第四章　正规矩阵与矩阵的分解

除特别说明, 本章讨论的矩阵都是复数矩阵.

引　言　矩阵如何快速计算?

在第一章中, 我们已经看到如果将一个小秩矩阵分解为两个满秩矩阵的乘积, 则可以快速地计算该矩阵的高次幂. 实际上, 利用初等变换求可逆矩阵 A 的逆矩阵, 其本质就是将矩阵 A 分解为若干较为简单的矩阵 (即初等矩阵) 的乘积. 解线性方程组的Gauss消元法其实质也是矩阵分解. 回忆两个 $m\times n$ 矩阵 A 与 B 等价是指存在可逆矩阵 P 与 Q 使得 $B=PAQ$. 若 $r(A)=r$, 则 A 等价于分块矩阵 $\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}$. 现考虑线性方程组 $Ax=b$ 的同解方程组 $PAx=Pb$, 如果令 $x=Qy$, 则得

$$\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix} y = Pb. \tag{4.0.1}$$

于是由 (4.0.1) 的解即可得到原方程组的解. 这里的实质也是矩阵分解, 即将 A 分解成 $A=P^{-1}\begin{pmatrix} I_r & 0\\ 0 & 0\end{pmatrix}Q^{-1}$.

近几十年来, 随着计算机的不断更新换代, 计算技术得到了迅猛发展并促使矩阵分解快速成为解决众多工程问题的有力手段. 因此本章集中介绍几种常用的矩阵分解, 包括谱分解和Schur三角化分解, Cholesky分解与 LU 分解, 正交三角分解 (又称为 QR 分解) 以及奇异值分解. 由于所有分解的目的不外乎简化计算或深化理论, 因此都要涉及一些特殊矩阵, 故我们首先介绍这些“好矩阵”.

第一节　正规矩阵

由Schur三角化定理, 任何一个矩阵都可以酉三角化, 因此一个“好矩阵”当能够酉对角化. 但是以往判断一个矩阵能否 (酉) 对角化需要借助于特征值与特征向量, 这是极其不方便的, 因为我们知道寻找矩阵的特征值与特征向量常常是极为困难的工作. 本节的目的即是给出一类可以酉对角化的“好矩阵”一个直接的判断, 即下述

定理 4.1.1　矩阵 A 可以酉对角化 $\iff AA^*=A^*A$.

证　由Schur三角化定理, 存在酉矩阵 U 使得 $U^*AU=T$ 为上三角矩阵. 显然 $AA^*=A^*A \iff TT^*=T^*T$. 因此不妨设 A 是上三角矩阵.

必要性是显然的, 因为如果 A 可以酉对角化, 则存在酉矩阵 U 使得 $U^*AU=D$ 为对角矩阵, 因此

$$AA^*=(UDU^*)(UD^*U^*)=UDD^*U^*=(UD^*U^*)(UDU^*)=A^*A.$$

充分性. 记 $A=(a_{ij})$, 其中 $a_{ij}=0, i>j$. 因为 $AA^*=A^*A$, 故两端矩阵具有相同的对角元素, 因此

$$a_{11}\overline{a}_{11}+a_{12}\overline{a}_{12}+\cdots+a_{1n}\overline{a}_{1n}=\overline{a}_{11}a_{11},$$

故知 A 的第一行的非对角元素均为0. 于是利用归纳法即可知 A 是对角矩阵. □

通常将可以酉对角化的矩阵称为正规矩阵, 即有下面的定义:

定义 4.1.1 设 $A \in \mathbb{C}^{n\times n}$, 若 $AA^* = A^*A$, 则称 A 为正规矩阵.

实对称矩阵, 实反对称矩阵, 正交矩阵, Hermite 矩阵, 反 Hermite 矩阵, 酉矩阵等都是正规矩阵. 另外, 若 A 为正规矩阵, 则与 A 酉相似的矩阵仍为正规矩阵.

例 4.1.1 矩阵 $\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$ 是正规矩阵.

由 定理 4.1.1 的证明可以得到下面的

引理 4.1.1 设 A 为正规矩阵, 若 A 又为三角矩阵, 则 A 为对角矩阵.

例 4.1.2 矩阵 $\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ 可以相似对角化, 但由 引理 4.1.1, 它不能酉对角化. 注意, 该矩阵是一个幂等矩阵.

将正规矩阵 A 酉对角化的酉矩阵的每一列都是 A 的特征向量, 由酉矩阵的构造可得 (细节见习题 4)

定理 4.1.2 设 $A \in \mathbb{C}^{n\times n}$, 则 A 为正规矩阵 $\iff$ A 有 n 个两两正交的单位特征向量.

推论 4.1.1 正规矩阵属于不同特征值的特征向量是相互正交的.

正规矩阵有许多良好的数字特性, 比如下面的

定理 4.1.3 设 $A = (a_{ij})_{n\times n}$ 是复矩阵, $\lambda_1, \lambda_2, \cdots, \lambda_n$ 为 A 的 n 个特征值. 则
(1) (**Schur 不等式**) $\sum\limits_{i=1}^{n} |\lambda_i|^2 \le \sum\limits_{i,j=1}^{n} |a_{ij}|^2$;
(2) A 为正规矩阵 $\iff$ $\sum\limits_{i=1}^{n} |\lambda_i|^2 = \sum\limits_{i,j=1}^{n} |a_{ij}|^2$.

证 此处仅给出证明轮廓, 细节见习题 9. 由 Schur 酉三角化定理可知 $U^*AU = B$ 是上三角矩阵, 利用下述等式 (见第一章, 命题 1.1.1(5))

$$\operatorname{tr}(AA^*) = \sum_{i,j=1}^{n} |a_{ij}|^2$$

并比较 $\operatorname{tr}(AA^*)$ 与 $\operatorname{tr}(BB^*)$. □

例 4.1.3 设 A 为正规矩阵且幂零, 则 $A = 0$.
由第三章例 3.1.11 知 A 的特征值均为 0, 再由 定理 4.1.3 即知 $A = 0$.

例 4.1.4 实数矩阵 $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ 可以酉对角化, 故它是正规矩阵, 但它不能正交对角化 (为什么?).

为了更好地描述实正规矩阵, 我们引入以下的定义.

定义 4.1.2 设 a 与 b 是实数且 $b \neq 0$. 则称2阶实矩阵

$$\begin{pmatrix} a & b \\ -b & a \end{pmatrix} \tag{4.1.1}$$

为一个 **Schur 型**.

注. 1831 年, Gauss将复数看作平面上的点, 并得出了模长为1的复数相当于平面上的旋转变换这一结论. 式 (4.1.1) 中的Schur型正是复数 $a+b\mathrm{i}$ 的矩阵表示, 源自A.Cayley(1845年). 易知, 一个Schur型的特征值正是复数 $a \pm b\mathrm{i}$.

例 4.1.5 记式 (4.1.1) 中的Schur型为 A. 由于 $b \neq 0$, A 具有非实特征值 $a \pm b\mathrm{i}$, 且酉相似于对角矩阵 $(a+b\mathrm{i}) \oplus (a-b\mathrm{i})$. $A^*A = AA^* = (a^2+b^2) \oplus (a^2+b^2)$. 因此每个Schur型都是正规矩阵, 但不能正交对角化 (见本节思考题4).

实际上在第三章习题5中, 我们已经看到了实正规矩阵在正交变换下的最简形式如下

定理 4.1.4 (实正规矩阵) 设 A 是 n 阶实矩阵, 则 A 是正规矩阵 $\iff$ 存在正交矩阵 Q 使得

$$Q^TAQ = A_1 \oplus A_2 \oplus \cdots \oplus A_s \tag{4.1.2}$$

其中每个 A_i 或者是1阶实矩阵, 或者是一个Schur型.

Hermite矩阵 (特别, 实对称矩阵) 是最重要的正规矩阵. 在线性代数课程中, 我们知道实对称矩阵必定可以正交对角化 (或见第一章), 而实反对称矩阵的特征值为0或纯虚数. 它们的逆命题成立吗? 由公式 (4.1.2), 可以知道这些逆命题确实成立, 见下面的推论.

推论 4.1.2 设 A 是 n 阶实矩阵.

(1) A 是对称矩阵 $\iff$ 存在正交矩阵 Q 使得 Q^TAQ 是对角矩阵 $\iff$ 对任意正整数 k 有 $\mathrm{tr}(A^k) = \sum_{i=1}^n \lambda_i^k$, 其中 $\lambda_1, \cdots, \lambda_n$ 是 A 的所有特征值;

(2) A 是反对称矩阵 $\iff$ 存在正交矩阵 Q 使得

$$Q^TAQ = 0 \oplus A_1 \oplus A_2 \oplus \cdots \oplus A_s$$

其中每个 $A_i = \begin{pmatrix} 0 & b_i \\ -b_i & 0 \end{pmatrix}$, 从而反对称矩阵的非零特征值为纯虚数;

(3) A 是正交矩阵 $\iff$ 存在正交矩阵 Q 使得

$$Q^TAQ = I_s \oplus (-I_t) \oplus A_1 \oplus A_2 \oplus \cdots \oplus A_m$$

其中每个 A_i 是2阶Givens旋转矩阵, 从而正交矩阵的特征值的模均为1.

设 B 是 n 阶复矩阵.

(4) B 是Hermite矩阵 $\iff$ 存在酉矩阵 U 使得 U^*BU 是实对角矩阵;

(5) A 是反Hermite矩阵 $\iff$ 存在酉矩阵 U 使得 U^*BU 是纯虚数对角矩阵 (即实部为0);

(6) A 是酉矩阵 $\iff$ 存在酉矩阵 U 使得 U^*BU 是对角元素的模均为1的对角矩阵, 从而酉矩阵的特征值的模均为1;

(7) Hermite矩阵 A 正定 $\iff$ A 的所有顺序主子式均大于0.

推论 4.1.2 的证明见习题 11, 其中 (7) 的证明可利用下面的引理.

引理 4.1.2 Hermite 阵或实对称矩阵 A 在某一个 k 维子空间上正定 $\iff$ A 至少有 k 个特征值 (包括重数) 大于零.

证 充分性的证明较为简单, 见习题 12. 下证必要性. 设 A 在一个 k 维子空间 V_k 上正定. 假设 $\lambda_1, \lambda_2, \cdots, \lambda_m$ 为 A 的所有大于零的特征值, $\lambda_{m+1}, \lambda_{m+2}, \cdots, \lambda_n$ 为所有非正的特征值. 设 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 分别为 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 对应的两两正交的单位特征向量. 记

$$W = \mathrm{Span}\{\alpha_{m+1}, \alpha_{m+2}, \cdots, \alpha_n\}$$

为 $\alpha_{m+1}, \alpha_{m+2}, \cdots, \alpha_n$ 生成的子空间. 则对 $\forall x \in W$, 有 $x^*Ax \leq 0$(为什么?).

