1}C_r(B))\\&=\det A\sum_{k=0}^n\mathrm{tr}(\det A^{-1}\mathrm{adj}_k(A)C_r(B))=\sum_{k=0}^n\mathrm{tr}(\mathrm{adj}_k(A)C_r(B))\end{aligned}$$

(f) 证明恒等式(0.8.12.3).

2.3.P13 考虑 $A=\begin{bmatrix}-2 & 5\\ -1 & 2\end{bmatrix}$. (a)证明 $\pm i$ 是 A 的特征值，并说明为什么 A 与 $B=\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$ 是实相似的. (b)说明为什么 A 与 B 不是实正交相似的.

2.3.P14 设 $A=[a_{ij}]\in M_n$. (a)设 $V=[v_{ij}]\in M_n$ 是酉矩阵. 说明为什么 $|\mathrm{tr}VA|=\left|\sum_{i,j}v_{ij}a_{ji}\right|\leqslant\sum_{i,j}|a_{ji}|$. (b)设 $\lambda_1,\cdots,\lambda_n$ 是 A 的特征值. 证明 $\sum_i|\lambda_i|\leqslant\sum_{i,j}|a_{ji}|$.

进一步的阅读参考 有关上三角化(2.3.1)的一个改进的结果，见(3.4.3.1). (2.3.P5)中给出的(2.3.3)的一个更强形式的证明见 Y. P. Hong 以及 R. A. Horn, On simultaneous reduction of families of matrices to triangular or diagonal form by unitary congruences, *Linear Multilinear Algebra* 17(1985)271-288.

2.4 Schur 三角化定理的推论

从 Schur 的酉三角化定理可以收获一批结果. 在这一节里我们研究其中的几个.

2.4.1 迹与行列式

假设 $A\in M_n$ 有特征值 $\lambda_1,\cdots,\lambda_n$. 在(1.2)中我们曾利用特征多项式证明了 $\sum_{i=1}^n\lambda_i=\mathrm{tr}A$, $\sum_{i=1}^n\prod_{j\neq i}^n\lambda_j=\mathrm{tr}(\mathrm{adj}A)$ 以及 $\det A=\prod_{i=1}^n\lambda_i$，但是这些恒等式以及其他一些结果都可以从对(2.3.1)中的三角型的核查直接得出.

对任何非奇异的 $S\in M_n$，我们有 $\mathrm{tr}(S^{-1}AS)=\mathrm{tr}(ASS^{-1})=\mathrm{tr}A$；$\mathrm{tr}(\mathrm{adj}(S^{-1}AS))=\mathrm{tr}((\mathrm{adj}S)(\mathrm{adj}A)(\mathrm{adj}S^{-1}))=\mathrm{tr}((\mathrm{adj}S)(\mathrm{adj}A)(\mathrm{adj}S)^{-1})=\mathrm{tr}(\mathrm{adj}A)$；以及 $\det(S^{-1}AS)=(\det S^{-1})(\det A)(\det S)=(\det S)^{-1}(\det A)(\det S)=\det A$. 这样一来，$\mathrm{tr}A$，$\mathrm{tr}(\mathrm{adj}A)$ 以及 $\det A$ 都可以用任何与 A 相似的矩阵来计算. (2.3.1)中的上三角矩阵 $T=[t_{ij}]$ 对这一目的来说是方便的，因为它的主对角线元素 $t_{11},\cdots t_{nn}$ 是 A 的特征值，$\mathrm{tr}T=\sum_{i=1}^n t_{ii}$, $\det T=\prod_{i=1}^n t_{ii}$，而 $\mathrm{adj}T$ 的主对角线元素是 $\prod_{j\neq 1}^n t_{jj},\cdots,\prod_{j\neq n}^n t_{jj}$.

2.4.2 A 的多项式的特征值

假设 $A\in M_n$ 有特征值 $\lambda_1,\cdots,\lambda_n$，并设 $p(t)$ 是一个给定的多项式. 在(1.1.6)中我们证明了，对每一个 $i=1,\cdots,n$，$p(\lambda_i)$ 都是 $p(A)$ 的特征值，又如果 μ 是 $p(A)$ 的一个特征值，那么就存在某个 $i\in\{1,\cdots,n\}$，使得 $\mu=p(\lambda_i)$. 这些结论将 $p(A)$ 的不同的特征值(即它的谱(1.1.4))辨识出来，但没有给出它们的重数. Schur 定理 2.3.1 揭示出它们的重数. 108

设 $A=UTU^*$，其中 U 是酉矩阵，而 $T=[t_{ij}]$ 是上三角矩阵，其主对角元素是 $t_{11}=\lambda_1$，$t_{22}=\lambda_2$，$\cdots$，$t_{nn}=\lambda_n$. 这样就有 $p(A)=p(UTU^*)=Up(T)U^*$ (1.3.P2). $p(T)$ 的主对角元素是 $p(\lambda_1)$，$p(\lambda_2)$，$\cdots$，$p(\lambda_n)$，故而这些就是 $p(T)$(从而也就是 $p(A)$)的特征值(计入重数). 特别地，对每个 $k=1,2,\cdots$，A^k 的特征值是 $\lambda_1^k,\cdots,\lambda_n^k$，且

$$\mathrm{tr}A^k=\lambda_1^k+\cdots+\lambda_n^k \tag{2.4.2.1}$$

习题 如果 $T\in M_n$ 是严格上三角的，证明：T^p 的主对角线以及前 $p-1$ 条超对角线上的所有元素都是零，$p=1$，…，n，特别地，有 $T^n=0$. ◀

假设 $A\in M_n$. 我们知道(1.1.P6)：如果对某个正整数 k 有 $A^k=0$，那么 $\sigma(A)=\{0\}$，所以 A 的特征多项式是 $p_A(t)=t^n$. 现在我们可以证明它的逆命题，且还能略微得到更多一点结果. 如果 $\sigma(A)=\{0\}$，那么存在一个酉矩阵 U 以及一个**严格**上三角矩阵 T，使得 $A=UTU^*$；上一个习题告诉我们有 $T^n=0$，所以 $A^n=UT^nU^*=0$. 于是，如下结论对 $A\in M_n$是等价的：(a)A 是幂零的；(b)$A^n=0$；以及(c)$\sigma(A)=\{0\}$.

2.4.3 Cayley-Hamilton 定理

每个复方阵都满足它自己的特征方程这一事实可由 Schur 定理以及关于有特殊零元素模式的三角矩阵的乘法的一个结论推出.

引理 2.4.3.1 假设 $R=[r_{ij}]$，$T=[t_{ij}]\in M_n$ 是上三角矩阵，且 $r_{ij}=0$，$1\leqslant i$，$j\leqslant k<n$，以及 $t_{k+1,k+1}=0$. 设 $S=[s_{ij}]=RT$. 那么 $s_{ij}=0$，$1\leqslant i$，$j\leqslant k+1$.

证明 这些假设条件描述了形如

$$R=\begin{bmatrix}0_k & R_{12}\\ 0 & R_{22}\end{bmatrix},\quad T=\begin{bmatrix}T_{11} & T_{12}\\ 0 & T_{22}\end{bmatrix},\quad T_{11}\in M_k$$

的分块矩阵 R 和 T，其中 R_{22}，T_{11} 以及 T_{22} 都是上三角的，且 T_{22} 的第一列为零. 乘积 RT 必定是上三角的. 我们必须证明它在左上角有一个为零的 $k+1$ 阶主子矩阵. 分划 $T_{22}=[0\quad Z]$ 以显示出它的第一列，作分块乘法

$$RT=\begin{bmatrix}0_kT_{11}+R_{12}0 & 0_kT_{12}+R_{12}[0\quad Z]\\ 0T_{11}+R_{22}0 & 0T_{12}+R_{22}[0\quad Z]\end{bmatrix}=\begin{bmatrix}0_k & [0\quad R_{12}Z]\\ 0 & [0\quad R_{22}Z]\end{bmatrix}$$

它显示出所希望的位于左上角的那个 $k+1$ 阶零主子矩阵. □

定理 2.4.3.2(Cayley-Hamilton) 设 $p_A(t)$是 $A\in M_n$ 的特征多项式. 那么 $p_A(A)=0$.

证明 如同在(1.2.6)中那样分解 $p_A(t)=(t-\lambda_1)(t-\lambda_2)\cdots(t-\lambda_n)$，利用(2.3.1)将 A 记为 $A=UTU^*$，其中 U 是酉矩阵，T 是上三角矩阵，而 T 的主对角元素是 λ_1，…，λ_n.

109 计算

$$\begin{aligned}p_A(A)&=p_A(UTU^*)=Up_A(T)U^*\\&=U[(T-\lambda_1I)(T-\lambda_2I)\cdots(T-\lambda_nI)]U^*\end{aligned}$$

只要证明 $p_A(T)=0$ 就足够了. $T-\lambda_1I$ 的左上角的 1×1 分块是零，而 $T-\lambda_2I$ 位于(2，2)处的元素是零，所以上一个引理确保$(T-\lambda_1I)(T-\lambda_2I)$的左上角 2×2 主子矩阵是零. 假设$(T-\lambda_1I)\cdots(T-\lambda_kI)$的左上角 $k\times k$ 主子矩阵是零. $(T-\lambda_{k+1}I)$的位于 $k+1$，$k+1$ 处的元素是零，故而再次借用该引理我们就知道，$(T-\lambda_1I)\cdots(T-\lambda_{k+1}I)$的左上角的 $k+1$ 阶主子矩阵也是零. 根据归纳法，我们得出$((T-\lambda_1I)\cdots(T-\lambda_{n-1}I))(T-\lambda_nI)=0$. □

习题 下面的推理中何处有错误？“由于对 $A\in M_n$ 的每一个特征值 λ_i 有 $p_A(\lambda_i)=0$，又因为 $p_A(A)$的特征值是 $p_A(\lambda_1)$，…，$p_A(\lambda_n)$，故而 $p_A(A)$的所有特征值都是 0. 这样一来就有 $p_A(A)=0$.”给出一个例子来说明推理中的错误. ◀

习题 下面的推理中何处有错误？“由于 $p_A(t)=\det(tI-A)$，我们就有 $p_A(A)=\det(AI-A)=\det(A-A)=\det 0=0$. 这样一来就有 $p_A(A)=0$.” ◀

Cayley-Hamilton 定理常常被解释成“每个方阵都满足它自己的特征方程”(1.2.3)，不过这必须要仔细加以理解：纯量多项式 $p_A(t)$ 首先是作为 $p_A(t)=\det(tI-A)$ 来计算的；然后才是通过代换 $t\to A$ 来计算矩阵 $p_A(A)$.

我们已经对元素为复数的矩阵证明了 Cayley-Hamilton 定理，从而它必定对元素取自复数的任意子域(例如实数或者有理数)的矩阵也成立. 事实上，Cayley-Hamilton 定理是一个完全形式的结果，它对元素取自任何域(更一般地，是对任何交换环)的矩阵都成立；见(2.4.P3).

Cayley-Hamilton 定理的一项重要用途是将 $A\in M_n$ 的幂 A^k(对 $k\geqslant n$)写成 I，A，A^2，…，A^{n-1} 的线性组合.

例 2.4.3.3 设 $A=\begin{bmatrix}3 & 1\\ -2 & 0\end{bmatrix}$. 那么 $p_A(t)=t^2-3t+2$，所以 $A^2-3A+2I=0$. 从而 $A^2=3A-2I$；$A^3=A(A^2)=3A^2-2A=3(3A-2I)-2A=7A-6I$；$A^4=7A^2-6A=15A-14I$，如此等等. 我们还可以将非奇异矩阵 A 的负数次幂表示成 A 与 I 的线性组合. 将 $A^2-3A+2I=0$ 写成 $2I=-A^2+3A=A(-A+3I)$，或者 $I=A\left[\frac{1}{2}(-A+3I)\right]$. 从而 $A^{-1}=-\frac{1}{2}A+\frac{3}{2}I=\begin{bmatrix}0 & -1/2\\ 1 & 3/2\end{bmatrix}$，$A^{-2}=\left(-\frac{1}{2}A+\frac{3}{2}I\right)^2=\frac{1}{4}A^2-\frac{3}{2}A+\frac{9}{4}I=\frac{1}{4}(3A-2I)-\frac{3}{2}A+\frac{9}{4}I=-\frac{3}{4}A+\frac{7}{4}I$，如此等等.

推论 2.4.3.4 假设 $A\in M_n$ 是非奇异的，且 $p_A(t)=t^n+a_{n-1}t^{n-1}+\cdots+a_1t+a_0$. 令 $q(t)=-(t^{n-1}+a_{n-1}t^{n-2}+\cdots+a_2t+a_1)/a_0$. 那么 $A^{-1}=q(A)$ 是 A 的多项式.

证明 将 $p_A(A)=0$ 写成 $A(A^{n-1}+a_{n-1}A^{n-2}+\cdots+a_2A+a_1I)=-a_0I$，这就是 $Aq(A)=I$. 110

□

习题 如果 A，$B\in M_n$ 是相似的，而 $g(t)$ 是任意一个给定的多项式，证明 $g(A)$ 与 $g(B)$ 相似，且 A 满足的任何多项式方程也被 B 满足. 对其逆命题给出某种见解：满足同一个多项式方程就蕴含相似——这一结论是真还是假？ ◀

例 2.4.3.5 我们已经证明了：每一个 $A\in M_n$ 都满足一个 n 次多项式方程，例如它的特征方程. 然而对 $A\in M_n$ 来说，它有可能满足一个次数低于 n 的多项式方程. 考虑

$$A=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 1\\ 0 & 0 & 1\end{bmatrix}\in M_3$$

其特征多项式是 $p_A(t)=(t-1)^3$，而且确有 $(A-I)^3=0$. 但是 $(A-I)^2=0$，所以 A 满足一个 2 次多项式方程. 不存在一次多项式 $h(t)=t+a_0$，使得 $h(A)=0$，这是因为对所有 $a_0\in\mathbf{C}$ 都有 $h(A)=A+a_0I\neq 0$.

习题 假设一个可以对角化的矩阵 $A\in M_n$ 有 $d\leqslant n$ 个不同的特征值 λ_1，…，λ_d. 设 $q(t)=(t-\lambda_1)\cdots(t-\lambda_d)$. 证明 $q(A)=0$，所以 A 满足一个 d 次多项式方程. 为什么不存在次数严格小于 d 的多项式 $g(t)$，使得 $g(A)=0$？考虑上一个例子中的矩阵，证明：能被一个不可对角化的矩阵满足的多项式方程的最小次数可能严格大于它的不同的特征值的个数. ◀

2.4.4 关于线性矩阵方程的 Sylvester 定理

与交换性有关的方程 $AX-XA=0$ 是线性矩阵方程 $AX-XB=C$ 的一个特例，通常称为 **Sylvester 方程**. 对每个给定的 C，下面的定理对 Sylvester 方程有唯一解给出了一个必要且充分的条件. 它依赖于 Cayley-Hamilton 定理以及如下的结论：如果 $AX=XB$，那么 $A^2X=A(AX)=A(XB)=(AX)B=(XB)B=XB^2$，$A^3X=A(A^2X)=A(XB^2)=(AX)B^2=XB^3$，如此等等. 于是，根据标准的约定俗成，我们总是将 A^0 视为单位矩阵，我们就有

$$\left(\sum_{k=0}^{m} a_k A^k\right)X = \sum_{k=0}^{m} a_k A^k X = \sum_{k=0}^{m} a_k X B^k = X\left(\sum_{k=0}^{m} a_k B^k\right)$$

我们将这个结论总结成如下的引理.

引理 2.4.4.0 设 $A\in M_n$，$B\in M_m$ 以及 $X\in M_{n,m}$. 如果 $AX-XB=0$，那么对任何多项式 $g(t)$ 都有 $g(A)X-Xg(B)=0$.

定理 2.4.4.1(Sylvester) 设给定 $A\in M_n$ 以及 $B\in M_m$. 对每个给定的 $C\in M_{n,m}$，方程 $AX-XB=C$ 有唯一解 $X\in M_{n,m}$，当且仅当 $\sigma(A)\cap\sigma(B)=\varnothing$，也就是说，当且仅当 A 与 B 没有公共的特征值. 特别地，如果 $\sigma(A)\cap\sigma(B)=\varnothing$，那么满足 $AX-XB=0$ 的唯一的解就是 $X=0$. 如果 A 与 B 是实的，那么对每个给定的 $C\in M_{n,m}(\mathbf{R})$，$AX-XB=C$ 有唯一的解
111 $X\in M_{n,m}(\mathbf{R})$.

证明 考虑由 $T(X)=AX-XB$ 定义的线性变换 $T: M_{n,m}\to M_{n,m}$. 为确保对每个给定的 $C\in M_{n,m}$，方程 $T(X)=C$ 有唯一解 X，只需要证明 $T(X)=0$ 的唯一解是 $X=0$ 就足够了；见(0.5). 如果 $AX-XB=0$，由上面的讨论我们知道 $p_B(A)X-Xp_B(B)=0$. Cayley-Hamilton 定理确保有 $p_B(B)=0$，所以 $p_B(A)X=0$.

设 $\lambda_1,\cdots,\lambda_n$ 是 B 的特征值，所以 $p_B(t)=(t-\lambda_1)\cdots(t-\lambda_n)$，且 $p_B(A)=(A-\lambda_1 I)\cdots(A-\lambda_n I)$. 如果 $\sigma(A)\cap\sigma(B)=\varnothing$，则每一个因子 $A-\lambda_j I$ 都是非奇异的，从而 $p_B(A)$ 是非奇异的，且 $p_B(A)X=0$ 的唯一解就是 $X=0$. 反过来，如果 $p_B(A)X=0$ 有一个非平凡的解，那么 $p_B(A)$ 必定是奇异的，从而它的某个因子 $A-\lambda_j I$ 是奇异的，因而某个 λ_j 就是 A 的一个特征值.

如果 A 与 B 是实的，考虑由 $T(X)=AX-XB$ 定义的线性变换 $T: M_{n,m}(\mathbf{R})\to M_{n,m}(\mathbf{R})$. 同样的讨论表明：实矩阵 $p_B(A)$ 是非奇异的，当且仅当 $\sigma(A)\cap\sigma(B)=\varnothing$(即便当 B 的某些特征值 λ_i 不为实数时也依然成立). □

形如 $AX=XB$ 的矩阵恒等式称为**缠绕关系**(intertwining relation). 交换性方程 $AB=BA$ 似乎是最为熟悉的缠绕关系；其他的例子是反交换性方程 $AB=-BA$，$AB=BA^T$，$AB=B\overline{A}$，以及 $AB=BA^*$. Sylvester 定理的如下推论常用来证明：一个矩阵是分块对角的，如果它满足某种类型的缠绕关系.

推论 2.4.4.2 设 $B, C\in M_n$ 是分块对角的，且共形地分划成 $B=B_1\oplus\cdots\oplus B_k$ 以及 $C=C_1\oplus\cdots\oplus C_k$. 假设只要 $i\neq j$ 就有 $\sigma(B_i)\cap\sigma(C_j)=\varnothing$. 如果 $A\in M_n$ 且 $AB=CA$，那么 A 是与 B 以及 C 分块对角共形的，也就是说，$A=A_1\oplus\cdots\oplus A_k$，且对每个 $i=1,\cdots,k$ 有 $A_iB_i=C_iA_i$.

证明 与 B 以及 C 共形地分划 $A=[A_{ij}]$. 那么 $AB=CA$，当且仅当 $A_{ij}B_j=C_iA_{ij}$. 如

果 $i\neq j$，那么(2.4.4.1)就保证有 $A_{ij}=0$.

一个值得记住的基本原则是：如果 $AX=XB$，且如果对于 A 和 B 的结构存在某种特殊性，那么关于 X 的结构也就可能存在某种特殊性. 通过用标准型代替 A 与 B，并研究所产生的与标准型以及被变换的 X 有关的缠绕关系，我们或许能够发现特殊的结构是什么. 下面的推论是这种类型的结果的一个例子. □

推论 2.4.4.3 设 A，$B\in M_n$. 假设存在一个非奇异的 $S\in M_n$，使得 $A=S(A_1\oplus\cdots\oplus A_d)S^{-1}$，其中每个 $A_j\in M_{n_j}(j=1,\cdots,d)$，且只要 $i\neq j$，就有 $\sigma(A_i)\cap\sigma(A_j)=\varnothing$. 那么，$AB=BA$ 当且仅当 $B=S(B_1\oplus\cdots\oplus B_d)S^{-1}$，其中每个 $B_j\in M_{n_j}(j=1,\cdots,d)$，又对每个 $i=1,\cdots,d$，有 $A_iB_i=B_iA_i$.

证明 如果 A 与 B 可交换，那么$(S^{-1}AS)(S^{-1}BS)=(S^{-1}BS)(S^{-1}AS)$，所以结论中 $S^{-1}BS$ 的直和分解即由上面的推论得出. 而其逆则由计算得出.

在上一结果的通常应用中，每一个矩阵 A_i 都有单独一个特征值，且它常常是一个纯量矩阵：$A_i=\lambda_iI_{n_i}$. □ 112

2.4.5 Schur 三角化定理中的唯一性

对给定的 $A\in M_n$，(2.3.1)中描述的那种可以通过酉相似得到的上三角型 T 不一定是唯一的. 也就是说，有相同主对角线的不同的上三角矩阵可能是酉相似的.

如果 T，$T'\in M_n$ 是上三角的，且有相同的主对角线，主对角线上相同的元素归并在一起，关于使得 $T'=WTW^*$(也就是 $WT=T'W$)成立的酉矩阵 $W\in M_n$，能说些什么呢？下面的定理说的是：W 必定是分块对角的，而且在关于 T 的超对角线元素的某种假设之下，W 必定是对角矩阵，甚至是一个纯量矩阵. 在后一种情形有 $T=T'$.

定理 2.4.5.1 设 n，d，n_1，$\cdots$，n_d 是正整数，它们满足 $n_1+\cdots+n_d=n$. 设 $\Lambda=\lambda_1I_{n_1}\oplus\cdots\oplus\lambda_dI_{n_d}\in M_n$，其中 $\lambda_i\neq\lambda_j$(如果 $i\neq j$). 设 $T=[t_{ij}]\in M_n$ 以及 $T'=[t'_{ij}]\in M_n$ 是与 Λ 有相同主对角线的上三角矩阵. 与 Λ 共形地分划 $T=[T_{ij}]_{i,j=1}^d$，$T'=[T'_{ij}]_{i,j=1}^d$ 以及 $W=[W_{ij}]_{i,j=1}^d\in M_n$. 假设 $WT=T'W$.

(a) 如果 $i>j$ 就有 $W_{ij}=0$，也就是说，W 是与 Λ 共形的分块上三角矩阵.

(b) 如果 W 是酉矩阵，那么它是与 Λ 共形的分块对角矩阵：$W=W_{11}\oplus\cdots\oplus W_{dd}$.

(c) 假设每一个块 T_{11}，$\cdots$，T_{dd} 的第一条超对角线的每个元素都不是零. 那么 W 是上三角的. 如果 W 是酉矩阵，那么它是对角矩阵：$W=\mathrm{diag}(w_1,\cdots,w_n)$.

(d) 如果 W 是酉矩阵，且对每个 $i=1,\cdots,n-1$ 有 $t_{i,i+1}>0$ 以及 $t'_{i,i+1}>0$，那么 W 是一个纯量酉矩阵：$W=wI$. 在此情形有 $T=T'$.

证明 (a)如果 $d=1$，就没有什么要证明的了，故而可以假设 $d\geq2$. 我们的策略是研究恒等式 $WT=T'W$ 两边对应的块的相等. WT 的位于$(d,1)$处的分块是 $W_{d1}T_{11}$，而 $T'W$ 的位于$(d,1)$处的分块是 $T'_{dd}W_{d1}$. 由于 $\sigma(T_{11})$与 $\sigma(T'_{dd})$是不相交的，(2.4.4.1)就确保 $W_{d1}=0$ 是 $W_{d1}T_{11}=T'_{dd}W_{d1}$ 的仅有的解. 如果 $d=2$，我们就此停止. 如果 $d>2$，那么 WT 的位于$(d,2)$处的分块是 $W_{d2}T_{22}$(由于 $W_{d1}=0$)，而 $T'W$ 的位于$(d,2)$处的分块是 $T'_{dd}W_{d2}$，我们就有 $W_{d2}T_{22}=T'_{dd}W_{d2}$. (2.4.4.1)再次确保 $W_{d2}=0$，这是因为 $\sigma(T_{22})$与 $\sigma(T'_{dd})$是不相交的. 按照此程序做下去，越过 $TW=WT'$ 的第 d 个分块行，我们发现 W_{d1}，$\cdots$，

$W_{d,d-1}$全都是零．现在，对于$k=1$，…，$d-2$，让$WT=T'W$的位于$(d-1, k)$，位置上的分块相等，我们就用同样的方法推断出$W_{d-1,1}$，…，$W_{d-1,d-2}$全都是零．从左到右，一直向上做完$WT=T'W$的分块行，我们就得出结论：对所有$i>j$都有$W_{ij}=0$．

(b) 现在假设W是酉矩阵．分划$W=\begin{bmatrix} W_{11} & X \\ 0 & \hat{W} \end{bmatrix}$并由(2.1.10)推断有$X=0$．由于$\hat{W}$也是分块上三角的，且还是酉矩阵，归纳法就导致结论：$W=W_{11}\oplus\cdots\oplus W_{dd}$，见(2.5.2)．

(c) 我们有d个恒等式$W_{ii}T_{ii}=T'_{ii}W_{ii}$，$i=1$，…，$d$，又我们假设，每个$T_{ij}$的第一条超对角线的所有元素都不是零．这样的话，只需要考虑$d=1$的情形就够了：$T=[t_{ij}]\in M_n$与$T'=[t'_{ij}]\in M_n$是上三角的，对所有$i=1$，…，n有$t_{ii}=t'_{ii}=\lambda$，而对所有$i=1$，…，$n-1$，$t_{i,i+1}$都不为零，又有$WT=T'W$．如同在(a)中那样，让恒等式$WT=T'W$对应的元素相等：在位置$(n, 2)$我们有$w_{n1}t_{12}+w_{n2}\lambda=\lambda w_{n2}$或者$w_{n1}t_{12}=0$；由于$t_{12}\neq 0$，由此推出 $w_{n1}=0$．由此往前直至越过$WT=T'W$的第n行，我们就得到一列恒等式$w_{ni}t_{i,i+1}+w_{n,i+1}\lambda=\lambda w_{n,i+1}$ $(i=1, \cdots, n-1)$，由此推出$w_{ni}=0$(对所有$i=1$，…，$n-1$)．以这种方式从左向右顺着$WT=T'W$的行一直往上做下去，我们发现对所有$i>j$都有$w_{ij}=0$．从而W是上三角的，如果它是酉矩阵，那么(b)中的论证确保它是对角的．

(d) 假设条件与(c)确保$W=\mathrm{diag}(w_1, \cdots, w_n)$是对角的酉矩阵．让$WT=T'W$处于位置$(i, i+1)$处的元素相等，我们就有$w_i t_{i,i+1}=t'_{i,i+1}w_{i+1}$，所以$t_{i,i+1}/t'_{i,i+1}=w_{i+1}/w_i$，它是正的实数，且它的模为1．我们得出结论：对每个$i=1$，…，$n-1$都有$w_{i+1}/w_i=1$，从而$w_1=\cdots=w_n$以及$W=w_{11}I$． □

习题 假设A，$B\in M_n$是酉相似的(通过纯量酉矩阵)．说明为什么$A=B$． ◀

习题 假设T，$T'\in M_n$是上三角的，且有同样的主对角线(主对角线上的元素各不相同)，又它们通过一个酉矩阵$U\in M_n$而相似．说明为什么U必定是对角的．如果T与T'的第一条超对角线的所有元素都是正的实数，说明为什么$T=T'$． ◀

习题 如果$A=[a_{ij}]\in M_n$，且对每一个$i=1$，…，$n-1$有$a_{i,i+1}\neq 0$．证明：存在一个对角酉矩阵D，使得DAD^*的第一条超对角线上的元素是正的实数．提示：考虑$D=\mathrm{diag}(1, a_{12}/|a_{12}|, a_{12}a_{23}/|a_{12}a_{23}|, \cdots)$． ◀

2.4.6 每一个方阵都可以分块对角化

(2.3.1)的如下的应用以及拓广是通向Jordan标准型的重要一步，我们将在下一章里来讨论Jordan标准型．

定理 2.4.6.1 设$A\in M_n$的不同的特征值是λ_1，…，λ_d，重数分别为n_1，…，n_d．定理2.3.1确保A与一个$d\times d$分块上三角矩阵$T=[T_{ij}]_{i,j=1}^d$酉相似，其中每一个分块T_{ij}是$n_i\times n_j$的，如果$i>j$，则有$T_{ij}=0$，且每一个对角分块T_{ii}都是上三角的，其对角元素为λ_i，也就是说，每一个$T_{ii}=\lambda_i I_{n_i}+R_i$，且$R_i\in M_{n_i}$是严格上三角的．这样$A$就相似于

$$\begin{bmatrix} T_{11} & & & 0 \\ & T_{22} & & \\ & & \ddots & \\ 0 & & & T_{dd} \end{bmatrix} \qquad (2.4.6.2)$$

如果 $A\in M_n(\mathbf{R})$，且它所有的特征值都是实的，那么将 A 化简为特殊的上三角型 T 的酉相似以及将 T 化简为分块对角型(2.4.6.2)的相似矩阵这两者都可以取为实的.

证明 将 T 分划为

$$T=\begin{bmatrix} T_{11} & Y \\ 0 & S_2 \end{bmatrix}$$

其中 $S_2=[T_{ij}]_{i,j=2}^d$. 注意 T_{11} 的仅有的特征值是 λ_1，而 S_2 的特征值是 λ_2，…，λ_n. Sylvester 定理 2.4.4.1 保证了方程 $T_{11}X-XS=-Y$ 有一个解 X，用它来构造 114

$$M=\begin{bmatrix} I_{n_1} & X \\ 0 & I \end{bmatrix} \quad 以及其逆 \quad M^{-1}=\begin{bmatrix} I_{n_1} & -X \\ 0 & I \end{bmatrix}$$

那么

$$M^{-1}TM=\begin{bmatrix} I_{n_1} & -X \\ 0 & I \end{bmatrix}\begin{bmatrix} T_{11} & Y \\ 0 & S_2 \end{bmatrix}\begin{bmatrix} I_{n_1} & X \\ 0 & I \end{bmatrix}=\begin{bmatrix} T_{11} & T_{11}X-XS_2+Y \\ 0 & S_2 \end{bmatrix}=\begin{bmatrix} T_{11} & 0 \\ 0 & S_2 \end{bmatrix}$$

如果 $d=2$，这就是所要的分块对角化. 如果 $d>2$，重复这一化简过程来证明 S_2 与 $T_{22}\oplus S_3$ 相似，其中 $S_3=[T_{ij}]_{i,j=3}^d$. 经过 $d-1$ 次化简，我们就得知 T 相似于 $T_{11}\oplus\cdots\oplus T_{dd}$.

