高等工程數學（9）

廈門大學
XIAMEN UNIVERSITY
信息学院
（国家示范性软件学院）
School of Informatics
黃 煒
博士，副教授
Dr. Wei Huang

方阵的酉相似化简

矩阵论（9）

厦門大學 XIAMEN UNIVERSITY
信息学院（国家示范性软件学院）School of Informatics
黄 炜 博士·副教授 Dr. Wei Huang

2.3 方阵的酉相似化简

在 2.1 节中，我们论证了复数域上任何方阵都相似于 Jordan 矩阵，它是特殊的上三角矩阵. 这说明复数域上每个方阵必可与上三角矩阵相似.

上(下)三角矩阵有一些重要的性质，例如，上(下)三角矩阵的主对角线元素是它的特征值；上(下)三角矩阵的逆矩阵仍是上(下)三角矩阵；上(下)三角矩阵的乘积仍是上(下)三角矩阵，等等. 特别是，求解系数矩阵为上(下)三角矩阵的线性方程组的算法简单，因此在数值计算中往往要化一个方阵为上(下)三角矩阵.

另一方面，酉(正交)矩阵比一般的可逆矩阵有显著的优点，例如，酉(正交)矩阵作用于一个复(实)向量不改变其长度；酉(正交)矩阵的逆矩阵是它的共轭转置(转置)矩阵；n 阶酉(正交)矩阵的列向量组构成酉空间 $\mathbf{C}^n$(欧氏空间 $\mathbf{R}^n$)的一个标准正交基，等等. 在第三章还将指出，酉(正交)矩阵的条件数小，从而在数值计算中有较好的数值稳定性，因此，一般都尽可能用酉(正交)矩阵解决有关的数值计算问题.

于是就提出了一些问题：复数域上一个方阵能否酉相似于上三角矩阵？怎样的方阵才能酉相似于对角矩阵？本节的任务是回答这些问题.

定理 2.3-1(Schur) 任一 n 阶方阵都酉相似于一个上三角矩阵，即存在酉矩阵 $\boldsymbol{U}$ 使

$$\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{R}, \tag{2.3-1}$$

其中 $\boldsymbol{R}$ 是 n 阶上三角矩阵，其主对角线元素是 $\boldsymbol{A}$ 的特征值.

2.3 方阵的酉相似化简

证 对方阵的阶数 n 用数学归纳法. 当 $n=1$ 时，定理显然成立. 假设定理对 $n-1$ 阶方阵成立，即每个 $n-1$ 阶方阵 $\boldsymbol{A}_1$，都存在 $n-1$ 阶酉矩阵 $\boldsymbol{U}_1$ 使 $\boldsymbol{U}_1^{-1}\boldsymbol{A}_1\boldsymbol{U}_1=\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}_1\boldsymbol{U}_1=\boldsymbol{R}_1$ 为上三角矩阵.

在复数域上 $\boldsymbol{A}$ 必有 n 个特征值，任取一个特征值 λ_1，属于 λ_1 的一个单位特征向量为 $\boldsymbol{\varepsilon}_1$，从而有 $\boldsymbol{A}\boldsymbol{\varepsilon}_1=\lambda_1\boldsymbol{\varepsilon}_1$. 以 $\boldsymbol{\varepsilon}_1$ 为第 1 列构造一个酉矩阵 $\boldsymbol{U}_2$，记 $\boldsymbol{U}_2$ 的其余各列为 $\boldsymbol{\varepsilon}_2,\cdots,\boldsymbol{\varepsilon}_n$，因而 $\boldsymbol{U}_2=[\boldsymbol{\varepsilon}_1 \quad \boldsymbol{\varepsilon}_2 \quad \cdots \quad \boldsymbol{\varepsilon}_n]$. 由于