如果 $m < k$, 则由维数定理 (第二章定理 2.1.2) 以及 $\dim(V_k + W) \leq \dim \mathbb{C}^n = n$, 知

$$\dim(V_k \cap W) = \dim V_k + \dim W - \dim(V_k + W) \geq k + (n - m) - n = k - m > 0.$$

即 $V_k \cap W \neq \{0\}$. 任取 $0 \neq x \in V_k \cap W$, 则 $x^*Ax > 0$. 矛盾! 所以必须 $m \geq k$, 也就是说 A 至少有 k 个大于零的特征值. □

例 4.1.6 (Hermite 矩阵与二次型) 设 A 为 n 阶 Hermite 矩阵, x^*Ax 是相应的复二次型, 则 A 可以酉对角化等价于该复二次型可以通过酉变换 (即等距变换) 化为

$$\lambda_1|y_1|^2 + \lambda_2|y_2|^2 + \cdots + \lambda_n|y_n|^2. \tag{4.1.3}$$

由于诸特征值 λ_i 均为实数, 所以 Hermite 矩阵对应的复二次型的值总是实数. 若 A 还是可逆矩阵, 则诸特征值均非零, 它们的倒数正是主轴平方 (在 $\mathbb{R}^2$ 中即是椭圆或双曲线的半轴长的平方). 由此还可知, 重特征值意味着对称轴的不确定性, 比如圆的情形, 或在 $\mathbb{R}^3$ 中有两个半轴长相等的椭球

$$\frac{x^2}{4} + \frac{y^2}{4} + \frac{z^2}{9} = 1,$$

其在 xoy- 平面的对称轴有无穷多对.

例 4.1.7 (2 阶实正规矩阵的几何意义) 由 定理 4.1.4 可知, 2 阶实正规矩阵 A 一定正交相似于对角矩阵或一个 Schur 型 $\begin{pmatrix} a & b \\ -b & a \end{pmatrix}$, 其中 $b \neq 0$. 由于 2 阶正交矩阵 (对应的线性变换) 是旋转变换或者反射变换的复合, 对角矩阵则是在两个正交方向的伸缩, 而 Schur 型则是系数为 $\sqrt{a^2 + b^2}$ 的位似与一个旋转的合成, 因此 2 阶实正规矩阵对应的平面线性变换是正交变换与一个在两个正交方向的伸缩的合成. 一般将正规矩阵对应的线性变换称为**正规变换**, 则平面上的可逆线性变换 σ 是正规变换 $\iff$ σ 将某个正方形伸缩 为矩形 (因此非正规的可逆线性变换不可能将任何正方形伸缩 为矩形) 或者将所有正方形均变为正方形 (证明见习题 13).

思考题

1. 复对称矩阵是否是正规矩阵?
2. 正规矩阵的和与积是否为正规矩阵?
3. 相似变换是否保持矩阵的正规性?
4. 为什么说 例 4.1.5 中的矩阵可以酉对角化但不能正交对角化?
5. 研究 例 4.1.6, 讨论 2 阶与 3 阶实对称矩阵的特征值 (包括零) 的几何意义.

第二节 正规矩阵的谱分解

设 A 为正规矩阵, 则由 定理4.1.1 知, 存在酉矩阵 U 使得 $U^*AU = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. 因而 $A = U\operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)U^*$. 令 $U = (\alpha_1, \alpha_2, \cdots, \alpha_n)$, 则

$$\begin{aligned} A &= (\alpha_1, \alpha_2, \cdots, \alpha_n)\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}\begin{pmatrix} \alpha_1^* \\ \alpha_2^* \\ \vdots \\ \alpha_n^* \end{pmatrix} \\ &= \lambda_1\alpha_1\alpha_1^* + \lambda_2\alpha_2\alpha_2^* + \cdots + \lambda_n\alpha_n\alpha_n^*. \end{aligned} \tag{4.2.1}$$

由于 $\lambda_1, \lambda_2, ..., \lambda_n$ 为 A 的特征值, $\alpha_1, \alpha_2, ..., \alpha_n$ 为对应的两两正交的单位特征向量, 故公式 (4.2.1) 称为正规矩阵 A 的**谱分解** 或**特征 (值) 分解**. 若把公式 (4.2.1) 中系数相同的放在一起 (0特征值对应的项去掉), 然后把系数提出来, 则公式 (4.2.1) 就变成

$$A = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_s P_s, \tag{4.2.2}$$

其中 $\lambda_1, \lambda_2, \cdots, \lambda_s$ 为 A 的互不相同的非零特征值. 由于

$$\begin{aligned} &(\alpha_i\alpha_i^*)^* = \alpha_i\alpha_i^*, && 1 \le i \le n, \\ &(\alpha_i\alpha_i^*)(\alpha_j\alpha_j^*) = 0, && 1 \le i \ne j \le n, \\ &(\alpha_i\alpha_i^*)^2 = \alpha_i\alpha_i^*, && 1 \le i \le n, \end{aligned}$$

所以

$$P_i^* = P_i, \quad P_i^2 = P_i, \quad P_iP_j = 0, \quad 1 \le i \ne j \le s. \tag{4.2.3}$$

从第二章幂等矩阵与投影变换的对应关系可知, P_i 是某正交投影变换 (在某基下) 的矩阵, 故常称为**正交投影矩阵** .

例 4.2.1 (谱分解的几何意义) 如果2阶实正规矩阵 A 有两个相同的特征值 λ, 则 $A = \lambda I$ 就是它的谱分解. 如果 A 有两个不同的特征值 λ_1 与 λ_2, 则其谱分解为 $A = \lambda_1 P_1 + \lambda_2 P_2$. 因此, 对任意 $\alpha \in \mathbb{R}^2$, 有

$$A\alpha = \lambda_1 P_1\alpha + \lambda_2 P_2\alpha. \tag{4.2.4}$$

计算内积可得 $(P_1\alpha, P_2\alpha) = (P_1\alpha)^T P_2\alpha = \alpha^T P_1^T P_2\alpha = 0$, 所以 $\lambda_1 P_1\alpha$ 与 $\lambda_2 P_2\alpha$ 是正交的向量. 所以公式 (4.2.4) 将 $A\alpha$ 分解成了两个正交向量的和. 因此, 二维正规矩阵的谱分解实际上是平面的正交投影变换的推广. 对任意 n 阶正规矩阵的谱分解公式 (4.2.2) 有类似的解释.

例 4.2.2 设 $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. 则 A 的谱分解为

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \mathrm{i}\begin{pmatrix} \frac{1}{2} & -\frac{\mathrm{i}}{2} \\ \frac{\mathrm{i}}{2} & \frac{1}{2} \end{pmatrix} - \mathrm{i}\begin{pmatrix} \frac{1}{2} & \frac{\mathrm{i}}{2} \\ -\frac{\mathrm{i}}{2} & \frac{1}{2} \end{pmatrix},$$

于是

$$A\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \mathrm{i}\begin{pmatrix} x_1 - \mathrm{i}x_2 \\ \mathrm{i}x_1 + x_2 \end{pmatrix} - \mathrm{i}\begin{pmatrix} x_1 + \mathrm{i}x_2 \\ -\mathrm{i}x_1 + x_2 \end{pmatrix},$$

容易看出, 上式右端的两个复向量是正交的.

例 4.2.3 已知正规矩阵

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

求 A 的谱分解.

解

$$|\lambda I-A|=\begin{vmatrix} \lambda-8 & -4 & 1 \\ -4 & \lambda+7 & -4 \\ 1 & -4 & \lambda-8 \end{vmatrix}=(\lambda-9)^2(\lambda+9).$$

所以 A 的特征值为 $\lambda_1=9$(二重), $\lambda_2=-9$. 相应的相互正交的单位特征向量为:

$$\alpha_1=\frac{1}{3}\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix},\quad \alpha_2=\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix},\quad \alpha_3=\frac{1}{3\sqrt{2}}\begin{pmatrix} 1 \\ -4 \\ 1 \end{pmatrix}.$$

则

$$P_1=\alpha_1\alpha_1^*+\alpha_2\alpha_2^*=\frac{1}{3}\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}(\frac{2}{3},\frac{1}{3},\frac{2}{3})+\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}(\frac{1}{\sqrt{2}},0,-\frac{1}{\sqrt{2}})$$

$$=\frac{1}{9}\begin{pmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{pmatrix}+\frac{1}{2}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}=\frac{1}{18}\begin{pmatrix} 17 & 4 & -1 \\ 4 & 2 & 4 \\ -1 & 4 & 17 \end{pmatrix}$$

$$P_2=\alpha_3\alpha_3^*=\frac{1}{3\sqrt{2}}\begin{pmatrix} 1 \\ -4 \\ 1 \end{pmatrix}(\frac{\sqrt{2}}{6},-\frac{2\sqrt{2}}{3},\frac{\sqrt{2}}{6})=\frac{1}{18}\begin{pmatrix} 1 & -4 & 1 \\ -4 & 16 & -4 \\ 1 & -4 & 1 \end{pmatrix}.$$

所以 A 的谱分解为

$$A=9P_1-9P_2.$$

例 4.2.4 设 $A=\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$, 求 A^TA 与 AA^T 的谱分解.