如果 A 是实的且有实特征值，那么它与一个刚刚考虑过的实的分块上三角矩阵实正交相似. 每一步的化简都可以用实相似来实现. □

习题 假设 $A\in M_n$ 与一个 $d\times d$ 分块上三角矩阵 $T=[T_{ij}]_{i,j=1}^d$ 酉相似. 如果满足 $j>i$ 的任何分块 T_{ij} 都不是零，利用(2.2.2)说明为什么 T 不与 $T_{11}\oplus\cdots\oplus T_{dd}$ 酉相似. ◀

上一定理有两个拓广的结果，一是对交换族，二是对同时相似(但不一定是酉相似)，这大大改善了(2.3.3)中得到的分块结构.

定理 2.4.6.3 设 $\mathcal{F}\subset M_n$ 是一个交换族，令 A_0 是 $\mathcal{F}$ 中任意一个给定的矩阵，又假设 A_0 有 d 个不同的特征值 λ_1，…，λ_d，重数分别为 n_1，…，n_d. 那么就存在一个非奇异的 $S\in M_n$，使得

(a) $\hat{A}_0=S^{-1}A_0S=T_1\oplus\cdots\oplus T_d$，其中每个 $T_i\in M_{n_i}$ 都是上三角的，且它所有的对角元素都是 λ_i；

(b) 对每个 $A\in\mathcal{F}$，$S^{-1}AS$ 都是上三角的，且是与 $\hat{A}_0$ 共形的分块对角矩阵.

证明 首先利用(2.4.6.1)选取一个非奇异的 S_0，使得 $S_0^{-1}A_0S_0=R_1\oplus\cdots\oplus R_d=\widetilde{A}_0$，其中每一个 $R_i\in M_{n_i}$ 都以 λ_i 作为它仅有的特征值. 设 $S_0^{-1}\mathcal{F}S_0=\{S_0^{-1}AS_0: A\in\mathcal{F}\}$，它也是一个交换族. 将任意给定的 $B\in S_0^{-1}\mathcal{F}S_0$ 与 $\widetilde{A}_0$ 共形地分划成 $B=[B_{ij}]_{i,j=1}^d$. 那么 $[R_i \quad B_{ij}]=\widetilde{A}_0B=B\widetilde{A}_0=[B_{ij} \quad R_j]$，所以对所有 i，$j=1$，…，d 都有 $R_iB_{ij}=B_{ij}R_j$. 现在 Sylvester 定理 2.4.4.1 确保对所有 $i\neq j$ 都有 $B_{ij}=0$，这是因为 R_i 与 R_j 没有共同的特征值. 从而 $S_0^{-1}\mathcal{F}S_0$ 是一个由全都与 $\widetilde{A}_0$ 共形的分块对角矩阵组成的交换族. 对每一个 $i=1$，…，d，考虑族 $\mathcal{F}_i\subset M_{n_i}$，它是由 $S_0^{-1}\mathcal{F}S_0$ 中每一个矩阵的第 i 个对角分块组成的；注意到对每个 $i=1$，…，d，有 $R_i\in\mathcal{F}_i$. 每一个 $\mathcal{F}_i$ 都是一个交换族，所以(2.3.3)确保存在一个酉矩阵 $U_i\in M_{n_i}$，使得 $U_i^*\mathcal{F}_iU_i$ 是一个上三角族. $U_i^*R_iU_i$ 的主对角元素是它的特征值，它们全都等于 λ_i. 设 $U=U_1\oplus\cdots\oplus U_d$ 并注意到 $S=S_0U$ 就完成了结论中所断言的化简，其中 $T_i=U_i^*R_iU_i$. □

推论 2.4.6.4 设 $\mathcal{F}\subset M_n$ 是一个交换族. 则存在一个非奇异的 $S\in M_n$ 以及正整数 k， 115 n_1，…，n_k，使得 $n_1+\cdots+n_k=n$，且对每个 $A\in\mathcal{F}$，$S^{-1}AS=A_1\oplus\cdots\oplus A_k$(其中 $A_i\in M_{n_i}$)都是分块对角的(对每个 $i=1$，…，k)．此外，每个对角分块 A_i 都是上三角的且恰好只有一个特征值．

证明 如果$\mathcal{F}$中每个矩阵都仅有一个特征值，运用(2.3.3)并停止．如果$\mathcal{F}$中某个矩阵至少有两个不同的特征值，设 $A_0\in\mathcal{F}$是在$\mathcal{F}$的所有矩阵之中任意一个有最大数目的不同特征值的矩阵．如同在上一个定理中那样，构造一个同时分块对角的上三角化，并注意得到的每一个对角分块的大小都严格小于 A_0 的大小．与 A_0 的化简的形式中的每一个对角块相伴的是一个交换的矩阵族．在那个族的成员之中，要么(a)每一个矩阵都只有一个特征值(无须进一步的化简)，要么(b)某个矩阵至少有两个不同的特征值，在此情形，我们选取任何一个有最大个数不同特征值的矩阵，并再次化简它，以得到一组严格更小的对角分块．循环重复这一化简，它必定在有限多步之后终止，一直到任何交换族中没有任何成员能有多于一个特征值为止． □

2.4.7 每个方阵都是几乎可以对角化的

Schur 的结果的另一个用途是用两种可能的解释来使得每个复方阵都是“几乎可以对角化的”这一说法变得明晰．第一种解释是说存在一个可对角化的矩阵，它任意接近给定的矩阵；第二种解释是说任意给定的矩阵都相似于一个上三角矩阵，这个上三角矩阵位于对角线之外的元素任意地小．

定理 2.4.7.1 设 $A=[a_{ij}]\in M_n$．对每个 $\varepsilon>0$，存在一个矩阵 $A(\varepsilon)=[a_{ij}(\varepsilon)]\in M_n$，它有 n 个不同的特征值(于是它可以对角化)，且使得 $\sum_{i,j=1}^{n}|a_{ij}-a_{ij}(\varepsilon)|^2<\varepsilon$．

证明 设 $U\in M_n$ 是酉矩阵，它使得 $U^*AU=T$ 是上三角的．设 $E=\mathrm{diag}(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n)$，其中 ε_1，ε_2，…，ε_n 这样来选取，使得 $|\varepsilon_i|<\left(\dfrac{\varepsilon}{n}\right)^{1/2}$，所以对所有 $i\neq j$ 都有 $t_{ii}+\varepsilon_i\neq t_{jj}+\varepsilon_j$．(考虑一下可以看出这可以做到．)这样一来 $T+E$ 就有 n 个不同的特征值 $t_{11}+\varepsilon_1$，…，$t_{nn}+\varepsilon_n$，故而 $A+UEU^*$ 亦然，而后者与 $T+E$ 相似．设 $A(\varepsilon)=A+UEU^*$，所以 $A-A(\varepsilon)=-UEU^*$，从而(2.2.2)确保有 $\sum_{i,j}|a_{ij}-a_{ij}(\varepsilon)|^2=\sum_{i=1}^{n}|\varepsilon_i|^2<n\left(\dfrac{\varepsilon}{n}\right)=\varepsilon$． □

习题 证明：(2.4.6)中的条件 $\sum_{i,j}|a_{ij}-a_{ij}(\varepsilon)|^2<\varepsilon$ 可以用 $\max_{i,j}|a_{ij}-a_{ij}(\varepsilon)|<\varepsilon$ 来代替．提示：用 ε^2 代替 ε 来应用定理并完成之，如果一个平方和小于 ε^2，那么其中的每一项之绝对值必定小于 ε．

定理 2.4.7.2 设 $A\in M_n$．对每个 $\varepsilon>0$，存在一个非奇异的矩阵 $S_\varepsilon\in M_n$，使得 $S_\varepsilon^{-1}AS_\varepsilon=T_\varepsilon=[t_{ij}(\varepsilon)]$是上三角的，且对所有满足 $i<j$ 的 i，$j\in\{1,\cdots,n\}$，都有 $|t_{ij}(\varepsilon)|\leqslant\varepsilon$．

证明 首先应用 Schur 定理作出一个酉矩阵 $U\in M_n$ 以及一个上三角矩阵 $T\in M_n$，使得 $U^*AU=T$．对一个非零的纯量 α，定义 $D_\alpha=\mathrm{diag}(1,\alpha,\alpha^2,\cdots,\alpha^{n-1})$，并置 $t=\max_{i<j}|t_{ij}|$．假设 $\varepsilon<1$，这是由于这肯定足以用来证明在此情形下的命题．如果 $t\leqslant1$，令 $S_\varepsilon=UD_\varepsilon$；如果
116 $t>1$，则令 $S_\varepsilon=UD_{1/t}D_\varepsilon$．在随便哪种情形，合适的 S_ε 都说明定理的结论是正确的．如果 $t\leqslant1$，计算表明 $t_{ij}(\varepsilon)=t_{ij}\varepsilon^{-i}\varepsilon^{j}=t_{ij}\varepsilon^{j-i}$，其绝对值不大于 ε^{j-i}，如果 $i<j$，这自然不大于 ε．如果$t>1$，通过 $D_{1/t}$所作的相似对矩阵预处理，就得到一个矩阵，它的位于对角线之外的元素的

绝对值均不大于 1. □

习题 证明(2.4.7.2)的如下变形：如果 $A\in M_n$ 且 $\varepsilon>0$，则存在一个非奇异的 $S_\varepsilon\in M_n$，使得 $S_\varepsilon^{-1}AS_\varepsilon=T_\varepsilon=[t_{ij}(\varepsilon)]$是上三角的，且 $\sum\limits_{j>i}|t_{ij}(\varepsilon)|\leqslant\varepsilon$. 提示：应用(2.4.7)，在其中用$[2/n(n-1)]\varepsilon$ 代替 ε. ◀

2.4.8 交换族以及同时三角化

现在要利用 Schur 定理的交换族形式(2.3.3)来证明：对于可交换的矩阵，其特征值可以按照某种秩序“相加”以及“相乘”.

定理 2.4.8.1 假设 $A, B\in M_n$ 可交换. 那么就存在 A 的特征值的一个排序 $\alpha_1, \cdots, \alpha_n$ 以及 B 的特征值的一个排序 $\beta_1, \cdots, \beta_n$，使得 $A+B$ 的特征值是 $\alpha_1+\beta_1, \cdots, \alpha_n+\beta_n$，且 AB 的特征值是 $\alpha_1\beta_1, \cdots, \alpha_n\beta_n$. 特别地，$\sigma(A+B)\subseteq\sigma(A)+\sigma(B)$，$\sigma(AB)\subseteq\sigma(A)\sigma(B)$.

证明 由于 A 与 B 可交换，(2.3.3)确保存在一个酉矩阵 $U\in M_n$，使得 $U^*AU=T=[t_{ij}]$以及 $U^*BU=R=[r_{ij}]$这两者都是上三角的. 上三角矩阵 $T+R=U^*(A+B)U$ 的主对角元素(因此也就是其特征值)就是 $t_{11}+r_{11}, \cdots, t_{nn}+r_{nn}$，这些就是 $A+B$ 的特征值，这是因为 $A+B$ 与 $T+R$ 相似. 上三角矩阵 $TR=U^*(AB)U$ 的主对角元素(因此也就是其特征值)是 $t_{11}r_{11}, \cdots, t_{nn}r_{nn}$，这些就是 AB 的特征值，它是与 TR 相似的. □

习题 假设 $A, B\in M_n$ 可交换. 说明为什么 $\rho(A+B)\leqslant\rho(A)+\rho(B)$以及 $\rho(AB)\leqslant\rho(A)+\rho(B)$，所以谱半径函数对于可交换的矩阵是**次加性的**(subadditive)且是**次积性的**(submultiplicative). ◀

例 2.4.8.2 即使 A 与 B 可交换，它们的特征值的每一对相应的和也不一定都是 $A+B$ 的特征值. 考虑对角矩阵

$$A=\begin{bmatrix}1 & 0\\ 0 & 2\end{bmatrix}\quad 以及\quad B=\begin{bmatrix}3 & 0\\ 0 & 4\end{bmatrix}$$

由于 $1+4=5\notin\{4, 6\}=\sigma(A+B)$，我们看到 $\sigma(A+B)$包含在 $\sigma(A)+\sigma(B)$中，但并不与之相等.

例 2.4.8.3 如果 A 与 B **不可**交换，很难说 $\sigma(A+B)$与 $\sigma(A)$以及 $\sigma(B)$有何种关联. 特别地，$\sigma(A+B)$不一定包含在 $\sigma(A)+\sigma(B)$中. 设

$$A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\quad 以及\quad B=\begin{bmatrix}0 & 0\\ 1 & 0\end{bmatrix}$$

这样就有 $\sigma(A+B)=\{-1, 1\}$，而 $\sigma(A)=\sigma(B)=\{0\}$.

习题 考虑上一个例子中的矩阵. 说明为什么 $\rho(A+B)>\rho(A)+\rho(B)$，所以 M_n 上的谱半径函数不是次加性的. ◀ 117

例 2.4.8.4 是否有(2.4.8.1)的逆命题存在？如果将 A 与 B 的特征值按照某种次序相加，A 与 B 必须可交换吗？答案是否定的，即使对所有纯量 α 和 β，按照某种次序将 αA 与 βB 的特征值相加亦然如此. 这是一个有意思的现象，刻画这样一对矩阵的特征是一个尚未解决的问题！考虑不可交换的矩阵

$$A=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & -1\\ 0 & 0 & 0\end{bmatrix}\quad 以及\quad B=\begin{bmatrix}0 & 0 & 0\\ 1 & 0 & 0\\ 0 & 1 & 0\end{bmatrix}$$

对它们有 $\sigma(A)=\sigma(B)=\{0\}$. 此外，$p_{\alpha A+\beta B}(t)=t^3$，所以，对所有 α，$\beta\in\mathbf{C}$ 都有 $\sigma(\alpha A+\beta B)=\{0\}$，且特征值可以相加. 如果 A 与 B 可以同时上三角化，则(2.4.8.1)的证明表明：AB 的特征值就是 A 与 B 的特征值按照某种次序的乘积. 然而，$\sigma(AB)=\{-1, 0, 1\}$，它不包含在 $\sigma(A)\cdot\sigma(B)=\{0\}$中，故而 A 与 B 不可以同时三角化.

推论 2.4.8.5 假设 A，$B\in M_n$ 可交换，$\sigma(A)=\{\alpha_1, \cdots, \alpha_{d_1}\}$，而 $\sigma(B)=\{\beta_1, \cdots, \beta_{d_2}\}$. 如果对所有 i，j 都有 $\alpha_i\neq-\beta_j$，那么 $A+B$ 是非奇异的.

习题 利用(2.4.8.1)验证(2.4.8.5). ◀

习题 假设 $T=[t_{ij}]$以及 $R=[r_{ij}]$是同样大小的 $n\times n$ 上三角矩阵，又设 $p(s, t)$是关于两个非交换变量的多项式，也就是说，它是关于两个非交换变量的字的一个任意的线性组合. 说明为什么 $p(T, R)$是上三角的，且它的主对角元素(即它的特征值)是 $p(t_{11}, t_{11})$，…，$p(t_{nn}, r_{nn})$. ◀

对复矩阵而言，同时三角化与同时酉三角化是等价的概念.

定理 2.4.8.6 设给定 A_1，…，$A_m\in M_n$. 则存在一个非奇异的 $S\in M_n$，使得对每一个 $i=1, \cdots, m$，$S^{-1}A_iS$ 都是上三角的，当且仅当存在一个酉矩阵 $U\in M_n$，使得对所有 $i=1, \cdots, m$，U^*A_iU 都是上三角的.

证明 利用(2.1.14)记 $S=QR$，其中 Q 是酉矩阵，而 R 是上三角矩阵. 那么 $T_i=S^{-1}A_iS=(QR)^{-1}A_i(QR)=R^{-1}(Q^*A_iQ)R$ 是上三角矩阵，所以，作为三个上三角矩阵的乘积，$Q^*A_iQ=RT_iR^{-1}$ 也是上三角矩阵. □

用相似性对 m 个矩阵同时上三角化的特征已经由下面 McCoy 的定理给出了完全的刻画. 该定理涉及一个有 m 个非交换变量的多项式 $p(t_1, \cdots, t_m)$，它是这些变量的幂的乘积的线性组合，也就是说，它是关于 m 个非交换变量的字的线性组合. 关键点掌控在上一个习题中：如果 T_1，…，T_m 是上三角的，那么 $p(T_1, \cdots, T_m)$亦然，而且 T_1，…，T_m 与 $p(T_1, \cdots, T_m)$的主对角线展现出它们的特征值的特定的次序. 对每个 $k=1, \cdots, n$，$p(T_1, \cdots, T_m)$的第 k 个主对角元素(它是 $p(T_1, \cdots, T_m)$的一个特征值)分别是关于

118 T_1，…，T_m的第 k 个主对角元素的同样的多项式.

定理 2.4.8.7(McCoy) 设给定 $m\geqslant 2$ 以及 A_1，…，$A_m\in M_n$. 则下面的命题是等价的.

(a) 对关于 m 个非交换变量的每个多项式 $p(t_1, \cdots, t_m)$以及每一个 k，$\ell=1, \cdots, m$，$p(A_1, \cdots, A_m)(A_kA_\ell-A_\ell A_k)$都是幂零的.

(b) 存在一个酉矩阵 $U\in M_n$，使得 U^*A_iU 对每个 $i=1, \cdots, m$ 都是上三角的.

(c) 每个矩阵 A_i 的特征值都存在一个排序 $\lambda_1^{(i)}$，…，$\lambda_n^{(i)}$ $(i=1, \cdots, m)$，使得关于 m 个非交换变量的任何一个多项式 $p(t_1, \cdots, t_m)$，$p(A_1, \cdots, A_m)$的特征值都是 $p(\lambda_i^{(1)}, \cdots, \lambda_i^{(m)})$(对 $i=1, \cdots, n$).

证明 (b)⇒(c)：设 $T_k=U^*A_kU=[t_{ij}^{(k)}]$是上三角的，并设 $\lambda_1^{(k)}=t_{11}^{(k)}$，…，$\lambda_n^{(k)}=t_{nn}^{(k)}$. 那么 $p(A_1, \cdots, A_m)=p(UT_1U^*, \cdots, UT_mU^*)=Up(T_1, \cdots, T_m)U^*$ 的特征值是 $p(T_1, \cdots, T_m)$的主对角元素，即是 $p(\lambda_i^{(1)}, \cdots, \lambda_i^{(m)})$(对 $i=1, \cdots, n$).

(c)⇒(a)：对关于 m 个非交换变量的任意一个给定的多项式 $p(t_1, \cdots, t_m)$，考虑关于 m 个非交换变量的多项式 $q_{k\ell}(t_1, \cdots, t_m)=p(t_1, \cdots, t_m)(t_kt_\ell-t_\ell t_k)$(对 k，$\ell=1$，…，

m). 根据(c)$q_{k\ell}(A_1, \cdots, A_m)$的特征值是 $q_{k\ell}(\lambda_i^{(1)}, \cdots, \lambda_i^{(m)}) = p(\lambda_i^{(1)}, \cdots, \lambda_i^{(m)})(\lambda_i^{(k)}\lambda_i^{(\ell)} - \lambda_i^{(\ell)}\lambda_i^{(k)}) = p(\lambda_i^{(1)}, \cdots, \lambda_i^{(m)}) \times 0 = 0$(对所有 $i=1, \cdots, n$). 于是，每一个矩阵 $p(A_1, \cdots, A_m)(A_kA_\ell - A_\ell A_k)$都是幂零的，见(2.4.2).

(a)⇒(b)：假设(见下面的引理)A_1，…，A_m 有公共的单位特征向量 x. 在此假设下，我们如同在(2.3.3)的证明中那样用归纳法来进行. 设 U_1 是任意一个以 x 为其第一列的酉矩阵. 利用 U_1 依照同样的方法来压缩每一个$\mathcal{A}_i$：

$$\mathcal{A}_i = U_1^* A_i U_1 = \begin{bmatrix} \lambda_1^{(i)} & \bigstar \\ 0 & \widetilde{A}_i \end{bmatrix}, \quad \widetilde{A}_i \in M_{n-1}, i = 1, \cdots, m \qquad (2.4.8.8)$$

设 $p(t_1, \cdots, t_m)$是关于 m 个非交换变量的任意一个给定的多项式. 这样(a)就确保矩阵

$$U^* p(A_1, \cdots, A_m)(A_kA_\ell - A_\ell A_k)U = p(\mathcal{A}_1, \cdots, \mathcal{A}_m)(\mathcal{A}_k \mathcal{A}_\ell - \mathcal{A}_\ell \mathcal{A}_k) \qquad (2.4.8.9)$$

对每个 k，$\ell=1, \cdots, m$ 都是幂零的. 与(2.4.8.8)共形地分划每一个矩阵(2.4.8.9)，并注意到位置(1，1)上的元素是零，而它的右下方的分块是 $p(\widetilde{A}_1, \cdots, \widetilde{A}_m)(\widetilde{A}_k \widetilde{A}_\ell - \widetilde{A}_\ell \widetilde{A}_k)$，它必定是幂零的. 从而，矩阵$\widetilde{A}_1$，…，$\widetilde{A}_m \in M_{n-1}$也具有性质(a)。如同在(2.3.3)中那样，由归纳法就得出(b). □

我们知道交换矩阵总是有一个公共的特征向量(1.3.19). 如果上一个定理中的矩阵 A_1，…，A_m可交换，则条件(a)显然满足，这是因为对所有 k，$\ell=1, \cdots, m$ 都有 $p(A_1, \cdots, A_m)(A_kA_\ell - A_\ell A_k) = 0$. 下面的引理表明：条件(a)(它弱于交换性)已足以保证公共特征向量的存在.

引理 2.4.8.10 设给定 $A_1, \cdots, A_m \in M_n$. 假设对关于 $m \geqslant 2$ 个非交换变量的每一个多项式 $p(t_1, \cdots, t_m)$，以及对每个 k，$\ell=1, \cdots, m$，矩阵 $p(A_1, \cdots, A_m)(A_kA_\ell - A_\ell A_k)$中的每一个都是幂零的. 那么，对每个给定的非零向量 $x \in \mathbf{C}^n$，都存在关于 m 个非交换变量的一个多项式 $q(t_1, \cdots, t_m)$，使得 $q(A_1, \cdots, A_m)x$ 是 A_1，…，A_m 的一个公共特征向量. 119

证明 我们只考虑 $m=2$ 的情形，它展现了一般情形的所有特征. 设 $A, B \in M_n$，$C = AB - BA$，又假设 $p(A, B)C$ 对每一个关于两个非交换变量的多项式 $p(s, t)$都是幂零的. 设 $x \in \mathbf{C}^n$ 是任意一个给定的非零向量. 我们断言：存在关于两个非交换变量的多项式 $q(s, t)$，使得 $q(A, B)x$ 是 A 与 B 的公共特征向量.

从(1.1.9)开始，并且令 $g_1(t)$是一个多项式，它使得 $\xi_1 = g_1(A)x$ 是 A 的一个特征向量：$A\xi_1 = \lambda\xi_1$.

情形Ⅰ：假设对每个多项式 $p(t)$有 $Cp(B)\xi_1 = 0$，也就是说对**每个**多项式 $p(t)$有

$$ABp(B)\xi_1 = BAp(B)\xi_1 \qquad (2.4.8.11)$$

对 $p(t)=1$ 利用这个恒等式即给出 $AB\xi_1 = BA\xi_1$. 现在用归纳法：假设对某个 $k \geqslant 1$ 已经有 $AB^k\xi_1 = B^kA\xi_1$. 利用(2.4.8.11)以及归纳假设，我们计算

$$AB^{k+1}\xi_1 = AB \cdot B^k\xi_1 = BA \cdot B^k\xi_1 = B \cdot AB^k\xi_1 = B \cdot B^kA\xi_1 = B^{k+1}A\xi_1$$

我们断定，对每个 $k \geqslant 1$ 都有 $AB^k\xi_1 = B^kA\xi_1$. 从而对每个多项式 $p(t)$都有 $Ap(B)\xi_1 = p(B)A\xi_1 = p(B)\lambda\xi_1 = \lambda(p(B)\xi_1)$. 于是，如果 $p(B)\xi_1$ 不为零，它就是 A 的一个特征向量. 再次利用(1.1.9)选取一个多项式 $g_2(t)$，使得 $g_2(B)\xi_1 = g_2(B)g_1(A)x$ 是 B 的一个特征向量(一定不为零). 设 $q(s, t) = g_2(t)g_1(s)$. 我们就证明了 $q(A, B)x$ 是 A 与 B 的一个公共特

征向量，如所断言.

情形Ⅱ：假设存在**某个**多项式 $f_1(t)$，使得 $Cf_1(B)\xi_1\neq0$. 利用(1.1.9)来求一个多项式 $q_1(t)$，使得 $\xi_2=q_1(A)Cf_1(B)\xi_1$ 是 A 的一个特征向量. 如果对每个多项式 $p(t)$都有 $Cp(B)\xi_2=0$，那么情形Ⅰ使得我们可以构造出所想要的公共特征向量；否则就设 $f_2(t)$是使得 $Cf_2(B)\xi_2\neq0$ 的一个多项式，并令 $q_2(t)$是这样一个多项式，它使得 $\xi_3=q_2(A)Cf_2(B)\xi_2$ 是 A 的一个特征向量. 继续此法构造出 A 的一列特征向量

$$\xi_k=q_{k-1}(A)Cf_{k-1}(B)\xi_{k-1},\quad k=2,3,\cdots \tag{2.4.8.12}$$

直到要么(i)对某个多项式 $p(t)$有 $Cp(B)\xi_k=0$，或者(ii)$k=n+1$. 如果对某个 $k\leqslant n$ 有情形(i)出现. 情形Ⅰ允许我们构造出所想要的 A 与 B 的公共特征向量. 如果对每一个$k=1$，2，…，n，(i)都是错的，我们的构造就产生出 $n+1$ 个向量 ξ_1，…，ξ_{n+1}，它们必定是线性相关的，故而存在 $n+1$ 个不全为零的纯量 c_1，…，c_{n+1}，使得 $c_1\xi_1+\cdots+c_{n+1}\xi_{n+1}=0$. 设 $r=\min\{i:\ c_i\neq0\}$. 这样就有

$$\begin{aligned}-c_r\xi_r&=\sum_{i=r}^{n}c_{i+1}\xi_{i+1}=\sum_{i=r}^{n}c_{i+1}q_i(A)Cf_i(B)\xi_i\\&=c_{r+1}q_r(A)Cf_r(B)\xi_r+\sum_{i=r}^{n-1}c_{i+2}q_{i+1}(A)Cf_{i+1}(B)\xi_{i+1}\end{aligned} \tag{2.4.8.13}$$

利用(2.4.8.12)，(2.4.8.13)中满足 $i=r$ 的求和项就可以展开成表达式

$$c_{r+2}q_{r+1}(A)Cf_{r+1}(B)q_r(A)Cf_r(B)\xi_r$$

按照同样的方式，我们可以利用(2.4.8.12)将(2.4.8.13)中的每一个求和项(对 $i=r+1$，$r+2$，…，$n-1$)展开成形如 $h_i(A,\ B)Cf_r(B)\xi_r$ 的表达式，其中每一个 $h_i(A,\ B)$都是关于 A 与 B 的多项式. 以这种方法我们得到一个形如$-c_r\xi_r=p(A,\ B)Cf_r(B)\xi_r$ 的恒等式，其中 $p(s,\ t)$是关于两个非交换变量的多项式. 这就意味着 $f_r(B)\xi_r$ 是 $p(A,\ B)C$ 的与非零特征值$-c_r$ 相伴的特征向量，这与 $p(A,\ B)C$ 是幂零矩阵的假设矛盾. 这个矛盾表明(i)对某个 $k\leqslant n$ 为真，从而 A 与 B 有结论中那种形式的公共特征向量. □

我们已经陈述了关于复矩阵的 McCoy 定理 2.4.8.7，但是如果我们重新表述(b)仅要求同时相似(而不是同时酉相似)，那么该定理对 $\mathbf{C}$ 的包含所有矩阵 A_1，…，A_m 的特征值的任意一个子域上的矩阵和多项式都成立.

2.4.9 特征值的连续性

Schur 的酉三角化定理可以用来证明一个基本的然而却有广泛用途的事实：实方阵或者复方阵的特征值连续地依赖于它的元素. Schur 定理的两个方面(酉的和三角的)在证明中起着关键作用. 下面的引理囊括了所涉及的基本原理.

引理 2.4.9.1 设给定一个由矩阵组成的无穷序列 A_1，A_2，$\cdots\in M_n$，并假设 $\lim_{k\to\infty}A_k=A$(逐个按照元素收敛). 那么就存在由正整数组成的一个无穷序列 $k_1<k_2<\cdots$以及一列酉矩阵 $U_{k_i}\in M_n$(对 $i=1$，2，…)，使得

(a) 对所有 $i=1$，2，…，$T_i=U_{k_i}^*A_{k_i}U_{k_i}$ 都是上三角的；

(b) $U=\lim_{i\to\infty}U_{k_i}$ 存在且是酉矩阵；

(c) $T=U^*AU$ 是上三角的；

(d) $\lim_{i\to\infty} T_i = T$.

证明 利用(2.3.1)，对每一个 $k=1,2,\cdots$，设 $U_k\in M_n$ 是酉矩阵，且使得 $U_k^* A U_k$ 是上三角的. 引理 2.1.8 确保存在一个无穷子列 U_{k_1}，U_{k_2}，…以及一个酉矩阵 U，使得 $U_{k_i}\to U$（当 $i\to\infty$时）. 这样一来，它的三个因子中每一个的收敛性就确保乘积 $T_i=U_{k_i}^* A_{k_i} U_{k_i}$ 收敛于一个极限 $T=U^* AU$，它是上三角的，因为每一个 T_i 都是上三角的. □

在上面的推理中，每一个上三角矩阵 T，T_1，T_2，…的主对角线分别是 A，A_{k_1}，A_{k_2}，…的特征值的一个特别的表示(视之为一个 n 元向量). 逐个元素的收敛性 $T_i\to T$ 确保有直到 n！种不同的方式将每一个矩阵 A，A_{k_1}，A_{k_2}，…的特征值表示成为一个 n 元向量，对每一个矩阵，至少有一种表示使得其特征值所对应的向量分别收敛于一个向量，这个向量的元素包含 A 所有的特征值. 正是在这个意义上，在下面的定理里总结成结论：实方阵或者复方阵的特征值连续地依赖于它的元素. 121

定理 2.4.9.2 设给定一个无穷序列 A_1，A_2，…$\in M_n$，并假设 $\lim_{k\to\infty} A_k=A$(逐个按照元素收敛). 设 $\lambda(A)=[\lambda_1(A)\ \cdots\ \lambda_n(A)]^{\mathrm{T}}$ 以及 $\lambda(A_k)=[\lambda_1(A_k)\ \cdots\ \lambda_n(A_k)]^{\mathrm{T}}$ 分别是 A 与 A_k 的特征值的给定的表示(对 $k=1,2,\cdots$). 设 $S_n=\{\pi:\pi$ 是$\{1,2,\cdots,n\}$的一个排列$\}$. 那么对每个给定的 $\varepsilon>0$，存在一个正整数 $N=N(\varepsilon)$，使得

$$\min_{\pi\in S_n}\max_{i=1,\cdots,n}\{|\lambda_{\pi(i)}(A_k)-\lambda_i(A)|\}\leqslant\varepsilon,\quad 对所有\ k\geqslant N \tag{2.4.9.3}$$

证明 如果论断(2.4.9.3)不真，则存在一个 $\varepsilon_0>0$ 以及正整数组成的一个无穷序列 $k_1<k_2<\cdots$，使得对每个 $j=1,2,\cdots$，都有

$$\max_{i=1,\cdots,n}|\lambda_{\pi(i)}(A_{k_j})-\lambda_i(A)|>\varepsilon_0,\quad 对每个\ \pi\in S_n \tag{2.4.9.4}$$

然而，(2.4.9.1)确保存在一个无穷子列 $k_1\leqslant k_{j_1}<k_{j_2}<\cdots$，一列酉矩阵 U，$U_{k_{j_1}}$，$U_{k_{j_2}}$，…，一个上三角矩阵 $T=U^* AU$，以及 $T_p=U_{k_{j_p}}^* A_{k_{j_p}} U_{k_{j_p}}$（对 $p=1,2,\cdots$），使得当 $p\to\infty$时，T_p 的所有元素(特别是它的主对角线元素)都收敛于 T 的对应的元素. 由于 T，T_1，T_2，…的主对角元素组成的向量可以通过对它们的元素进行排列从而分别从特征值 $\lambda(A)$，$\lambda(A_{k_{j_1}})$，$\lambda(A_{k_{j_2}})$，…的给定的表示得到，我们已经看到的逐个元素的收敛性与(2.4.9.4)矛盾，这就证明了定理. □

上一个定理中的存在性结论“对每个给定的 $\varepsilon>0$，存在一个正整数 $N=N(\varepsilon)$”可以用显式的界限取代，见附录 D 中的(D2).