$$\begin{aligned}\boldsymbol{A}\boldsymbol{U}_2&=[\boldsymbol{A}\boldsymbol{\varepsilon}_1 \quad \boldsymbol{A}\boldsymbol{\varepsilon}_2 \quad \cdots \quad \boldsymbol{A}\boldsymbol{\varepsilon}_n]=[\lambda_1\boldsymbol{\varepsilon}_1 \quad \boldsymbol{A}\boldsymbol{\varepsilon}_2 \quad \cdots \quad \boldsymbol{A}\boldsymbol{\varepsilon}_n]\\&=[\boldsymbol{\varepsilon}_1 \quad \boldsymbol{\varepsilon}_2 \quad \cdots \quad \boldsymbol{\varepsilon}_n]\begin{bmatrix}\lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & \boldsymbol{A}_1 & \\ 0 & & & \end{bmatrix},\end{aligned}$$

所以有
$$\boldsymbol{U}_2^{-1}\boldsymbol{A}\boldsymbol{U}_2=\boldsymbol{U}_2^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}_2=\begin{bmatrix}\lambda_1 & * & \cdots & * \\ 0 & & & \\ \vdots & & \boldsymbol{A}_1 & \\ 0 & & & \end{bmatrix},$$

其中 $\boldsymbol{A}_1$ 是 $n-1$ 阶方阵. 根据归纳假设，存在 $\boldsymbol{U}_1$ 使 $\boldsymbol{U}_1^{-1}\boldsymbol{A}_1\boldsymbol{U}_1=\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}_1\boldsymbol{U}_1=\boldsymbol{R}_1$ 为上三角矩阵.

2.3 方阵的酉相似化简

令

$$U_3=\left[\begin{array}{c:c}1 & \\ \hdashline & U_1\end{array}\right],$$

则 U_3 是 n 阶酉矩阵. 再令 $U=U_2U_3$，它是一个酉矩阵，且有

$$U_3^{-1}U_2^{-1}AU_2U_3=U_3^{H}U_2^{H}AU_2U_3$$

$$=\left[\begin{array}{c:c}1 & \\ \hdashline & U_1^{H}\end{array}\right]\left[\begin{array}{c:ccc}\lambda_1 & * & \cdots & * \\ \hdashline 0 & & & \\ \vdots & & A_1 & \\ 0 & & & \end{array}\right]\left[\begin{array}{c:c}1 & \\ \hdashline & U_1\end{array}\right]=\left[\begin{array}{c:ccc}\lambda_1 & *' & \cdots & *' \\ \hdashline 0 & & & \\ \vdots & & R_1 & \\ 0 & & & \end{array}\right]=R,$$

亦即 $U^{-1}AU=U^{H}AU=R$ 是上三角矩阵.

因 A 相似于 R，故它们有相同的特征值，所以 R 的主对角线元素是 A 的特征值. ∎

Schur 定理有许多应用，下面举两个例子说明.

2.3 方阵的酉相似化简

例 1 证明：若 n 阶方阵 $\boldsymbol{A}=[a_{ij}]$ 酉相似于 $\boldsymbol{B}=[b_{ij}]$，则 $\boldsymbol{A}$、$\boldsymbol{B}$ 诸元素之模的平方和相等，即有

$$\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2=\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij}|^2. \tag{2.3-2}$$

证 因 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 酉相似，故有酉矩阵 $\boldsymbol{U}$ 使

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{B},$$

从而 $(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U})^{\mathrm{H}}=\boldsymbol{B}^{\mathrm{H}}$，即 $\boldsymbol{U}^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{B}^{\mathrm{H}}$.

于是 $\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}\cdot\boldsymbol{U}^{-1}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{U}=\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}$.

这就是说，$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 与 $\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}$ 酉相似，从而 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 与 $\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}}$ 有相同的特征值. 因此 $\mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})=\mathrm{tr}(\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}})$.