解 直接计算可知

$$A^TA=\begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 1 \end{pmatrix},\qquad |\lambda I-A^TA|=\lambda(\lambda-1)(\lambda-3).$$

只需计算非零特征值的特征向量, 得: 属于特征值 1 与 3 的特征向量 (必定正交) 分别为 $(1,0,-1)^T$ 与 $(1,2,1)^T$. 因此 A^TA 的谱分解为

$$A^TA=1\bullet[\frac{1}{2}\begin{pmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{pmatrix}]+3\bullet[\frac{1}{6}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix}].$$

类似地, 可得 AA^T 的谱分解为

$$AA^T = 1 \bullet [\frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}] + 3 \bullet [\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}]. \tag{4.2.5}$$

例 4.2.5 如果 A 是可逆 Hermite 矩阵, 则可以利用 A 的谱分解来求其逆矩阵. 设 A 的谱分解为

$$A = \sum_{i=1}^{n} \lambda_i \alpha_i \alpha_i^*,$$

则 (证明见习题 20)

$$A^{-1} = \sum_{i=1}^{n} \frac{1}{\lambda_i} \alpha_i \alpha_i^*. \tag{4.2.6}$$

比如 例 4.2.4 中的矩阵 AA^T 是可逆对称的, 故由其谱分解公式 (4.2.5) 及公式 (4.2.6) 可知其逆矩阵为

$$(AA^T)^{-1} = 1 \bullet [\frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}] + \frac{1}{3} \bullet [\frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}] = \frac{1}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

对于 n 阶可对角化矩阵 A(这样的矩阵称为**单纯矩阵**), 也可以类似于正规矩阵定义 A 的谱分解. 设 A 有 s 个不同的特征值 $\lambda_1, \lambda_2, \cdots, \lambda_s$, 其重数分别为 $k_1, k_2, \cdots, k_s$. 设 $\alpha_{i1}, \alpha_{i2}, \cdots, \alpha_{i,k_i}$ 为对应于 λ_i 的线性无关的特征向量. 令

$$P = (\alpha_{11}, \alpha_{12}, \cdots, \alpha_{1,k_1}, \alpha_{21}, \alpha_{22}, \cdots, \alpha_{2,k_2}, \cdots, \alpha_{s1}, \alpha_{s2}, \cdots, \alpha_{s,k_s}),$$

则

$$P^{-1}AP = \operatorname{diag}(\overbrace{\lambda_1, \cdots, \lambda_1}^{k_1}, \overbrace{\lambda_2, \cdots, \lambda_2}^{k_2}, , \cdots, \overbrace{\lambda_s, \cdots, \lambda_s}^{k_s}),$$

即

$$A = P \operatorname{diag}(\overbrace{\lambda_1, \cdots, \lambda_1}^{k_1}, \overbrace{\lambda_2, \cdots, \lambda_2}^{k_2}, , \cdots, \overbrace{\lambda_s, \cdots, \lambda_s}^{k_s}) P^{-1}.$$

令

$$P^{-1} = \begin{pmatrix} \beta_{11}^T \\ \vdots \\ \beta_{1,k_1}^T \\ \vdots \\ \beta_{s1}^T \\ \vdots \\ \beta_{s,k_s}^T \end{pmatrix}.$$

则

$$A = \sum_{i=1}^{s} \lambda_i (\alpha_{i1}\beta_{i1}^T + \alpha_{i2}\beta_{i2}^T + \cdots + \alpha_{i,k_i}\beta_{i,k_i}^T) \tag{4.2.7}$$

令

$$P_i = \alpha_{i1}\beta_{i1}^T + \alpha_{i2}\beta_{i2}^T + \cdots + \alpha_{i,k_i}\beta_{i,k_i}^T, \quad i = 1, 2, \cdots, s. \tag{4.2.8}$$

则

$$A = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_s P_s. \tag{4.2.9}$$

公式 (4.2.9) 称为矩阵 A 的谱分解. 由于 $PP^{-1} = I$, 所以

$$\sum_{i=1}^{s} (\alpha_{i1}\beta_{i1}^T + \alpha_{i2}\beta_{i2}^T + \cdots + \alpha_{i,\,k_i}\beta_{i,\,k_i}^T) = I,$$

即 $\sum_{i=1}^{s} P_i = I$. 又因 $P^{-1}P = I$, 所以 $\beta_{ij}^T \alpha_{kl} = \delta_{ik}\delta_{jl}$. 因此 $P_i P_j = 0, \forall\, i \neq j$, 且 $(\alpha_{ij}\beta_{ij}^T)^2 = \alpha_{ij}\beta_{ij}^T$, 因此 $P_i^2 = P_i$. 由于

$$A^T = (P^{-1})^T \operatorname{diag}(\overbrace{\lambda_1, \cdots, \lambda_1}^{k_1}, \overbrace{\lambda_2, \cdots, \lambda_2}^{k_2}, , \cdots, \overbrace{\lambda_s, \cdots, \lambda_s}^{k_s}) P^T,$$

所以,

$$A^T (P^{-1})^T = (P^{-1})^T \operatorname{diag}(\overbrace{\lambda_1, \cdots, \lambda_1}^{k_1}, \overbrace{\lambda_2, \cdots, \lambda_2}^{k_2}, , \cdots, \overbrace{\lambda_s, \cdots, \lambda_s}^{k_s}),$$

由此可知 $(P^{-1})^T$ 的各列为 A^T 的对应于特征值 $\lambda_1, \cdots, \lambda_1, \lambda_2, \cdots, \lambda_2, \cdots, \lambda_s, \cdots, \lambda_s$ 的线性无关的特征向量. 而

$$(P^{-1})^T = (\beta_{11}, \cdots, \beta_{1,\,k_1}, \beta_{21}, \cdots, \beta_{2,\,k_2}, \cdots, \beta_{s1}, \cdots, \beta_{s,\,k_s}),$$

因此

$$A^T \beta_{ij} = \lambda_i \beta_{ij}, \quad \text{或} \quad \beta_{ij}^T A = \lambda_i \beta_{ij}^T, \quad i = 1, 2, \cdots, s; \;\; j = 1, 2, \cdots, k_i. \tag{4.2.10}$$

所以, β_{ij}^T 也常被称为矩阵 A 的**左特征向量**(相应地, α_{ij} 被称为 A 的**右特征向量**).

因 $\sum\limits_{i=1}^{s} P_i = I$, 所以

$$n = r(I) \leq \sum_{i=1}^{s} r(P_i) \leq k_1 + k_2 + \cdots + k_s = n,$$

其中 k_i 为 λ_i 的重数, $i = 1, 2, \cdots, s$, 而上式的第二个不等式由式 (4.2.8) 推出. 因此 $r(P_i) = k_i$. 综上所述, 我们得到如下结论 (唯一性的证明见习题 19):

定理 4.2.1 (谱分解定理) 设 A 为一个 n 阶可对角化矩阵, A 的谱为 $\sigma(A) = \{\lambda_1, \lambda_2, \cdots, \lambda_s\}$, 其中 λ_i 的重数为 k_i. 则存在唯一一组 s 个 n 阶方阵 $P_1, P_2, \cdots, P_s$ 满足:

$$(1)\;\; A = \sum_{i=1}^{s} \lambda_i P_i; \qquad (2)\;\; P_i^2 = P_i; \qquad (3)\;\; P_i P_j = 0 (\, i \neq j\,);$$

$$(4)\;\; \sum_{i=1}^{s} P_i = I; \qquad (5)\;\; r(P_i) = k_i.$$

这些矩阵 P_i 称为矩阵 A 的 (谱分解的) 成分矩阵或**主幂等矩阵**. 注意, 与正规矩阵相比, 一般矩阵的谱分解中的成分矩阵不一定是Hermite矩阵. 因此, $Ax = \lambda_1 P_1 x + \lambda_2 P_2 x + \cdots + \lambda_s P_s x$ 中的诸向量 $P_i x$ 未必是正交的.

例 4.2.6 求矩阵的谱分解:

$$A = \begin{pmatrix} -1 & 3 & -1 \\ -3 & 5 & -1 \\ -3 & 3 & 1 \end{pmatrix}.$$

解

$$|\lambda I - A| = \begin{vmatrix} \lambda+1 & -3 & 1 \\ 3 & \lambda-5 & 1 \\ 3 & -3 & \lambda-1 \end{vmatrix} = (\lambda-1)(\lambda-2)^2.$$

所以 A 有特征值 $\lambda_1 = 1$, $\lambda_2 = 2$(二重). 通过解齐次线性方程组 $(\lambda_i I - A)x = 0$, 可得对应于 λ_1, λ_2 的线性无关的特征向量分别为

$$\alpha_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}; \quad \alpha_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \alpha_3 = \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix}.$$

令 $P = (\alpha_1, \alpha_2, \alpha_3)$, 则可求出

$$P^{-1} = \begin{pmatrix} 3 & -3 & 1 \\ -3 & 4 & -1 \\ -1 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \beta_1^T \\ \beta_2^T \\ \beta_3^T \end{pmatrix}.$$

因此,

$$P_1 = \alpha_1 \beta_1^T = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} (3, -3, 1) = \begin{pmatrix} 3 & -3 & 1 \\ 3 & -3 & 1 \\ 3 & -3 & 1 \end{pmatrix},$$

$$\begin{aligned} P_2 = \alpha_2 \beta_2^T + \alpha_3 \beta_3^T &= \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} (-3, 4, -1) + \begin{pmatrix} -1 \\ 0 \\ 3 \end{pmatrix} (-1, 1, 0) \\ &= \begin{pmatrix} -3 & 4 & -1 \\ -3 & 4 & -1 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 0 \\ -3 & 3 & 0 \end{pmatrix} = \begin{pmatrix} -2 & 3 & -1 \\ -3 & 4 & -1 \\ -3 & 3 & 0 \end{pmatrix}. \end{aligned}$$

故 A 的谱分解为 $A = P_1 + 2P_2$. 注意, 矩阵 P_1 与 P_2 的第一列不正交, 故向量 $P_1 e_1$ 与 $P_2 e_1$ 不正交.

推论 4.2.1 设 $A = \sum_{i=1}^s \lambda_i P_i$ 是单纯矩阵 A 的谱分解, 则

$$A^m = \sum_{i=1}^s \lambda_i^m P_i \tag{4.2.11}$$

从而对任意多项式 $f(x)$ 有 $f(A) = \sum_{i=1}^s f(\lambda_i) P_i$.

例 4.2.7 设 A 如 例4.2.6, 求 $e^A=\sum_{n=0}^{\infty}\frac{1}{n!}A^n$.

解 利用 定理4.2.1 的 (2),(3) 和 (4), 我们有 $A^n=(P_1+2P_2)^n=P_1+2^nP_2, \forall n\geq 0$. 从而

$$e^A=\sum_{n=0}^{\infty}\frac{1}{n!}A^n=eP_1+e^2P_2.$$

我们将在第五章把 推论4.2.1 推广到一般函数, 从而利用谱分解给出一个求矩阵函数的简便方法.

思考题

1. 试讨论非正规矩阵的谱分解的几何意义.
2. 设单纯矩阵 A 仅有一个非零特征值 λ, 则 A 的谱分解是什么?
3. 两个 n 阶矩阵 A 与 B 何时满足条件 $AB=BA=0$?
4. 研究单纯矩阵的谱分解, 说明为什么不定义非单纯矩阵的谱分解.

第三节 矩阵的三角分解与 Cholesky 分解

我们知道如果线性方程组 $Ax=b$ 的系数矩阵 A 是可逆上 (下) 三角矩阵, 则容易利用反向 (顺向) 代入法求解. 一般地, 如果能将矩阵 A 分解成下三角矩阵与上三角矩阵之积, 即 $A=LU$, 其中 L 与 U 分别为下三角矩阵与上三角矩阵, 则令 $y=Ux$ 即可将原方程化为 $Ly=b$ 与 $Ux=y$ 两个简单的线性方程组. 上面的做法显然也适合线性矩阵方程 $AX=B$ 的求解. 因此需要研究任意矩阵是否有这样的分解. 注意, 如果这样的分解是存在的, 则可以将 L(或 U) 的对角元素均变为 1 (这样的矩阵称为单位三角矩阵, 证明见习题 21).