2.4.10 秩 1 摄动的特征值

知道一个矩阵的任意一个特征值都能通过秩 1 摄动来任意移动而不干扰其他的特征值，常常是有用的.

定理 2.4.10.1(A. Brauer) 假设 $A\in M_n$ 有特征值 λ，λ_2，…，λ_n，又设 x 是一个非零向量，它使得 $Ax=\lambda x$. 则对任何 $v\in\mathbf{C}^n$，$A+xv^*$ 的特征值是 $\lambda+v^*x$，λ_2，…，λ_n.

证明 设 $\xi=x/\|x\|_2$，并令 $U=[\xi\ u_2\ \cdots\ u_n]$是酉矩阵. 那么(2.3.1)的证明表明

$$U^* AU=\begin{bmatrix}\lambda & \bigstar\\ 0 & A_1\end{bmatrix}$$

其中 $A_1\in M_{n-1}$有特征值 λ_2，…，λ_n. 又有

$$U^*xv^*U=\begin{bmatrix}\xi^*x\\u_2^*x\\\vdots\\u_n^*x\end{bmatrix}v^*U=\begin{bmatrix}\|x\|_2\\0\\\vdots\\0\end{bmatrix}[v^*\xi\quad v^*u_2\quad\cdots\quad v^*u_n]=\begin{bmatrix}\|x\|_2v^*\xi&\bigstar\\0&0\end{bmatrix}=\begin{bmatrix}v^*x&\bigstar\\0&0\end{bmatrix}$$

这样一来，

$$U^*(A+xv^*)U=\begin{bmatrix}\lambda+v^*x&\bigstar\\0&A_1\end{bmatrix}$$

就有特征值 $\lambda+v^*x$，λ_2，$\cdots\lambda_n$. □

对此结果的一种不同的论证方法，见(1.2.8).

2.4.11 双正交的完备原理

双正交原理(principle biorthogonality)是说与不同特征值相伴的左右特征向量是正交的，见(1.4.7(a)). 我们现在来对左右特征向量阐述所有的可能性.

定理 2.4.11.1 设给定 $A\in M_n$，单位向量 x，$y\in\mathbf{C}^n$，以及 λ，$\mu\in\mathbf{C}$.

(a) 如果 $Ax=\lambda x$，$y^*A=\mu y^*$，且 $\lambda\neq\mu$，这样就有 $y^*x=0$. 设 $U=[x\quad y\quad u_3\quad\cdots\quad u_n]\in M_n$ 是酉矩阵. 那么

$$U^*AU=\begin{bmatrix}\lambda&\bigstar&\bigstar\\0&\mu&0\\0&\bigstar&A_{n-2}\end{bmatrix},\quad A_{n-2}\in M_{n-2}\tag{2.4.11.2}$$

(b) 假设 $Ax=\lambda x$，$y^*A=\lambda y^*$，且 $y^*x=0$. 设 $U=[x\quad y\quad u_3\quad\cdots\quad u_n]\in M_n$ 是酉矩阵. 那么

$$U^*AU=\begin{bmatrix}\lambda&\bigstar&\bigstar\\0&\lambda&0\\0&\bigstar&A_{n-2}\end{bmatrix},\quad A_{n-2}\in M_{n-2}\tag{2.4.11.3}$$

且 λ 的代数重数至少为 2.

(c) 假设 $Ax=\lambda x$，$y^*A=\lambda y^*$，且 $y^*x\neq0$. 设 $S=[x\quad S_1]\in M_n$，其中 S_1 的列是 y 的正交补的任意一组给定的基. 这样 S 就是非奇异的，S^{-*} 的第一列是 y 的非零的纯量倍数，且 $S^{-1}AS$ 有分块形式

$$\begin{bmatrix}\lambda&0\\0&A_{n-1}\end{bmatrix},\quad A_{n-1}\in M_{n-1}\tag{2.4.11.4}$$

如果 λ 的几何重数为 1，那么它的代数重数也为 1. 反之，如果 A 与形如(2.4.11.4)的分块矩阵相似，那么它就有一对与特征值 λ 相伴的非正交的左以及右特征向量.

(d) 假设 $Ax=\lambda x$，$y^*A=\lambda y^*$，且 $x=y$，这样的 x 称为**正规特征向量**(normal eigenvector). 设 $U=[x\quad U_1]\in M_n$ 是酉矩阵. 那么 U^*AU 有分块形式(2.4.11.4).

证明 (a) 与(2.3.1)中的化简比较，(2.4.11.2)的第二行额外多出来的零来自左特征向量 $y^*Au_i=\mu y^*u_i=0$(对 $i=3$，$\cdots$，n).

(b) (2.4.11.3)中零的样式与(2.4.11.2)中的相同，其原因相同. 有关其代数重数的结论，见(1.2.P14).

（c）见（1.4.7）以及（1.4.12）. 123

（d）与（2.3.1）中的化简比较，（2.4.11.4）的第一行中出现额外多出来的零，是因为 x 也是一个左特征向量：$x^*AU_1=\lambda x^*U_1=0$. □

问题

2.4.P1 假设 $A=[a_{ij}]\in M_n$ 有 n 个不同的特征值．利用（2.4.9.2）证明：存在一个 $\varepsilon>0$，使得满足 $\sum_{i,j=1}^{n}|a_{ij}-b_{ij}|^2<\varepsilon$ 的每一个 $B=[b_{ij}]\in M_n$ 都有 n 个不同的特征值．导出结论：有不同特征值的矩阵组成的集合是 M_n 的一个**开子集**.

2.4.P2 为什么一个上三角矩阵的秩至少与它的不为零的主对角元素的个数一样大？设 $A=[a_{ij}]\in M_n$，又假设 A 恰好有有 $k\geqslant 1$ 个非零的特征值 $\lambda_1,\cdots,\lambda_k$．记 $A=UTU^*$，其中 U 是酉矩阵，而 $T=[t_{ij}]$ 是上三角矩阵．证明 $\operatorname{rank}A\geqslant k$，其中的等式当 A 可对角化时成立．说明为什么

$$\left|\sum_{i=1}^{k}\lambda_i\right|^2\leqslant k\sum_{i=1}^{k}|\lambda_i|^2=k\sum_{i=1}^{n}|t_{ii}|^2\leqslant k\sum_{i,j=1}^{n}|t_{ij}|^2=k\sum_{i,j=1}^{k}|a_{ij}|^2$$

并推出结论 $\operatorname{rank}A\geqslant|\operatorname{tr}A|^2/(\operatorname{tr}A^*A)$，其中的等式当且仅当对某个非零的 $a\in\mathbf{C}$ 有 $T=aI_k\oplus 0_{n-k}$ 时成立.

2.4.P3 我们对（2.4.3.2）的证明依赖于复矩阵有特征值这一事实，但是与特征多项式的定义无关，也与涉及特征值或者复数域的任何特殊性质的代换 $p_A(t)\to p_A(A)$ 无关．事实上，Cayley-Hamilton 定理对元素来自**有单位元的交换环**（commutative ring with unit）的矩阵是正确的，以某个整数 k 为模的整数环（当且仅当 k 为素数时它是一个域）以及关于一个或者多个形式不定元的复系数多项式环就是这样的例子．对于下面所给出的（2.4.3.2）的证明给出证明细节．注意：此证明中用到的代数运算涉及加法和乘法，但不涉及除法运算以及求多项式方程的根.

（a）从基本恒等式 $(tI-A)(\operatorname{adj}(tI-A))=\det(tI-A)I=p_A(t)I$（0.8.2）出发，并记

$$p_A(t)I=It^n+a_{n-1}It^{n-1}+a_{n-2}It^{n-2}+\cdots+a_1It+a_0I \tag{2.4.12}$$

它是关于 t 的以矩阵为系数的 n 次多项式，它的每一个系数都是一个纯量矩阵.

（b）说明为什么 $\operatorname{adj}(tI-A)$ 是一个以关于 t 的至多 $n-1$ 次多项式为其元素的矩阵，从而它可以写成形式

$$\operatorname{adj}(tI-A)=A_{n-1}t^{n-1}+A_{n-2}t^{n-2}+\cdots+A_1t+A_0 \tag{2.4.13}$$

其中 $A_0=(-1)^{n-1}\operatorname{adj}A$，而每一个 A_k 都是一个 $n\times n$ 矩阵，该矩阵的元素都是 A 的元素的多项式函数.

（c）利用（2.4.13）将乘积 $(tI-A)(\operatorname{adj}(tI-A))$ 作为

$$A_{n-1}t^n+(A_{n-2}-AA_{n-1})t^{n-1}+\cdots+(A_0-AA_1)t-AA_0 \tag{2.4.14}$$ 124

来计算.

（d）将（2.4.12）与（2.4.14）对应的系数等同起来，从而得到 $n+1$ 个方程

$$\begin{aligned}A_{n-1}&=I\\A_{n-2}-AA_{n-1}&=a_{n-1}I\\&\vdots\\A_0-AA_1&=a_1I\\-AA_0&=a_0I\end{aligned} \tag{2.4.15}$$

（e）对每个 $k=1,\cdots,n$，用 A^{n-k+1} 左乘（2.4.15）中的第 k 个方程，并把全部 $n+1$ 个方程相加，就得到 Cayley-Hamilton 定理 $0=p_A(A)$.

（f）对每个 $k=1,\cdots,n-1$，用 A^{n-k} 左乘（2.4.15）中的第 k 个方程，仅把前 n 个方程相加，就得到恒等式

$$\operatorname{adj}A = (-1)^{n-1}(A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2 A + a_1 I) \tag{2.4.16}$$

于是，adj$A$ 是 A 的多项式，其系数(除 $a_0=(-1)^n\det A$ 之外)与在 $p_A(t)$ 中的系数相同，不过是按照相反的次序.

(g) 利用(2.4.15)证明：(2.4.13)的右边的矩阵系数是 $A_{n-1}=I$ 以及

$$A_{n-k-1} = A^k + a_{n-1}A^{k-1} + \cdots + a_{n-k+1}A + a_{n-k}I \tag{2.4.17}$$

(对 $k=1$，…，$n-1$).

2.4.P4 设 A，$B\in M_n$，并假设 A 与 B 可交换．说明为什么 B 与 adj$A$ 可交换，又为什么 adj$A$ 与 adj$B$ 可交换．如果 A 是非奇异的，推导出 B 与 A^{-1} 可交换.

2.4.P5 考虑矩阵 $\begin{bmatrix}0 & \varepsilon\\ 0 & 0\end{bmatrix}$，并说明为什么有可能存在不可对角化的矩阵，它们可以任意接近于一个给定的可对角化矩阵．利用(2.4.P1)说明为什么在给定的矩阵有不同的特征值时，这种情形不可能发生.

2.4.P6 证明：对

$$A=\begin{bmatrix}1&0&0\\0&2&0\\0&0&3\end{bmatrix} \quad 以及 \quad B=\begin{bmatrix}-2&1&2\\-1&-2&-1\\1&1&1\end{bmatrix}$$

有 $\sigma(aA+bB)=\{a-2b, 2a-2b, 3a+b\}$(对所有纯量 a，$b\in\mathbf{C}$)，但是 A 与 B 不同时与上三角矩阵相似．AB 的特征值是什么？

2.4.P7 利用(2.3.P6)中的判别法证明：(2.4.8.4)中的两个矩阵不可能同时上三角化．对(2.4.P6)中的两个矩阵应用同样的判别法.

2.4.P8 对 McCoy 定理本质的观察在证明两个矩阵**并非**酉相似这方面有时是有用的．设 $p(t, s)$是关于两个非交换变量的复系数多项式，又设 A，$B\in M_n$ 是酉相似的，其中 $A=UBU^*$(对某个酉矩阵 $U\in M_n$)．说明为什么 $p(A, A^*)=Up(B, B^*)U^*$．导出结论：如果 A 与 B 是酉相似的，那么对关于两个非交换变量的每一个复多项式 $p(t, s)$，都有 tr$p(A, A^*)$ = tr$p(B, B^*)$．这与

125 (2.2.6)有何关系？

2.4.P9 设 $p(t)=t^n+a_{n-1}t^{n-1}+\cdots+a_1t+a_0$ 是一个给定的 n 次首 1 多项式，其零点为 λ_1，…，λ_n．设 $\mu_k=\lambda_1^k+\cdots+\lambda_n^k$ 表示零点的 k 阶矩，$k=0$，1，…(取 $\mu_0=n$)．对 Newton 恒等式

$$ka_{n-k} + a_{n-k+1}\mu_1 + a_{n-k+2}\mu_2 + \cdots + a_{n-1}\mu_{k-1} + \mu_k = 0 \tag{2.4.18}$$

(对 $k=1$，2，…，$n-1$)以及

$$a_0\mu_k + a_1\mu_{k+1} + \cdots + a_{n-1}\mu_{n+k-1} + \mu_{n+k} = 0 \tag{2.4.19}$$

(对 $k=0$，1，2，…)的如下证明提供细节：首先证明，如果 $|t|>R=\max\{|\lambda_i|: i=1, \cdots, n\}$，那么$(t-\lambda_i)^{-1}=t^{-1}+\lambda_i t^{-2}+\lambda_i^2 t^{-3}+\cdots$，从而

$$f(t) = \sum_{i=1}^{n}(t-\lambda_i)^{-1} = nt^{-1} + \mu_1 t^{-2} + \mu_2 t^{-3} + \cdots, \quad |t|>R$$

现在证明 $p'(t)=p(t)f(t)$并比较系数．Newton 恒等式表明，一个 n 次首 1 多项式的零点的前 n 个矩唯一地确定它的系数．有关 Newton 恒等式的矩阵分析方法，见(3.3.P18).

2.4.P10 证明：A，$B\in M_n$ 有同样的特征多项式(从而有同样的特征值)当且仅当对所有 $k=1$，2，…，n 都有 trA^k=trB^k．导出结论：A 是幂零的，当且仅当对所有 $k=1$，2，…，n 都有 $A^k=0$.

2.4.P11 设给定 A，$B\in M_n$ 并考虑它们的**换位子**(commutator)$C=AB-BA$．证明(a)tr$C=0$．(b)考虑 $A=\begin{bmatrix}0&0\\1&0\end{bmatrix}$与 $B=\begin{bmatrix}0&1\\0&0\end{bmatrix}$，并证明换位子未必是幂零的；也就是说，换位子可能有一些非零的特征值，但是它们的和必定为零．(c)如果 rank$C\leqslant 1$，证明 C 是幂零的．(d)如果 rank$C=0$，

说明为什么 A 和 B 可以同时酉三角化．(e)如果 $\operatorname{rank}C=1$，**Laffey 定理**说的是 A 与 B 可以通过相似同时三角化．对于下面给出的 Laffey 定理的证明概述加入细节．我们可以假设 A 是奇异的(如果需要，就用 $A-\lambda I$ 代替 A)．如果 A 的零空间是 B -不变的，那么它就是一个非零的公共不变子空间，所以 A 与 B 同时相似于一个形如(1.3.17)的分块矩阵．如果 A 的零空间**不是** B -不变的，令 $x\neq 0$ 使得 $Ax=0$ 以及 $ABx\neq 0$．那么 $Cx=ABx$，所以存在一个 $z\neq 0$，使得 $C=ABxz^{\mathrm{T}}$．对任何 y，有 $(z^{\mathrm{T}}y)ABx=Cy=ABy-BAy$，$BAy=AB(y-(z^{\mathrm{T}}y)x)$，从而 $\operatorname{range}BA\subset\operatorname{range}AB\subset\operatorname{range}A$，$\operatorname{range}A$ 是 B -不变的，而 A 与 B 同时相似于一个形如(1.3.17)的分块矩阵．现在假设 $A=\begin{bmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{bmatrix}$，$B=\begin{bmatrix}B_{11} & B_{12}\\ 0 & B_{22}\end{bmatrix}$，$A_{11}$，$B_{11}\in M_k$，$1\leqslant k<n$，而 $C=\begin{bmatrix}A_{11}B_{11}-B_{11}A_{11} & X\\ 0 & A_{22}B_{22}-B_{22}A_{22}\end{bmatrix}$的秩为 1．$C$ 的对角分块中至少有一个为零，故我们可以求助于(2.3.3)．如果有一个对角分块的秩为 1，而它的阶大于 1，就重复这一化简．1×1 对角分块的秩不可能为 1．

2.4.P12 设 A，$B\in M_n$，又设 $C=AB-BA$．这个问题检查了 C 与 A 或者与 B、或者与它们两者都可交换这一假设条件的若干推论．(a)如果 C 与 A 可交换，说明为什么对所有 $k=2,\cdots,n$ 都有 $\operatorname{tr}C^k=\operatorname{tr}(C^{k-1}(AB-BA))=\operatorname{tr}(AC^{k-1}B-C^{k-1}BA)=0$．从(2.4.P10)推导出 **Jacobson 引理**：C 是幂零的，如果它或者与 A 或者与 B 可交换．(b)如果 $n=2$，证明：C 与 A 以及 B 两者都可交换，当且仅当 $C=0$，也即当且仅当 A 与 B 可交换．(c)如果 A 可以对角化，证明：C 与 A 可交换，当且仅当 $C=0$．(d)A 与 B 称为**拟交换的**(quasicommute)，如果它们两者都与 C 可交换．如果 A 与 B 是拟交换的，且 $p(s,t)$是关于两个非交换变量的任意一个多项式，证明 $p(A,B)$ 与 C 可交换，借助于(2.4.8.1)，并用(a)证明 $p(A,B)C$ 是幂零的．(e)如果 A 与 B 是拟交换的，利用(2.4.8.7)证明 A 与 B 可以同时三角化．这个结论称为**小 McCoy 定理**(little McCoy's theorem)．(f)设 $n=2$，如果 C 与 A 可交换，则(3.2.P32)确保 A 与 B(从而 B 与 C 亦如此)是可以同时三角化的．证明：A 与 B 可以同时三角化，当且仅当 $C^2=0$．(g)这种情形与 $n=3$ 不同．考虑 126

$$A=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}\quad\text{以及}\quad B=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix}$$

证明：(i)A 与 C 可交换，所以 A 与 C 可同时三角化；(ii)B 与 C 不可交换；以及(iii)B 与 C(从而 A 与 B 亦然)不可同时三角化．(h)设 $n=3$．Laffey 的另一个定理是说：A 与 B 可以同时三角化，当且仅当 C，AC^2，BC^2，以及 A^2C^2，ABC^2 与 B^2C^2 中至少有一个是幂零的．由此定理推导出结论：A 与 B 可以同时三角化，如果 $C^2=0$．给出一个例子来说明：条件 $C^3=0$ **并不蕴含** A 与 B 可以同时三角化．

2.4.P13 对有关线性矩阵方程 $AX-XB=C$ 的结果(2.4.4.1)的另一种证明提供细节：假设 $A\in M_n$ 与 $B\in M_m$ 没有公共的特征值．考虑由 $T_1(X)=AX$ 以及 $T_2(X)=XB$ 所定义的线性变换 T_1，T_2：$M_{n,m}\to M_{n,m}$．证明 T_1 与 T_2 可交换，并由(2.4.8.1)推出结论：$T=T_1-T_2$ 的特征值是 T_1 与 T_2 的特征值之差．讨论：λ 是 T_1 的特征值，当且仅当存在一个非零的 $X\in M_{n,m}$，使得 $AX-\lambda X=0$，此情形当且仅当 λ 是 A 的一个特征值(而 X 的每一个非零的列都是它一个对应的特征向量)时才会发生．这样一来，T_1 与 A 的谱就是相同的，且对 T_2 与 B 也有类似的结论．从而，如果 A 与 B 没有公共的特征值，那么 T 就是非奇异的．如果 x 是 A 的与特征值 λ 相伴的特征向量，而 y 则是 B 的与特征值 μ 相伴的左特征向量，考虑 $X=xy^*$，证明 $T(X)=(\lambda-\mu)X$，并推出结论：T 的谱由 A 与 B 的特征值的所有可能的差组成．

2.4.P14 设 $A\in M_n$，并假设 $\operatorname{rank}A=r$. 证明 A 与一个上三角矩阵酉相似，这个上三角矩阵的前 r 行是线性无关的，而后 $n-r$ 行都是零.

2.4.P15 设 A，$B\in M_n$，并考虑由 $p_{A,B}(s,t)=\det(tB-sA)$ 所定义的关于两个复变量的多项式. (a)假设 A 与 B 可以同时三角化，其中 $A=S\mathcal{A}S^{-1}$，$B=S\mathcal{B}S^{-1}$，这里的$\mathcal{A}$与$\mathcal{B}$是上三角矩阵，diag $\mathcal{A}=(\alpha_1,\cdots,\alpha_n)$，而 diag $\mathcal{B}=(\beta_1,\cdots,\beta_n)$. 证明 $p_{A,B}(s,t)=\det(t\mathcal{B}-s\mathcal{A})=\prod_{i=1}^{n}(t\beta_i-s\alpha_i)$. (b)现在假设 A 与 B 可交换. 导出结论

127

$$p_{A,B}(B,A)=\prod_{i=1}^{n}(\beta_i A-\alpha_i B)=S\Big(\prod_{i=1}^{n}(\beta_i\mathcal{A}-\alpha_i\mathcal{B})\Big)S^{-1}$$

说明为什么上三角矩 $\beta_i\mathcal{A}-\alpha_i\mathcal{B}$的位于$(i,i)$处的元素是零. (c)利用引理 2.4.3.1 证明 $p_{A,B}(B,A)=0$，如果 A 与 B 可交换. 说明为什么这个恒等式是 Cayley-Hamilton 定理的一个两变量推广. (d)假设 A，$B\in M_n$ 可交换. 对 $n=2$，证明 $p_{A,B}(B,A)=(\det B)A^2-(\operatorname{tr}(A\operatorname{adj}B))AB+(\det A)B^2$. 对 $n=3$，证明 $p_{A,B}(B,A)=(\det B)A^3-(\operatorname{tr}(A\operatorname{adj}B))A^2B+\operatorname{tr}((B\operatorname{adj}A))AB^2-(\det A)B^3$. 对 $B=I$ 这些恒等式是什么？(e)对例 2.4.8.3 以及例 2.4.8.4 中的矩阵，计算 $\det(tB-sA)$；并加以讨论. (f)为什么我们在(b)中而不在(a)中假设交换性？

2.4.P16 设 λ 是 $A=\begin{bmatrix}a & b\\ c & d\end{bmatrix}\in M_2$ 的一个特征值. (a)说明为什么 $\mu=a+d-\lambda$ 是 A 的特征值. (b)说明为什么$(A-\lambda I)(A-\mu I)=(A-\mu I)(A-\lambda I)=0$. (c)导出结论：$\begin{bmatrix}a-\lambda & b\\ c & d-\lambda\end{bmatrix}$的任何非零的列都是 A 的与 μ 相伴的特征向量，且其任何非零的行都是与 λ 相伴的一个左特征向量的共轭转置. (d)推出结论：$\begin{bmatrix}\lambda-d & b\\ c & \lambda-a\end{bmatrix}$的任何非零的列都是 A 的与 μ 相伴的特征向量，而其任何非零的行都是与 λ 相伴的一个左特征向量的共轭转置.

2.4.P17 设给定 A，$B\in M_n$ 并考虑$\mathcal{A}(A,B)$，它是 M_n 的由 A 与 B 生成的子代数(见(1.3.P36)). 那么 $\mathcal{A}(A,B)$是 M_n 的一个子空间，所以 $\dim\mathcal{A}(A,B)\leqslant n^2$. 考虑 $n=2$，$A=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$，以及 $B=A^{\mathrm{T}}$，在此情形证明 $\dim\mathcal{A}(A,B)=n^2$. 利用 Cayley-Hamilton 定理证明：对任何 $A\in M_n$ 有 $\dim\mathcal{A}(A,I)\leqslant n$. **Gerstenhaber 定理**说的是：如果 A，$B\in M_n$ 可交换，那么 $\dim\mathcal{A}(A,B)\leqslant n$.

2.4.P18 假设 $A=\begin{bmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{bmatrix}\in M_n$，$A_{11}\in M_k$，$1\leqslant k<n$，$A_{22}\in M_{n-k}$. 证明：$A$ 是幂零的，当且仅当 A_{11}与 A_{22}两者都是幂零的.

2.4.P19 设给定 $n\geqslant 3$ 以及 $k\in\{1,\cdots,n-1\}$. (a)假设 $A=\begin{bmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{bmatrix}\in M_n$，$B=\begin{bmatrix}B_{11} & B_{12}\\ 0 & B_{22}\end{bmatrix}\in M_n$，$A_{11}$，$B_{11}\in M_k$，以及 A_{22}，$B_{22}\in M_{n-k}$. 证明 A 与 B 可以同时上三角化，当且仅当以下两个条件同时成立：(i)A_{11}与 B_{11}可同时上三角化以及(ii)A_{22}与 B_{22}可同时上三角化. (b)如果 $m\geqslant 3$，$\mathcal{F}=\{A_1,\cdots,A_m\}\subset M_n$，且每个 $A_j=\begin{bmatrix}A_{j1} & A_{j2}\\ 0 & A_{j3}\end{bmatrix}(A_{j1}\in M_k)$，证明：$\mathcal{F}$可同时上三角化，当且仅当$\{A_{11},\cdots,A_{m1}\}$与$\{A_{13},\cdots,A_{m3}\}$中的每一个都(分别)是可以同时上三角化的.

2.4.P20 假设 A，$B\in M_n$ 且 $AB=0$，所以 $C=AB-BA=-BA$. 令 $p(s,t)$是关于两个非交换变量的多项式. (a)如果 $p(0,0)=0$，证明 $Ap(A,B)B=0$，因此$(p(A,B)C)^2=0$. (b)证明 $C^2=0$. (c)利用(2.4.8.7)证明 A 与 B 可以同时上三角化. (d)$\begin{bmatrix}-3 & 3\\ -4 & 4\end{bmatrix}$与$\begin{bmatrix}2 & -1\\ 2 & -1\end{bmatrix}$是同时可上三角化的吗？

2.4.P21 设 $A\in M_n$ 有特征值 λ_1，…，λ_n. Hankel 矩阵 $K=[\operatorname{tr}A^{i+j-2}]_{i,j=1}^n$ 称为是与 A 相伴的**矩矩阵**(moment matrix). 我们恒取 $A^0=I$，所以 $\operatorname{tr}A^0=n$. (a)证明 $K=VV^{\mathrm{T}}$，其中 $V\in M_n$ 是 Vandermonde 矩阵(0.9.11.1)，这个矩阵的第 j 列是$[1、\lambda_j、\lambda_j^2、\cdots、\lambda_j^{m-1}]^{\mathrm{T}}$ $(j=1,\ \cdots,\ n)$. (b)说明为什么 $K=(\det V)^2=\prod_{i<j}(\lambda_j-\lambda_i)^2$；这个乘积就是 A 的**判别式**(discriminant). (c)推出结论：A 的特征值是不同的，当且仅当它的矩矩阵非奇异. (d)说明为什么 K(从而 A 的判别式亦然)在 A 的相似之下是不变的. (e)计算 $A=\begin{bmatrix} a & b \\ c & d\end{bmatrix}\in M_2$ 的矩矩阵的行列式；验证它是 A 的判别式，如同在包含(1.2.4b)的习题中计算的那样. (f)考虑实矩阵 128

$$A=\begin{bmatrix} a & b & 0 \\ 0 & 0 & c \\ d & -e & 0\end{bmatrix},a,b,c,d\ 是正数 \tag{2.4.20}$$

它的零元素是指定的，而剩下的元素仅仅符号的样式是指定的. A 的矩矩阵是

$$K=\begin{bmatrix} 3 & a & a^2-2ce \\ a & a^2-2ce & a^3+3bcd \\ a^2-2ce & a^3+3bcd & a^4+4bdac+2e^2c^2\end{bmatrix}$$

而 $\det K=-27b^2c^2d^2-4c^3e^3-4a^4ce-8a^2c^2e^2-4a^3bcd-36abc^2de$. 说明为什么 A 恒有三个不同的特征值.

2.4.P22 假设 $A\in M_n$ 有 d 个不同的特征值 μ_1，…，μ_d，它们分别有重数 v_1，…，v_d. 矩阵 $K_m=[\operatorname{tr}A^{i+j-2}]_{i,j=1}^m$ 是与 A 相伴的 m 阶矩矩阵，$m=1$，2，…；如果 $m\leqslant n$，它就是上一个问题中的矩矩阵 K 的**前主子矩阵**(leading principal submatrix). 设 $v_j^{(m)}=[1、\mu_j、\mu_j^2、\cdots、\mu_j^{m-1}]^{\mathrm{T}}$，$j=1$，…，$d$，并构造 $m\times d$ 矩阵 $V_m=[v_1^{(m)}\ \cdots\ v_d^{(m)}]$. 设 $D=\operatorname{diag}(v_1,\ \cdots,\ v_d)\in M_d$. 证明：(a) V_m 有行秩 m，如果 $m\leqslant d$；它有列秩 d，如果 $m\geqslant d$；(b) $K_m=V_mDV_m^{\mathrm{T}}$；(c)如果 $1\leqslant p<q$，K_p 就是 K_q 的一个前主子矩阵；(d) K_d 是非奇异的；(e) $\operatorname{rank}K_m=d$，如果 $m\geqslant d$；(f) $d=\max\{m\geqslant 1:K_m$ 非奇异$\}$，但是 K_p 有可能对某个 $p<d$ 是奇异的；(g) K_d 是非奇异的，且 K_{d+1}，…，K_n，K_{n+1} 中每一个全都是奇异的；(h) $K_n=K$，即是上一问题中的矩矩阵；(i) $\operatorname{rank}K$ 恰好是 A 的不同特征值的个数.

2.4.P23 假设 $T=[t_{ij}]\in M_n$ 是上三角的. 证明 $\operatorname{adj}T=[\tau_{ij}]$是上三角的，且有主对角元素 $\tau_{ii}=\prod_{j\neq i}t_{jj}$.

2.4.P24 设 $A\in M_n$ 有特征值 λ_1，…，λ_n. 证明 $\operatorname{adj}A$ 的特征值是 $\prod_{j\neq i}\lambda_j$，$i=1$，…，n.