由于 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的主对角线元素

$$(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})_{ii}=\sum_{j=1}^{n}a_{ij}\bar{a}_{ij}=\sum_{j=1}^{n}|a_{ij}|^2\ (i=1,2,\cdots,n),$$

故 $$\mathrm{tr}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})=\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2.$$

同理有 $$\mathrm{tr}(\boldsymbol{B}\boldsymbol{B}^{\mathrm{H}})=\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij}|^2.$$

于是(2.3-2)式成立. ∎

2.3 方阵的酉相似化简

例 2 证明：设 $\lambda_1,\cdots,\lambda_n$ 是 n 阶方阵 $\boldsymbol{A}$ 的特征值，则

$$\sum_{i=1}^{n}|\lambda_i|^2\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2. \tag{2.3-3}$$

证 由 Schur 定理，存在酉矩阵 $\boldsymbol{U}$，使 $\boldsymbol{U}^{-1}\boldsymbol{AU}=\boldsymbol{R}=[r_{ij}]$为上三角矩阵，且 $\boldsymbol{R}$ 的主对角线元素是 $\boldsymbol{A}$ 的特征值。又由例 1 知，$\boldsymbol{A}$ 的诸元素之模的平方和等于 $\boldsymbol{R}$ 的诸元素之模的平方和，因此有

$$\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2=\sum_{i=1}^{n}\sum_{j=1}^{n}|r_{ij}|^2\geqslant\sum_{i=1}^{n}|r_{ii}|^2=\sum_{i=1}^{n}|\lambda_i|^2.$$

(2.3-3)式称为 Schur 不等式，它在方阵的特征值整体估计中有应用。

定义 2.3-1 若方阵 $\boldsymbol{A}$ 满足 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$，则称 $\boldsymbol{A}$ 为**正规矩阵**。

容易验证，Hermite 矩阵、酉矩阵、反 Hermite 矩阵（即 $\boldsymbol{A}^{\mathrm{H}}=-\boldsymbol{A}$）等都是正规矩阵。实对称矩阵、正交矩阵、反对称的实矩阵（即 $\boldsymbol{A}$ 是实方阵，且满足 $\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{A}$）等也都是正规矩阵。

定理 2.3-2 一个方阵酉相似于对角矩阵的充要条件是它为正规矩阵。

证 不妨以 n 阶方阵 $\boldsymbol{A}$ 证明这个定理。

充分性：设 $\boldsymbol{A}$ 是正规矩阵，即 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$，由 Schur 定理，存在酉矩阵 $\boldsymbol{U}$ 使 $\boldsymbol{U}^{-1}\boldsymbol{AU}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AU}=\boldsymbol{R}$，这里 $\boldsymbol{R}$ 是一个上三角矩阵：

2.3 方阵的酉相似化简

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}.$$

于是，由 $A^HA=AA^H$ 得

$$RR^H=U^HAU\cdot U^HA^HU=U^HAA^HU=U^HA^HAU=U^HA^HU\cdot U^HAU=R^HR,$$

即

$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}\begin{bmatrix} \bar{r}_{11} & & & \\ \bar{r}_{12} & \bar{r}_{22} & & \\ \vdots & \vdots & \ddots & \\ \bar{r}_{1n} & \bar{r}_{2n} & \cdots & \bar{r}_{nn} \end{bmatrix}=\begin{bmatrix} \bar{r}_{11} & & & \\ \bar{r}_{12} & \bar{r}_{22} & & \\ \vdots & \vdots & \ddots & \\ \bar{r}_{1n} & \bar{r}_{2n} & \cdots & \bar{r}_{nn} \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ & r_{22} & \cdots & r_{2n} \\ & & \ddots & \vdots \\ & & & r_{nn} \end{bmatrix}.$$

由此可知当 $i\neq j$ 时有 $r_{ij}=0$. 因此 R 是一个对角矩阵，从而 A 酉相似于对角矩阵.

必要性：设 A 酉相似于对角矩阵，即存在酉矩阵 U 使 $U^{-1}AU=D$，这里 D 是对角矩阵. 因此，

$$A=UDU^{-1}=UDU^H,A^H=UD^HU^H.$$

于是，由 $DD^H=D^HD$ 得

$$AA^H=UDU^H\cdot UD^HU^H=UDD^HU^H=UD^HDU^H=UD^HU^H\cdot UDU^H=A^HA,$$

即 A 是正规矩阵. ∎

2.3 方阵的酉相似化简

这一定理表明，n 阶方阵 A 是正规矩阵，当且仅当 A 有 n 个两两正交的单位特征向量.