定义 4.3.1 设 A 是 n 阶矩阵, 如果存在上三角矩阵 U 与单位下三角矩阵 L 使得

$$A=LU \tag{4.3.1}$$

则称 A 有三角分解或 LU 分解[46], (4.3.1) 称为 A 的一个**三角分解** 或 **LU 分解**.

例 4.3.1 (三角分解未必存在) 可逆矩阵 $A=\begin{pmatrix}0&1\\1&0\end{pmatrix}$ 不存在三角分解, 证明见习题 22.

虽然三角分解未必存在, 但却有下面的唯一性定理.

定理 4.3.1 (LU 分解的唯一性) 设 A 是 n 阶可逆矩阵, 并且 A 有三角分解 $A=LU$. 则该分解是唯一的, 且 $|A|=|U|=u_{11}u_{22}\cdots u_{nn}$.

证 设 $A=LU=L'U'$ 是 A 的两个三角分解, 则由于 A 可逆, 故所有的矩阵均可逆, 而且 $L^{-1}L'=U(U')^{-1}$ 既是下三角矩阵又是上三角矩阵, 故是对角矩阵. 但 L^{-1} 与 L' 均是单位下三角矩阵, 故知对角矩阵 $L^{-1}L'$ 是单位矩阵. 因此 $L=L', U=U'$. □

下面的定理表明, 实正定矩阵一定存在三角分解, 而且还可以使两个三角矩阵互为转置.

[46]矩阵的 LU 分解由著名英国数学家、逻辑学家、密码专家、计算机先驱 Alan Mathison Turing (图灵) 于 1948 年提出.

定理 4.3.2 (Cholesky[47]分解) 实正定矩阵 A 必有三角分解 $A = LU$, 且存在唯一的对角元素均为正的下三角矩阵 G 使得 $A = GG^T$(此称为 **Cholesky 分解**), 矩阵 G 称为 **Cholesky 三角**.

证 对 A 的阶数 n 作归纳. 当 $n = 1$ 时定理显然成立. 因为 A 正定, 故 $a_{11} > 0$, 因此令

$$G_1 = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\frac{a_{21}}{a_{11}} & 1 & 0 & \cdots & 0 \\ -\frac{a_{31}}{a_{11}} & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \\ -\frac{a_{n1}}{a_{11}} & 0 & 0 & \cdots & 1 \end{pmatrix}$$

即可得

$$G_1 A G_1^T = \begin{pmatrix} a_{11} & 0 \\ 0 & A_1 \end{pmatrix},$$

其中 A_1 是 $n-1$ 阶正定矩阵 (为什么?). 由归纳假设即可知存在下三角矩阵 G_2 使得

$$G_2 A G_2^T = \mathrm{diag}\,(a_1, a_2, \cdots, a_n).$$

由于 A 正定, 从而诸 $a_i > 0$. 因此

$$A = G_2^{-1}\mathrm{diag}\,(a_1, \cdots, a_n) G_2^{-T} = G_2^{-1}\mathrm{diag}\,(\sqrt{a_1}, \cdots, \sqrt{a_n})\mathrm{diag}\,(\sqrt{a_1}, \cdots, \sqrt{a_n}) G_2^{-T},$$

故知 $A = GG^T$, 其中 $G = G_2^{-1}\mathrm{diag}\,(\sqrt{a_1}, \cdots, \sqrt{a_n})$ 是下三角矩阵.

唯一性的证明见习题 23. □

例 4.3.2 求正定矩阵 $A = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ 的 Cholesky 分解.

方法一. 由 定理 4.3.2 的证明可知, 可以利用三角矩阵将 A 化为正定的对角矩阵, 即令

$$G_1 = \begin{pmatrix} 1 & 0 \\ -\frac{1}{2} & 1 \end{pmatrix},$$

则

$$G_1 A G_1^T = \begin{pmatrix} 2 & 0 \\ 0 & \frac{3}{2} \end{pmatrix}.$$

因此

$$G = G_1^{-1} \begin{pmatrix} \sqrt{2} & 0 \\ 0 & \sqrt{\frac{3}{2}} \end{pmatrix} = \begin{pmatrix} \sqrt{2} & 0 \\ \frac{\sqrt{2}}{2} & \sqrt{\frac{3}{2}} \end{pmatrix}.$$

方法二. 设 G 为对角元素均为正的下三角矩阵, 直接比较 $A = GG^T$ 的两端可知

$$a_{11} = g_{11}^2, a_{12} = g_{11}g_{21}, a_{22} = g_{21}^2 + g_{22}^2,$$

于是

$$g_{11} = \sqrt{2}, g_{21} = \frac{\sqrt{2}}{2}, g_{22} = \sqrt{\frac{3}{2}}.$$

[47] André-Louis Cholesky(1875-1918), 法国大地测量学家, 在法军服役时死于战争.

上例中的方法二称为**Cholesky 算法**. 一般地, 设 A 为 n 阶正定矩阵, 比较 $A=GG^T$ 的两端可得

$$a_{ij}=\sum_{k=1}^{j} g_{ik}g_{jk}, i\geq j$$

因此

$$g_{jj}g_{ij}=a_{ij}-\sum_{k=1}^{j-1} g_{ik}g_{jk} \tag{4.3.2}$$

记上式右端为 $v(i)$, 则可由 (按列) 递推的办法求出诸 $v(i)$. 特别地, 在公式 (4.3.2) 中令 $i=j$ 可得 $g_{jj}^2=v(j)$, 因此

$$g_{ij}=\frac{v(i)}{g_{jj}}=\frac{v(i)}{\sqrt{v(j)}} \tag{4.3.3}$$

公式 (4.3.2) 与 (4.3.3) 合称为 Cholesky 算法或**平方根法** , 因为 Cholesky 三角矩阵 G 可以看作是 A 的平方根, 记作 $G=\sqrt{A}$.

例 4.3.3 设正定矩阵 A 的 Cholesky 分解为 $A=GG^T$, 则其逆矩阵 A^{-1} 可由下式求得

$$A^{-1}=G^{-T}G^{-1} \tag{4.3.4}$$

由 定理 4.3.2 的证明可以看出下面的定理成立, 证明见习题 25.

定理 4.3.3 设 n 阶矩阵 A 可逆. 则 A 存在三角分解 $\iff$ A 的所有顺序主子式均非 0. 此时, 唯一地存在一对单位下三角矩阵 L' 和单位上三角矩阵 U' 与对角矩阵 D, 使得 $A=L'DU'$, 其中 D 与 A 具有完全相同的顺序主子式.

实际上, 上述定理可以推广为

推论 4.3.1 设 n 阶矩阵 A 的前 $r(A)$ 个顺序主子式均非 0, 则 A 存在三角分解.

思考题

1. 如果一个矩阵有 LU 分解, 它是否一定有 UL(即上三角在左, 下三角在右) 分解?
2. 设一个矩阵既有 LU 分解也有 UL 分解, 试比较正定矩阵的这两种分解在计算上的差异?
3. 半正定矩阵有无类似的 Cholesky 分解? 负定矩阵和不定矩阵呢?
4. 如果去掉对角元素均为正的条件, 正定矩阵的 Cholesky 分解是否具有唯一性?
5. 可逆矩阵未必有三角分解. 能否设计一种方法以比较有三角分解的可逆矩阵与没有三角分解的可逆矩阵的数量?

第四节 矩阵的 QR 分解

我们在上一节看到, 即使可逆矩阵也可能不存在三角分解, 因此需要寻找其它类型的矩阵分解, 矩阵的正交三角分解即是一种对任何可逆矩阵均存在的理想分解, 其原理就是 Gram-Schmidt 的正交化方法.

定理 4.4.1 设 $A\in\mathbb{C}^{n\times n}$, 且 A 为满秩的, 则存在唯一的酉矩阵 U 和对角线元素都大于零的上三角矩阵 R 满足

$$A=UR. \tag{4.4.1}$$

证 设 α_j 为 A 的第 j 个列向量, $j=1,2,\cdots,n$. 则 $A=(\alpha_1,\alpha_2,\cdots,\alpha_n)$. 因 A 满秩, $\alpha_1,\alpha_2,\cdots,\alpha_n$ 线性无关. 我们知道由下述 Gram-Schmidt 正交化过程可以将 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 化成两两正交的向量 $\eta_1,\eta_2,\cdots,\eta_n$:

$$\begin{cases}\eta_1=\alpha_1,\\ \eta_2=\alpha_2-\dfrac{(\alpha_2,\eta_1)}{(\eta_1,\eta_1)}\eta_1,\\ \eta_3=\alpha_3-\dfrac{(\alpha_3,\eta_1)}{(\eta_1,\eta_1)}\eta_1-\dfrac{(\alpha_3,\eta_2)}{(\eta_2,\eta_2)}\eta_2,\\ \cdots\cdots\cdots\cdots\\ \eta_n=\alpha_n-\dfrac{(\alpha_n,\eta_1)}{(\eta_1,\eta_1)}\eta_1-\dfrac{(\alpha_n,\eta_2)}{(\eta_2,\eta_2)}\eta_2-\cdots-\dfrac{(\alpha_n,\eta_{n-1})}{(\eta_{n-1},\eta_{n-1})}\eta_{n-1}.\end{cases}$$

于是

$$\begin{cases}\alpha_1=\eta_1,\\ \alpha_2=\dfrac{(\alpha_2,\eta_1)}{(\eta_1,\eta_1)}\eta_1+\eta_2,\\ \alpha_3=\dfrac{(\alpha_3,\eta_1)}{(\eta_1,\eta_1)}\eta_1+\dfrac{(\alpha_3,\eta_2)}{(\eta_2,\eta_2)}\eta_2+\eta_3,\\ \cdots\cdots\cdots\cdots\\ \alpha_n=\dfrac{(\alpha_n,\eta_1)}{(\eta_1,\eta_1)}\eta_1+\dfrac{(\alpha_n,\eta_2)}{(\eta_2,\eta_2)}\eta_2+\cdots+\dfrac{(\alpha_n,\eta_{n-1})}{(\eta_{n-1},\eta_{n-1})}\eta_{n-1}+\eta_n.\end{cases}$$