2.4.P25 设 A，$B\in M_2$，并假设 λ_1，λ_2 是 A 的特征值. (a)证明 A 与 $\begin{bmatrix}\lambda_1 & x \\ 0 & \lambda_2\end{bmatrix}$ 酉相似，其中 $x\geqslant 0$，且 $x^2=\operatorname{tr}AA^*-|\lambda_1|^2-|\lambda_2|^2$. (b)证明：$A$ 与 B 酉相似，当且仅当 $\operatorname{tr}A=\operatorname{tr}B$，$\operatorname{tr}A^2=\operatorname{tr}B^2$，以及 $\operatorname{tr}AA^*=\operatorname{tr}BB^*$.

2.4.P26 设 $B\in M_{n,k}$ 以及 $C\in M_{k,n}$. 证明：对任意多项式 $p(t)$ 有 $BCp(BC)=Bp(CB)C$.

2.4.P27 设给定 $A\in M_n$. (a)如果 $A=BC$ 且 B，$C^{\mathrm{T}}\in M_{n,k}$，利用(2.4.3.2)证明：存在一个次数最多为 $k+1$的多项式 $q(t)$，使得 $q(A)=0$.

2.4.P28 假设 $A\in M_n$ 是奇异的，又令 $r=\operatorname{rank}A$. 证明存在一个次数至多为 $r+1$ 的多项式 $p(t)$，使得 $p(A)=0$.

2.4.P29 设 $A\in M_n$，并假设 x，$y\in\mathbf{C}^n$ 是满足 $Ax=\lambda x$ 以及 $y^*A=\lambda y^*$ 的非零向量. 如果 λ 是 A 的一个单重特征值，证明 $A-\lambda I+\kappa xy^*$ 对所有 $\kappa\neq 0$ 都是非奇异的.

2.4.P30 对于(2.4.3.3)中所描述的计算有一个系统的解决方法．设给定 $A\in M_n$，并假定 $p(t)$是一个次
129 数大于 n 的多项式．利用 Euclid 算法(多项式长除法)将它表示成 $p(t)=h(t)p_A(t)+r(t)$，其中 $r(t)$的次数严格小于 n(有可能为零)．说明为什么 $p(A)=r(A)$.

2.4.P31 利用(2.4.3.2)证明：如果 $A\in M_n$ 的所有的特征值都为零，那么 $A^n=0$.

2.4.P32 设 A，$B\in M_n$，又令 $C=AB-BA$．说明为什么不可能有 $\operatorname{tr}C\neq 0$．特别地，如果 $c\neq 0$，则不可能有 $C=cI$.

2.4.P33 设 A，$B\in M_n$，并设 p 是一个正整数．假设 $A=\begin{bmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{bmatrix}$，其中 $A_{11}\in M_k$ 与 $A_{22}\in M_{n-k}$ 没有公共的特征值．如果 $B^p=A$，证明：B 是与 A 共形的分块上三角矩阵，$B=\begin{bmatrix}B_{11} & B_{12}\\ 0 & B_{22}\end{bmatrix}$，$B_{11}^p=A_{11}$，且 $B_{22}^p=A_{22}$.

2.4.P34 设 $A=\begin{bmatrix}a & b\\ c & d\end{bmatrix}\in M_2$．通过显式计算，对 A 来验证 Cayley-Hamilton 定理，也就是验证 $A^2-(a+d)A+(ad-bc)I_2=0$.

2.4.P35 设 $A\in M_n(\mathbf{F})$($\mathbf{F}=\mathbf{R}$ 或者 $\mathbf{C}$)．利用(2.4.4.2)证明 A 与 $M_n(\mathbf{F})$中的每个酉矩阵可交换，当且仅当 A 是一个纯量矩阵.

注记以及进一步的阅读参考 有关同时三角化的详尽的阐述，见 Radjavi 以及 P. Rosenthal(2000)的专著．定理 2.4.8.7 及其推广由 N. McCoy，On the characteristic roots of matric polynomials，*Bull. Amer. Math. Soc.* 42 (1936) 592-600 给出了证明．我们对(2.4.8.7)给出的证明取自 M. P. Drazin，J. W. Dungey 以及 K. W. Gruenberg，Some theorems on commutative matrices，*J. London Math. Soc.* 26(1951)221-228，它包含了对 $m\geqslant 2$ 的一般情形(2.4.8.10)的一个证明．特征值与线性组合之间的关系在 T. Motzkin 以及 O. Taussky，Pairs of matrices with property L，*Trans. Amer. Math. Soc.* 73(1952)108-114 中进行了讨论．一对对所有 a，$b\in\mathbf{C}$ 都满足 $\sigma(aA+bB)=\{a\alpha_j+b\beta_{i_j}：j=1,\cdots,n\}$ 的 A，$B\in M_n$ 称为有**性质** L(property$L$)；条件(2.4.8.7(c))称为**性质** P(property$P$)．对所有 $n=2$，3，…，性质 P 蕴含性质 L；仅当 $n=2$ 时性质 L 蕴含性质 P．性质 L 还没有得到充分的研究和了解，不过已经知道：一对正规矩阵有性质 L，当且仅当它们可交换．见 N. A. Wiegmann，A note on pairs of normal matrices with property L，*Proc. Amer. Math. Soc.* 4(1953)35-36．对于(2.4.P10)中的结论，有一个引人关注的近似表述形式：非奇异矩阵 A，$B\in M_n$ 有同样的特征多项式(从而有同样的特征值)，当且仅当 $|\operatorname{tr}A^k-\operatorname{tr}B^k|\leqslant 1$(对所有 $k=\pm 1,\pm 2,\cdots$)，见(2.2)末尾提及的 Marcoux，Mastnak 以及 Radjavi 的论文．Jacobson 引理(2.4.P12)是 N. Jacobson，Rational methods in the theory of Lie algebras，*Ann. of Math.* (2)36 (1935)875-881 中的引理 2．拟交换性这一概念(2.4.P12)是在量子力学中出现的：位置以及动量算子 x 与 p_x(线性算子，但不是有限维的，见(2.4.P32))满足恒等式 $xp_x-p_xx=i\hbar I$．这个恒等式(它蕴含关于位置和动量的 Heisenberg 测不准原理)确保 x 与 p_x 这两者都与它们的换位子可交换．(2.4.P12g)中的例子属于 Gérald Bourgeois．(2.4.P11)以及(2.4.P12)中提到的 Laffey 定理的证明在 T. J. Laffey，Simultaneous triangularization of a pair of matrices-low rank cases and the nonderogatory case，*Linear Multilinear Algebra* 6(1978)269-306 以及 T. J. Laffey，Simultaneous quasidiagonalization of a pair of 3×3 complex
130 matrices，*Rev. Roumaine Math. Pures Appl.* 23(1978)1047-1052 中给出．矩阵(2.4.20)是要求不同的特征值的符号模式矩阵的一个例子．这一令人极感兴趣的性质的讨论，见 Z. Li 以及 L. Harris，Sign patterns that require all distinct eigenvalues，*JP J. Algebra Number Theory Appl.* 2(2002)161-179，这篇论文包含了一列具有此项性质的所有不可约的 3×3 符号模式矩阵．(2.4.P34)中所要求的显式计算发表在 A. Cayley，A memoir on the theory of matrices，*Philos. Trans. R. Soc. London* 148(1858)17-37 中；见该文

第 23 页．在第 24 页上，Cayley 说他还验证了 3×3 的情形(可能由其他的人作了显式计算，而计算并未出现在该论文中)，但是他“并不认为有必要花力气来对任意阶矩阵的一般情形给出定理的正式的证明．”在一篇发表于 1878 年的论文中，F. G. Frobenius 对如下结论给出了一个严格的证明：每一个复方阵满足它自己的特征方程，但是他的方法与我们对(2.4.3.2)的证明迥然不同．Frobenius 首先定义了矩阵的极小多项式(这是他本人创造的一个新概念，见(3.3))，然后他证明了：极小多项式整除特征多项式，见 Ferdinand Georg Frobenius：*Gesammelte Abhandlungen*(Frobenius 全集)，第 1 卷第 355 页(由 J. P. Serre 编辑，Springer 出版社，Berlin，1968)．(2.4.P3)中概述的 Cayley-Hamilton 定理的证明取自 A. Buchheim，Mathematical notes，*Messenger Math*. 13(1884)62-66.

2.5 正规矩阵

正规矩阵类(它在酉相似那一部分自然出现)在整个矩阵分析中都是重要的，它包括酉矩阵、Hermite 矩阵、斜 Hermite 矩阵、实正交矩阵、实对称矩阵以及实的斜对称矩阵.

定义 2.5.1 矩阵 $A\in M_n$ 称为是**正规的**(normal)，如果 $AA^*=A^*A$，也就是，如果 A 与它的共轭转置可交换.

习题 如果 $A\in M_n$ 是正规的，且 $\alpha\in\mathbf{C}$，证明 αA 是正规的．给定大小的正规矩阵类在用复纯量作的乘法运算下是封闭的. ◀

习题 如果 $A\in M_n$ 是正规的，又如果 B 与 A 酉相似，证明 B 是正规的．给定大小的正规矩阵类在酉相似之下是封闭的. ◀

习题 如果 $A\in M_n$ 与 $B\in M_m$ 是正规的，证明 $A\oplus B\in M_{n+m}$ 是正规的．正规矩阵类在直和运算之下是封闭的. ◀

习题 如果 $A\in M_n$ 且 $B\in M_m$，又如果 $A\oplus B\in M_{n+m}$ 是正规的，证明 A 与 B 是正规的. ◀

习题 如果给定 a，$b\in\mathbf{C}$．证明 $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$ 是正规的且有特征值 $a\pm ib$. ◀

习题 证明：每个酉矩阵都是正规的. ◀

习题 证明：每个 Hermite 矩阵或者斜 Hermite 矩阵都是正规的. ◀

习题 验证：$A=\begin{bmatrix} 1 & e^{i\pi/4} \\ -e^{i\pi/4} & 1 \end{bmatrix}$ 是正规的，但是 A 的任何纯量倍数都不是酉矩阵、不是 Hermite 矩阵或者斜 Hermite 矩阵. ◀

习题 说明为什么每个对角矩阵都是正规的．如果一个对角矩阵是 Hermite 的，为什么它必定是实的? ◀

习题 证明：酉矩阵、Hermite 矩阵以及斜 Hermite 矩阵组成的每一个类在酉相似之下都是封闭的．如果 A 是酉矩阵，且 $|\alpha|=1$，证明 αA 是酉矩阵．如果 A 是 Hermite 矩阵且 α 是实数，证明 αA 是 Hermite 矩阵．如果 A 是斜 Hermite 矩阵且 α 是实数，证明 αA 是斜 Hermite 矩阵. ◀

习题 证明：Hermite 矩阵的主对角元素是实数．证明：斜 Hermite 矩阵的主对角元素是纯虚数．一个实的斜对称矩阵的主对角元素是什么样的? ◀

习题 检查(1.3.7)的证明，并导出结论：$A\in M_n$ 是可以酉对角化的，当且仅当 $\mathbf{C}^n$ 中存在一组 n 个标准正交的向量，它们每一个都是 A 的一个特征向量. ◀

在理解并利用正规矩阵的定义恒等式 $AA^*=A^*A$ 时，在脑子里保留一个几何的解释会是有助益的. 按照它们的列来分划 $A=[c_1\ \ \cdots\ \ c_n]$以及 $A^{\mathrm{T}}=[r_1\ \ \cdots\ \ r_n]$；向量 c_j 是 A 的列，而向量 r_i^{T} 是 A 的行. 审视定义恒等式 $AA^*=A^*A$ 揭示出：A 是正规的，当且仅当对所有 i，$j=1$，…，n 都有 $c_i^*c_j=\overline{r_i^*r_j}$. 特别地，$c_i^*c_i=\|c_i\|_2^2=\|r_i\|_2^2=r_i^*r_i$，所以 A 的每一列与它对应的行一样，有同样的 Euclid 范数；一列为零当且仅当对应的行为零.

如果 $A\in M_n(\mathbf{R})$是实的正规矩阵，那么对所有 i 与 j 都有 $c_i^{\mathrm{T}}c_j=\langle c_i,\ c_j\rangle=\langle r_i,\ r_j\rangle=r_i^{\mathrm{T}}r_j$. 如果 i 列与 j 列不是零，那么 i 行与 j 行不是零，且恒等式

$$\frac{\langle c_i,c_j\rangle}{\|c_i\|_2\|c_j\|_2}=\frac{\langle r_i,r_j\rangle}{\|r_i\|_2\|r_j\|_2}$$

告诉我们：A 的 i 列与 j 列的向量之间的角度与 A 的 i 行与 j 行的向量之间的角度是相同的，见(0.6.3.1).

对于正规矩阵，不仅为零的行或者列，而且关于它的某种为零的分块有某些特殊之处.

引理 2.5.2 设 $A\in M_n$ 被分划成 $A=\begin{bmatrix}A_{11} & A_{12}\\ 0 & A_{22}\end{bmatrix}$，其中 A_{11} 与 A_{22} 是方阵. 那么 A 是正规的，当且仅当 A_{11} 与 A_{22} 是正规的，且 $A_{12}=0$. 分块上三角矩阵是正规的，当且仅当它位于对角线之外的分块是零且每一个对角线上的分块都是正规的. 特别地，上三角矩阵是正规的，当且仅当它是对角的.

证明 如果 A_{11} 与 A_{22} 是正规的，且 $A_{12}=0$，那么 $A=A_{11}\oplus A_{22}$ 就是正规矩阵的直和，

132 故而它是正规的.

反之，如果 A 是正规的，那么

$$AA^*=\begin{bmatrix}A_{11}A_{11}^*+A_{12}A_{12}^* & \bigstar\\ \bigstar & \bigstar\end{bmatrix}=\begin{bmatrix}A_{11}^*A_{11} & \bigstar\\ \bigstar & \bigstar\end{bmatrix}=A^*A$$

所以 $A_{11}^*A_{11}=A_{11}A_{11}^*+A_{12}A_{12}^*$，它蕴含

$$\begin{aligned}\mathrm{tr}A_{11}^*A_{11}&=\mathrm{tr}(A_{11}A_{11}^*+A_{12}A_{12}^*)\\&=\mathrm{tr}A_{11}A_{11}^*+\mathrm{tr}A_{12}A_{12}^*=\mathrm{tr}A_{11}^*A_{11}+\mathrm{tr}A_{12}A_{12}^*\end{aligned}$$

从而 $\mathrm{tr}A_{12}A_{12}^*=0$. 由于 $\mathrm{tr}A_{12}A_{12}^*$ 是 A_{12} 的元素的绝对值的平方和(0.2.5.1)，由此推出 $A_{12}=0$. 这样 $A=A_{11}\oplus A_{22}$ 就是正规的，所以 A_{11} 与 A_{22} 都是正规的.

假设 $B=[B_{ij}]_{i,j=1}^k\in M_n$ 是正规的，且是分块上三角的，也就是说，$B_{ii}\in M_{n_i}$（对 $i=1$，…，k）以及 $B_{ij}=0$（对 $i>j$）. 将它分划成 $B=\begin{bmatrix}B_{11} & X\\ 0 & \tilde{B}\end{bmatrix}$，其中 $X=[B_{12}\ \ \cdots\ \ B_{1k}]$，而 $\tilde{B}=[B_{ij}]_{i,j=2}^k$是分块上三角的. 那么 $X=0$，且$\tilde{B}$是正规的，故而有限归纳法允许我们得出 B 是分块对角的结论. 关于其逆，在上一个习题中我们已经注意到：正规矩阵的直和仍是正规矩阵. □

习题 设 $A\in M_n$ 是正规的，又令 $\alpha\in\{1,\ \cdots,\ n\}$是给定的指标集. 如果 $A[\alpha,\ \alpha^c]=0$，

证明 $A[\alpha^c, \alpha]=0$. ◀

接下来我们对有关正规矩阵的基本结果进行整理分类.

定理 2.5.3 设 $A=[a_{ij}]\in M_n$ 有特征值 $\lambda_1, \cdots, \lambda_n$，则下述诸命题等价.

(a) A 是正规的.

(b) A 可以酉对角化.

(c) $\sum_{i,j=1}^{n}|a_{ij}|^2=\sum_{i=1}^{n}|\lambda_i|^2$.

(d) A 有 n 个标准正交的特征向量.

证明 利用(2.3.1)将 A 写成 $A=UTU^*$，其中 $U=[U_1 \quad \cdots \quad U_n]$是酉矩阵，而 $T=[t_{ij}]\in M_n$ 是上三角矩阵.

如果 A 是正规的，则 T 亦然(与每一个与 A 酉相似的矩阵一样). 上一个引理确保 T 实际上是对角矩阵，故而 A 可以酉对角化.

如果存在一个酉矩阵 V，使得 $A=V\Lambda V^*$，且 $\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$，那么由(2.2.2)有 $\mathrm{tr}A^*A=\mathrm{tr}\Lambda^*\Lambda$，这就是(c)中的结论.

T 的对角元素是 $\lambda_1, \cdots, \lambda_n$(按照某种次序排列)，从而 $\mathrm{tr}A^*A=\mathrm{tr}T^*T=\sum_{i=1}^{n}|\lambda_i|^2+\sum_{i<j}|t_{ij}|^2$. 于是，(c)蕴含 $\sum_{i<j}|t_{ij}|^2=0$，所以 T 是对角的. 分解式 $A=UTU^*$ 等价于等式 $AU=UT$，该等式是说，对每一个 $i=1, \cdots, n$ 都有 $Au_i=t_{ii}u_i$. 从而，U 的 n 列是 A 的标准正交的特征向量.

最后，标准正交组是线性无关的，所以(d)确保 A 可以对角化，且可以选择标准正交化的列作为对角化相似矩阵(1.3.7). 这就意味着 A 与一个对角矩阵(从而是酉矩阵)酉相似，所以 A 是正规的. □ 133

正规矩阵 $A\in M_n$ 表示成 $A=U\Lambda U^*$(其中 U 是酉矩阵，而 Λ 是对角矩阵)的表示法称为 A 的**谱分解**(spectral decomposition).

习题 说明为什么正规矩阵是无亏的，也就是说，它的每一个特征值的几何重数与其代数重数相同. ◀

习题 如果 $A\in M_n$ 是正规的，证明：$x\in \mathbf{C}^n$ 是 A 的一个与 A 的特征值 λ 相伴的右特征向量，当且仅当 x 是 A 的与 λ 相伴的左特征向量；也就是说，如果 A 是正规的，那么 $Ax=\lambda x$ 等价于 $x^*A=\lambda x^*$. 提示：将 x 标准化，并记 $A=U\Lambda U^*$，其中以 x 作为 U 的第一列. 那么 A^* 是什么？A^*x 呢？另一个证明见(2.5.P20). ◀

习题 如果 $A\in M_n$ 是正规的，又如果 x 与 y 是 A 的与不同的特征值相伴的特征向量，利用上一个习题以及双正交原理证明 x 与 y 是正交的. ◀

一旦已知一个正规矩阵 $A\in M_n$ 的不同的特征值 $\lambda_1, \cdots, \lambda_d$，就能通过如下概念性的处置方法将它酉对角化：对每一个特征空间$\{x\in \mathbf{C}^n: Ax=\lambda x\}$，确定一组基并将它标准正交化以得到一组标准正交基. 这些特征空间是相互正交的，且每一个特征空间的维数等于其对应的特征值的重数(A 的正规性是这两者成立的理由)，所以这些基的并集就是关于 $\mathbf{C}^n$ 的

一组标准正交基．将这些基向量排列成一个矩阵 U 的列，就得到一个酉矩阵，它使得 U^*AU是对角矩阵．

然而，特征空间总会有多于一组标准正交基，故而按照上面的概念性处理方法所构造的对角化酉矩阵从来都不是唯一的．如果 $X, Y\in M_{n,k}$ 都有标准正交的列(即 $X^*X=I_k=Y^*Y$)，又如果 range$X$=range$Y$，那么 X 的每一列都是 Y 的列的线性组合，这就是说，对某个 $G\in M_k$ 有 $X=YG$．这样就有 $I_k=X^*X=(YG)^*(YG)=G^*(Y^*Y)G=G^*G$，所以 G 必定是酉矩阵．这一结果就对如下唯一性定理的第一部分给出一个几何的解释．

定理 2.5.4 设 $A\in M_n$ 是正规的，且有不同的特征值 $\lambda_1, \cdots, \lambda_d$，其重数分别为 $n_1, \cdots, n_d$．设 $\Lambda=\lambda_1 I_{n_1}\oplus\cdots\oplus\lambda_d I_{n_d}$，并假设 $U\in M_n$ 是酉矩阵以及 $A=U\Lambda U^*$．

(a) 对某个酉矩阵 $V\in M_n$ 有 $A=V\Lambda V^*$，当且仅当存在酉矩阵 $W_1, \cdots, W_d$，每个 $W_i\in M_{n_i}$，使得 $U=V(W_1\oplus\cdots\oplus W_d)$．

(b) 两个正规矩阵是酉相似的，当且仅当它们有同样的特征值．

证明 (a)如果 $U\Lambda U^*=V\Lambda V^*$，那么 $\Lambda U^*V=U^*V\Lambda$，所以 $W=U^*V$ 是酉矩阵，且与 Λ 可交换，(2.4.4.2)确保 W 是与 Λ 共形的分块对角矩阵．反之，如果 $U=VW$ 以及 $W=W_1\oplus\cdots\oplus W_d$，其中每一个 $W_i\in M_{n_i}$，这样 W 就与 Λ 可交换，且 $U\Lambda U^*=VW\Lambda W^*V^*=$

134 $V\Lambda WW^*V^*=V\Lambda V^*$．

(b)如果对某个酉矩阵 V 有 $B=V\Lambda V^*$，那么 $(UV^*)B(UV^*)^*=(UV^*)V\Lambda V^*\times(UV^*)^*=U\Lambda U^*=A$．反之，如果 B 与 A 相似，那么它们有相同的特征值；如果 B 与一个正规矩阵酉相似，那么它也是正规的． □

接下来注意可交换的正规矩阵可以同时酉对角化．

定理 2.5.5 设$\mathcal{N}\subseteq M_n$ 是一个非空的正规矩阵族．那么$\mathcal{N}$是一个交换族，当且仅当它是可以同时酉对角化的族．对任一给定的 $A_0\in\mathcal{N}$以及对 A_0 的特征值的任意一个给定的排序 $\lambda_1, \cdots, \lambda_n$，都存在一个酉矩阵 $U\in M_n$，使得 $U^*A_0U=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$，且对每个$B\in\mathcal{N}$，$U^*BU$ 都是对角矩阵．

习题 利用(2.3.3)以及正规的三角矩阵必定是对角矩阵这一事实来证明(2.5.5)．关于 A_0 的最后那个结论可以如同在(1.3.21)的证明中那样得出，这是因为每一个置换矩阵都是酉矩阵． ◀

(2.5.3)应用于 Hermite 矩阵这种特殊情形就会得到一个基本的结果，它称之为关于**Hermite 矩阵的谱定理**(spectral theorem for Hermitian matrices)．

定理 2.5.6 设 $A\in M_n$ 是 Hermite 矩阵且有特征值 $\lambda_1, \cdots, \lambda_n$．设 $\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)$．

(a) $\lambda_1, \cdots, \lambda_n$ 是实数．

(b) A 可以酉对角化．

(c) 存在一个酉矩阵 $U\in M_n$，使得 $A=U\Lambda U^*$．

证明 Hermite 对角矩阵的对角元素必定是实数，所以(a)就从(b)以及如下事实推出：Hermite 矩阵的集合在酉相似之下是封闭的．命题(b)由(2.5.3)推出，因为 Hermite 矩阵是正规的．命题(c)是(b)的复述并加入了如下信息：Λ 的对角元素必定是 A 的特征值． □

与第1章里的有关对角化的讨论相比，在(2.5.4)以及(2.5.6)中没有理由假设特征值各不相同，在(2.5.5)中不需要假设可对角化．特征向量的基(事实上是标准正交基)在结构上由正规性加以保证．这就是 Hermite 矩阵以及正规矩阵如此重要且有如此令人赏心悦目的性质的原因．

现在我们转过来讨论**实的**正规矩阵．它们可以通过**复的酉**相似对角化．但是通过**实正交**相似可以得到何种特殊的形式呢？由于实正规矩阵可能有非实的特征值，有可能无法用实相似使之对角化．然而，每一个实矩阵都与一个实的拟三角矩阵实正交相似，这个拟三角矩阵必定还是拟对角的，如果它是正规的．

引理 2.5.7 假设 $A=\begin{bmatrix} a & b \\ c & d \end{bmatrix}\in M_2(\mathbf{R})$ 是正规的，且有一对共轭的非实特征值．那么 $c=-b\neq 0$ 且 $d=a$．

证明 计算表明：$AA^{\mathrm{T}}=A^{\mathrm{T}}A$ 当且仅当 $b^2=c^2$ 且 $ac+bd=ab+cd$．如果 $b=c$，则 A 是 Hermite 的(因为它是实对称的)，所以上面的定理保证它有两个实的特征值．这样一来，我们必有 $b=-c\neq 0$ 以及 $b(d-a)=b(a-d)$，这就蕴含 $a=d$． □ 135

定理 2.5.8 设 $A\in M_n(\mathbf{R})$ 是正规的．

(a) 存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$，使得 $Q^{\mathrm{T}}AQ$ 是实的拟对角矩阵

$$A_1\oplus\cdots\oplus A_m\in M_n(\mathbf{R})\text{，每个 }A_i\text{ 都是 }1\times 1\text{ 或 }2\times 2\text{ 的}\tag{2.5.9}$$

它满足下述性质：(2.5.9)中的那些 1×1 直和项给出 A 所有实的特征值．(2.5.9)中每一个 2×2 直和项有特殊的形式

$$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}\tag{2.5.10}$$

其中 $b>0$．该矩阵是正规的且有特征值 $a\pm \mathrm{i}b$．

(b) (2.5.9)中的直和项由 A 的特征值完全决定，它们可以按照任意预先指定的次序出现．

(c) 两个实的 $n\times n$ 正规矩阵是实正交相似的，当且仅当它们有同样的特征值．

证明 (a) 定理2.3.4b确保 A 与一个实的上拟三角矩阵实正交相似，它的每一个 2×2 对角分块有一对非实的共轭特征值．由于这个上拟三角矩阵是正规的，(2.5.2)就保证了它实际上是拟对角的，且它的每一个 2×2 直和项都是正规的，且有一对共轭的非实特征值．上一个引理告诉我们：这些 2×2 直和项中的每一个都有特殊的形式(2.5.10)，其中 $b\neq 0$．如果必要，我们可以通过用矩阵 $\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$ 做成的相似来确保 $b>0$．(b)(2.5.9)中的直和项给出了 A 所有的特征值，且通过置换相似还能使得这些直和项按照所希望的任何次序排列．

(c) 两个有同样特征值的实的 $n\times n$ 正规矩阵与同一个形如(2.5.9)的直和实正交相似． □

上一个定理揭示出实正规矩阵在实正交相似下的标准型．它引导到实对称矩阵、实的斜对称矩阵以及实正交矩阵在实正交相似之下的标准型．

推论 2.5.11 设 $A\in M_n(\mathbf{R})$．

(a) $A=A^{\mathrm{T}}$，当且仅当存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$，使得

$$Q^{\mathrm{T}}AQ = \mathrm{diag}(\lambda_1,\cdots,\lambda_n) \in M_n(\mathbf{R}) \tag{2.5.12}$$

A 的特征值是 λ_1，…，λ_n. 两个实对称矩阵是实正交相似的，当且仅当它们有相同的特征值.

(b) $A=-A^{\mathrm{T}}$，当且仅当存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$和一个非负整数 p，使得 $Q^{\mathrm{T}}AQ$有形式

$$0_{n-2p} \oplus b_1\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \oplus \cdots \oplus b_p\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad 所有\ b_j > 0 \tag{2.5.13}$$

如果 $A\neq 0$，它的非零特征值是 $\pm \mathrm{i}b_1$，…，$\pm \mathrm{i}b_p$. 两个实的斜对称矩阵是实正交相似的，当且仅当它们有同样的特征值.

(c) $AA^{\mathrm{T}}=I$ 当且仅当存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$和一个非负整数 p，使得 $Q^{\mathrm{T}}AQ$ 有形式

$$\Lambda_{n-2p} \oplus \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ -\sin\theta_1 & \cos\theta_1 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} \cos\theta_p & \sin\theta_p \\ -\sin\theta_p & \cos\theta_p \end{bmatrix} \tag{2.5.14}$$

其中 $\Lambda_{n-2p}=\mathrm{diag}(\pm 1, \cdots, \pm 1)\in M_{n-2p}(\mathbf{R})$，而每一个 $\theta_j\in(0, \pi)$. A 的特征值是 Λ_{n-2p} 的对角元素添上 $\mathrm{e}^{\pm \mathrm{i}\theta_1}$，…，$\mathrm{e}^{\pm \mathrm{i}\theta_p}$. 两个实正交矩阵是实正交相似的，当且仅当它们有同样的特征值.

证明 每一个假设条件都确保 A 是实的且是正规的，所以它与一个形如(2.5.9)的拟对角矩阵实正交相似. 只要考虑每一个假设对于(2.5.9)中的直和项蕴含何种意义就足够了. 如果 $A=A^{\mathrm{T}}$，有可能不存在形如(2.5.10)的直和项. 如果 $A=-A^{\mathrm{T}}$，则每一个 1×1 直和项都是零，而任何一个 2×2 直和项的对角元素都是零. 如果 $AA^{\mathrm{T}}=I$，则每一个 1×1 直和项形如$[\pm 1]$，而任何 2×2 分块(2.5.10)有行列式± 1，所以 $a^2+b^2=1$，且存在某个 $\theta\in(0, \pi)$，使得 $a=\cos\theta$ 以及 $b=\sin\theta$，也就是 $a\pm \mathrm{i}b=\mathrm{e}^{\pm \mathrm{i}\theta}$. □

习题 设 $A_1=\begin{bmatrix} a & b \\ -b & a \end{bmatrix}\in M_2$，$A_2=\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix}\in M_2$，并假设 $b\neq 0$. 证明 A_1 与 A_2 可交换，当且仅当 $\alpha=\delta$ 以及 $\gamma=-\beta$. ◀

习题 设 a，$b\in\mathbf{C}$. 说明为什么$\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$与$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$是实正交相似的. 提示：考虑通过$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$所作的相似. ◀

下面的定理是(2.5.5)的实正规矩阵表述形式.

定理 2.5.15 设$\mathcal{N}\in M_n(\mathbf{R})$是由实正规矩阵组成的非空的交换族. 则存在一个实正交矩阵 Q 以及一个非负整数 q，使得对每个 $A\in\mathcal{N}$，$Q^{\mathrm{T}}AQ$ 都是一个形如

$$\Lambda(A) \oplus \begin{bmatrix} a_1(A) & b_1(A) \\ -b_1(A) & a_1(A) \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} a_q(A) & b_q(A) \\ -b_q(A) & a_q(A) \end{bmatrix} \tag{2.5.15a}$$

的实拟对角矩阵，其中每一个 $\Lambda(A)\in M_{n-2q}(\mathbf{R})$都是对角的；对所有 $A\in\mathcal{N}$以及所有 $j=1$，…，q，参数 $a_j(A)$与 $b_j(A)$都是实的；又对每一个 $j\in\{1, \cdots, q\}$，都存在某个 $A\in\mathcal{N}$，对于它有 $b_j(A)>0$.