推论 n 阶方阵 A 是正规矩阵的充要条件是(2.3-3)式为等式，即

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

请读者自证此推论.

例 3 设 $A=\begin{bmatrix} 3 & 2 \\ -1 & 1 \end{bmatrix}$，求酉矩阵 U 使 $U^{-1}AU$ 为上三角矩阵.

解 A 的特征多项式是

$$|A-\lambda I| = \begin{vmatrix} 3-\lambda & 2 \\ -1 & 1-\lambda \end{vmatrix} = 5-4\lambda+\lambda^2,$$

它的一个特征值是 $\lambda_1 = 2+\mathrm{i}$（另一个特征值是 $2-\mathrm{i}$）.

对 $\lambda_1 = 2+\mathrm{i}$，解齐次线性方程组

$$(A-\lambda_1 I)x=0,$$

2.3 方阵的酉相似化简

即解
$$\begin{cases}(1-\mathrm{i})x_1+2x_2=0,\\ -1\cdot x_1-(1+\mathrm{i})x_2=0,\end{cases}$$
得它的一个解为
$$\boldsymbol{x}_1=\begin{bmatrix}1+\mathrm{i}\\ -1\end{bmatrix},$$
单位化后得
$$\boldsymbol{\varepsilon}_1=\frac{1}{\sqrt{3}}\begin{bmatrix}1+\mathrm{i}\\ -1\end{bmatrix}.$$

以 $\boldsymbol{\varepsilon}_1$ 为第 1 列的一个酉矩阵可取为
$$\boldsymbol{U}=\frac{1}{\sqrt{3}}\begin{bmatrix}1+\mathrm{i} & 1\\ -1 & 1-\mathrm{i}\end{bmatrix}.$$
于是
$$\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}=\frac{1}{\sqrt{3}}\begin{bmatrix}1-\mathrm{i} & -1\\ 1 & 1+\mathrm{i}\end{bmatrix}\begin{bmatrix}3 & 2\\ -1 & 1\end{bmatrix}\cdot\frac{1}{\sqrt{3}}\begin{bmatrix}1+\mathrm{i} & 1\\ -1 & 1-\mathrm{i}\end{bmatrix}=\begin{bmatrix}2+\mathrm{i} & 1-2\mathrm{i}\\ 0 & 2-\mathrm{i}\end{bmatrix}.$$ ■

2.3 方阵的酉相似化简

例 4 设 $\boldsymbol{A}$ 是 n 阶正规矩阵，证明

(1) 对于任一 n 维向量 $\boldsymbol{x}$，$\boldsymbol{Ax}$ 与 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}$ 的长度相等；

(2) $\boldsymbol{A}$ 的任一特征向量都是 $\boldsymbol{A}^{\mathrm{H}}$ 的特征向量.

证 (1)
$$\begin{aligned}|\boldsymbol{Ax}|^2 &= \langle \boldsymbol{Ax},\boldsymbol{Ax}\rangle=(\boldsymbol{Ax})^{\mathrm{T}}(\overline{\boldsymbol{Ax}})=(\boldsymbol{Ax})^{\mathrm{H}}\boldsymbol{Ax}=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{Ax}\\ &=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{AA}^{\mathrm{H}}\boldsymbol{x}.\end{aligned}$$

另一方面，
$$\begin{aligned}|\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}|^2 &= \langle \boldsymbol{A}^{\mathrm{H}}\boldsymbol{x},\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}\rangle=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x})^{\mathrm{T}}(\overline{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}})\\ &=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x})^{\mathrm{H}}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x})=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{AA}^{\mathrm{H}}\boldsymbol{x}.\end{aligned}$$