令 $\beta_i=\frac{1}{\sqrt{(\eta_i,\eta_i)}}\eta_i$, 则 $U=(\beta_1,\beta_2,\cdots,\beta_n)$ 为酉阵, 且

$$\begin{cases}\alpha_1=\sqrt{(\eta_1,\eta_1)}\beta_1,\\ \alpha_2=(\alpha_2,\beta_1)\beta_1+\sqrt{(\eta_2,\eta_2)}\beta_2,\\ \alpha_3=(\alpha_3,\beta_1)\beta_1+(\alpha_3,\beta_2)\beta_2+\sqrt{(\eta_3,\eta_3)}\beta_3,\\ \cdots\cdots\cdots\cdots\\ \alpha_n=(\alpha_n,\beta_1)\beta_1+(\alpha_n,\beta_2)\beta_2+\cdots+(\alpha_n,\beta_{n-1})\beta_{n-1}+\sqrt{(\eta_n,\eta_n)}\beta_n.\end{cases}$$

因而 $A=UR$, 其中 $U=(\beta_1,\beta_2,\cdots,\beta_n)$, 以及

$$R=\begin{pmatrix}\sqrt{(\eta_1,\eta_1)} & (\alpha_2,\beta_1) & \cdots & (\alpha_{n-1},\beta_1) & (\alpha_n,\beta_1)\\ & \sqrt{(\eta_2,\eta_2)} & \cdots & (\alpha_{n-1},\beta_2) & (\alpha_n,\beta_2)\\ & & \ddots & \vdots & \vdots\\ & & & \sqrt{(\eta_{n-1},\eta_{n-1})} & (\alpha_n,\beta_{n-1})\\ & & & & \sqrt{(\eta_n,\eta_n)}\end{pmatrix}.$$

设 $A=U_1R_1$ 为另一分解, 其中 U_1 为酉矩阵, R_1 为对角线元素大于零的上三角矩阵. 则从 $UR=U_1R_1$ 得 $RR_1^{-1}=U^*U_1$ 为上三角的酉阵, 故为正规矩阵, 从而 RR_1^{-1} 为对角矩阵, 且对角线元素都大于零. 这样的酉阵必须是单位矩阵 (因酉阵的特征值都是模为 1 的). 因此 $U=U_1$, $R=R_1$, 唯一性得证. □

公式 (4.4.1) 称为矩阵 A 的**正交三角分解**, 也叫 UR 分解. 当 A 为实满秩矩阵时, 上面定理中的 U 为正交矩阵, 往往记为 Q, 而 R 为实上三角矩阵, 因此 UR 分解也称为**QR 分解** .

推论 4.4.1 设可逆矩阵 A 的正交三角分解为 $A=UR$, 则上三角矩阵 R 的行列式的模等于矩阵 A 的行列式的模.

实际上所有列满秩的矩阵均存在类似的正交三角分解, 即有下面的

定理 4.4.2 设 $A\in\mathbb{C}^{n\times r}$, 且 A 是列满秩的, 则

$$A=UR,$$

其中 $U\in\mathbb{C}^{n\times r}$ 的 r 个列向量构成一组标准正交向量组, $R\in\mathbb{C}^{r\times r}$ 为对角线元素大于零的上三角矩阵. 此分解是唯一的.

证 我们只证明唯一性, 存在性的证明与 定理 4.4.1 类似, 见习题 26. 设 $A=UR=U_1R_1$, 其中 U,U_1 是列向量为标准正交组的矩阵, R,R_1 是对角元素为正的上三角矩阵. 因为 $U_1^*U_1=I_r$, 故 $U_1^*U=R_1R^{-1}$, 容易验证 (请验证!)U_1^*U 是酉矩阵, 所以对角元素为正数的上三角矩阵 R_1R^{-1} 是单位矩阵! □

定理 4.4.2 可以作如下的变化, 即将 U 取为 n 阶酉矩阵, 而将 R 取为 $n\times r$ 阶准上三角矩阵, 即 $R=\begin{pmatrix} R_1 \\ 0 \end{pmatrix}$, 其中 r 阶上三角矩阵 R_1 具有正对角元素. 通常称此种分解为矩阵 A 的正交三角分解, 而将 定理 4.4.2 中的分解称为矩阵 A 的**薄 QR 分解** .

注意在 定理 4.4.2 的证明中我们利用了 $U^*U=I_r$, 实际上可以将酉矩阵的概念作下述推广, 即若矩阵 U 满足条件 $U^*U=I$, 则称 U 是列正交矩阵, 那么 定理 4.4.1 与 定理 4.4.2 就可以合起来写了. 容易证明 (见习题 27) 一个列正交矩阵必是列满秩的. 类似地, 可以定义行正交矩阵, 则酉矩阵既是列正交矩阵又是行正交矩阵. (列正交矩阵与行正交矩阵常统称为**半正交矩阵**.)

例 4.4.1 设列满秩矩阵 $A\in\mathbb{C}^{m\times n}$ 的正交三角分解为 $A=UR$, 则 $A^*A=R^*R$, 因此下三角矩阵 R^* 恰好是正定矩阵 A^*A 的 Cholesky 三角.

思考题

1. 可逆矩阵是否存在“三角正交分解”即“$A=RU$”, 其中 R,U 同正交三角分解? 又, 能否将上三角矩阵变为下三角矩阵?
2. 对行满秩矩阵如何定义正交三角分解?
3. 对不可逆矩阵能否定义类似的分解?
4. 由 $U^*U=I$ 是否可以推出 $UU^*=I$?

第五节 矩阵的奇异值分解与极分解

我们已经知道, 正规矩阵可以酉对角化, 因此其对应的线性变换具有优良的性质 (旋转伸缩再反转), 非正规矩阵当然不具有这样的性质, 但能否有类似的分解呢? 正规矩阵 A 的酉对角化导出的分解为 $A=UDU^*$, 注意其中两个酉矩阵互为共轭转置, 放弃此条是否就可以使任意矩阵具有同样的分解呢? 答案是肯定的, 这就是矩阵的奇异值分解, 是 Beltrami [48] 1873 年在研

[48] Eugenio Beltrami(1835-1900), 意大利数学家, 对非欧几何学, 数学物理, 力学, 大地测量学有重要贡献.

究双线性函数 $f(x,y)=x^*Ay$ 时发现的. 作坐标变换 $x=U\xi, y=V\eta$, 其中 U,V 均是酉矩阵, 则 $f(x,y)=\xi^*U^*AV\eta$, 令 $D=U^*AV$, 则可得 $A=UDV^*$. 下面的奇异值分解定理表明, 的确可以使 D 为对角矩阵.

设 $m\geq n$, 我们以符号 $\mathrm{diag}\,(d_1,d_2,\cdots,d_n)_{m\times n}$ 表示下面的 $m\times n$ 矩阵 (称为 $m\times n$ 阶对角矩阵)

$$\begin{pmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & d_3 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & d_n \\ 0 & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}.$$

类似地, 若 $m\leq n$, 则以符号 $\mathrm{diag}\,(d_1,d_2,\cdots,d_m)_{m\times n}$ 表示 $m\times n$ 阶对角矩阵.

定理 4.5.1 (奇异值分解定理) 设 $A=(a_{ij})\in\mathbb{C}^{m\times n}(m\geq n)$, 且 $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_r>0$, 则存在 m 阶和 n 阶的酉矩阵 U 与 V, 使得

$$A=UDV^* \tag{4.5.1}$$

其中 $D=\mathrm{diag}\,(\sigma_1,\sigma_2,\cdots,\sigma_r,0,\cdots,0)_{m\times n}$.

公式 (4.5.1) 称为矩阵 A 的**奇异值分解**, 简称为SVD, 而 $\sigma_1,\cdots,\sigma_r,0,\cdots,0$(共 n 个) 称为 A 的**奇异值**.

证 我们知道, 对任意矩阵 A, A^*A 总是半正定的, 因此可设 $\sigma(A^*A)=\{\sigma_1^2,\sigma_2^2,\cdots,\sigma_n^2\}$, 其中 $\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_r>0=\sigma_{r+1}=\cdots=\sigma_n$. 由于 A^*A 是正规矩阵, 故可设 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 是由 A^*A 的特征向量构成的 $\mathbb{C}^n$ 的一组标准正交基. 令

$$V_1=(\alpha_1,\alpha_2,\cdots,\alpha_r), V_2=(\alpha_{r+1},\cdots,\alpha_n), \Lambda=\mathrm{diag}\,(\sigma_1,\sigma_2,\cdots,\sigma_r).$$

则

$$A^*AV_1=V_1\Lambda^2.$$

因为 $V_1^*V_1=I_r$, 故

$$\Lambda^{-1}V_1^*A^*AV_1\Lambda^{-1}=I. \tag{4.5.2}$$

注意到 $A^*AV_2=0$ 故有 $AV_2=0$(为什么?). 记 $U_1=AV_1\Lambda^{-1}$, 则公式 (4.5.2) 就是 $U_1^*U_1=I$, 因此 U_1 的列向量可以扩充成一组标准正交基, 记此标准正交基构成的酉矩阵为 $U=(U_1,U_2)$. 令 $V=(V_1,V_2)$, 则

$$U^*AV=\begin{pmatrix} U_1^*AV_1 & U_1^*AV_2 \\ U_2^*AV_1 & U_2^*AV_2 \end{pmatrix}=\begin{pmatrix} \Lambda & 0 \\ U_2^*U_1\Lambda & 0 \end{pmatrix}=\begin{pmatrix} \Lambda & 0 \\ 0 & 0 \end{pmatrix}=D,$$

所以 $A=UDV^*$. □

显然, 对 $m<n$ 依然有完全相同的奇异值分解 (但零奇异值的个数有可能不同).

一般, 将矩阵 A 的最大奇异值与最小奇异值分别记为 $\sigma_{\max}(A)$ 与 $\sigma_{\min}(A)$ 或 σ_1 与 σ_n.

由奇异值分解定理的证明可知,

$$Av_i=\begin{cases} \sigma_iu_i, & 1\leq i\leq r \\ 0, & r+1\leq i\leq n \end{cases} \tag{4.5.3}$$

与

$$u_i^* A = \begin{cases} \sigma_i v_i^*, & 1 \le i \le r \\ 0, & r+1 \le i \le m \end{cases} \tag{4.5.4}$$

因此矩阵 V 与 U 的列向量分别称为矩阵 A 的右奇异向量和左奇异向量, 而 V 与 U 分别称为 A 的右奇异向量矩阵和左奇异向量矩阵.

奇异值分解的计算方法. 如果矩阵 A 的阶数较小, 则公式 (4.5.3) 与 (4.5.4) 实际上给出了计算奇异值分解的一种方法. 先求 A^*A 的一组标准正交特征向量 v_i, 然后计算 $\sigma_i u_i = Av_i, 1 \le i \le n$(由公式 (4.5.3) 与 (4.5.4) 可知这样的 u_i 必是 AA^* 的特征向量, 且满足公式 (4.5.3) 与 (4.5.4)), 其余向量 $u_i, i > n$ 可由Hermite矩阵的特征向量的正交性获得 (显然不唯一).