证明 定理2.3.6b确保存在一个实正交矩阵Q以及一个拟对角矩阵$D=J_{n_1}\oplus\cdots\oplus J_{n_m}$，使得对每个$A\in\mathcal{N}$，$Q^TAQ$都是一个形如(2.3.6.1)的上拟三角矩阵，它与D共形地加以分划．此外，如果$n_j=2$，则对某个$A\in\mathcal{N}$，$A_j(A)$有一对共轭的非实特征值．由于每一个上拟三角矩阵Q^TAQ都是正规的，(2.5.2)确保它实际上是拟对角的，也就是说，对每一个$A\in\mathcal{N}$，$Q^TAQ=A_1(A)\oplus\cdots\oplus A_m(A)$都是与$D$共形地进行分划的，且每一个直和项$A_j(A)$都是正规的．如果每个$n_j=1$，就没什么要进一步证明的了．假设$n_j=2$就考虑交换族$\mathcal{F}=\{A_j(A):A\in\mathcal{N}\}$．由于$\mathcal{F}$中某个矩阵有一对共轭的非实特征值，(2.5.7)告诉我们：它有特别的形式(2.5.10)，其中$b\neq0$；如果必要，我们可以通过矩阵$\begin{bmatrix}1&0\\0&-1\end{bmatrix}$来执行一个相似运算以确保$b>0$．现在上一个习题告诉我们：$\mathcal{F}$中每个矩阵都有(2.5.10)的形式．执行最后的同时置换相似，使能得到分划矩阵(2.5.15a)中所展现的直和排列的次序． □ 137

如果$A,B\in M_n$是正规的(可以是复的或者实的)，且满足一个缠绕关系，Fuglede-Putnam **定理**说的是A^*与B^*满足同样的缠绕关系．这一结果证明的关键在于如下事实：对$a,b\in\mathbf{C}$，$ab=0$当且仅当$a\bar{b}=0$时成立．不同的证明参见(2.5.P26)．

定理2.5.16(Fuglede-Putnam) 设$A\in M_n$与$B\in M_m$是正规的，又设给定$X\in M_{n,m}$．那么$AX=XB$当且仅当$A^*X=XB^*$．

证明 设$A=U\Lambda U^*$以及$B=VMV^*$是谱分解．其中$\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，而$M=\mathrm{diag}(\mu_1,\cdots,\mu_m)$．设$U^*XV=[\xi_{ij}]$，那么，$AX=XB\Leftrightarrow U\Lambda U^*X=XVMV^*\Leftrightarrow\Lambda(U^*XV)=(U^*XV)M\Leftrightarrow\lambda_i\xi_{ij}=\xi_{ij}\mu_j$(对所有$i,j$)$\Leftrightarrow\xi_{ij}(\lambda_i-\mu_j)=0$(对所有$i,j$)$\Leftrightarrow\xi_{ij}\overline{(\lambda_i-\mu_j)}=0$(对所有$i,j$)$\Leftrightarrow\bar{\lambda}_i\xi_{ij}=\xi_{ij}\bar{\mu}_j$(对所有$i,j$)$\Leftrightarrow\bar{\Lambda}(U^*XV)=(U^*XV)\bar{M}\Leftrightarrow U\bar{\Lambda}U^*X=XV\bar{M}V^*\Leftrightarrow A^*X=XB^*$． □

上面两个定理引导到与它们的转置(或者等价地说，与它们的复共轭)可交换的正规矩阵的一种有用的表示．

习题 假设$A\in M_n$是正规的，$\bar{A}A=A\bar{A}$，且$A=B+\mathrm{i}C$，其中B与C是实的．说明为什么B与C是正规的且可交换．提示：$B=(A+\bar{A})/2$． ◀

习题 设给定$A=\begin{bmatrix}a&b\\-b&a\end{bmatrix}\in M_2(\mathbf{C})$，其中$b\neq0$，令$A_0=\begin{bmatrix}1&\mathrm{i}\\-\mathrm{i}&1\end{bmatrix}$，且$Q=\begin{bmatrix}1&0\\0&-1\end{bmatrix}$．证明：(a)$A$是非奇异的，当且仅当$A=c\begin{bmatrix}\alpha&\beta\\-\beta&\alpha\end{bmatrix}$，$c,\beta\neq0$，且$\alpha^2+\beta^2=1$；(b)$A$是奇异的非零矩阵，当且仅当$A$是$A_0$或者$\overline{A_0}$的一个纯量倍数；(c)$Q$是实正交矩阵且$\overline{A_0}=QA_0Q^T$． ◀

定理2.5.17 设$A\in M_n$是正规矩阵．则下面三个命题等价．

(a) $\bar{A}A=A\bar{A}$．

(b) $A^TA=AA^T$．

(c) 存在一个实正交矩阵Q，使得Q^TAQ是按照任意预先指定次序的分块的直和，其中每一个分块要么是零块，要么是

$$[1],\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix},\begin{bmatrix}a & b\\ -b & a\end{bmatrix}\text{或者}\begin{bmatrix}1 & \mathrm{i}\\ -\mathrm{i} & 1\end{bmatrix}(a,b\in\mathbf{C}) \tag{2.5.17.1}$$

 中某一个的非零的纯量倍数，其中 $a\neq 0\neq b$，且 $a^2+b^2=1$.

反之，如果 A 与形如(2.5.17.1)的分块的复的纯量倍数的某个直和实正交相似，那么 A 是正规的，且有 $A\overline{A}=\overline{A}A$.

证明 (a)与(b)的等价性由上一个定理推出：$\overline{A}A=A\overline{A}$，当且仅当 $A^{\mathrm{T}}A=(\overline{A})^*A=A(\overline{A})^*=AA^{\mathrm{T}}$.

设 $A=B+\mathrm{i}C$，其中 B 与 C 是实的．(2.5.16)后面的习题表明$\{B,C\}$是一个实正规的交换族，故而(2.5.15)确保存在一个实正交矩阵 Q 和一个非负整数 q，使得

$$Q^{\mathrm{T}}BQ=\Lambda(B)\oplus\begin{bmatrix}a_1(B) & b_1(B)\\ -b_1(B) & a_1(B)\end{bmatrix}\oplus\cdots\oplus\begin{bmatrix}a_q(B) & b_q(B)\\ -b_q(B) & a_q(B)\end{bmatrix}$$

以及

$$Q^{\mathrm{T}}CQ=\Lambda(C)\oplus\begin{bmatrix}a_1(C) & b_1(C)\\ -b_1(C) & a_1(C)\end{bmatrix}\oplus\cdots\oplus\begin{bmatrix}a_q(C) & b_q(C)\\ -b_q(C) & a_q(C)\end{bmatrix}$$

成立，其中 $\Lambda(B)$，$\Lambda(C)\in M_{n-2q}$ 中的每一个都是对角的，且对每一个 $j\in\{1,\cdots,q\}$，$b_j(B)$与 $b_j(C)$中至少有一个是正的．这样就有

$$\begin{aligned}Q^{\mathrm{T}}AQ&=Q^{\mathrm{T}}(B+\mathrm{i}C)Q\\&=\Lambda(A)\oplus\begin{bmatrix}\alpha_1(A) & \beta_1(A)\\ -\beta_1(A) & \alpha_1(A)\end{bmatrix}\oplus\cdots\oplus\begin{bmatrix}\alpha_q(A) & \beta_q(A)\\ -\beta_q(A) & \alpha_q(A)\end{bmatrix}\end{aligned} \tag{2.5.17.2}$$

其中 $\Lambda(A)=\Lambda(B)+\mathrm{i}\Lambda(C)$，每一个 $\alpha_j(A)=\alpha_j(B)+\mathrm{i}\alpha_j(C)$，且每一个 $\beta_j(A)=b_j(B)+\mathrm{i}b_j(C)\neq 0$．上一个习题表明，(2.5.17.2)中每一个非奇异的 2×2 分块都是$\begin{bmatrix}0 & 1\\ -1 & 0\end{bmatrix}$或者是$\begin{bmatrix}a & b\\ -b & a\end{bmatrix}$(其中 $a\neq 0\neq b$ 且 $a^2+b^2=1$)的非零纯量倍数．这也表明(2.5.17.2)中每一个奇异的 2×2 分块或者是$\begin{bmatrix}1 & \mathrm{i}\\ -\mathrm{i} & 1\end{bmatrix}$的非零纯量倍数，或者与它的一个非零纯量倍数实正交相似． □

上一定理的两种特殊情形在下一节里起着重要的作用：对称的酉矩阵或者是斜对称的酉矩阵．

习题 证明：(2.5.17.1)中的头两个分块是酉矩阵；第三个分块是复正交的但不是酉矩阵；而第四个分块则是奇异的，所以它既不是酉矩阵，也不是复正交矩阵． ◀

推论 2.5.18 设 $U\in M_n$ 是酉矩阵．

(a) 如果 U 是对称的，那么就存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$以及实数 $\theta_1,\cdots,\theta_n\in[0,2\pi)$，使得

$$U=Q\mathrm{diag}(\mathrm{e}^{\mathrm{i}\theta_1},\cdots,\mathrm{e}^{\mathrm{i}\theta_n})Q^{\mathrm{T}} \tag{2.5.19.1}$$

(b) 如果 U 是斜对称的，那么 n 是偶数且存在一个实正交矩阵 $Q\in M_n(\mathbf{R})$以及实数 $\theta_1,\cdots,\theta_{n/2}\in[0,2\pi)$，使得

$$U=Q\left(\mathrm{e}^{\mathrm{i}\theta_1}\begin{bmatrix}0&1\\-1&0\end{bmatrix}\oplus\cdots\oplus\mathrm{e}^{\mathrm{i}\theta_{n/2}}\begin{bmatrix}0&1\\-1&0\end{bmatrix}\right)Q^{\mathrm{T}}\tag{2.5.19.2}$$

反之，任何一个形如(2.5.19.1)的矩阵都是酉矩阵且是对称的，而任何一个形如(2.5.19.2)的矩阵都是酉矩阵且是斜对称的. 139

证明 或者对称或者斜对称的酉矩阵 U 满足恒等式 $UU^{\mathrm{T}}=U^{\mathrm{T}}U$，所以(2.5.17)确保存在一个实正交矩阵 Q，使得 $Q^{\mathrm{T}}AQ$ 是从(2.5.17.1)的四种类型中选取的分块的非零纯量倍数之直和.

(a) 如果 U 是对称的，从(2.5.17.1)中只可以选取对称的分块，所以 $Q^{\mathrm{T}}UQ$ 是形如 $c[1]$的分块之直和，其中 $|c|=1$，因为 U 是酉矩阵.

(b) 如果 U 是斜对称的，从(2.5.17.1)中只可以选取斜对称的分块，所以 $Q^{\mathrm{T}}UQ$ 是形如 $c\begin{bmatrix}0&1\\-1&0\end{bmatrix}$的分块之直和，其中 $|c|=1$，因为 U 是酉矩阵. 由此推出 n 是偶数. □

我们最后一个定理在酉相似这个框架下是与(1.3.29)类似的结果：实矩阵是酉相似的，当且仅当它们是实正交相似的. 为此我们给出如下的习题和推论作为预备知识.

习题 设给定 $U\in M_n$. 说明为什么 U 既是酉矩阵又是复正交矩阵的充分必要条件是：它是实正交矩阵. 提示：$U^{-1}=U^*=U^{\mathrm{T}}$. ◀

推论 2.5.20 设 $U\in M_n$ 是酉矩阵.

(a) 如果 U 是对称的，则存在一个对称的酉矩阵 V，使得 $V^2=U$，且 V 是关于 U 的多项式. 此外，V 与任何与 U 可交换的矩阵都是可交换的.

(b) (酉矩阵的 QS 分解)存在一个实正交矩阵 Q 以及一个对称的酉矩阵 S，使得 $U=QS$，且 S 是 $U^{\mathrm{T}}U$ 的多项式. 此外，S 以及任何与 $U^{\mathrm{T}}U$ 可交换的矩阵都是可交换的.

证明 (a) 利用上一个推论来分解 $U=P\mathrm{diag}(\mathrm{e}^{\mathrm{i}\theta_1},\cdots,\mathrm{e}^{\mathrm{i}\theta_n})P^{\mathrm{T}}$，其中 P 是实正交矩阵，而 $\theta_1,\cdots,\theta_n\in[0,2\pi)$是实数. 设 $p(t)$是这样一个多项式，对每一个 $j=1,\cdots,n$ 有 $p(\mathrm{e}^{\mathrm{i}\theta_j})=\mathrm{e}^{\mathrm{i}\theta_j/2}$(0.9.11.4)，又设 $V=p(U)$. 那么

$$\begin{aligned}V=p(U)&=p(P\mathrm{diag}(\mathrm{e}^{\mathrm{i}\theta_1},\cdots,\mathrm{e}^{\mathrm{i}\theta_n})P^{\mathrm{T}})\\&=Pp(\mathrm{diag}(\mathrm{e}^{\mathrm{i}\theta_1},\cdots,\mathrm{e}^{\mathrm{i}\theta_n}))P^{\mathrm{T}}=P\mathrm{diag}(p(\mathrm{e}^{\mathrm{i}\theta_1}),\cdots,p(\mathrm{e}^{\mathrm{i}\theta_n}))P^{\mathrm{T}}\\&=P\mathrm{diag}(\mathrm{e}^{\mathrm{i}\theta_1/2},\cdots,\mathrm{e}^{\mathrm{i}\theta_n/2})P^{\mathrm{T}}\end{aligned}$$

所以 V 是酉矩阵，且是对称的，而 $V^2=P(\mathrm{diag}(\mathrm{e}^{\mathrm{i}\theta_1/2},\cdots,\mathrm{e}^{\mathrm{i}\theta_n/2})^2)P^{\mathrm{T}}=P\mathrm{diag}(\mathrm{e}^{\mathrm{i}\theta_1},\cdots,\mathrm{e}^{\mathrm{i}\theta_n})P^{\mathrm{T}}=U$. 最后那个结论由(2.4.4.0)推出.

(b) (a)这部分的结论确保存在一个对称的酉矩阵 S，使得 $S^2=U^{\mathrm{T}}U$，且 S 是 $U^{\mathrm{T}}U$ 的多项式. 考虑酉矩阵 $Q=US^*$，它有如下性质：$QS=US^*S=U$. 计算 $Q^{\mathrm{T}}Q=S^*U^{\mathrm{T}}US^*=S^*S^2S^*=I$. 从而 Q 既是正交矩阵也是酉矩阵，故而上一个习题确保它是实正交的. □

习题 如果 $U\in M_n$ 是酉矩阵，说明为什么存在一个实正交矩阵 Q 以及一个对称的酉矩阵 S，使得 $U=SQ$，且 S 是 UU^{T} 的多项式. ◀ 140

定理 2.5.21 设$\mathcal{F}=\{A_\alpha:\alpha\in\mathcal{I}\}\subset M_n(\mathbf{R})$与$\mathcal{G}=\{B_\alpha:\alpha\in\mathcal{I}\}\subset M_n(\mathbf{R})$是给定的实矩阵族. 如果存在一个酉矩阵 $U\in M_n$，使得对每个 $\alpha\in\mathcal{I}$都有 $A_\alpha=UB_\alpha U^*$，那么就存在一个实

正交矩阵 $Q\in M_n(\mathbf{R})$，使得对每个 $\alpha\in\mathcal{I}$ 都有 $A_\alpha=QB_\alpha Q^{\mathrm{T}}$. 特别地，两个酉相似的实矩阵也是实正交相似的.

证明 由于每一个 A_α 与 B_α 都是实的，$A_\alpha=UB_\alpha U^*=\overline{U}B_\alpha U^{\mathrm{T}}=\overline{A}_\alpha$，所以对每个 $\alpha\in\mathcal{I}$ 都有 $U^{\mathrm{T}}UB_\alpha=B_\alpha U^{\mathrm{T}}U$. 上一个推论确保存在一个对称的酉矩阵 S 与一个实正交矩阵 Q，使得 $U=QS$ 且 S 与 B_α 可交换. 于是对每个 $\alpha\in\mathcal{I}$ 都有 $A_\alpha=UB_\alpha U^*=QSB_\alpha S^*Q^{\mathrm{T}}=QB_\alpha SS^*Q^{\mathrm{T}}=QB_\alpha Q^{\mathrm{T}}$. □

问题

2.5.P1 证明：$A\in M_n$ 是正规的，当且仅当对所有 $x\in\mathbf{C}^n$ 都有 $(Ax)^*(Ax)=(A^*x)^*(A^*x)$，也即对所有 $x\in\mathbf{C}^n$ 都有 $\|Ax\|_2=\|A^*x\|_2$.

2.5.P2 证明：正规矩阵是酉矩阵，当且仅当它所有特征值的绝对值都为 1.

2.5.P3 证明：正规矩阵是 Hermite 矩阵，当且仅当它所有特征值都是实的.

2.5.P4 证明：正规矩阵是斜 Hermite 矩阵，当且仅当它所有特征值都是纯虚数(即实部等于零).

2.5.P5 如果 $A\in M_n$ 是斜 Hermite(Hermite)矩阵，证明 $\mathrm{i}A$ 是 Hermite(斜 Hermite)矩阵.

2.5.P6 证明：$A\in M_n$ 是正规的，当且仅当它与某个有不同特征值的正规矩阵可交换.

2.5.P7 如同在(2.1.9)中那样，考虑形如 $A=B^{-1}B^*$ 的矩阵 $A\in M_n$(对某个非奇异的 $B\in M_n$). (a)证明：A 是酉矩阵当且仅当 B 是正规矩阵. (b)如果 B 有形式 $B=HNH$，其中 N 是正规矩阵，而 H 是 Hermite 矩阵(且两者都是非奇异的)，证明 A 与一个酉矩阵相似.

2.5.P8 将 $A\in M_n$ 写成 $A=H(A)+\mathrm{i}K(A)$，其中 $H(A)$与 $K(A)$是 Hermite 矩阵，见(0.2.5). 证明：A 是正规的，当且仅当 $H(A)$与 $K(A)$可交换.

2.5.P9 将 $A\in M_n$ 写成 $A=H(A)+\mathrm{i}K(A)$，其中 $H(A)$与 $K(A)$是 Hermite 矩阵. 如果 $H(A)$的每个特征向量都是 $K(A)$的一个特征向量，证明 A 是正规的. 关于其逆命题你有何结论？考虑 $A=\begin{bmatrix}1 & \mathrm{i}\\ -\mathrm{i} & 1\end{bmatrix}$.

2.5.P10 假设 $A,B\in M_n$ 两者均为正规矩阵. 如果 A 与 B 可交换，证明 AB 与 $A\pm B$ 全都是正规矩阵. 关于其逆有什么结论？验证 $A=\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}$，$B=\begin{bmatrix}-1 & 1\\ 1 & 1\end{bmatrix}$，$AB$ 以及 BA 全都是正规的，但是 A 与 B 不可交换.

2.5.P11 对任何复数 $z\in\mathbf{C}$，证明存在 $\theta,\tau\in\mathbf{R}$，使得 $\bar{z}=e^{\mathrm{i}\theta}z$ 以及 $|z|=e^{\mathrm{i}\tau}z$. 注意$[e^{\mathrm{i}\theta}]\in M_1$ 是酉矩阵. 对角酉矩阵 $U\in M_n$ 看起来像什么样？

2.5.P12 推广(2.5.P11)来证明：如果 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)\in M_n$，则存在对角的酉矩阵 U 与 V，使得 $\overline{\Lambda}=U\Lambda=\Lambda U$，且 $|\Lambda|=\mathrm{diag}(|\lambda_1|,\cdots,|\lambda_n|)=V\Lambda=\Lambda V$.

2.5.P13 利用(2.5.P12)证明：$A\in M_n$ 是正规的，当且仅当存在一个酉矩阵 $V\in M_n$，使得 $A^*=AV$. 如果 A 是正规的，导出结论：如果 A 是正规的，那么 $\mathrm{range}A=\mathrm{range}A^*$.

2.5.P14 设给定 $A\in M_n(\mathbf{R})$. 说明为什么 A 是正规的，且它所有的特征值是实的，当且仅当 A 是对称的.

2.5.P15 证明：两个正规矩阵是相似的，当且仅当它们有同样的特征多项式. 如果去掉两个矩阵均为正规的假设，结论还成立吗？考虑 $\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix}$与$\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$.

2.5.P16 如果 $U,V,\Lambda\in M_n$，且 U 与 V 是酉矩阵，证明 $U\Lambda U^*$ 与 $V\Lambda V^*$ 是酉相似的. 导出结论：两个正规矩阵相似，当且仅当它们是酉相似的. 给出两个可对角化的矩阵的例子：它们是相似的，但不是酉相似的.

2.5.P17 如果 $A\in M_n$ 是正规的，且 $p(t)$是给定的多项式，利用(2.5.1)证明 $p(A)$是正规的. 利用(2.5.3)对此事实给出另外一个证明.

2.5.P18 如果 $A\in M_n$ 且存在一个非零的多项式 $p(t)$，使得 $p(A)$是正规的，可以推出 A 是正规的吗?

2.5.P19 设给定 $A\in M_n$ 以及 $a\in\mathbf{C}$. 利用定义(2.5.1)证明：A 是正规的，当且仅当 $A+aI$ 是正规的. 不得借助于谱定理(2.5.3).

2.5.P20 设 $A\in M_n$ 是正规的，并假设 $x\in\mathbf{C}^n$ 是 A 的与特征值 λ 相伴的右特征向量. 利用(2.5.P1)以及(2.5.P19)证明：x 是 A 的与同一个特征值 λ 相伴的左特征向量.

2.5.P21 假设 $A\in M_n$ 是正规的. 利用上一个问题证明：$Ax=0$ 当且仅当 $A^*x=0$，这就是说，A 的零空间与 A^* 的零空间相同. 考虑 $B=\begin{bmatrix}0&1\\0&1\end{bmatrix}$来证明：即便 B 可对角化，非正规矩阵 B 的零空间也不一定与 B^* 的零空间相同.

2.5.P22 利用(2.5.6)证明：复 Hermite 矩阵的特征多项式的系数是实数.

2.5.P23 矩阵$\begin{bmatrix}1&\mathrm{i}\\\mathrm{i}&1\end{bmatrix}$与$\begin{bmatrix}\mathrm{i}&\mathrm{i}\\\mathrm{i}&-1\end{bmatrix}$两者都是对称的. 证明：其中一个是正规的，而另一个不是. 这就是实对称矩阵与复对称矩阵之间的一个重要区别.

2.5.P24 如果 $A\in M_n$ 既是正规的又是幂零的，证明 $A=0$.

2.5.P25 假设 $A\in M_n$ 与 $B\in M_m$ 是正规的，并设给定 $X\in M_{n,m}$. 说明为什么 $\overline{B}$ 是正规的，并导出结论：$AX=X\overline{B}$当且仅当 $A^*X=XB^T$.

2.5.P26 设给定 $A\in M_n$. (a)如果存在一个多项式 $p(t)$，使得 $A^*=p(A)$，证明 $A\in M_n$ 是正规的. (b)如果 A 是正规的，证明存在一个次数至多为 $n-1$ 的多项式 $p(t)$，使得 $A^*=p(A)$. (c)如果 A 是实的正规矩阵，证明存在一个实系数且次数至多为 $n-1$ 的多项式 $p(t)$，使得 $A^T=p(A)$. (d)如果 A 是正规矩阵，证明存在一个实系数且次数至多为 $2n-1$ 的多项式 $p(t)$，使得 $A^*=p(A)$. (e)如果 A 是正规的，且 $B\in M_m$ 也是正规的，证明：存在一个次数至多为 $n+m-1$ 的多项式 $p(t)$，使得 $A^*=p(A)$且 $B^*=p(B)$. (f)如果 A 是正规的，且 $B\in M_m$ 也是正规的， 142
证明：存在一个实系数且次数至多为 $2n+2m-1$ 的多项式 $p(t)$，使得 $A^*=p(A)$且 $B^*=p(B)$. (g)利用(e)以及(2.4.4.0)证明 Fuglede-Putnam 定理(2.5.16). (h)利用(f)证明(2.5.P25)中的结论.

2.5.P27 (a)设 $A,B\in M_{n,m}$. 如果 AB^* 与 B^*A 两者都是正规的，证明 $BA^*A=AA^*B$. (b)设 $A\in M_n$. 证明 $A\overline{A}$是正规的(这样一个矩阵称为是**相合正规的**(congruence normal))，当且仅当 $AA^*A^T=A^TA^*A$. (c)如果 $A\in M_n$ 是相合正规的，证明三个正规矩阵 $A\overline{A}$，$\overline{A^*A}$以及 AA^* 可交换，从而可同时酉对角化.

2.5.P28 设给定 Hermite 矩阵 $A,B\in M_n$，并假设 AB 是正规的. (a)为什么 BA 是正规的? (b)证明 A 与 B^2 可交换，且 B 与 A^2 可交换. (c)如果存在一个多项式 $p(t)$，使得有 $A=p(A^2)$或者 $B=p(B^2)$，证明 A 与 B 可交换，且 AB 实际上是 Hermite 矩阵. (d)说明为什么(c)中的条件是满足的，如果 A 或者 B 有如下性质：只要 λ 是一个非零的特征值，那么$-\lambda$ 就不会也是一个特征值. 例如，如果 A 或者 B 所有的特征值均不是负数，这个条件就满足. (d)讨论例子 $A=\begin{bmatrix}0&1\\1&0\end{bmatrix}$，$B=\begin{bmatrix}0&\mathrm{i}\\-\mathrm{i}&0\end{bmatrix}$.

2.5.P29 设 $A=\begin{bmatrix}a&b\\c&d\end{bmatrix}\in M_2$ 并假设 $bc\neq0$. (a)证明：A 是正规的当且仅当存在某个 $\theta\in\mathbf{R}$，使得 $c=e^{i\theta}b$ 以及 $a-d=e^{i\theta}b(\overline{a}-\overline{d})/\overline{b}$. 特别地，如果 A 是正规的，则必定有 $|c|=|b|$. (b)设 $b=|b|e^{i\phi}$. 如果 A 是正规的，且 $c=be^{i\theta}$，证明 $e^{-i(\phi+\theta/2)}(A-aI)$是 Hermite 的. 反之，如果存在一个 $\gamma\in\mathbf{C}$

使得 $A-\gamma I$ 是本性 Hermite 的，证明 A 是正规的.(c)如果 A 是实的，从(a)导出结论：A 是正规的，当且仅当或者 $c=b(A=A^{\mathrm{T}})$，或者 $c=-b$ 且 $a=d(AA^{\mathrm{T}}=(a^2+b^2)I$ 以及 $A=-A^{\mathrm{T}}$，如果 $a=0)$.

2.5.P30 证明：一个给定的 $A\in M_n$ 是正规的，当且仅当对所有 x，$y\in\mathbf{C}^n$ 都有 $(Ax)^*(Ay)=(A^*x)^*\times(A^*y)$. 如果 A，x 以及 y 是实的，这就意味着 Ax 以及 Ay 之间的角度与 $A^{\mathrm{T}}x$ 以及 $A^{\mathrm{T}}y$ 之间的角度总是相同的. 与(2.5.P1)比较. 如果我们取 $x=e_i$ 以及 $y=e_j$(标准 Euclid 基向量)，这个条件的含义是什么？如果对所有 i，$j=1$，…，n 都有 $(Ae_i)^*(Ae_j)=(A^*e_i)^*(A^*e_j)$. 证明 A 是正规的.

2.5.P31 设 $A\in M_n(\mathbf{R})$是实正规矩阵，即 $AA^{\mathrm{T}}=A^{\mathrm{T}}A$. 如果 AA^{T} 有 n 个不同的特征值，证明 A 是对称的.

2.5.P32 如果 $A\in M_3(\mathbf{R})$是实正交的，注意到 A 有一个或者三个实的特征值. 如果它有正的行列式，利用(2.5.11)证明它正交相似于$[1]\in M_1$ 与一个平面旋转的直和. 讨论其几何解释，视之为绕着某个经过 $\mathbf{R}^3$ 中原点的固定轴转动 θ 角的一个旋转. 这是力学中 Euler 定理的一部分：刚体的每一个运动都是一个平移以及绕某个轴的一个旋转的复合.

2.5.P33 如果$\mathcal{F}\subseteq M_n$ 是正规矩阵组成的一个交换族，证明存在单独一个 Hermite 矩阵 B，使得对每个 $A_\alpha\in\mathcal{F}$，存在一个次数至多为 $n-1$ 的多项式 $p_\alpha(t)$，使得 $A_\alpha=p_\alpha(B)$. 注意：对$\mathcal{F}$中所有矩阵，B 都是固定不变的，但是多项式可能与$\mathcal{F}$的元素有关.

2.5.P34 设 $A\in M_n$，并设 $x\in\mathbf{C}^n$ 不为零. 我们称 x 是 A 的一个**正规特征向量**(normal eigenvector)，如果它同时是 A 的右特征向量以及左特征向量.(a)如果 $Ax=\lambda x$ 且 $x^*A=\mu x^*$，证明 $\lambda=\mu$.(b)如果 x 是 A 的与特征值 λ 相伴的正规特征向量，证明 A 与$[\lambda]\oplus A_1$ 酉相似，其中 $A_1\in M_{n-1}$是上三角的.(c)证明：A 是正规的，当且仅当它的每一个特征向量都是正规特征向量.

2.5.P35 设 x，$y\in\mathbf{C}^n$ 是给定的非零向量.(a)证明：$xx^*=yy^*$ 当且仅当存在某个实数 θ，使得 $x=e^{i\theta}y$.(b)证明如下诸结论对于秩 1 矩阵 $A=xy^*$ 是等价的：(i)A 是正规的；(ii)存在一个正实数 r 以及一个实数 $\theta\in[0,2\pi)$，使得 $x=re^{i\theta}y$；(iii)A 是本性 Hermite 的.

2.5.P36 对任何 $A\in M_n$，证明：$\begin{bmatrix}A & A^*\\ A^* & A\end{bmatrix}\in M_{2n}$是正规的. 于是，任何方阵都能成为一个正规矩阵的主子矩阵(每一个 $A\in M_n$ 都有一个成为正规矩阵的**膨体**(dilation)). 任何方阵都能是一个 Hermite 矩阵的主子矩阵吗？对酉矩阵呢？

2.5.P37 设 $n\geqslant 2$，并假设 $A=\begin{bmatrix}a & x^*\\ y & B\end{bmatrix}\in M_n$ 是正规的，其中 $B\in M_{n-1}$，而 x，$y\in\mathbf{C}^{n-1}$.(a)证明 $\|x\|_2=\|y\|_2$ 以及 $xx^*-yy^*=BB^*-B^*B$.(b)说明为什么对每个复方阵 F 都有 $\operatorname{rank}(FF^*-F^*F)\neq 1$.(c)说明为什么存在两种互相排斥的可能性：或者(i)主子矩阵 B 是正规的，或者(ii)$\operatorname{rank}(BB^*-B^*B)=2$.(d)说明为什么 B 是正规的，当且仅当对某个实数 θ 有 $x=e^{i\theta}y$.(e)讨论例子 $B=\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}$，$x=[-\sqrt{2}\quad 1]^{\mathrm{T}}$，$y=[1\quad -\sqrt{2}]^{\mathrm{T}}$ 以及 $a=1-\sqrt{2}$.