因此，$|\boldsymbol{Ax}|^2=|\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}|^2$，两边开方即得
$$|\boldsymbol{Ax}|=|\boldsymbol{A}^{\mathrm{H}}\boldsymbol{x}|.$$

(2) 设 $\boldsymbol{x}$ 是 $\boldsymbol{A}$ 的任一特征向量，对应于 $\boldsymbol{x}$ 的特征值是 λ_1，将 $\boldsymbol{x}$ 单位化后取为 $\boldsymbol{\varepsilon}_1$. 由于 $\boldsymbol{A}$ 是正规矩阵，所以存在以 $\boldsymbol{\varepsilon}_1$ 为第 1 列的酉矩阵 $\boldsymbol{U}=[\boldsymbol{\varepsilon}_1\quad \boldsymbol{\varepsilon}_2\quad \cdots\quad \boldsymbol{\varepsilon}_n]$，使
$$\boldsymbol{U}^{-1}\boldsymbol{AU}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AU}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n),$$
这里 $\lambda_i(i=1,2,\cdots,n)$是 $\boldsymbol{A}$ 的特征值. 从而

2.3 方阵的酉相似化简

$$A = U\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)U^H,$$

$$AU = U\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n).$$

另一方面，由于 $A^H = (U\mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)U^H)^H = U\mathrm{diag}(\bar{\lambda}_1, \bar{\lambda}_2, \cdots, \bar{\lambda}_n)U^H$，故

$$A^H U = U\mathrm{diag}(\bar{\lambda}_1, \bar{\lambda}_2, \cdots, \bar{\lambda}_n). \tag{2.3-4}$$

由(2.3-4)式得 $$A^H \varepsilon_1 = \bar{\lambda}_1 \varepsilon_1,$$

即 ε_1 是 A^H 的属于特征值 $\bar{\lambda}_1$ 的单位特征向量，从而 x 也是 A^H 的特征向量(为什么)． ∎

2.4 实方阵的正交相似化简

本节讨论实数域上实方阵的正交相似化简问题. 由于在实数域上，实方阵的特征值不一定存在，所以任一实方阵未必能正交相似于上三角矩阵. 但作为 Schur 定理的一个推论，我们有下述定理.

定理 2.4-1（Schur） 设 n 阶实方阵 $\boldsymbol{A}$ 的 n 个特征值均是实数，则 $\boldsymbol{A}$ 正交相似于一个上三角实矩阵. ■

请读者自证.

例 1 设 $\boldsymbol{A}=\begin{bmatrix}1 & 1 & -1\\ 1 & 1 & 1\\ 0 & -1 & 2\end{bmatrix}$，验证定理 2.4-1.

解 由于 $\boldsymbol{A}$ 的特征多项式是

$$\det(\boldsymbol{A}-\lambda\boldsymbol{I})=(2-\lambda)(1-\lambda)^2,$$

故 $\boldsymbol{A}$ 的特征值都是实数.

$\boldsymbol{A}$ 的属于特征值 2 的一个单位特征向量是 $\boldsymbol{\varepsilon}_1=\left[\frac{\sqrt{2}}{2},0,-\frac{\sqrt{2}}{2}\right]^{\mathrm{T}}$，以 $\boldsymbol{\varepsilon}_1$ 为第 1 列的一个正交矩阵可取为

2.4 实方阵的正交相似化简

$$Q=[\boldsymbol{\varepsilon}_1 \quad \boldsymbol{\varepsilon}_2 \quad \boldsymbol{\varepsilon}_3]=\begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{bmatrix}.$$

于是

$$AQ=\begin{bmatrix} 1 & 1 & -1 \\ 1 & 1 & 1 \\ 0 & -1 & 2 \end{bmatrix}\begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{bmatrix}=\begin{bmatrix} \sqrt{2} & 1 & 0 \\ 0 & 1 & \sqrt{2} \\ -\sqrt{2} & -1 & \sqrt{2} \end{bmatrix}$$

$$=\begin{bmatrix} \frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ -\frac{\sqrt{2}}{2} & 0 & \frac{\sqrt{2}}{2} \end{bmatrix}\begin{bmatrix} 2 & \sqrt{2} & -1 \\ 0 & 1 & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}=Q\begin{bmatrix} 2 & \sqrt{2} & -1 \\ 0 & 1 & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix}.$$