例 4.5.1 求矩阵的奇异值分解:

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

解 在实际计算中, 应该求 A^*A 和 AA^* 中阶数较小的矩阵的特征值. 因为

$$AA^* = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

故 AA^* 的特征多项式为

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

因此 AA^* 的特征值为 $\lambda_1 = 1$, $\lambda_2 = 3$. 它们相应的单位特征向量分别为

$$\beta_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \beta_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

则 A^*A 的三个特征值分别为 $\lambda_1 = 1$, $\lambda_2 = 3$, $\lambda_3 = 0$, 且属于它们的特征向量分别为

$$\alpha_1 = \frac{1}{\sqrt{\lambda_1}} A^* \beta_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \ \alpha_2 = \frac{1}{\sqrt{\lambda_2}} A^* \beta_2 = \frac{1}{\sqrt{6}} \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \ \alpha_3 = \frac{1}{\sqrt{3}} \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix},$$

其中 α_3 可由Hermite矩阵的属于不同特征值的特征向量彼此正交得到. 令

$$V = (\alpha_1, \alpha_2, \alpha_3), \quad U = (\beta_1, \beta_2), \quad D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3} & 0 \end{pmatrix},$$

则

$$A = UDV^* = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{3} & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{3}} \end{pmatrix}.$$

例 4.5.2 设 A 是正规矩阵, 则可以由 A 的谱分解导出 A 的一个奇异值分解. 设 $A = U^*\Lambda U$, 其中 $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$. 则存在对角酉矩阵 W 使得

$$D = \Lambda W^* = \operatorname{diag}(|\lambda_1|, |\lambda_2|, \cdots, |\lambda_n|),$$

于是得 $A = U^*\Lambda W^*WU = U^*DWU = U^*DV$, 其中矩阵 $V = WU$ 是酉矩阵 (为什么?), 即 A 的一个奇异值分解为 $A = U^*DV$.

例 4.5.3 (奇异值分解的几何意义) 将 $A_{m\times n} = UDV^*$ 看作是从 $\mathbb{C}^n$ 到 $\mathbb{C}^m$ 的线性变换, 则该线性变换首先 (在 $\mathbb{C}^n$ 内) 将向量 x 做一旋转而得到向量 V^*x, 然后再 (将 $\mathbb{C}^n$ 中的向量 V^*x) 沿前 $r = r(A)$ 个坐标做伸缩 (其余坐标变为 0) 而得到 ($\mathbb{C}^m$ 的) 向量 DV^*x, 最后再 (在 $\mathbb{C}^m$ 内) 做一旋转而得到向量 UDV^*x. 比较正规矩阵的酉对角化可知, 此处的两次旋转可能不是互逆的.

例 4.5.4 (奇异值分解的几何意义) 研究单位圆 $S^1 : x^2 + y^2 = 1$ 在矩阵 $A = \begin{pmatrix} 2 & 2 \\ -1 & 1 \end{pmatrix}$ 作用下的变化. 由于 $A^TA = \begin{pmatrix} 5 & 3 \\ 3 & 5 \end{pmatrix}$ 的特征值为 $2, 8$, 易得 A 的奇异值分解为

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

因此, 矩阵 A 在单位圆 S^1 上的作用被分解为三步: 第一步, 旋转, 这不会改变 S^1; 第二步, 伸缩, S^1 变为椭圆, 第三步, 再次旋转 (本次的旋转使原来的坐标轴互换).

将 A 的奇异值分解 (4.5.1) 展开可得

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^* \tag{4.5.5}$$

上式称为矩阵 A 的**并向量分解**或奇异值分解展开. 由于 $\sigma_j = 0$, $j > r$, 所以公式 (4.5.5) 可以写为

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^* = U_r D_r V_r^* \tag{4.5.6}$$

其中 U_r, V_r 分别是 U, V 的前 r 列构成的矩阵, 而 $D_r = \operatorname{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r)$. 上式称为矩阵 A 的**截尾奇异值分解**或**薄奇异值分解**. 矩阵的并向量分解与截尾奇异值分解具有重要的应用, 见本章第六节.

由奇异值分解定理立即可得 (证明见习题36并比较Sylvester降幂公式即第三章, 命题 3.1.1)

推论 4.5.1 设 $A = (a_{ij})_{m\times n}$ 的秩为 r, 则 AA^* 与 A^*A 有完全相同的非零特征值 (相同的按重数计算) $\sigma_1^2, \sigma_2^2, \cdots, \sigma_r^2$, 其中 $\sigma_1, \sigma_2, \cdots, \sigma_r$ 是 A 的非零奇异值. 因此, A 与 A^* 具有相同的奇异值.

例 4.5.5 设可逆矩阵 A 的奇异值分解为 $A = UDV^*$, 则其逆的的奇异值分解为 $A^{-1} = VD^{-1}U^*$. 因此, 若 A 的奇异值为 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$, 则 A^{-1} 的奇异值为 $1/\sigma_n \geq 1/\sigma_{n-1} \geq \cdots \geq 1/\sigma_1 > 0$. 矩阵计算中的重要概念 – 矩阵 A 的 (谱) **条件数**, 记为 $Cond(A)$, 可以表示为

$$Cond(A) = \sigma_1(A)/\sigma_n(A). \tag{4.5.7}$$

设 $A = U_1DV^*$ 是 A 的奇异值分解, 令 $P = U_1DU_1^*$, $U = U_1V^*$, 即可得到矩阵的另一种有趣分解 – 极分解.

定理 4.5.2 设 $A \in \mathbb{C}^{n\times n}$, 则存在酉矩阵 U 和唯一的半正定矩阵 P 使得

$$A = PU \tag{4.5.8}$$

上式称为矩阵 A 的**极分解**. 矩阵 P 与 U 分别称为 A 的Hermite因子与酉因子.

容易证明 (见习题37)A 的极分解中的半正定矩阵 $P = \sqrt{AA^*}$, 即满足条件 $P^2 = AA^*$ 的矩阵 (称为矩阵 AA^* 的平方根, 也常记为 $(AA^*)^{1/2}$). 当矩阵 A 可逆时, 极分解中的酉矩阵 U 也是唯一的.

例 4.5.6 (极分解的几何意义) 矩阵的极分解是仿照复数的极形式 $z = re^{\mathrm{i}\theta}$ 做出的, 因为若记 $r = |P|, e^{\mathrm{i}\theta} = |U|$, 则行列式 $|A| = re^{\mathrm{i}\theta}$ 恰好是复数 $|A|$ 的极分解. 由于半正定矩阵是正规矩阵, 故矩阵的极分解的几何意义是先旋转然后再沿着一组正交的方向做伸缩. 回忆复数的极分解的几何意义恰好是旋转角度 θ, 再伸缩 r 倍.

由一个矩阵的极分解可以判断其正规性, 即有下述命题 (证明见习题38).

命题 4.5.1 设 $A = PU$ 是矩阵 A 的极分解, 则 A 是正规矩阵 $\iff PU = UP$.

奇异值分解具有许多良好的性质, 我们将这些性质分成下面几个命题, 所有的证明均比较简单, 见习题40-43.

命题 4.5.2 (正规矩阵的奇异值分解) 正规矩阵的奇异值是其特征值的模. 特别地, 半正定矩阵的特征值与奇异值相同.

命题 4.5.3 (奇异值分解与矩阵的四个子空间) 设 $A = UDV^*$ 是 $m \times n$ 矩阵 A 的一个奇异值分解, $r = r(A)$, 则

(1) 酉矩阵 U 的前 r 列是 A 的列空间的一组标准正交基;
(2) 酉矩阵 V 的前 r 列是 A 的行空间的一组标准正交基;
(3) U 的后 $m - r$ 列是 A^* 的零空间的一组标准正交基;
(4) V 的后 $n - r$ 列是 A 的零空间的一组标准正交基.

命题 4.5.4 (奇异值与特征值) 设 λ 是 n 阶矩阵 A 的一个特征值, 则 $\sigma_{\max}(A) \geq |\lambda| \geq \sigma_{\min}(A)$. 换言之, 矩阵的最大奇异值与最小奇异值是其特征值的模的上下界.

例 4.5.7 设 $A=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$, 则 A 的特征值为 1, 但奇异值为 $\sqrt{(3\pm\sqrt{5})/2}$.

例 4.5.8 (奇异值与矩阵的迹) 设 $A\in\mathbb{C}^{m\times n}$, 则 $\mathrm{tr}(A^*A)=\sum_{i=1}^r\sigma_i^2$, 证明见习题 43. 请对照 Schur 不等式.

本节最后, 我们解释奇异值这个词的来历.

命题 4.5.5 (奇异值与奇异矩阵) 矩阵 A 列满秩 $\iff$ A 的奇异值均非 0. 特别地, 方阵 A 非奇异 $\iff$ A 的奇异值均非 0.

例 4.5.9 设 $A\in\mathbb{C}^{m\times n}$. 则 $A=0\iff A^*A=0$. 换言之, 方阵 $A=0\iff$ 它的奇异值均为 0.

矩阵的奇异值较之其特征值的一个优点是非零奇异值的个数恰好是该矩阵的秩, 而矩阵的非零特征值的个数一般比其秩小 (比如幂零矩阵无非零特征值). 因此常常利用此点来计算矩阵的秩.

思考题

1. 矩阵的奇异值分解不唯一, 但是否可以确定到某种程度?
2. 能否将极分解中的顺序改变? 即是否存在酉矩阵 U 和半正定矩阵 P 使得 $A=UP$?
3. 不是方阵的矩阵可否定义极分解? 唯一性如何?
4. 可否以满足条件 $B^2=A$ 的矩阵 B 来定义 $\sqrt{A}$? 更一般地, 可否以满足条件 $B^m=A$ 的矩阵 B 来定义 $A^{1/m}$?

第六节 应用: 最小二乘法, 图像压缩, 子空间的交

一. QR 分解的应用：最小二乘解与 QR 方法

设矩阵 A 的正交三角分解为 $A=UR$, 将此代入方程 $Ax=b$ 的正规化方程 $A^*Ax=A^*b$ 可得, $R^*Rx=R^*U^*b$, 即 $Rx=U^*b$, 所以正规化方程的解为 $x=R^{-1}U^*b$, 此即原方程的最小二乘解. 如果 $A=QR$ 是实矩阵, 则 $x=R^{-1}Q^Tb$.