2.5.P38 设 $A=[a_{ij}]\in M_n$ 以及 $C=AA^*-A^*A$.(a)说明为什么 C 是 Hermite 矩阵以及为什么 C 是幂零的，当且仅当 $C=0$.(b)证明：A 是正规的，当且仅当它与 C 可交换.(c)证明 $\operatorname{rank}C\neq 1$.(d)说明为什么 A 是正规的，当且仅当 $\operatorname{rank}C\leqslant 1$，即只有两种可能性：$\operatorname{rank}C=0$($A$ 是正规的)以及 $\operatorname{rank}C\geqslant 2$($A$ 不是正规的). 我们称 A 是**几乎正规的**(nearly normal)，如果 $\operatorname{rank}C=2$.(d)假设 A 是三对角的 Toeplitz 矩阵. 证明 $C=\operatorname{diag}(\alpha,0,\cdots,0,-\alpha)$，其中 $\alpha=|a_{12}|^2-|a_{21}|^2$. 推出结论：$A$ 是正规的，当且仅当 $|a_{12}|=|a_{21}|$；如若不然，A 就是几乎正规的.

2.5.39 假设 $U\in M_n$ 是酉矩阵，故而它所有的特征值的模均为 1.(a)如果 U 是对称的，证明它的特征值

唯一地确定它的表示(2.5.19.1)，相差不过是对角元素的排列次序. (b)如果 U 是斜对称的，说明(2.5.19.2)中的纯量 $e^{i\theta_j}$ 是如何与它的特征值联系在一起的. 为什么 U 的特征值必定按照 ± 成对地出现？证明：U 的特征值唯一地决定它的表示(2.5.19.2)，所相差的不过是其直和项的排列次序.

2.5.P40 设 $A=\begin{bmatrix}0 & B\\ 0 & 0\end{bmatrix}\in M_4$，其中 $B=\begin{bmatrix}1 & i\\ -i & 1\end{bmatrix}$. 验证 A 与 A^{T} 可交换，A 与 $\bar{A}$ 可交换，但 A 与 A^* 不可交换，即 A 不是正规的.

2.5.P41 设 $z\in\mathbf{C}^n$ 是非零的，并记 $z=x+iy$，其中 x，$y\in\mathbf{R}^n$. (a)证明下面三个命题是等价的：(1)$\{z, \bar{z}\}$是线性相关的；(2)$\{x, y\}$是线性相关的；(3)存在一个单位向量 $u\in\mathbf{R}^n$ 和一个非零的 $c\in\mathbf{C}$，使得 $z=cu$. (b)证明以下诸命题等价：(1)$\{z, \bar{z}\}$是线性无关的；(2)$\{x, y\}$是线性无关的；(3)存在标准正交的实向量 v，$w\in\mathbf{R}^n$，使得在 $\mathbf{C}$ 上有 $\mathrm{span}\{z, \bar{z}\}=\mathrm{span}\{v, w\}$.

2.5.P42 设 A，$B\in M_n$，并假设 λ_1，…，λ_n 是 A 的特征值. 函数 $\Delta(A)=\mathrm{tr}A^*A-\sum_{i=1}^{n}|\lambda_i|^2$ 称为 A **偏离正规性的亏量**(defect of A from normality). Schur 不等式(2.3.2a)是说 $\Delta(A)\geqslant 0$，(2.5.3(c))则保证：A 是正规的，当且仅当 $\Delta(A)=0$. (a)如果 A，B 以及 AB 都是正规的，证明 144
$\mathrm{tr}((AB)^*(AB))=\mathrm{tr}((BA)^*(BA))$，并说明为什么 BA 是正规的. (b)假设 A 是正规的，A 与 B 有同样的特征多项式，且 $\mathrm{tr}A^*A=\mathrm{tr}B^*B$. 证明 B 是正规的，且与 A 酉相似. 与(2.5.P15)比较.

2.5.P43 设 $A=[a_{ij}]\in M_n$ 是正规的. (a)分划 $A=[A_{ij}]_{i,j=1}^{k}$，其中每一个 A_{ii} 都是方阵. 假设 A 的特征值就是 A_{11}，A_{22}，…以及 A_{kk} 的特征值(计入重数)；例如，我们可以假设 $p_A(t)=p_{A_{11}}(t)\cdots p_{A_{kk}}(t)$. 证明 A 是分块对角的，即对所有 $i\neq j$ 都有 $A_{ij}=0$，且每个对角分块 A_{ii} 都是正规的. (b)如果 A 的每个对角元素 a_{ii} 都是 A 的特征值，说明为什么它是对角矩阵.

2.5.P44 (a)证明：$A\in M_n$ 是 Hermite 的，当且仅当 $\mathrm{tr}A^2=\mathrm{tr}A^*A$. (b)证明：Hermite 矩阵 A，$B\in M_n$ 可交换，当且仅当 $\mathrm{tr}(AB)^2=\mathrm{tr}(A^2B^2)$.

2.5.P45 设 $\mathcal{F}\subseteq M_n(\mathbf{R})$是由实对称矩阵组成的一个交换族. 证明：存在单独一个实正交矩阵 Q，使得对每一个 $A\in\mathcal{F}$，Q^TAQ 都是对角矩阵.

2.5.P46 利用(2.3.1)证明：实矩阵的任何非实的特征值都必定成对共轭出现.

2.5.P47 设 $A\in M_n$ 是正规的，且有特征值 λ_1，…，λ_n. 证明(a)adj$A$ 是正规的，且有特征值 $\prod_{j\neq i}\lambda_j$，$i=1$，…，n；(b)adj$A$ 是 Hermite 的，如果 A 是 Hermite 的；(c)adj$A$ 有正的(非负的)特征值，如果 A 有正的(非负的)特征值；(d)adj$A$ 是酉矩阵，如果 A 是酉矩阵.

2.5.P48 设 $A\in M_n$ 是正规的，并假设 $\mathrm{rank}A=r>0$. 利用(2.5.3)记 $A=U\Lambda U^*$，其中 $U\in M_n$ 是酉矩阵，且 $\Lambda=\Lambda_r\oplus 0_{n-r}$是对角矩阵. (a)说明为什么 $\det\Lambda_r\neq 0$. 设 $\det\Lambda_r=|\det\Lambda_r|e^{i\theta}$，其中 $\theta\in[0, 2\pi)$. (b)分划 $U=[V\quad U_2]$，其中 $V\in M_{n,r}$. 说明为什么 $A=V\Lambda_rV^*$. 这是 A 的一个满秩分解. (c)设 α，$\beta\subseteq\{1, …, n\}$是基数为 r 的指标集，并令 $V[\alpha\quad\varnothing^c]=V_\alpha$. 说明为什么$A[\alpha\quad\beta]=V_\alpha\Lambda_rV_\beta^*$；$\det A[\alpha]\det A[\beta]=\det A[\alpha\quad\beta]\det A[\beta\quad\alpha]$；以及 $\det A[\alpha]=|\det V_\alpha|^2\det\Lambda_r$. (d)说明为什么 A 的每个 r 阶主子式都位于复平面的射线$\{se^{i\theta}: s\geqslant 0\}$上，且这些主子式中至少有一个不为零. (e)导出结论：A 是主秩的. 如果 A 是 Hermite 矩阵，此结果的一种表述形式见(4.2.P30)，也见(3.1.P20).

2.5.P49 假设 $A\in M_n$ 是上三角的，且可以对角化. 证明它可以通过一个上三角相似来对角化.

2.5.P50 反序矩阵 K_n(0.9.5.1)是实对称的. 检验它也是实正交的，并说明为什么它的特征值只能是±1. 验证：如果 n 是偶数，则 $\mathrm{tr}K_n=0$；如果 n 是奇数，则 $\mathrm{tr}K_n=1$. 说明为什么当 n 为偶

数时，K_n 的特征值为 ± 1，每一个特征值的重数均为 $n/2$；而当 n 为奇数时，K_n 的特征值为 $+1$（其重数为 $(n+1)/2$）以及 -1（其重数为 $(n-1)/2$）.

2.5.P51 设 $A\in M_n$ 是正规矩阵，令 $A=U\Lambda U^*$ 是谱分解，其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，设 $x\in\mathbf{C}^n$ 是任意给定的单位向量，又设 $\xi=[\xi_i]=Ux$. 说明为什么 $x^*Ax=\sum_{i=1}^{n}|\xi_i|^2\lambda_i$，为什么 x^*Ax 位于 A 的特征值组成的凸包中，又为什么位于 A 的特征值的凸包中的每一个复数都等于 x^*Ax（对某个单位向量 x）. 于是，如果 A 是正规的，那么对每个单位向量 x 都有 $x^*Ax\neq 0$ 的充分必要条件是：0 不在 A 的特征值的凸包中.

2.5.P52 设 A，$B\in M_n$ 是非奇异的. 矩阵 $C=ABA^{-1}B^{-1}$ 称为 A 与 B 的**积性换位子**（multiplicative commutator）. 说明为什么 $C=I$ 成立的充分必要条件是 A 与 B 可交换. 假设 A 与 C 是正规的，且 0 不在 B 的特征值的凸包之中. 下面给出了 A 与 C 可交换当且仅当 A 与 B 可交换（这就是 **Marcus-Thompson 定理**）这一命题的证明概述，请补充证明的细节：设 $A=U\Lambda U^*$ 以及 $C=UMU^*$ 是谱分解，其中 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，而 $M=\mathrm{diag}(\mu_1,\cdots,\mu_n)$. 设 $\mathcal{B}=U^*BU=[\beta_{ij}]$. 那么所有 $\beta_{ii}\neq 0$，且 $M=U^*CU=\Lambda\mathcal{B}\Lambda^{-1}\mathcal{B}^{-1}\Rightarrow M\mathcal{B}=\Lambda\mathcal{B}\Lambda^{-1}\Rightarrow\mu_i\beta_{ii}=\beta_{ii}\Rightarrow M=I\Rightarrow C=I$. 与（2.4.P12(c)）比较.

2.5.P53 设 U，$V\in M_n$ 是酉矩阵，并假设 V 的所有的特征值都落在单位圆的长度为 π 的一段开弧上；这样一个矩阵称为**受限的酉矩阵**（cramped unitary matrix）. 令 $C=UVU^*V^*$ 是 U 与 V 的积性换位子. 利用上一个问题证明 **Frobenius 定理**：U 与 C 可交换，当且仅当 U 与 V 可交换.

2.5.P54 如果 A，$B\in M_n$ 是正规的，证明：(a) A 的零空间与 A 的值域正交；(b) A 与 A^* 的值域相同；(c) A 的零空间包含在 B 的零空间中，当且仅当 A 的值域包含 B 的值域.

2.5.P55 验证（2.2.8）对正规矩阵的如下改进：如果 A，$B\in M_n$ 是正规的，那么 A 与 B 酉相似，当且仅当 $\mathrm{tr}A^k=\mathrm{tr}B^k$，$k=1,2\cdots,n$.

2.5.P56 设给定 $A\in M_n$ 以及一个整数 $k\geqslant 2$，又设 $\omega=\mathrm{e}^{2\pi\mathrm{i}/(k+1)}$. 证明 $A^k=A^*$，当且仅当 A 是正规的，且它的谱包含在集合 $\{0,1,\omega,\omega^2,\cdots,\omega^k\}$ 中. 如果 $A^k=A^*$，且 A 是非奇异的，说明为什么它是酉矩阵.

2.5.P57 设给定 $A\in M_n$. 证明：(a) A 是正规的且是对称的，当且仅当存在一个实正交矩阵 $Q\in M_n$ 以及一个对角矩阵 $\Lambda\in M_n$，使得 $A=Q\Lambda Q^{\mathrm{T}}$；(b) A 是正规的，且是斜对称的，当且仅当存在一个实正交矩阵 $Q\in M_n$，使得 $Q^{\mathrm{T}}AQ$ 是一个零块与形如 $\begin{bmatrix}0 & z_j\\ -z_j & 0\end{bmatrix}$（$z_j\in\mathbf{C}$）的分块的直和.

2.5.P58 设 $A\in M_n$ 是正规的. 那么 $A\overline{A}=0$，当且仅当 $AA^{\mathrm{T}}=A^{\mathrm{T}}A=0$. (a) 利用（2.5.17）证明之. (b) 对另一种证明提供细节：$A\overline{A}=0\Rightarrow 0=A^*A\overline{A}=0=AA^*\overline{A}\Rightarrow\overline{A}A^{\mathrm{T}}A=0\Rightarrow(A^{\mathrm{T}}A)^*(A^{\mathrm{T}}A)=0\Rightarrow A^{\mathrm{T}}A=0$（0.2.5.1）.

2.5.P59 设 A，$B\in M_n$. 假设 A 是正规的且有不同的特征值. 如果 $AB=BA$，证明 B 是正规的. 与（1.3.P3）比较.

2.5.P60 设给定 $x=[x_i]\in\mathbf{C}^n$. (a) 说明为什么 $\max_i|x_i|\leqslant\|x\|_2$（0.6.1）. (b) 设 $e=e_1+\cdots+e_n\in\mathbf{C}^n$ 是所有元素都是 $+1$ 的向量. 如果 $x^{\mathrm{T}}e=0$，证明 $\max_i|x_i|\leqslant\sqrt{\frac{n-1}{n}}\|x\|_2$，其中等式当且仅当对某个 $c\in M_n$ 以及某个指标 j 有 $x=c(ne_j-e)$ 时成立.

2.5.P61 设给定的 $A\in M_n$ 的特征值是 $\lambda_1,\cdots,\lambda_n$. (a) 证明

$$\max_{i=1,\cdots,n}\left|\lambda_i-\frac{\mathrm{tr}A}{n}\right|\leqslant\sqrt{\frac{n-1}{n}}\left(\sum_{i=1}^{n}|\lambda_i|^2-\frac{|\mathrm{tr}A|^2}{n}\right)^{1/2}$$

并推导出

$$\max_{i=1,\cdots,n}\left|\lambda_i-\frac{\mathrm{tr}A}{n}\right|\leqslant\sqrt{\frac{n-1}{n}}\left(\mathrm{tr}A^*A-\frac{|\mathrm{tr}A|^2}{n}\right)^{1/2}$$

其中等式当且仅当 A 是正规的且有特征值 $(n-1)c$，$-c$，…，$-c$(对某个 $c\in\mathbf{C}$)时成立．(b)关于 A 的特征值从几何上说有何结论？如果 A 是 Hermite 矩阵呢？(c)量 $\mathrm{spread}A=\max\{|\lambda-\mu|:\lambda,$ 146
$\mu\in\sigma(A)\}$ 表示 A 的两个特征值之间的最大距离．说明为什么 $\mathrm{spread}A\leqslant 2\sqrt{\frac{n-1}{n}}\times\left(\mathrm{tr}A^*A-\frac{|\mathrm{tr}A|^2}{n}\right)^{1/2}$，又如果 A 是 Hermite 矩阵，则 $\mathrm{spread}A\leqslant 2\sqrt{\frac{n-1}{n}}\left(\mathrm{tr}A^2-\frac{|\mathrm{tr}A|^2}{n}\right)^{1/2}$．关于 $\mathrm{spread}A$ 的下界，见(4.3.P16)．

2.5.P62 如果 $A\in M_n$ 恰有 k 个非零的特征值，我们知道 $\mathrm{rank}A\geqslant k$．如果 A 是正规的，为什么 $\mathrm{rank}A=k$？

2.5.P63 假设 $A=[a_{ij}]\in M_n$ 是三对角的．如果 A 是正规的，证明：对每个 $i=1$，…，$n-1$ 有 $|a_{i,i+1}|=|a_{i+1,i}|$．关于其逆有何结论？与(2.5.P38(d))作比较．

2.5.P64 设给定 $A\in M_n$．(a)证明 $\mathrm{rank}(AA^*-A^*A)\neq 1$．

2.5.P65 假设 $A\in M_n$ 是正规的，又令 $r\in\{1,\cdots,n\}$．说明为什么复合矩阵 $C_r(A)$ 是正规的．

2.5.P66 设 $A\in M_n$．我们称 A 是**平方正规的**(squared normal)，如果 A^2 是正规的．已知：A 是平方正规的，当且仅当 A 与这样的分块之直和酉相似，其中的每一分块都有如下之形状：

$$[\lambda]\ \text{或者}\ \tau\begin{bmatrix}0&1\\ \mu&0\end{bmatrix},\text{其中}\ \tau\in\mathbf{R},\lambda,\mu\in\mathbf{C},\tau>0\ \text{且}\ |\mu|<1.\qquad(2.5.22a)$$

这个直和是由 A 唯一决定的，相差不过是各个分块的排列次序．利用(2.2.8)证明(2.5.22a)中每一个 2×2 分块都与一个形如 $\begin{bmatrix}\nu&r\\0&-\nu\end{bmatrix}$ 的分块酉相似，其中 $\nu=\tau\sqrt{\mu}\in\mathcal{D}_+$，$r=\tau(1-|\mu|)$，且 $\mathcal{D}_+=\{z\in\mathbf{C}:\ \mathrm{Re}\,z>0\}\cup\{it:\ t\in\mathbf{R}\ \text{且}\ t\geqslant 0\}$．导出结论：$A^2$ 是正规的，当且仅当 A 与这样的分块之直和酉相似，其中每一分块都有如下之形状：

$$[\lambda]\ \text{或者}\begin{bmatrix}\nu&r\\0&-\nu\end{bmatrix},\text{其中}\ \lambda,\mu\in\mathbf{C},r\in\mathbf{R},r>0\ \text{且}\ \nu\in\mathcal{D}_+.\qquad(2.5.22b)$$

说明为什么这个直和由 A 唯一决定(相差至多是分块的排列)．

2.5.P67 设 A，$B\in M_n$ 是正规的．证明：$AB=0$ 当且仅当 $BA=0$．

2.5.P68 设 A，$B\in M_n$．假设 B 是正规的且 A 的零空间中的每个向量都是正规的特征向量．证明：$AB=0$ 当且仅当 $BA=0$．

2.5.P69 考虑一个 $k\times k$ 分块矩阵 $M_A=[A_{ij}]_{i,j=1}^k\in M_{kn}$，其中 $A_{ij}=0$(如果 $i\geqslant j$)以及 $A_{ij}=I_n$(如果 $j=i+1$)．类似地定义 $M_B=[B_{ij}]_{i,j=1}^k$．设 $W=[W_{ij}]_{i,j=1}^k\in M_{kn}$ 与 M_A 以及 M_B 共形地分划．(a)如果 $M_AW=WM_B$，证明 W 是分块上三角的，且 $W_{11}=\cdots=W_{kk}$．(b)如果 W 是酉矩阵，且 $M_AW=WM_B$(即 $M_A=WM_BW^*$，故而 M_A 与 M_B 通过 W 酉相似)，证明 W 是分块对角的，$W_{11}=U$ 是酉矩阵，$W=U\oplus\cdots\oplus U$，且对所有 i，j 都有 $A_{ij}=UB_{ij}U^*$．分块矩阵 M_A 以及 M_B 的进一步的性质，见(4.4.P46)以及(4.4.P47)．

2.5.P70 设给定 $n\times n$ 复矩阵对 (A_1,B_1)，…，(A_m,B_m)．我们称这些矩阵对是**同时酉相似的**(simultaneously unitarily similar)，如果存在一个酉矩阵 $U\in M_n$，使得对每个 $j=1$，…，m，都有 $A_j=UB_jU^*$．考虑 $(m+2)\times(m+2)$ 分块矩阵 $N_A=[N_{ij}]_{i,j=1}^k$，其中 $N_{ij}=I_n$(如果 $j=i+1$)，$N_{ij}=A_i$(如果 $j=i+2$)以及 $N_{ij}=0$(如果 $j-i\notin\{1,2\}$)．以类似的方式定义 N_B．(a)说明为什么 N_A 与 N_B 酉相似的充分必要条件是矩阵对 (A_1,B_1)，…，(A_m,B_m) 同时酉相似．(b)描述具有这种令人愉悦的性质的其他的分块矩阵．(c)说明怎样通过有限多次计算对有限多对同阶复矩阵对是否同时酉相似加以验证或者证伪．

2.5.P71　矩阵 $\begin{bmatrix} a & b \\ -b & a \end{bmatrix} \in M_2(\mathbf{R})$ 在我们关于实正规矩阵的讨论中起着重要的作用．作为(3.1.P20)中研究过的实表示 $R_1(A)$ 的一个特例，讨论这个矩阵的性质．

2.5.P72　考虑矩阵 $A_1=\begin{bmatrix} \mathrm{i} & 0 \\ 0 & -\mathrm{i} \end{bmatrix}$ 以及 $A_2=\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$．(a)证明每一个矩阵都是正规的且与它的复共轭可交换．(b)每一个矩阵的特征值是什么？(c)说明为什么 A_1 与 A_2 酉相似．(d)证明 A_1 不与 A_2 实正交相似．(e)设 $A\in M_n$ 是正规的且满足(2.5.17)的条件(a)或者(b)中的某一个．那么 A 与一个直和实正交相似，这个直和是由一个零矩阵以及(2.5.17.1)中四种类型的分块中的一个或者多个的非零纯量倍数作成的．为了确定哪些分块以及何种纯量倍数会在直和中出现，说明为什么我们必须除了知晓特征值之外还需对 A 知道得更多．

2.5.P73　设 $A\in M_n(\mathbf{R})$ 是正规的，并令 λ，x 是 A 的一个特征对，其中 $\lambda=a+\mathrm{i}b$ 不是实数．(a)证明 $\bar{\lambda}$，$\bar{x}$ 是 A 的特征对，且 $x^{\mathrm{T}}x=0$．(b)设 $x=u+\mathrm{i}v$，其中 u 与 v 是实向量．证明 $u^{\mathrm{T}}u=v^{\mathrm{T}}v\neq 0$ 以及 $u^{\mathrm{T}}v=0$．(c)设 $q_1=u/\sqrt{u^{\mathrm{T}}u}$ 以及 $q_2=v/\sqrt{v^{\mathrm{T}}v}$，并设 $Q=[q_1\quad q_2\quad Q_1]\in M_n(\mathbf{R})$ 是实正交矩阵．证明 $Q^{\mathrm{T}}AQ=\begin{bmatrix} B & * \\ 0 & * \end{bmatrix}$，其中 $B=\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$．(d)给(2.5.8)另一个不依赖于(2.3.4b)的证明．

2.5.P74　设 A，B，$X\in M_n$．如果 $AX=XB$ 且 X 是正规的，$AX^*=X^*B$ 是正确的吗？与 Fuglede-Putnam 定理(2.5.16)作比较．

2.5.P75　设 A，B，$X\in M_n$．(a)证明 $AX=XB$ 且 $XA=BX$ 成立的充分必要条件是 $\begin{bmatrix} 0 & X \\ X & 0 \end{bmatrix}$ 与 $\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}$ 可交换．(b)如果 X 是正规的，$AX=XB$ 且 $XA=BX$，证明 $AX^*=X^*B$ 以及 $X^*A=BX^*$．

2.5.P76　假设 $A\in M_n(\mathbf{R})$ 的每个元素都是 0 或者 1，设 $e\in\mathbf{R}^n$ 是全 1 向量，又设 $J\in M_n(\mathbf{R})$ 是全 1 矩阵．令 $A=[c_1\quad\cdots\quad c_n]$ 以及 $A^{\mathrm{T}}=[r_1\quad\cdots\quad r_n]$．(a)说明为什么 Ae 的元素既是行和，又是 A 的行的 Euclid 范数之平方．用类似的方式解释 $A^{\mathrm{T}}e$ 的元素．(b)证明：A 是正规的，当且仅当 $Ae=A^{\mathrm{T}}e$ 以及 $c_i^{\mathrm{T}}c_j=r_i^{\mathrm{T}}r_j$(对所有 $i\neq j$)．(c)证明：A 是正规的，当且仅当补余 0-1 矩阵 J-A 是正规的．

注记以及进一步的阅读参考　有关正规性的 89 种特征刻画的讨论，见 R. Grone，C. R. Johnson，E. Sa，以及 H. Wolkowicz，Normal matrices，*Linear Algebra Appl*. 87(1987) 213-225 以及 L. Elsner 与 Kh. Ikramov，Normal matrices：an update，*Linear Algebra Appl*. 285(1998) 291-303．尽管在(2.5.P72)中提出了这一事项，(2.5.17)中所描述的表示实际上就是在实正交相似之下的标准型，相差仅限于直和项的排列次序以及直和项可以用其转置来代替(除了特征值之外，还存在其他的实正交相似不变量)；见 G. Goodson 以及 R. A. Horn，Canonical forms for normal matrices that commute with their complex conjugate，*Linear Algebra Appl*. 430(2009)1025-1038．有关平方正规矩阵的标准型(2.5.22a,b)的一个证明，见 R. A. Horn 以及 V. V. Sergeichuk，Canonical forms for unitary congruence and *congruence，*Linear Multilinear Algebra* 57(2009)777-815．问题 4.4.P38 包含(2.5.20(b))的一个有深远影响的推广：每一个非奇异的复方阵(以及某些奇异的矩阵)都有 QS 分解．

2.6　酉等价与奇异值分解

假设给定的矩阵 A 是 n 维复向量空间上一个线性变换 T：$V\rightarrow V$ 关于一组给定的标准正交基的基表示．酉相似 $A\rightarrow UAU^*$ 与从给定的这组基到另一组标准正交基的改变相对应，酉矩阵 U 是基变化矩阵．

如果 $T: V_1 \to V_2$ 是从一个 n 维复向量空间到一个 m 维复向量空间的线性变换，又如果 $A \in M_{m,n}$ 是它关于 V_1 与 V_2 的给定的标准正交基的基表示，那么酉等价 $A \to UAW^*$ 对应于 V_1 与 V_2 从一组给定的标准正交基到另一组标准正交基的基的改变.

酉等价 $A \to UAV$ 涉及**两个**可以独立选取的酉矩阵. 这个附加的灵活性允许我们将其化简到特殊的形式，而这种特殊的形式或许用酉相似无法达到.

为了确保能用相同的酉相似将 A，$B \in M_n$ 化为上三角型，必须对它们设定某种条件(如交换性). 然而，我们可以用相同的酉等价将**任意**两个给定的矩阵化为上三角型.

定理 2.6.1 设 A，$B \in M_n$. 则存在酉矩阵 V，$W \in M_n$，使得 $A = VT_AW^*$，$B = VT_BW^*$，且 T_A 和 T_B 两者都是上三角的. 如果 B 是非奇异的，$T_B^{-1}T_A$ 的主对角线元素就是 $B^{-1}A$ 的特征值.

证明 假设 B 是非奇异的，利用(2.3.1)记 $B^{-1}A = UTU^*$，其中 U 是酉矩阵，而 T 是上三角的. 利用 QR 分解(2.1.14)来记 $BU = QR$，其中 Q 是酉矩阵，而 R 是上三角的. 那么 $A = BUTU^* = Q(RT)U^*$ 就是上三角的，RT 是上三角的，且 $B = QRU^*$. 此外，$B^{-1}A = UR^{-1}Q^*QRTU^* = UTU^*$ 的特征值是 T 的主对角线元素.

如果 A 与 B 两者都是奇异的，则存在一个 $\delta > 0$，使得只要 $0 < \varepsilon < \delta$，$B_\varepsilon = B + \varepsilon I$ 就是非奇异的，见(1.2.17). 对任何满足这一限制条件的 ε，我们已经证明了存在酉矩阵 V_ε，$W_\varepsilon \in M_n$，使得 $V_\varepsilon^* AW_\varepsilon$ 与 $V_\varepsilon^* BW_\varepsilon$ 两者都是上三角的. 选取一列非零的纯量 ε_k，使得 $\varepsilon_k \to 0$ 且 $\lim_{k\to\infty} V_{\varepsilon_k} = V$ 与 $\lim_{k\to\infty} W_{\varepsilon_k} = W$ 这两者都存在. 极限 V 与 W 中的每一个都是酉矩阵，见(2.1.8). 这样，$\lim_{k\to\infty} V_{\varepsilon_k}^* AW_{\varepsilon_k} = V^*AW = T_A$ 与 $\lim_{k\to\infty} V_{\varepsilon_k}^* BW_{\varepsilon_k} = V^*BW = T_B$ 中每一个都是上三角的. 我们就得出结论 $A = VT_AW^*$ 与 $B = VT_BW^*$，如所断言. □

这个定理还有一个实的形式，它利用了如下的事实.

习题 假设 A，$B \in M_n$，A 是上三角的，而 B 是上拟三角的. 证明：AB 是与 B 共形的上拟三角矩阵.

定理 2.6.2 设 A，$B \in M_n(\mathbf{R})$. 则存在实正交矩阵 V，$W \in M_n$，使得 $A = VT_AW^{\mathrm{T}}$，$B = VT_BW^{\mathrm{T}}$，T_A 是实的且是上拟三角的，而 T_B 则是实的且是上三角的. 149

证明 如果 B 是非奇异的，就利用(2.3.4)记成 $B^{-1}A = UTU^{\mathrm{T}}$，其中 U 是实正交矩阵，而 T 是实的上拟三角矩阵. 利用(2.1.14d)记 $BU = QR$，其中 Q 是实正交矩阵，而 R 则是实的上三角矩阵. 这样 RU 就是上拟三角矩阵，$A = Q(RT)U^{\mathrm{T}}$，而 $B = QRU^{\mathrm{T}}$. 如果 A 与 B 两者都是奇异的，就可以利用上一个证明中的实形式的极限论证方法. □

尽管只有正规的方阵才可以用酉相似来使其对角化，任何复矩阵也可以用酉等价来对角化.

定理 2.6.3(奇异值分解) 设给定 $A \in M_{n,m}$，令 $q = \min\{m, n\}$ 并假设 $\operatorname{rank} A = r$.

(a) 存在酉矩阵 $V \in M_n$ 与 $W \in M_m$，以及一个对角方阵

$$\Sigma_q = \begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_q \end{bmatrix} \tag{2.6.3.1}$$

使得 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0 = \sigma_{r+1} = \cdots = \sigma_q$ 以及 $A = V\Sigma W^*$，其中

$$
\begin{aligned}
&\Sigma=\Sigma_q && \text{如果 } m=n \\
&\Sigma=[\Sigma_q \quad 0]\in M_{n,m} && \text{如果 } m>n \\
&\Sigma=\begin{bmatrix}\Sigma_q\\ 0\end{bmatrix}\in M_{n,m} && \text{如果 } n>m
\end{aligned}
\tag{2.6.3.2}
$$

(b) 参数 σ_1，…，σ_r 是 AA^* 的按照递减次序排列的非零特征值的正的平方根，它们与 A^*A 的按照递减次序排列的非零特征值的正的平方根是相同的.