若记

$$R=\begin{bmatrix} 2 & \sqrt{2} & -1 \\ 0 & 1 & \sqrt{2} \\ 0 & 0 & 1 \end{bmatrix},$$

则 $Q^{-1}AQ=Q^{T}AQ=R$，且 R 是上三角实矩阵. ∎

2.4 实方阵的正交相似化简

定义 2.4-1 n 阶方阵 $\boldsymbol{A}$ 称为上 Hessenberg 矩阵，如果 $i>j+1$ 时 $\boldsymbol{A}$ 的元素满足 $a_{ij}=0$，$j=1,2,\cdots,n-1$. 即形如

$$\begin{bmatrix} * & * & * & \cdots & * & * \\ * & * & * & \cdots & * & * \\ & * & * & \ddots & * & * \\ & & * & \ddots & \ddots & \vdots \\ & & & \ddots & * & * \\ & & & & * & * \end{bmatrix} \tag{2.4-1}$$

的方阵称为上 Hessenberg 矩阵. 类似可定义下 Hessenberg 矩阵.

例如

$$\begin{bmatrix} 1 & 2 & -1 \\ 3 & -1 & 4 \\ 0 & 1 & 2 \end{bmatrix},\quad \begin{bmatrix} 3 & 2 & 1 & 0 \\ -1 & 1 & 2 & -1 \\ 0 & 4 & -1 & 1 \\ 0 & 0 & 2 & 3 \end{bmatrix},\quad \begin{bmatrix} -1 & 2 & 3 & 1 & -1 \\ 1 & -1 & 4 & -7 & 0 \\ 0 & 1 & 2 & -1 & 2 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 2 & 3 \end{bmatrix}$$

都是上 Hessenberg 矩阵.

上 Hessenberg 矩阵与上三角矩阵的差别仅在于主对角线下面的次对角线（称它为下次对角线）元素可能不全是零.

上 Hessenberg 矩阵具有下述性质：

2.4 实方阵的正交相似化简

1）上 Hessenberg 矩阵与上三角矩阵的乘积是上 Hessenberg 矩阵(作为习题请读者证明)；

2）若上 Hessenberg 矩阵的下次对角线上出现零元素，则可按此零元素所在行的上方划横线、所在列的右边划纵线得到分块矩阵，此分块矩阵的两个对角子块都是上 Hessenberg 矩阵.

例如，对上面所列出的最后一个上 Hessenberg 矩阵来说，其下次对角线上出现 0，从而可将分块为

$$\left[\begin{array}{ccc:cc} -1 & 2 & 3 & 1 & -1 \\ 1 & -1 & 4 & -7 & 0 \\ 0 & 1 & 2 & -1 & 2 \\ \hdashline 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 2 & 3 \end{array}\right].$$

该分块矩阵的两个对角子块分别为

$$\begin{bmatrix} -1 & 2 & 3 \\ 1 & -1 & 4 \\ 0 & 1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & -1 \\ 2 & 3 \end{bmatrix},$$

它们显然分别是 3 阶与 2 阶的上 Hessenberg 矩阵.

2.4 实方阵的正交相似化简

下面引入一个非常重要的正交线性变换及其变换矩阵，即 Householder 变换和 Householder 矩阵.

如图 2.4-1，在平面上给定一单位向量 $\boldsymbol{\omega}$，以及与它正交且过原点的直线 l，把任一向量 $\boldsymbol{\alpha}$ 变换为与 l 对称的向量 $\boldsymbol{\beta}$ 的变换称为关于 l 的 Householder 变换.