例 4.6.1 用 QR 分解解线性方程组 $AX=b$, 其中

$$A=\begin{pmatrix}1 & 1 & 3\\ 2 & 1 & 2\\ 2 & -5 & 8\\ -1 & 3 & -7\end{pmatrix},\quad b=\begin{pmatrix}-1\\ 1\\ -11\\ 9\end{pmatrix}.$$

解 将 A 的三个列向量 $\alpha_1,\alpha_2,\alpha_3$ 正交化, 可得:

$$\begin{cases}\eta_1=\alpha_1=(1,2,2,-1)^T,\\ \eta_2=\alpha_2-\dfrac{(\alpha_2,\eta_1)}{(\eta_1,\eta_1)}\eta_1=(1,1,-5,3)^T-\dfrac{-10}{10}(1,2,2,-1)^T=(2,3,-3,2)^T,\\ \eta_3=\alpha_3-\dfrac{(\alpha_3,\eta_1)}{(\eta_1,\eta_1)}\eta_1-\dfrac{(\alpha_3,\eta_2)}{(\eta_2,\eta_2)}\eta_2\\ \quad=(3,2,8,-7)^T-\dfrac{30}{10}(1,2,2,-1)^T-\dfrac{-26}{26}(2,3,-3,2)^T=(2,-1,-1,-2)^T.\end{cases}$$

再单位化, 得

$$\beta_1=\frac{1}{\sqrt{10}}\begin{pmatrix}1\\2\\2\\-1\end{pmatrix},\quad \beta_2=\frac{1}{\sqrt{26}}\begin{pmatrix}2\\3\\-3\\2\end{pmatrix},\quad \beta_3=\frac{1}{\sqrt{10}}\begin{pmatrix}2\\-1\\-1\\-2\end{pmatrix}.$$

则 $Q=(\beta_1,\beta_2,\beta_3)$. 由于 $Q^TQ=I_3$, 所以由 $A=QR$ 可得

$$\begin{aligned}R=Q^TQR=Q^TA&=\begin{pmatrix}\frac{1}{\sqrt{10}}&\frac{2}{\sqrt{10}}&\frac{2}{\sqrt{10}}&-\frac{1}{\sqrt{10}}\\ \frac{2}{\sqrt{26}}&\frac{3}{\sqrt{26}}&-\frac{3}{\sqrt{26}}&\frac{2}{\sqrt{26}}\\ \frac{2}{\sqrt{10}}&-\frac{1}{\sqrt{10}}&-\frac{1}{\sqrt{10}}&-\frac{2}{\sqrt{10}}\end{pmatrix}\begin{pmatrix}1&1&3\\2&1&2\\2&-5&8\\-1&3&-7\end{pmatrix}\\ &=\begin{pmatrix}\sqrt{10}&-\sqrt{10}&3\sqrt{10}\\ &\sqrt{26}&-\sqrt{26}\\ &&\sqrt{10}\end{pmatrix}.\end{aligned}$$

所以

$$R^{-1}=\begin{pmatrix}\frac{1}{\sqrt{10}}&\frac{1}{\sqrt{26}}&-\frac{2}{\sqrt{10}}\\ &\frac{1}{\sqrt{26}}&\frac{1}{\sqrt{10}}\\ &&\frac{1}{\sqrt{10}}\end{pmatrix}.$$

因此

$$\begin{aligned}x=R^{-1}Q^Tb&=\begin{pmatrix}\frac{1}{\sqrt{10}}&\frac{1}{\sqrt{26}}&-\frac{2}{\sqrt{10}}\\ &\frac{1}{\sqrt{26}}&\frac{1}{\sqrt{10}}\\ &&\frac{1}{\sqrt{10}}\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{10}}&\frac{2}{\sqrt{10}}&\frac{2}{\sqrt{10}}&-\frac{1}{\sqrt{10}}\\ \frac{2}{\sqrt{26}}&\frac{3}{\sqrt{26}}&-\frac{3}{\sqrt{26}}&\frac{2}{\sqrt{26}}\\ \frac{2}{\sqrt{10}}&-\frac{1}{\sqrt{10}}&-\frac{1}{\sqrt{10}}&-\frac{2}{\sqrt{10}}\end{pmatrix}\begin{pmatrix}-1\\1\\-11\\9\end{pmatrix}\\ &=\begin{pmatrix}\frac{1}{\sqrt{10}}&\frac{1}{\sqrt{26}}&-\frac{2}{\sqrt{10}}\\ &\frac{1}{\sqrt{26}}&\frac{1}{\sqrt{10}}\\ &&\frac{1}{\sqrt{10}}\end{pmatrix}\begin{pmatrix}-2\sqrt{10}\\2\sqrt{26}\\-\sqrt{10}\end{pmatrix}=\begin{pmatrix}1\\1\\-1\end{pmatrix}.\end{aligned}$$

将 $x=(1,1-1)^T$ 代入原方程组成立, 所以它是原方程组的解.

例 4.6.2 方程组 $\begin{pmatrix}1\\1\end{pmatrix}x=\begin{pmatrix}1\\0\end{pmatrix}$ 显然无解, 但由 定理 4.4.2, 列满秩矩阵 $A=\begin{pmatrix}1\\1\end{pmatrix}$ 的正交三角分解为

$$A=\begin{pmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{pmatrix}\bullet\sqrt{2}=QR,$$

因此原方程两端同乘以 $R^{-1}Q^T=\frac{1}{\sqrt{2}}(\frac{1}{\sqrt{2}},\frac{1}{\sqrt{2}})$ 得

$$x=1.$$

显然这是原方程组的最小二乘解.

在数值分析中讨论的计算矩阵特征值的重要方法 $-QR$ 算法的基础是矩阵的 QR 分解. 此处我们仅作简要介绍.

设 $A = A_0 = Q_0R_0$ 是 n 阶矩阵 A 的 QR 分解. 归纳地定义 $A_{m+1} = R_mQ_m$. 如果 A 的特征值均不相同 (我们在第一章第三节即已思考过此问题, 答案是: 几乎所有的方阵都是这样的矩阵!), 则矩阵序列

$$A_0, A_1, \cdots, A_k, \cdots \tag{4.6.1}$$

收敛到一个上三角矩阵 R. 由于序列 (4.6.1) 中的每个矩阵均与 A 酉相似, 因此上三角矩阵 R 的对角元素就是 A 的全部特征值, 证明见习题 46.

二. 奇异值分解与图像压缩

人造卫星常常需要将一些照片发回地面控制中心. 大部分照片的规格是 512×512(像素), 即每幅照片实际上包含超过 260000 个数据. 因此将这些数据全部传输需要大量的计算并花费相当长的时间. 所以往往在传输之前必须对原始数据进行压缩. 以下我们简要叙述利用奇异值分解进行图像压缩的过程.

用 $n \times n$ 矩阵 A 表示要传输的原始数据. 设 $A = UDV^T$ 是 A 的一个奇异值分解, 其中对角矩阵 D 的对角元素 (即 A 的奇异值) 从大到小排列. 假定我们选择前 m 个大奇异值进行图像传输, 就是说仅传输奇异值 $\sigma_1, \cdots, \sigma_m$ 以及相对应的左右奇异向量 u_i 与 $v_i (1 \le i \le m)$, 则我们实际上传输了 $m + mn + mn = m(2n+1)$ 个数据, 而不是原来的 n^2 个数据. 比值 $n^2/(2mn+m)$ 称为图像的压缩比 (其倒数称为数据压缩率). 利用矩阵的截尾奇异值分解即可根据实际接收到的数据还原图像, 即

$$\hat{A} = \sum_{i=1}^{m} \sigma_i u_i v_i^T \tag{4.6.2}$$

显然, 较大的 m 可以获得保真度较高的还原数据, 较小的 m 可以获得较高的传输效率. 在实际应用时, 可以根据不同需要适当选择 m 以获得满意的还原数据. 比如, 如果卫星照片的第 51 个奇异值已经较小, 则可以选择 $m = 50$, 于是仅需要传输 $50(2 \times 512 + 1) = 51250$ 个数据, 图像压缩比为 5.115.

三. 奇异值分解与子空间的交

若干个子空间的和的基与维数较易求得, 但若干个子空间的交的维数与基的计算则往往比较困难 (尽管维数较小). 下面给出的办法即是利用矩阵的奇异值分解来求两个子空间的交的基与维数.

例 4.6.3 设 A, B 是两个同阶矩阵, 则它们的行空间的交可由级联矩阵 $C = \begin{pmatrix} A \\ B \end{pmatrix}$ 的奇异值分解求得. 设 $C = UDV^*$ 是一个奇异值分解, 将 U 做适当分块可得

$$\begin{pmatrix} A \\ B \end{pmatrix} = UDV^* = \begin{pmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{pmatrix} \begin{pmatrix} D_r & 0 \\ 0 & 0 \end{pmatrix} V^*,$$

现由 $U^* \begin{pmatrix} A \\ B \end{pmatrix} = DV^*$ 可知,

$$U_{12}^* A = -U_{22}^* B \tag{4.6.3}$$

按矩阵乘积的行结构可知, 上式左右两端的行分别属于 A 的行空间 $R(A^*)$ 与 B 的行空间 $R(B^*)$, 因此可以断言 (见习题 48) $R(A^*) \cap R(B^*) = R((U_{12}^* A)^*) = R(A^* U_{12})$, 从而 $R(A^*) \cap R(B^*)$ 的维数与基均可求出.

习　题　四

1. 判断下列矩阵能否酉对角化, 如能, 则求一个酉矩阵 U, 使 U^*AU 为对角形:

$$(1)\ A=\begin{pmatrix}-1 & \mathrm{i} & 0\\ -\mathrm{i} & 0 & -\mathrm{i}\\ 0 & \mathrm{i} & -1\end{pmatrix};\ (2)\ A=\begin{pmatrix}0 & \mathrm{i} & 1\\ -\mathrm{i} & 0 & 0\\ 1 & 0 & 0\end{pmatrix};\ (3)\ A=\begin{pmatrix}\mathrm{i} & \mathrm{i} & 0\\ \mathrm{i} & 1 & 0\\ 0 & \mathrm{i} & \mathrm{i}\end{pmatrix}.$$

2. 证明正规矩阵与其共轭转置具有相同的化零空间. 该结论一般地成立吗?

3. 证明两个正规矩阵相似的充要条件是特征多项式相同.

4. 详细证明 定理 4.1.2.

5. 设 A 是 n 阶正规矩阵, x 是任意复数. 证明
(1)$A-xI$ 也是正规矩阵;
(2) 对于任何向量 x, 向量 Ax 与 A^*x 的长度相同;
(3)A 的任一特征向量都是 A^* 的特征向量;
(4)A 的属于不同特征值的特征向量正交.