证明 首先假设 $m=n$. Hermite 矩阵 $AA^*\in M_n$ 与 $A^*A\in M_n$ 有同样的特征值(1.3.22)，从而它们是酉相似的(2.5.4(d))，于是存在一个酉矩阵 U，使得 $A^*A=U(AA^*)U^*$. 这样就有

$$(UA)^*(UA)=A^*U^*UA=A^*A=UAA^*U^*=(UA)(UA)^*$$

所以 UA 是正规的. 设 $\lambda_1=|\lambda_1|e^{i\theta_1}$，…，$\lambda_n=|\lambda_n|e^{i\theta_n}$ 是 UA 的按照次序 $|\lambda_1|\geqslant\cdots\geqslant|\lambda_n|$ 排列的特征值. 这样 $r=\mathrm{rank}A=\mathrm{rank}UA$ 就是正规矩阵 UA 的非零特征值的个数，所以 $|\lambda_r|>0$ 且 $\lambda_{r+1}=\cdots=\lambda_n=0$. 设 $\Lambda=\mathrm{diag}(\lambda_1,\cdots,\lambda_n)$，令 $D=\mathrm{diag}(e^{i\theta_1},\cdots,e^{i\theta_n})$，$\Sigma_q=\mathrm{diag}(|\lambda_1|,\cdots,|\lambda_n|)$，又设 X 是酉矩阵，它使得 $UA=X\Lambda X^*$. 那么 D 是酉矩阵且 $A=U^*X\Lambda X^*=U^*X\Sigma_qDX^*=(U^*X)\Sigma_q(DX^*)$ 就给出了所欲求之分解，其中 $V=U^*X$ 与 $W=XD^*$ 是酉矩阵，而 $\sigma_j=|\lambda_j|$，$j=1$，…，n.

现在假设 $m>n$. 这样就有 $r\leqslant n$，故而 A 的零空间的维数为 $m-r\geqslant m-n$. 设 x_1，…，x_{m-n} 是 A 的零空间中任意一组标准正交的向量，设 $X_2=[x_1 \quad \cdots \quad x_{m-n}]\in M_{m,m-n}$，又设 $X=[X_1 \quad X_2]\in M_m$ 是酉矩阵，即将给定的标准正交向量组扩展成为 $\mathbf{C}^m$ 的一组基. 那么就有 $AX=[AX_1 \quad AX_2]=[AX_1 \quad 0]$ 以及 $AX_1\in M_n$. 利用上一种情形，记 $AX_1=V\Sigma_qW^*$，

150 其中 V，$W\in M_n$ 是酉矩阵，而 Σ_q 有(2.6.3.1)的形式. 这就给出

$$A=[AX_1 \quad 0]X^*=[V\Sigma_qW^* \quad 0]X^*=V[\Sigma_q \quad 0]\left(\begin{bmatrix}W^* & 0\\ 0 & I_{m-n}\end{bmatrix}X^*\right)$$

这就是结论中所说的分解.

如果 $n>m$，将上面的情形应用于 A^*.

利用分解 $A=V\Sigma W^*$，注意 $\mathrm{rank}A=\mathrm{rank}\Sigma$(这是因为 V 与 W 是非奇异的). 但是 $\mathrm{rank}\Sigma$ 等于 Σ 的不为零的(从而是正的)对角元素的个数，如结论所说. 现在计算 $AA^*=V\Sigma W^*W\Sigma^TV^*=V\Sigma\Sigma^TV^*$，它与 $\Sigma\Sigma^T$ 酉相似. 如果 $n=m$，那么 $\Sigma\Sigma^T=\Sigma_q^2=\mathrm{diag}(\sigma_1^2,\cdots,\sigma_n^2)$. 如果 $m>n$，则 $\Sigma\Sigma^T=[\Sigma_q \quad 0][\Sigma_q \quad 0]^T=\Sigma_q^2=0_n=\Sigma_q^2$. 最后，如果 $n>m$，那么

$$\Sigma\Sigma^T=\begin{bmatrix}\Sigma_q\\ 0\end{bmatrix}[\Sigma_q \quad 0]=\begin{bmatrix}\Sigma_q^2 & 0\\ 0 & 0_{n-m}\end{bmatrix}$$

在每一种情形，AA^* 的非零特征值都是 σ_1^2，…，σ_r^2，如所断言. □

(2.6.3.2)中矩阵 Σ 的对角元素(即纯量 σ_1，…σ_q，它们是方阵 Σ_q 的对角元素)称为 A 的**奇异值**(singular value). A 的奇异值 σ 的**重数**(multiplicity)是 σ^2 作为 AA^* 的特征值的重数，或者等价地说，也就是 A^*A 的特征值的重数. A 的一个奇异值 σ 称为是**单重的**(simple)，如果 σ^2 是 AA^* 的单重特征值，或者等价地说是 A^*A 的单重特征值. A 的秩**等于**它的非零奇异值的个数，而 $\mathrm{rank}A$ **不小于**(有可能大于)它的非零特征值的个数.

A 的奇异值由 A^*A 的特征值唯一地决定(等价地说，是由 AA^* 的特征值唯一地决定)，

所以，A 的奇异值分解式中对角因子 Σ 除了对角元素的排列可能会有变化之外也是唯一确定的；为使得 Σ 唯一，习惯上选择让奇异值按照非增的次序排列，不过也可以采用其他的选择方法.

习题 设 $A\in M_{m,n}$. 说明为什么 A，$\bar{A}$，A^{T} 以及 A^* 有同样的奇异值. ◀

设 σ_1，…，σ_n 是 $A\in M_n$ 的奇异值. 说明为什么有

$$\sigma_1\cdots\sigma_n = |\det A| \quad \text{以及} \quad \sigma_1^2+\cdots+\sigma_n^2=\operatorname{tr}A^*A. \tag{2.6.3.3}$$

习题 证明：$A\in M_2$ 的两个平方的奇异值是 ◀

$$\sigma_1^2,\sigma_2^2=\frac{1}{2}\left((\operatorname{tr}A^*A)\pm\sqrt{(\operatorname{tr}A^*A)^2-4|\det A|^2}\right) \tag{2.6.3.4}$$

习题 说明为什么幂零矩阵

$$A=\begin{bmatrix}0 & a_{12} & & 0\\ & \ddots & \ddots & \\ & & \ddots & a_{n-1,n}\\ 0 & & & 0\end{bmatrix}\in M_n$$

151

的奇异值(除了在第一条超对角线上的某些非零元素之外，其他地方的元素处处都是零)是 0，$|a_{12}|$，…，$|a_{n-1,n}|$. ◀

下面的定理对如下的结论给出一个精确的总结：矩阵的奇异值连续地依赖于它的元素.

定理 2.6.4 设给定一个无穷序列 A_1，A_2，$\cdots\in M_{n,m}$，假设 $\lim_{k\to\infty}A_k=A$(逐个元素地收敛)，又设 $q=\min\{m,n\}$. 设 $\sigma_1(A)\geqslant\cdots\geqslant\sigma_q(A)$ 以及 $\sigma_1(A_k)\geqslant\cdots\geqslant\sigma_q(A_k)$ 分别是 A 与 A_k 按照非增次序排列的奇异值(对 $k=1$，2，…). 那么，对每个 $i=1$，…，q 都有 $\lim_{k\to\infty}\sigma_i(A_k)=\sigma_i(A)$.

证明 如果定理的结论不真，则存在某个 $\varepsilon_0>0$ 以及正整数的无穷序列 $k_1<k_2<\cdots$，使得对每个 $j=1$，2，…都有

$$\max_{i=1,\cdots,q}|\sigma_i(A_{k_j})-\sigma_i(A)|>\varepsilon_0 \tag{2.6.4.1}$$

对每个 $j=1$，2，…，设 $A_{k_j}=V_{k_j}\Sigma_{k_j}W_{k_j}^*$，其中 $V_{k_j}\in M_n$ 和 $W_{k_j}\in M_m$ 是酉矩阵，而 $\Sigma_{k_j}\in M_{n,m}$ 是非负对角矩阵，它满足 $\operatorname{diag}\Sigma_{k_j}=[\sigma_1(A_{k_j})\cdots\sigma_q(A_{k_j})]^{\mathrm{T}}$. 引理 2.1.8 确保存在一个无穷子列 $k_{j_1}<k_{j_2}<\cdots$ 以及酉矩阵 V 与 W，使得 $\lim_{\ell\to\infty}V_{k_{j_\ell}}=V$ 以及 $\lim_{\ell\to\infty}W_{k_{j_\ell}}=W$。那么

$$\lim_{\ell\to\infty}\Sigma_{k_{j_\ell}}=\lim_{\ell\to\infty}V_{k_{j_\ell}}^*A_{k_{j_\ell}}W_{k_{j_\ell}}=\left(\lim_{\ell\to\infty}V_{k_{j_\ell}}^*\right)\left(\lim_{\ell\to\infty}A_{k_{j_\ell}}\right)\left(\lim_{\ell\to\infty}V_{k_{j_\ell}}\right)=V^*AW$$

存在，且是对角元素按非增次序排列的非负对角矩阵；我们用 Σ 表示它，并注意 $A=V\Sigma W^*$. A 的奇异值的唯一性确保 $\operatorname{diag}\Sigma=[\sigma_1(A)\ \cdots\ \sigma_q(A)]^{\mathrm{T}}$，这与(2.6.4.1)矛盾，这就证明了定理. □

奇异值分解中的酉因子从来都不是唯一的. 例如，如果 $A=V\Sigma W^*$，我们可以用 $-V$ 代替 V，用 $-W$ 代替 W. 下面的定理以一种显明且有用的方式描述了这样一个事实：在奇异值分解中给定一对酉因子，就可以得到所有可能的酉因子对.

定理 2.6.5(Autonne 唯一性定理) 设给定 $A\in M_{n,m}$，其中 $\operatorname{rank}A=r$. 设 s_1，…，s_d

是 A 的不同的正的奇异值，它们按照任意次序排列，且重数分别为 n_1，…，n_d，又设 $\Sigma_d = s_1 I_{n_1} \oplus \cdots \oplus s_d I_{n_d} \in M_r$. 设 $A = V\Sigma W^*$ 是奇异值分解，其中 $\Sigma = \begin{bmatrix} \Sigma_d & 0 \\ 0 & 0 \end{bmatrix} \in M_{n,m}$，如同在(2.6.3.2)中那样，则 $\Sigma^T\Sigma = s_1^2 I_{n_1} \oplus \cdots \oplus s_d^2 I_{n_d} \oplus 0_{n-r}$，且 $\Sigma\Sigma^T = s_1^2 I_{n_1} \oplus \cdots \oplus s_d^2 I_{n_d} \oplus 0_{m-r}$（如果 A 是满秩的，就会有一个为零的直和项不出现；如果 A 是方阵且非奇异，那么这两个为零的直和项就都不出现）. 设 $\hat{V} \in M_n$ 与 $\hat{W} \in M_m$ 是酉矩阵. 那么 $A = \hat{V}\Sigma\hat{W}^*$ 当且仅当存在酉矩阵 $U_1 \in M_{n_1}$，…，$U_d \in M_{n_d}$，$\tilde{V} \in M_{n-r}$，以及 $\tilde{W} \in M_{m-r}$，使得

$$\hat{V} = V(U_1 \oplus \cdots \oplus U_d \oplus \tilde{V}) \text{ 以及 } \hat{W} = W(U_1 \oplus \cdots \oplus U_d \oplus \tilde{W}) \qquad (2.6.5.1)$$

如果 A 是实的且诸因子 V，W，$\hat{V}$，$\hat{W}$ 是实正交的，那么，矩阵 U_1，…，U_d，$\tilde{V}$ 以及都 $\tilde{W}$

152 可以取为实正交矩阵.

证明 Hermite 矩阵 A^*A 被表示成 $A^*A = (V\Sigma W^*)^*(V\Sigma W^*) = W\Sigma^T\Sigma W^*$，也可表示为 $A^*A = \hat{W}\Sigma^T\Sigma\hat{W}^*$. 定理 2.5.4 确保存在酉矩阵 W_1，…，W_d，W_{d+1}，其中 $W_i \in M_{n_i}$（对 $i=1$，…，d），使得 $\hat{W} = W(W_1 \oplus \cdots \oplus W_d \oplus W_{d+1})$. 我们又有 $AA^* = V\Sigma\Sigma^T V^* = \hat{V}\Sigma\Sigma^T\hat{V}^*$，所以(2.5.4)再次告诉我们：存在酉矩阵 V_1，…，V_d，V_{d+1} 其中 $V_i \in M_{n_i}$（对 $i=1$，…，d），使得 $\hat{V} = V(V_1 \oplus \cdots \oplus V_d \oplus V_{d+1})$. 由于 $A = V\Sigma W^* = \hat{V}\Sigma\hat{W}^*$，我们有 $\Sigma = (V^*\hat{V})\Sigma(\hat{W}^*W)$，也即对 $i=1$，…，$d+1$ 有 $s_i I_{n_i} = V_i(s_i I_{n_i})W_i^*$，或者说对每个 $i=1$，…，d 有 $V_i W_i^* = I_{n_i}$. 矩阵 $\tilde{V}$ 与 $\tilde{W}$，如果存在的话，也是任意的. 由此推出，对每个 $i=1$，…，d 有 $V_i = W_i$. 最后的结论由上面的讨论以及 $V^T\hat{V}$ 与 $W^T\hat{W}$ 均为实的这一事实推出. □

奇异值分解是矩阵分析以及在工程、数值计算、统计、图像压缩以及其他许多领域中难以计数的应用中非常重要的工具. 更详细的介绍请见 Horn 与 Johnson(1991)所著书中的第 7 章以及第 3 章.

我们用上面的唯一性定理的三个应用来结束这一章：对称矩阵或者斜对称矩阵的奇异值分解可以选取为酉相合，而且实矩阵有这样的奇异值分解，其中三个因子全都是实的.

推论 2.6.6 设 $A \in M_n$，并令 $r = \operatorname{rank} A$.

(a) (Autonne) $A = A^T$ 当且仅当存在一个酉矩阵 $U \in M_n$ 以及一个非负的对角矩阵 Σ，使得 $A = U\Sigma U^T$. Σ 的对角元素是 A 的奇异值.

(b) 如果 $A = -A^T$，那么 r 是偶数且存在一个酉矩阵 $U \in M_n$ 以及正实的纯量 s_1，…，$s_{r/2}$，使得

$$A = U\left(\begin{bmatrix} 0 & s_1 \\ -s_1 & 0 \end{bmatrix} \oplus \cdots \oplus \begin{bmatrix} 0 & s_{r/2} \\ -s_{r/2} & 0 \end{bmatrix} \oplus 0_{n-r}\right)U^T \qquad (2.6.6.1)$$

A 的非零的奇异值是 s_1，s_1，…，$s_{r/2}$，$s_{r/2}$. 反之，任何形如(2.6.6.1)的矩阵都是斜对称的.

证明 (a) 如果对一个酉矩阵 $U \in M_n$ 以及一个非负对角矩阵 Σ 有 $A = U\Sigma U^T$，那么 A 是对称矩阵，且 Σ 的对角元素是它的奇异值. 反之，设 s_1，…，s_d 是 A 的不同的正的奇异值，按照任意次序排列，其重数分别为 n_1，…，n_d，又设 $A = V\Sigma W^*$ 是奇异值分解，其中 V，$W \in M_n$ 是酉矩阵，而 $\Sigma = s_1 I_{n_1} \oplus \cdots \oplus s_d I_{n_d} \oplus 0_{n-r}$. 如果 A 是非奇异的，那么其中不出

现为零的分块. 我们有 $A=V\Sigma W^*=\overline{W}\Sigma\overline{V}^*=A$，所以上一定理确保存在酉矩阵 U_W 与 U_V，使得 $\overline{V}=WU_W$，$\overline{W}=VU_V$，$U_V=U_1\oplus\cdots\oplus U_d\oplus\widetilde{V}$，$U_W=U_1\oplus\cdots\oplus U_d\oplus\widetilde{W}$，且每一个 $U_i\in M_{n_i}$，$i=1,\cdots,d$. 这样就有 $U_W=W^*\overline{V}=(V^*\overline{W})^{\mathrm{T}}=U_V^{\mathrm{T}}$，它蕴含结论：每一个 $U_j=U_j^{\mathrm{T}}$，即每一个 U_j 都是酉矩阵，且是对称的. 推论 2.5.20a 确保存在对称的酉矩阵 R_j，使得对每个 $j=1,\cdots,d$ 都有 $R_j^2=U_j$. 设 $R=R_1\oplus\cdots\oplus R_d\oplus I_{n-r}$. 则 R 是对称的酉矩阵，且 $U_V\Sigma=R^2\Sigma=R\Sigma R$，所以 $A=\overline{W}\Sigma V^{\mathrm{T}}=VU_V\Sigma V^{\mathrm{T}}=VR\Sigma RV^{\mathrm{T}}=(VR)\Sigma(VR)^{\mathrm{T}}$ 是结论中所说的分解.

(b) 从恒等式 $V\Sigma W^*=-\overline{W}\Sigma V^{\mathrm{T}}=-\overline{W}\Sigma\overline{V}^*$ 出发，并完全按照(a)中那样去做，我们就有 $\overline{V}=WU_W$，$\overline{W}=-VU_V$，这也就是 $U_W=W^*\overline{V}$ 以及 $U_V=-V^*\overline{W}=-U_W^{\mathrm{T}}$. 特别地，对 $j=1,\cdots,d$ 有 $U_j=-U_j^{\mathrm{T}}$，即每一个 U_j 都是酉矩阵，且是斜对称的. 推论 2.5.18b 确保对每个 $j=1,\cdots,d$，n_j 都是偶数，且存在实正交矩阵 Q_j 以及实参数 $\theta_1^{(j)},\cdots,\theta_{n/2}^{(j)}\in[0,2\pi)$，使得 153

$$U_j=Q_j\left(e^{i\theta_1^{(j)}}\begin{bmatrix}0&1\\-1&0\end{bmatrix}\oplus\cdots\oplus e^{i\theta_{n_j/2}^{(j)}}\begin{bmatrix}0&1\\-1&0\end{bmatrix}\right)Q_j^{\mathrm{T}}$$

定义实正交矩阵 $Q=Q_1\oplus\cdots\oplus Q_d\oplus I_{n-r}$ 以及斜对称的酉矩阵

$$S_j=e^{i\theta_1^{(j)}}\begin{bmatrix}0&1\\-1&0\end{bmatrix}\oplus\cdots\oplus e^{i\theta_{n_j/2}^{(j)}}\begin{bmatrix}0&1\\-1&0\end{bmatrix},\quad j=1,\cdots,d$$

设 $S=S_1\oplus\cdots\oplus S_d\oplus 0_{n-r}$. 那么 $U_V\Sigma=QSQ^{\mathrm{T}}\Sigma=QS\Sigma Q$，所以 $A=-\overline{W}\Sigma V^{\mathrm{T}}=VU_V\Sigma V^{\mathrm{T}}=VQS\Sigma Q^{\mathrm{T}}V^{\mathrm{T}}=(VQ)S\Sigma(VQ)^{\mathrm{T}}$ 是结论中那种形式的分解，且 $\mathrm{rank}A=n_1+\cdots+n_d$ 是偶数. □

推论 2.6.7 设 $A\in M_{n,m}(\mathbf{R})$，并假设 $\mathrm{rank}A=r$. 那么 $A=P\Sigma Q^{\mathrm{T}}$，其中 $P\in M_n(\mathbf{R})$ 与 $Q\in M_m(\mathbf{R})$ 是实正交矩阵，而 $\Sigma\in M_{n,m}(\mathbf{R})$ 则是非负的对角矩阵，且有(2.6.3.1)或者(2.6.3.2)的形状.

证明 利用(2.6.4)中的记号，设 $A=V\Sigma W^*$ 是给定的奇异值分解，酉矩阵 V 与 W 不一定是实的. 我们有 $V\Sigma W^*=A=\overline{A}=V\Sigma\overline{W}$，所以 $V^{\mathrm{T}}V\Sigma=\Sigma W^{\mathrm{T}}W$. 定理 2.6.5 确保存在酉矩阵 $U_V=U_1\oplus\cdots\oplus U_d\oplus\widetilde{V}\in M_n$ 以及 $U_W=U_1\oplus\cdots\oplus U_d\oplus\widetilde{W}\in M_m$，使得 $\overline{V}=VU_V$ 以及 $\overline{W}=WU_W$. 这样 $U_V=V^*\overline{V}=\overline{V}^{\mathrm{T}}\overline{V}$ 以及 $U_W=\overline{W}^{\mathrm{T}}\overline{W}$ 是酉矩阵，且是对称的，所以 $\widetilde{V}$，$\widetilde{W}$ 亦如此，且每一个 U_i 也都是酉矩阵，且是对称的. 推论 2.5.20(a)确保存在对称的酉矩阵 $R_{\widetilde{V}}$，$R_{\widetilde{W}}$ 以及 $R_1,\cdots,R_d$，使得 $R_{\widetilde{V}}^2=\widetilde{V}$，$R_{\widetilde{W}}^2=\widetilde{W}$ 以及 $R_i^2=U_i$(对每个 $i=1,\cdots,d$). 设 $R_V=R_1\oplus\cdots\oplus R_d\oplus R_{\widetilde{V}}$ 以及 $R_W=R_1\oplus\cdots\oplus R_d\oplus R_{\widetilde{W}}$. 那么 R_V 与 R_W 是对称的，且是酉矩阵，$R_V^{-1}=R_V^*=\overline{R}_V$，$R_W^{-1}=R_W^*=\overline{R}_W$，$R_V^2=U_V$，$R_V^2=U_V$，以及 $R_V\Sigma\overline{R}_W=\Sigma$，所以

$$\begin{aligned}A&=\overline{V}\Sigma\overline{W}^*=VU_V\Sigma(WU_W)^*=VR_V^2\Sigma(WR_W^2)^*\\&=(VR_V)(R_V\Sigma\overline{R_W})(WR_W)^*=(VR_V)\Sigma(WR_W)^*\end{aligned}$$

我们用下面的观察结果来作为结束：我们有 $\overline{V}=VU_V=VR_V^2$ 以及 $\overline{W}=WU_W=WR_W^2$，所以有 $\overline{VR_V}=\overline{V}R_V^*=VR_V$ 以及 $\overline{WR_W}=\overline{W}R_W^*=WR_W$. 这就是说，$VR_V$ 与 WR_W 两者都是酉矩阵且是实的，故而两者都是实正交的. □

问题

2.6.P1 设 $A\in M_{n,m}$，其中 $n\geqslant m$. 证明：A 是列满秩的，当且仅当它所有的奇异值都是正数.

2.6.P2 假设 $A, B\in M_{n,m}$ 可以用酉等价同时对角化，也就是说，假设存在酉矩阵 $X\in M_n$ 以及 $Y\in M_m$，使得 $X^*AY=\Lambda$ 和 $X^*BY=M$ 中的每一个都是对角的(0.9.1). 证明：AB^* 与 B^*A 两者都是正
154 规的.

2.6.P3 何时 $A, B\in M_{n,m}$ 同时与对角矩阵酉等价？证明：AB^* 与 B^*A 两者都是正规的，当且仅当存在酉矩阵 $X\in M_n$ 以及 $Y\in M_m$，使得 $A=X\Sigma Y^*$；$B=X\Lambda Y^*$；$\Sigma, \Lambda\in M_{n,m}$ 是对角的；且 $\Sigma\in M_{n,m}$ 有(2.6.3.1)和(2.6.3.2)的形式.

2.6.P4 何时 $A, B\in M_{n,m}$ 与实对角矩阵或非负的实对角矩阵是同时酉等价的？(a)证明：AB^* 与 B^*A 两者都是 Hermite 矩阵，当且仅当存在酉矩阵 $X\in M_n$ 以及 $Y\in M_m$，使得 $A=X\Sigma Y^*$；$B=X\Lambda Y^*$；$\Sigma, \Lambda\in M_{n,m}(\mathbf{R})$ 是对角的；且 Σ 有(2.6.3.1)和(2.6.3.2)的形式. (b)如果 A 与 B 是实的，证明：AB^{T} 与 $B^{\mathrm{T}}A$ 两者都是实对称的，当且仅当存在实正交矩阵 $X\in M_n(\mathbf{R})$ 以及 $Y\in M_m(\mathbf{R})$，使得 $A=X\Sigma Y^{\mathrm{T}}$；$B=X\Lambda Y^{\mathrm{T}}$；$\Sigma, \Lambda\in M_{n,m}(\mathbf{R})$ 是对角的；且 Σ 有(2.6.3.1)和(2.6.3.2)的形式. (c)在(a)与(b)这两者中证明：Λ 可以选取成有非负对角元素，当且仅当 Hermite 矩阵 AB^* 与 B^*A 的所有特征值都是非负的.

2.6.P5 设给定 $A\in M_{n,m}$，并记 $A=B+\mathrm{i}C$，其中 $B, C\in M_{n,m}(\mathbf{R})$. 证明：存在实正交矩阵 $X\in M_n(\mathbf{R})$ 以及 $Y\in M_m(\mathbf{R})$，使得 $A=X\Lambda Y^{\mathrm{T}}$ 与 $\Lambda\in M_{n,m}(\mathbf{C})$ 是对角的，当且仅当 BC^{T} 与 $C^{\mathrm{T}}B$ 两者都是实对称的.

2.6.P6 设给定 $A\in M_n$，并令 $A=QR$ 是 QR 分解(2.1.14). (a)说明为什么 QR 是正规的，当且仅当 RQ 是正规的. (b)证明：A 是正规的，当且仅当 Q 与 R^* 可以用酉等价同时对角化.

2.6.P7 证明：两个同样大小的复矩阵是酉等价的，当且仅当它们有同样的奇异值.

2.6.P8 设给定 $A\in M_{n,k}$ 以及 $B\in M_{k,m}$. 利用奇异值分解证明 $\mathrm{rank}AB\leqslant\min\{\mathrm{rank}A, \mathrm{rank}B\}$.

2.6.P9 设给定 $A\in M_n$. 假设 $\mathrm{rank}A=r$，按照递减次序排列正的奇异值来构作 $\Sigma_1=\mathrm{diag}\{\sigma_1, \cdots, \sigma_r\}$，并设 $\Sigma=\Sigma_1\oplus 0_{n-r}$. 假设 $W\in M_n$ 是酉矩阵，且 $A^*A=W\Sigma^2W^*$. 证明：存在一个酉矩阵 $V\in M_n$，使得 $A=V\Sigma W^*$.

2.6.P10 设给定 $A, B\in M_n$，设 $\sigma_1\geqslant\cdots\geqslant\sigma_n\geqslant 0$ 是 A 的奇异值，又设 $\Sigma_1=\mathrm{diag}\{\sigma_1, \cdots, \sigma_r\}$. 证明下面三个命题是等价的：(a) $A^*A=B^*B$；(b)存在酉矩阵 $W, X, Y\in M_n$，使得 $A=X\Sigma W^*$ 以及 $B=Y\Sigma W^*$；(c)存在一个酉矩阵 $U\in M_n$，使得 $B=UA$. 作为它的推广，请见(7.3.11).

2.6.P11 设给定 $A\in M_{n,m}$ 以及一个正规矩阵 M_m. 证明 A^*A 与 B 可交换，当且仅当存在酉矩阵 $V\in M_n$ 与 $W\in M_m$，以及对角矩阵 $\Sigma\in M_{n,m}$ 与 $\Lambda\in M_m$，使得 $A=X\Sigma W^*$ 以及 $B=W\Lambda W^*$.

2.6.P12 设 $A\in M_n$ 有奇异值分解 $A=V\Sigma W^*$，其中 $\Sigma=\mathrm{diag}(\sigma_1, \cdots, \sigma_n)$，而 $\sigma_1\geqslant, \cdots, \geqslant\sigma_n$. (a)证明：$\mathrm{adj}A$ 有奇异值分解 $\mathrm{adj}A=X^*SY$，其中 $X=(\det W)(\mathrm{adj}W)$，$Y=(\det V)(\mathrm{adj}V)$，而 $S=\mathrm{diag}(s_1, \cdots, s_n)$，其中每一个 $s_i=\prod_{j\neq i}\sigma_j$. (b)利用(a)说明为什么 $\mathrm{adj}A=0$，如果 $\mathrm{rank}A\leqslant n-2$. (c)如果 $\mathrm{rank}A=n-1$，且 $v_n, w_n\in\mathbf{C}^n$ 分别是 V 与 W 的最后一列，证明 $\mathrm{adj}A=\sigma_1\cdots\sigma_{n-1}\mathrm{e}^{\mathrm{i}\theta}w_nv_n^*$，其中 $\det(VW^*)=\mathrm{e}^{\mathrm{i}\theta}$，$\theta\in\mathbf{R}$.

2.6.P13 设 $A\in M_n$，且令 $A=V\Sigma W^*$ 是奇异值分解. (a)证明：A 是酉矩阵当且仅当 $\Sigma=I$. (b)证明：A
155 是酉矩阵的纯量倍数的充分必要条件是，只要 $x, y\in\mathbf{C}^n$ 正交，Ax 与 Ay 就正交.

2.6.P14 设给定 $A\in M_n$. (a)假设 A 是正规的，并设 $A=U\Lambda U^*$ 是谱分解，其中 U 是酉矩阵，而 $\Lambda=\mathrm{diag}(\lambda_1, \cdots, \lambda_n)=\mathrm{diag}(\mathrm{e}^{\mathrm{i}\theta_1}|\lambda_1|, \cdots, \mathrm{e}^{\mathrm{i}\theta_n}|\lambda_n|)$. 令 $D=\mathrm{diag}(\mathrm{e}^{\mathrm{i}\theta_1}, \cdots, \mathrm{e}^{\mathrm{i}\theta_n})$ 以及 $\Sigma=\mathrm{diag}(|\lambda_1|, \cdots, |\lambda_n|)$. 说明为什么 $A=(UD)\Sigma U^*$ 是 A 的奇异值分解，以及为什么 A 的奇异值是它的特征值之绝对值. (b)设 $s_1, \cdots, s_d$ 是 A 的不同的奇异值，并令 $A=V\Sigma W^*$ 是奇异值分

解，其中 V，$W\in M_n$ 是酉矩阵，而 $\Sigma=s_1I_{n_1}\oplus\cdots\oplus s_dI_{n_d}$. 证明：$A$ 是正规的，当且仅当存在一个与 Σ 共形地分划的分块对角酉矩阵 $U=U_1\oplus\cdots\oplus U_d$，使得 $V=WU$. (c)如果 A 是正规的，且有不同的奇异值，又如果 $A=V\Sigma W^*$ 是奇异值分解，说明为什么 $V=WD$，其中的 D 是一个对角酉矩阵. 有不同奇异值的假设对 A 的特征值有何结论?

2.6.P15 设 $A=[a_{ij}]\in M_n$ 有特征值 $\lambda_1,\cdots,\lambda_n$，其排序满足 $|\lambda_1|\geqslant\cdots\geqslant|\lambda_n|$，又有奇异值 $\sigma_1,\cdots,\sigma_n$，其排序为 $\sigma_1\geqslant\cdots\geqslant\sigma_n$. 证明：(a) $\sum_{i,j=1}^{n}|a_{ij}|^2=\mathrm{tr}A^*A=\sum_{i,j=1}^{n}\sigma_i^2$；(b) $\sum_{i=1}^{n}|\lambda_i|^2\leqslant\sum_{i=1}^{n}\sigma_i^2$，其中的等式当且仅当 A 是正规矩阵时成立(Schur 不等式)；(c)对所有 $i=1,\cdots,n$ 都有 $\sigma_i=|\lambda_i|$ 成立之充分必要条件是：A 是正规矩阵；(d)如果对所有 $i=1,\cdots,n$ 都有 $|a_{ii}|=\sigma_i$，那么 A 是对角矩阵；(e)如果 A 是正规的，且对所有 $i=1,\cdots,n$ 都有 $|a_{ii}|=|\lambda_i|$，那么 A 是对角矩阵.