6. 设 A 是正规矩阵, 证明
(1) A 是 Hermite 矩阵 $\Longleftrightarrow$ A 的特征值全为实数;
(2) A 是酉阵 $\Longleftrightarrow$ A 的特征值的模都是 1;
(3) A 是幂等阵 $\Longleftrightarrow$ A 的特征值只能是 0 与 1;
(4) 若 A 的全部特征值为 $\lambda_1, \lambda_2, ..., \lambda_n$, 则 AA^* 与 A^*A 的全部特征值为 $|\lambda_1|^2, |\lambda_2|^2, ..., |\lambda_n|^2$. 此结论对非正规矩阵成立吗?

7. 设 A 是正规矩阵, 证明
(1) 若 A 是幂等阵, 则 A 是 Hermite 矩阵;
(2) 若 $A^3=A^2$, 则 $A^2=A$;
(3) 若 A 又是 Hermite 阵, 而且也是一个幂幺阵 (即 $A^k=I$), 则 A 是对合阵 (即 $A^2=I$).

8. 证明特征值的极大极小定理: 设 A 是 Hermite 矩阵, 其全部特征值为 $\lambda_1\le\lambda_2\le\cdots\le\lambda_n$, 则:

$$\lambda_k=\min_{\substack{w_i\in\mathbb{C}^n\\1\le i\le n-k}}\ \max_{\substack{0\ne x\perp w_i\\1\le i\le n-k}}\frac{x^*Ax}{x^*x}=\max_{\substack{w_i\in\mathbb{C}^n\\1\le i\le n-k}}\ \min_{\substack{0\ne x\perp w_i\\1\le i\le n-k}}\frac{x^*Ax}{x^*x}.$$

特别地,

$$\lambda_{\max}=\lambda_n=\max_{x\ne0,x\in\mathbb{C}^n}\frac{x^*Ax}{x^*x}=\max_{x^*x=1}x^*Ax,$$

$$\lambda_{\min}=\lambda_1=\min_{x\ne0,x\in\mathbb{C}^n}\frac{x^*Ax}{x^*x}=\min_{x^*x=1}x^*Ax.$$

9. 详细证明 定理 4.1.3.

10. 直接证明实对称矩阵与 (实) 正交矩阵可以酉对角化, 从而均为正规矩阵.

11. 证明 推论 4.1.2.

12. 证明 引理 4.1.2 的充分性.

13. (1) 证明 例 4.1.7 关于平面正规变换的结论;
(2) 计算 2 阶实正规矩阵 $\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}$ 将哪些正方形变为了矩形?
(3) 证明矩阵 $\begin{pmatrix}1 & 2\\ 0 & 3\end{pmatrix}$ 是非正规矩阵, 说明它不能将任何正方形变为矩形;
(4) 试给出 3 阶实正规矩阵的几何意义.

14. 设 P, Q 各为 m 阶及 n 阶方阵, 证明: 若 $m+n$ 阶方阵 $A=\begin{pmatrix}P & B\\ 0 & Q\end{pmatrix}$ 是酉矩阵, 则 P, Q 也酉矩阵, 且 B 是零矩阵.

15. 证明**Sylvester 惯性定律**, 即两个Hermite矩阵合同 $\Longleftrightarrow$ 它们具有相同的惯性指标, 即相同的正负特征值 (因此0特征值) 的个数.

16. 已知正交矩阵 $A = \frac{1}{3}\begin{pmatrix} 2 & 1 & -2 \\ 1 & 2 & 2 \\ 2 & -2 & 1 \end{pmatrix}$ 表示一个旋转, 求其旋转轴与旋转角.

17. 若 3×3 矩阵 S 表示一个反射, 则存在一个正交矩阵 C, 使得 $C^{-1}SC = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$. 当 $S = \frac{1}{9}\begin{pmatrix} 7 & 4 & -4 \\ 4 & 1 & 8 \\ -4 & 8 & 1 \end{pmatrix}$ 时, 求这样的矩阵 C.

18. 求习题1中所有正规矩阵的谱分解.

19. 证明谱分解定理 (定理4.2.1) 中的唯一性.

20. 证明 例4.2.5 并写出其实数形式.

21. 设 $A = LU$, 其中 L 与 U 分别为下三角矩阵与上三角矩阵, 证明存在单位下三角矩阵 L' 与上三角矩阵 U' 使得 $A = L'U'$.

22. 证明矩阵 $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ 不存在三角分解.

23. 证明Cholesky分解 (定理4.3.2) 的唯一性.

24. 试给出正定Hermite矩阵的Cholesky分解定理.

25. 证明 定理4.3.3.

26. 证明 定理4.4.2 中的存在性.

27. 设 $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix}$, $b = \begin{pmatrix} 12 \\ 6 \\ 8 \end{pmatrix}$.

(1) 求 $R(A)$ 的标准正交基;
(2) 写出 A 的QR分解;
(3) 求 $Ax = b$ 的最小二乘解;
(4) 证明 $u_1 = (0,1,0)^T$, $u_2 = (\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}})^T$, 也是 $R(A)$ 的标准正交基, 其中 $R(A)$ 为 A 的列空间.

28. 求下列矩阵的QR分解:

$$(1)\quad A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix};\quad (2)\quad A = \begin{pmatrix} 2 & 2 & 1 \\ 0 & 2 & 2 \\ 2 & 1 & 2 \end{pmatrix}; (3)\quad A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

29. 设 $A, B \in \mathbb{C}^{m\times n}$. 证明矩阵分解引理: $A^*A = B^*B \Longleftrightarrow$ 存在酉矩阵 U 使得 $B = UA$.

30. 计算28题中各矩阵的奇异值分解和相应的四个子空间.

31. 设 $A \in \mathbb{C}^{m\times n}$, 证明:

$$\sigma_{\min}(A) = \min\{(x^*A^*Ax)^{1/2}, x^*x = 1\}, \qquad \sigma_{\max}(A) = \max\{(x^*A^*Ax)^{1/2} : x^*x = 1\}.$$

32. 设变换 $\sigma : \sigma x = x - a(x,w)w, \forall x \in \mathbb{R}^n$, 其中 w 为长度为1的向量, 问 a 取何值时, σ 为正交变换? 如果 w 是任意向量, 你的结论又如何?

33. 设 $A \in \mathbb{C}^{m\times n}$ 的秩为 $r > 0$, A 的奇异值分解为 $A = U\,\mathrm{diag}(s_1, ..., s_r, 0, ..., 0)V^*$, 求矩阵 $B = \begin{pmatrix} A \\ A \end{pmatrix}$ 的奇异值分解.

34. 详细计算 例 4.5.4, 求出单位圆在 A 作用下的最终轨迹方程.

35. (1) 证明矩阵的极分解的唯一性;
(2) 计算 Jordan 块 $J_n(\lambda)$ 的极分解.

36. 证明 推论 4.5.1.

37. 证明任意 n 阶矩阵 A 均可表示成 $A = Pe^{\mathrm{i}H}$, 其中 P 是半正定矩阵, H 是 Hermite 矩阵. 研究这种分解的唯一性.

38. 证明 命题 4.5.1.

39. 试对任意矩阵定义其极分解, 并由此计算任意向量 $x \in \mathbb{C}^n$ 的极分解.

40. 证明 命题 4.5.2.

41. 证明 命题 4.5.3, 并由此计算 $A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 1 & -1 \end{pmatrix}$ 的四个子空间.

42. 证明 命题 4.5.4.

43. 证明 例 4.5.8.

44. 设 $x, y, \alpha, \beta \in \mathbb{C}^n$, 且 $x^*y = \alpha^*\beta = 0$. 设 $A = x\alpha^* + y\beta^*$, 求 A 的 F- 范数.

45. 证明矩阵 A 可以对角化 $\iff$ 存在 Hermite 正定矩阵 P 使得 $P^{-1}AP$ 是正规矩阵.

46. 证明矩阵序列 (4.6.1) 中的每一个矩阵均与 $A = A_0$ 酉相似, 并且当 A 的特征值均不相同时, 该序列收敛于一个与 A 酉相似的上三角矩阵. 如果 A 有重特征值, 此结论还成立吗?

47. 研究正交三角分解, 谱分解, 极分解和奇异值分解之间的关系.

48. 仔细研究 例 4.6.3 的计算与证明 (参考下题), 再求第二章第 9 题中的两个子空间的交.

49. 设 $A, B \in \mathbb{C}^{m\times n}$. 证明: $x \in N(A) \cap N(B) \iff \begin{pmatrix} A \\ B \end{pmatrix} x = 0$.

50. 证明奇异值的极大极小定理: 设 $A \in \mathbb{C}^{m\times n}$ 的奇异值为 $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$, 则:

$$\sigma_k = \min_{\substack{w_i \in \mathbb{C}^n \\ 1\leq i\leq n-k}} \max_{\substack{0\neq x\perp w_i \\ 1\leq i\leq n-k}} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\substack{w_i \in \mathbb{C}^n \\ 1\leq i\leq n-k}} \min_{\substack{0\neq x\perp w_i \\ 1\leq i\leq n-k}} \frac{\|Ax\|_2}{\|x\|_2}.$$

特别地,

$$\sigma_{\max} = \sigma_1 = \max_{x\neq 0, x\in\mathbb{C}^n} \frac{\|Ax\|_2}{\|x\|_2} = \max_{x^*x=1} \|Ax\|_2,$$

$$\sigma_{\min} = \sigma_n = \min_{x\neq 0, x\in\mathbb{C}^n} \frac{\|Ax\|_2}{\|x\|_2} = \min_{x^*x=1} \|Ax\|_2.$$

51. 证明: 对任意同阶矩阵 A, B 均有 $\sigma_{\max}(A+B) \leq \sigma_{\max}(A) + \sigma_{\max}(B)$.

52. (矩阵的低秩近似) 设矩阵 $A_{m\times n}$ 的秩为 r, 其奇异值分解为 $A = UDV^*, U = (u_1, \cdots, u_m), V = (v_1, \cdots, v_n)$. 对任意 $k < r$, 定义矩阵

$$A_k = \sum_{i=1}^{k} \sigma_i u_i v_i^*, k < r.$$

证明:

$$\min_{r(B)=k} \|A - B\|_1 = \|A - A_k\|_1 = \sigma_{k+1}, k < r$$

以及

$$\min_{r(B)=k} \|A - B\|_F^2 = \|A - A_k\|_F^2 = \sigma_{k+1}^2 + \sigma_{k+1}^2 + \cdots + \sigma_r^2.$$

53. 利用上题的结果, 证明利用截尾奇异值分解压缩数据的合理性.

54. (同时奇异值分解) 设 A, B 是两个 $m \times n$ 矩阵. 证明存在酉矩阵 U, V 以及非负对角矩阵 D, Λ 使得 $A = UDV^*$, $B = U\Lambda V^* \iff A^*B$ 与 AB^* 均是正规矩阵. 该结论对三个或更多的矩阵成立吗?