2.6.P16 设 U，$V\in M_n$ 是酉矩阵. (a)证明总存在酉矩阵 X，$Y\in M_n$ 以及一个对角的酉矩阵 $D\in M_n$，使得 $U=XDY$ 以及 $V=Y^*DX^*$. (b)说明为什么 M_n 上的酉等价映射 $A\rightarrow UAV=XDYAY^*DX^*$ 是一个酉相似、一个对角的酉相合以及一个酉相似的合成.

2.6.P17 设 $A\in M_{n,m}$. 利用奇异值分解来说明为什么 $\mathrm{rank}A=\mathrm{rank}AA^*=\mathrm{rank}A^*A$.

2.6.P18 设 $A\in M_n$ 是一个射影矩阵，并假设 $\mathrm{rank}A=r$. (a)证明 A 与 $\begin{bmatrix}I_r & X\\ 0 & 0_{n-r}\end{bmatrix}$ 酉相似，见(1.1.P5). (b)设 $X=V\Sigma W^*$ 是奇异值分解. 证明：A 通过 $V\oplus W$ 与 $\begin{bmatrix}I_r & \Sigma\\ 0 & 0_{n-r}\end{bmatrix}$ 酉相似，从而 A 的奇异值是 $(I_r+\Sigma\Sigma^{\mathrm{T}})\oplus 0_{n-r}$ 的对角元素；设 $\sigma_1,\cdots,\sigma_g$ 是 A 的大于 1 的奇异值. (c)证明 A 与 $0_{n-r-g}\oplus I_{r-g}\oplus\begin{bmatrix}1 & (\sigma_1^2-1)^{1/2}\\ 0 & 0\end{bmatrix}\oplus\cdots\oplus\begin{bmatrix}1 & (\sigma_g^2-1)^{1/2}\\ 0 & 0\end{bmatrix}$ 酉相似.

2.6.P19 设 $U=\begin{bmatrix}U_{11} & U_{12}\\ U_{21} & U_{22}\end{bmatrix}\in M_{k+\ell}$ 是酉矩阵，其中 $U_{11}\in M_k$，$U_{22}\in M_\ell$，且 $k\leqslant\ell$. 证明：U 的分块的按照非增次序排列的奇异值由下述诸恒等式联系在一起：对每一个 $i=1,\cdots,k$ 有 $\sigma_i(U_{11})=\sigma_i(U_{22})$ 以及 $\sigma_i(U_{12})=\sigma_i(U_{21})=(1-\sigma_{k-j+1}^2(U_{11}))^{1/2}$；而对每个 $i=k+1,\cdots\ell$ 则有 $\sigma_i(U_{22})=1$. 特别地，有 $|\det U_{11}|=|\det U_{22}|$ 以及 $\det U_{12}U_{12}^*=\det U_{21}^*U_{21}$. 说明为什么这些结果蕴含(2.1.10).

2.6.P20 设 $A\in M_n$ 是对称矩阵. 假设当 A 是**非奇异**时，已知(2.6.6(a))中特殊的奇异值分解. 下面两种途径证明了：即便当 A 为奇异时此结论依然为真. 请对这两种方法补充证明的细节. (a)考虑 $A_\varepsilon=A+\varepsilon I$；利用(2.1.8)以及(2.6.4). (b)设 $U_1\in M_{n,v}$ 的列是 A 的零空间的一组标准正交基，又设 $U=[U_1\quad U_2]\in M_n$ 是酉矩阵. 令 $U^{\mathrm{T}}AU=[A_{ij}]_{i,j=1}^2$(与 U 共形地分划). 说明为什么 A_{11}，A_{12} 以及 A_{21} 是零矩阵，而 A_{22} 是非奇异的对称矩阵. 156

2.6.P21 设 A，$B\in M_n$ 是对称矩阵. 证明 $A\overline{B}$ 是正规的，当且仅当存在一个酉矩阵 $U\in M_n$，使得 $A=U\Sigma U^{\mathrm{T}}$，$B=U\Lambda U^{\mathrm{T}}$，$\Sigma$，$\Lambda\in M_n$ 是对角矩阵，而 Σ 的对角元素是非负的.

2.6.P22 设 A，$B\in M_n$ 是对称矩阵. (a)证明 $A\overline{B}$ 是 Hermite 矩阵，当且仅当存在一个酉矩阵 $U\in M_n$，使得 $A=U\Sigma U^{\mathrm{T}}$，$B=U\Lambda U^{\mathrm{T}}$，$\Sigma$，$\Lambda\in M_n(\mathbf{R})$是对角矩阵，而 Σ 的对角元素是非负的. (b)证明：$A\overline{B}$ 是 Hermite 矩阵且有非负的特征值，当且仅当存在一个酉矩阵 $U\in M_n$，使得 $A=U\Sigma U^{\mathrm{T}}$，$B=U\Lambda U^{\mathrm{T}}$，$\Sigma$，$\Lambda\in M_n(\mathbf{R})$是对角矩阵，而 Σ 与 Λ 的对角元素都是非负的.

2.6.P23 设给定 $A\in M_n$. 假设 $\mathrm{rank}A=r\geqslant 1$，又假设 A 是**自零化的**(self-annihilating)，也即 $A^2=0$. 下面给出 A 与

$$\sigma_1\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\oplus\cdots\oplus\sigma_r\begin{bmatrix}0 & 1\\ 0 & 0\end{bmatrix}\oplus 0_{n-2r} \tag{2.6.8}$$

酉**相似**的证明概述，请对此证明提供细节，上式中 $\sigma_1\geqslant\cdots\geqslant\sigma_r>0$ 是 A 的正的奇异值. (a) $\operatorname{range}\bar{A}\subseteq\operatorname{nullspace}A$，从而 $2r\leqslant n$. (b)设 $U_2\in M_{n,n-r}$ 的列是 A^* 的零空间的一组标准正交基，所以 $U_2^*A=0$. 设 $U=[U_1\quad U_2]\in M_n$ 是酉矩阵. 说明为什么 $U_1\in M_{n,r}$ 的列是 A 的值域的一组标准正交基，且 $AU_1=0$. (c) $U^*AU=\begin{bmatrix}0 & B\\ 0 & 0\end{bmatrix}$，其中 $B\in M_{r,n-r}$，且 $\operatorname{rank}B=r$. (d) $B=V[\Sigma_r\quad 0_{r,n-2r}]W^*$，其中 $V\in M_r$ 以及 $W\in M_{n-r}$ 是酉矩阵，且 $\Sigma_r=\operatorname{diag}(\sigma_1,\cdots,\sigma_r)$. (e)设 $Z=V\oplus W$. 那么 $Z^*(U^*AU)Z=\begin{bmatrix}0 & \Sigma_r\\ 0 & 0\end{bmatrix}\oplus 0_{n-2r}$，它通过一个置换矩阵与(2.6.8)相似.

2.6.P24 设给定 $A\in M_n$. 假设 $\operatorname{rank}A=r\geqslant 1$，且 A 是**共轭自零化的**(conjugate self-annihilating)：$A\bar{A}=0$. 下面给出了 A 与(2.6.8)酉**相合**的证明概述，请对此补充证明细节，其中 $\sigma\geqslant\cdots\geqslant\sigma_r>0$ 是 A 的正的奇异值. (a) $\operatorname{range}\bar{A}\subseteq\operatorname{nullspace}A$，从而 $2r\leqslant n$. (b)设 $U_2\in M_{n,n-r}$ 的列是 A^{T} 的零空间的一组标准正交基，所以 $U_2^{\mathrm{T}}A=0$. 设 $U=[U_1\quad U_2]\in M_n$ 是酉矩阵. 说明为什么 $U_1\in M_{n,r}$ 的列是 $\bar{A}$ 的值域的一组标准正交基，且 $AU_1=0$. (c) $U^{\mathrm{T}}AU=\begin{bmatrix}0 & B\\ 0 & 0\end{bmatrix}$，其中 $B\in M_{r,n-r}$ 且 $\operatorname{rank}B=r$. (d) $B=V[\Sigma_r\quad 0_{r,n-2r}]W^*$，其中 $V\in M_r$ 和 $W\in M_{n-r}$ 是酉矩阵，且 $\Sigma_r=\operatorname{diag}(\sigma_1,\cdots,\sigma_r)$. (e)设 $Z=\bar{V}\oplus W$. 那么 $Z^{\mathrm{T}}(U^{\mathrm{T}}AU)Z=\begin{bmatrix}0 & \Sigma_r\\ 0 & 0\end{bmatrix}\oplus 0_{n-2r}$，它通过一个置换矩阵与(2.6.8)酉相合. 不同的处理方法请参见(3.4.P5).

2.6.P25 设 $A\in M_n$，并假设 $\operatorname{rank}A=r<n$. 设 $\sigma_1\geqslant\cdots\geqslant\sigma_r>0$ 是 A 的正的奇异值，又设 $\Sigma_r=\operatorname{ding}(\sigma_1,\cdots,\sigma_r)$. 证明存在一个酉矩阵 $U\in M_n$ 以及矩阵 $k\in M_r$ 与 $L\in M_{r,n-r}$，使得

$$A=U\begin{bmatrix}\Sigma_rK & \Sigma_rL\\ 0 & 0_{n-r}\end{bmatrix}U^*,\quad KK^*+LL^*=I_r \tag{2.6.9}$$

2.6.P26 设 $A\in M_n$，并假设 $1\leqslant\operatorname{rank}A=r<n$，并考虑表示(2.6.9). 证明：(a) A 是正规的，当且仅当 $L=0$ 且 $\Sigma_rK=K\Sigma_r$；(b) $A^2=0$ 当且仅当 $K=0$(在此情形有 $LL^*=I_r$)；(c) $A^2=0$ 当且仅当 A 与一个形如(2.6.8)的直和酉相似.

157 2.6.P27 设 $A\in M_n$ 是斜对称的. 如果 $\operatorname{rank}A\leqslant 1$，说明为什么 $A=0$.

2.6.P28 矩阵 $A\in M_n$ 称为 ***EP* 矩阵**，如果 A 与 A^* 有相同的值域. 每一个正规矩阵都是 EP 矩阵(2.5.P54b)，且每一个非奇异的矩阵(正规矩阵或非正规矩阵)都是一个 EP 矩阵. (a)证明：A 是 EP 矩阵且 $\operatorname{rank}A=r$，当且仅当存在一个非奇异的 $B\in M_r$ 以及一个酉矩阵 $V\in M_n$，使得 $A=V\begin{bmatrix}B & 0\\ 0 & 0\end{bmatrix}V^*$. (b)说明为什么 EP 矩阵是主秩的.

2.6.P29 如果 $x\in\mathbf{C}^n$ 是 $A\in M_n$ 的与特征值 λ 相伴的正规特征向量，证明 $|\lambda|$ 是 A 的一个奇异值.

2.6.P30 利用奇异值分解对复矩阵验证(0.4.6f)：$A\in M_{m,n}$ 的秩为 r 的充分必要条件是：存在非奇异的矩阵 $S\in M_m$ 以及 $T\in M_n$，使得 $A=S\begin{bmatrix}I_r & 0\\ 0 & 0\end{bmatrix}T$.

2.6.P31 设 $A\in M_{m,n}$. (a)利用奇异值分解 $A=V\Sigma W^*$ 证明：Hermite 矩阵 $\mathcal{A}=\begin{bmatrix}0 & A\\ A^* & 0\end{bmatrix}\in M_{m+n}$ 与实矩阵 $\begin{bmatrix}0 & \Sigma\\ \Sigma^{\mathrm{T}} & 0\end{bmatrix}$ 是酉相似的. (b)如果 $m=n$ 且 $\Sigma=\operatorname{diag}(\sigma_1,\cdots,\sigma_n)$，说明为什么 $\mathcal{A}$ 的特征值是 $\pm\sigma_1,\cdots,\pm\sigma_n$.

2.6.P32 设 $A\in M_n$，并设 $\mathcal{A}=\begin{bmatrix}0 & A\\ A^{\mathrm{T}} & 0\end{bmatrix}\in M_{2n}$. 如果 σ_1，…，σ_n 是 A 的奇异值，证明：σ_1，σ_1，…，σ_n，σ_n 是$\mathcal{A}$的奇异值.

2.6.P33 设 $\sigma_1\geqslant\cdots\geqslant\sigma_n$ 是 $A\in M_n$ 的有序排列的奇异值，又设 $r\in\{1,\cdots,n\}$. 证明：复合矩阵 $C_r(A)$的奇异值是$\binom{n}{r}$个可能的乘积σ_{i_1}，…，σ_{i_r}，其中 $1\leqslant i_1<i_2<\cdots<i_r\leqslant n$. 说明为什么 $\mathrm{tr}(C_r(A)C_r(A)^*)=\mathrm{tr}C_r(AA^*)=S_r(\sigma_1^2,\cdots,\sigma_n^2)$是 $C_r(A)$的奇异值的平方之和；见(1.2.14). 特别地，$\mathrm{tr}C_2(AA^*)=\sum_{1\leqslant i<j\leqslant n}\sigma_i^2(A)\sigma_j^2(A)$. 说明为什么 $\sigma_1\cdots\sigma_r$ 是 $C_r(A)$的最大的奇异值. 对于与$C_r(A)$的特征值有关的结果，见(2.3.P12).

2.6.P34 用 $\lambda_1(A)$，…，$\lambda_n(A)$以及 $\lambda_1(A^2)$，…，$\lambda_n(A^2)$分别记 $A\in M_n$ 以及 A^2 的特征值；用 $\sigma_1(A)$，…，$\sigma_n(A)$以及 $\sigma_1(A^2)$，…，$\sigma_n(A^2)$分别记它们的奇异值. (a)将 Schur 不等式(2.3.2a)应用于 A^2，推导出不等式 $\sum_{i=1}^n|\lambda_i(A)|^4\leqslant\sum_{i=1}^n\sigma_i^2(A^2)$. (b)将 Schur 不等式应用于复合矩阵 $C_2(A)$，推导出不等式 $\sum_{1\leqslant i<j\leqslant n}|\lambda_i(A)\lambda_j(A)|^2\leqslant\sum_{1\leqslant i<j\leqslant n}\sigma_i^2(A)\sigma_j^2(A)$. (c)证明 $(\mathrm{tr}AA^*)^2=\sum_{i=1}^n\sigma_i^4(A)+2\sum_{1\leqslant i<j\leqslant n}\sigma_i^2(A)\sigma_j^2(A)$. (d)证明 $\mathrm{tr}((AA^*-A^*A)^2)=2\sum_{i=1}^n\sigma_i^4(A)-2\sum_{i=1}^n\sigma_i^2(A^2)$. (e)证明 $\left(\sum_{i=1}^n|\lambda_i(A)|^2\right)^2=\sum_{i=1}^n|\lambda_i(A)|^4+2\sum_{1\leqslant i<j\leqslant n}|\lambda_i(A)\lambda_j(A)|^2$. (f)导出结论

$$\sum_{i=1}^n|\lambda_i(A)|^2\leqslant\sqrt{(\mathrm{tr}AA^*)^2-\frac{1}{2}\mathrm{tr}((AA^*-A^*A)^2)}\qquad(2.6.10)$$

它加强了 Schur 不等式(2.3.2a). 为什么当 A 为正规矩阵时，(2.6.10)中的等式成立？说明为什么(2.6.10)中等式成立当且仅当 A^2 与 $C_2(A)$两者都是正规的.

2.6.P35 利用上一个问题中的记号，证明

$$\sum_{i=1}^n|\lambda_i(A)|^2\leqslant\sqrt{\left(\mathrm{tr}AA^*-\frac{1}{n}|\mathrm{tr}A|^2\right)^2-\frac{1}{2}\mathrm{tr}((AA^*-A^*A)^2)}+\frac{1}{n}|\mathrm{tr}A|^2\qquad(2.6.11)$$

进而证明：(2.6.11)中的上界小于或者等于(2.6.10)中的上界，其中的等式当且仅当 $\mathrm{tr}A=0$ 或者 A 是正规矩阵时成立. 158

2.6.P36 设 $A\in M_n$ 的秩为 r，设 s_1，…，s_d 是 A 的不同的正的奇异值，按任意次序排列，其重数分别为 n_1，…，n_d，又令 $A=V\Sigma W^*$ 是奇异值分解，其中 V，$W\in M_n$ 是酉矩阵，而 $\Sigma=s_1I_{n_1}\oplus\cdots\oplus s_dI_{n_d}\oplus 0_{n-r}$. (a)说明为什么 A 是对称的充分必要条件是 $V=\overline{W}(S_1\oplus\cdots\oplus S_d\oplus\widetilde{W})$，其中 $\widetilde{W}\in M_{n-r}$是酉矩阵，而每一个 $S_j\in M_{n_j}$ 都是酉矩阵，且是对称的. (b)如果 A 的奇异值各不相同(即如果 $d\geqslant n-1$)，说明为什么 A 是对称的，当且仅当 $V=\overline{W}D$，其中 $D\in M_n$ 是一个对角酉矩阵.

2.6.P37 假设 $A\in M_n$ 有不同的奇异值. 设 $A=V\Sigma W^*$ 以及 $A=\hat{V}\Sigma\hat{W}^*$ 是奇异值分解. (a)如果 A 是非奇异的，说明为什么存在一个对角酉矩阵 D，使得$\hat{V}=VD$ 以及$\hat{W}=WD$. (b)如果 A 是奇异的，说明为什么存在至多相差一个对角元素的对角酉矩阵 D 以及$\widetilde{D}$，使得有$\hat{V}=VD$ 以及$\hat{W}=W\widetilde{D}$.

2.6.P38 设 $A\in M_n$ 是非奇异的，并设 σ_n 是 $A+A^{-*}$ 的最小的奇异值. 证明 $\sigma_n\geqslant 2$. 对于等式成立的情形有何结论？

2.6.P39 设 $A\in M_n$ 是共轭对合的，所以 A 是非奇异的，且 $A=\overline{A}^{-1}$. 说明为什么 A 的不等于 1 的奇异

值出现在倒数对中.

2.6.P40 利用(2.4.5.1)的记号，假设 T 与 T' 是酉相似的. (a)说明：对于每个 i，$j=1$，…，d，为什么 T_{ij} 与 T'_{ij} 的奇异值都必定是相同的. (b)当 $n=2$ 时这个必要条件说的是什么？在此情形为什么它既是必要的，又是充分的？(c)设 $n=4$ 以及 $d=2$. 考虑例子 $T_{11}=T'_{11}=\begin{bmatrix}1&1\\0&1\end{bmatrix}$，$T_{22}=T'_{22}=\begin{bmatrix}2&2\\0&2\end{bmatrix}$，$T_{12}=\begin{bmatrix}3&0\\0&4\end{bmatrix}$以及 $T'_{11}=\begin{bmatrix}0&4\\3&0\end{bmatrix}$；说明为什么(a)中的必要条件不一定是充分的.

注记以及进一步的阅读参考 复对称矩阵的特殊的奇异值分解(2.6.6a)是由 L. Autonne 于 1915 年发表的，自从那时起，它又多次被人重新发现. Autonne 的证明用到了(2.6.4)的一种形式，不过他的方法要求矩阵是非奇异的. (2.6.P20)指出了怎样从非奇异的情形推导出奇异的情形. 有关奇异值分解的历史，包括对 Autonne 的贡献的介绍，参见 Horn 与 Johnson(1991)所著书中的 3.0 节.

2.7 CS 分解

CS 分解是分划的酉矩阵在分划的酉等价之下的标准型. 它的证明涉及奇异值分解、*QR* 分解以及如下习题中的结果.

习题 设 Γ，$L\in M_p$. 假设 $\Gamma=\mathrm{diag}(\gamma_1,\cdots,\gamma_p)$，其中 $0\leqslant\gamma_1\leqslant\cdots\leqslant\gamma_p\leqslant1$，$L=[\ell_{ij}]$ 是下三角的，且$[\Gamma\quad L\quad 0]\in M_{p,2p+k}$的行是标准正交的. 说明为什么 L 是对角的，$L=\mathrm{diag}(\gamma_1,\cdots,\gamma_p)$，且 $|\lambda_j|^2=1-\gamma_j^2$，$j=1$，…，$p$. 提示：如果 $\gamma_1=1$，为什么 $L=0$？如果
159 $\gamma_1<1$，那么$|\ell_{11}|^2=1-\gamma_1^2$. 为什么正交性能确保有 $\ell_{21}=\cdots=\ell_{p1}=0$？对行着手去做. ◀

定理 2.7.1(*CS* 分解) 设 p，q 与 n 是给定的整数，其中 $1<p\leqslant q<n$ 且 $p+q=n$. 设 $U=\begin{bmatrix}U_{11}&U_{12}\\U_{21}&U_{22}\end{bmatrix}\in M_n$ 是酉矩阵，其中 $U_{11}\in M_p$ 且 $U_{22}\in M_q$. 则存在酉矩阵 V_1，$W_1\in M_p$ 以及 V_2，$W_2\in M_q$，使得

$$\begin{bmatrix}V_1&0\\0&W_1\end{bmatrix}\begin{bmatrix}U_{11}&U_{12}\\U_{21}&U_{22}\end{bmatrix}\begin{bmatrix}V_2&0\\0&W_2\end{bmatrix}=\begin{bmatrix}C&S&0\\-S&C&0\\0&0&I_{q-p}\end{bmatrix}\tag{2.7.1.2}$$

其中 $C=\mathrm{diag}(\sigma_1,\cdots,\sigma_p)$，$\sigma_1\geqslant\cdots\geqslant\sigma_p$ 是 U_{11} 的按照非增次序排列的奇异值，而 $S=\mathrm{diag}\left((1-\sigma_1^2)^{1/2},\cdots,(1-\sigma_p^2)^{1/2}\right)$.

证明 我们的策略是做出一系列分划的酉等价，它们一步一步将 U 化简为具有所需要的形式的分块矩阵. 第一步是利用奇异值分解：记 $U_{11}=V\Sigma W=(VK_p)(K_p\Sigma K_p)(K_pW)=\tilde{V}\Gamma\tilde{W}$，其中 V，$W\in M_p$ 是酉矩阵，K_p 是 $p\times p$ 反序矩阵(0.9.5.1). $\tilde{V}=VK_p$，$\tilde{W}=K_pW$，$\Sigma=\mathrm{diag}(\sigma_1,\cdots,\sigma_p)$，其中 $\sigma_1\geqslant\cdots\geqslant\sigma_p$，且 $\Gamma=K_p\Sigma K_p=\mathrm{diag}(\sigma_p,\cdots,\sigma_1)$. 计算

$$\begin{bmatrix}\tilde{V}^*&0\\0&I_q\end{bmatrix}\begin{bmatrix}U_{11}&U_{12}\\U_{21}&U_{22}\end{bmatrix}\begin{bmatrix}\tilde{W}^*&0\\0&I_q\end{bmatrix}=\begin{bmatrix}\Gamma&\hat{V}^*U_{12}\\U_{21}\tilde{W}^*&U_{22}\end{bmatrix}$$

这个矩阵是酉矩阵(它是三个酉矩阵的乘积)，所以每一列的 Euclid 范数均为 1，这就意味着 $\sigma_1=\gamma_p\leqslant1$. 现在利用 *QR* 分解(2.1.14)以及它的变形(2.1.15b)来记$\tilde{V}^*U_{12}=[L\quad 0]\tilde{Q}$

以及 $U_{21}\widetilde{W}^*=Q\begin{bmatrix}R\\0\end{bmatrix}$，其中 $\widetilde{Q}$，$Q\in M_q$ 是酉矩阵，$L=[\ell_{ij}]\in M_p$ 是下三角矩阵，而 $R=[r_{ij}]\in M_p$ 是上三角矩阵．计算

$$\begin{bmatrix}I_p & 0\\0 & Q^*\end{bmatrix}\begin{bmatrix}\Gamma & \widetilde{V}^*U_{12}\\U_{21}\widetilde{W}^* & U_{22}\end{bmatrix}\begin{bmatrix}I_p & 0\\0 & \widetilde{Q}^*\end{bmatrix}=\begin{bmatrix}\Gamma & [L\quad 0]\\\begin{bmatrix}R\\0\end{bmatrix} & Q^*U_{22}\widetilde{Q}^*\end{bmatrix}$$

上一习题中的论证方法表明：L 与 R 两者都是对角的，且对每个 $i=1,\cdots,p$ 有 $|r_{ii}|=|\ell_{ii}|=\sqrt{1-\gamma_i^2}$．设 $M=\mathrm{diag}(\sqrt{1-\gamma_1^2},\cdots,\sqrt{1-\gamma_p^2})$，并令 $t=\max\{i:\gamma_i<1\}$．（我们可以假设 $t\geqslant 1$，因为如果 $\gamma_1=1$，那么 $\Gamma=I_p$，且 $M=0$，所以就对 $C=I_p$ 以及 $S=0_p$ 有(2.7.1.2)成立．）则存在对角酉矩阵 D_1，$D_2\in M_p$ 使得 $D_1R=-M$ 以及 $LD_2=M$，所以通过 $I_p\oplus D_1\oplus I_{n-2p}$ 在左边作成的酉相合与通过 $I_p\oplus D_2\oplus I_{n-2p}$ 在右边作成的酉相合产生出一个形如

$$\begin{bmatrix}\Gamma & [M\quad 0]\\\begin{bmatrix}-M\\0\end{bmatrix} & Z\end{bmatrix}=\begin{bmatrix}\Gamma_1 & 0 & M_1 & 0 & 0\\0 & I_{p-t} & 0 & 0_{p-t} & 0\\-M_1 & 0 & Z_{11} & Z_{12} & Z_{13}\\0 & 0_{p-t} & Z_{21} & Z_{22} & Z_{23}\\0 & 0 & Z_{31} & Z_{32} & Z_{33}\end{bmatrix}$$

的酉矩阵，其中有分划的矩阵 $\Gamma=\Gamma_1\oplus I_{p-t}$ 以及 $M=M_1\oplus 0_{p-t}$，所以 M_1 是非奇异的．第一和第三分块列的正交性(以及 M_1 的非奇异性)就蕴含 $Z_{11}=\Gamma_1$，因此要求每一行和每一列都是单位向量就确保了 Z_{12}，Z_{13}，Z_{21} 以及 Z_{31} 全都是零分块．从而我们有

$$\begin{bmatrix}\Gamma_1 & 0 & M_1 & 0 & 0\\0 & I_{p-t} & 0 & 0_{p-t} & 0\\-M_1 & 0 & \Gamma_1 & 0 & 0\\0 & 0_{p-t} & 0 & Z_{22} & Z_{23}\\0 & 0 & 0 & Z_{32} & Z_{33}\end{bmatrix}$$

其右下角的分块 $\widetilde{Z}=\begin{bmatrix}Z_{22} & Z_{23}\\Z_{32} & Z_{33}\end{bmatrix}\in M_{q-t}$ 是酉矩阵的一个直和项，故而它是酉矩阵，于是对某个酉矩阵 $\hat{V}$，$\hat{W}\in M_{q-1}$ 有 $\widetilde{Z}=\hat{V}I_{q-t}\hat{W}$．通过 $I_{p+t}\oplus\hat{V}^*$ 在左边作出的酉等价以及通过 $I_{p+t}\oplus\hat{W}^*$ 在右边作出的酉等价产生出分块矩阵 160

$$\begin{bmatrix}\Gamma_1 & 0 & M_1 & 0 & 0\\0 & I_{p-t} & 0 & 0_{p-t} & 0\\-M_1 & 0 & \Gamma_1 & 0 & 0\\0 & 0_{p-t} & 0 & I_{p-t} & 0\\0 & 0 & 0 & 0 & I_{q-p}\end{bmatrix}$$

最后，通过 $K_p\oplus K_p\oplus I_{q-p}$ 作出的酉相似产生出一个酉矩阵，它具有所要求的构造(2.7.1.2)． □

CS 分解是与 $I_p\oplus I_q$ 共形地加以分划且阶为 $n=p+q$(为方便起见，设 $p\leqslant q$，但这不

是本质的要求)的所有酉矩阵 $U=\begin{bmatrix}U_{11} & U_{12}\\ U_{21} & U_{22}\end{bmatrix}\in M_n$ 组成的集合的一种参数化的描述．这些参数是四个更小的任意的酉矩阵 V_1，$W_1\in M_p$ 以及 V_2，$W_2\in M_q$，以及任意 p 个实数 σ_1，…，σ_p，$1\geqslant\sigma_1\geqslant\cdots\geqslant\sigma_p\geqslant 0$．这四个分块的参数化是

$$U_{11}=V_1CW_1,\quad U_{12}=V_1[S\quad 0]W_2, \tag{2.7.1.3}$$

$$U_{21}=V_2\begin{bmatrix}-S\\ 0\end{bmatrix}W_1,\quad \text{以及 } U_{22}=V_2\begin{bmatrix}C & 0\\ 0 & I_{q-p}\end{bmatrix}W_2$$

其中 $C=\mathrm{diag}(\sigma_1,\cdots,\sigma_p)$，而 $S=\mathrm{diag}((1-\sigma_1^2)^{1/2},\cdots,(1-\sigma_p^2)^{1/2})$．CS 分解是一种用途广泛的工具，特别是在与子空间之间的距离以及角度有关的问题中．

问题

利用 CS 分解求解如下的每一个问题，即使这些问题有可能用其他方法解决．一个给定的 $A\in M_{n,m}$ 称作一个**短缩**(contraction)，如果它最大的奇异值小于或等于 1．

2.7.P1　证明：酉矩阵的每一个子矩阵都是一个短缩；这里的子矩阵不一定是主子矩阵，也不一定是方阵．

2.7.P2　对上一个问题有一个有意思的逆命题．设 $A\in M_{m,n}$ 是一个短缩，并假设它的奇异值中恰好有 ν 个

161 是严格小于 1 的．证明：存在矩阵 B，C 以及 D，使得 $U=\begin{bmatrix}D & B\\ C & A\end{bmatrix}\in M_{\max\{m,n\}+|m,n|+\nu}$ 是酉矩阵．这样一个矩阵 U 称为短缩 A 的**酉膨体**(unitary dilation)．

2.7.P3　如果 A 是一个 $n\times n$ 酉矩阵的 $k\times k$ 子矩阵，且如果 $2k>n$，证明 A 的某个奇异值等于 1．

2.7.P4　设 $U=\begin{bmatrix}U_{11} & U_{12}\\ U_{21} & U_{22}\end{bmatrix}$ 是酉矩阵，其中 $U_{11}\in M_p$ 且 $U_{22}\in M_q$．证明：U_{11} 与 U_{22} 的零空间相同，而 U_{12} 与 U_{21} 的零空间相同．与(0.7.5)比较．

2.7.P5　证明(2.6.P19)中的结论．

2.7.P6　设 $U=\begin{bmatrix}u_{11} & u_{12}\\ u_{21} & u_{22}\end{bmatrix}\in M_2$ 是酉矩阵．(a)证明 $|u_{11}|=|u_{22}|$ 以及 $|u_{21}|=|u_{12}|=(1-|u_{11}|^2)^{1/2}$．(b)证明：$U$ 对角酉相似于一个复的对称矩阵．

进一步的阅读参考　关于历史综述以及 CS 分解的一种形式(它包含酉矩阵 $U=[U_{ij}]\in M_n$ 的任意的 2×2 分划，这里 $U_{ij}\in M_{r_i,c_j}$，i，$j=1$，2，$r_1+r_2=n=c_1+c_2$)，见 C. Paige 以及 M. Wei，History and

162 generality of the CS decomposition，*Linear Algebra Appl*. 208/209(1994)303-326．