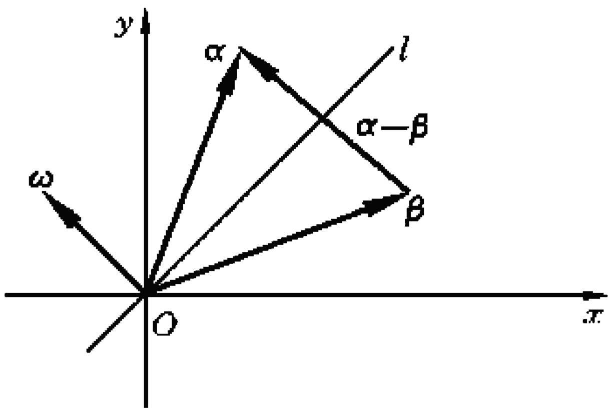

图 2.4-1

由图 2.4-1 可知，

$$\boldsymbol{\alpha}-\boldsymbol{\beta}=2(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\alpha})\boldsymbol{\omega}=2\boldsymbol{\omega}(\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\alpha})=2(\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})\boldsymbol{\alpha},$$

因此 $\boldsymbol{\beta}=\boldsymbol{\alpha}-2(\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})\boldsymbol{\alpha}=(\boldsymbol{I}-2\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})\boldsymbol{\alpha}.$

（在上述推导中，利用了 $\boldsymbol{\omega}\cdot\boldsymbol{\alpha}=\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\alpha}$ 和矩阵乘法满足结合律，行向量、列向量分别看成特殊的一行、一列矩阵.）

一般地，在欧氏空间 $\mathbf{R}^n$ 中，定义 **Householder 变换** $\mathscr{H}_\omega$ 为

$$\mathscr{H}_\omega\boldsymbol{\alpha}=(\boldsymbol{I}-2\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})\boldsymbol{\alpha},\quad \forall\,\boldsymbol{\alpha}\in\mathbf{R}^n. \tag{2.4-2}$$

Householder 变换 $\mathscr{H}_\omega$ 的变换矩阵 $\boldsymbol{I}-2\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}$ 称为 **Householder 矩阵**，记作 $\boldsymbol{H}$（当要强调 $\boldsymbol{\omega}$ 时，记为 $\boldsymbol{H}_\omega$），即

$$\boldsymbol{H}=\boldsymbol{I}-2\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}, \tag{2.4-3}$$

2.4 实方阵的正交相似化简

它是 n 阶对称正交矩阵，事实上，

$$\boldsymbol{H}^{\mathrm{T}}=(\boldsymbol{I}-2\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})^{\mathrm{T}}=\boldsymbol{I}^{\mathrm{T}}-2(\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})^{\mathrm{T}}=\boldsymbol{I}-2\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}=\boldsymbol{H};$$

$$\begin{aligned}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}&=\boldsymbol{H}^2=(\boldsymbol{I}-2\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})(\boldsymbol{I}-2\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})=\boldsymbol{I}-4\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}+4(\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})(\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}})\\&=\boldsymbol{I}-4\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}+4\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}=\boldsymbol{I}.\end{aligned}$$

因此，Householder 变换 $\mathscr{H}_\omega$ 是正交线性变换，它把 $\mathbf{R}^n$ 中任一向量 $\boldsymbol{\alpha}$ 变换为与某个 $n-1$ 维超平面(其指向 $\boldsymbol{\alpha}$ 一侧的单位法向量为 $\boldsymbol{\omega}$)对称的向量 $\boldsymbol{\beta}=\mathscr{H}_\omega\boldsymbol{\alpha}$，从而 $|\boldsymbol{\beta}|=|\boldsymbol{\alpha}|$.

反之，对于任意两个非零向量 $\boldsymbol{x},\boldsymbol{y}\in\mathbf{R}^n$，只要 $|\boldsymbol{x}|=|\boldsymbol{y}|$，则一定存在 Householder 矩阵 $\boldsymbol{H}$ 使 $\boldsymbol{y}=\boldsymbol{H}\boldsymbol{x}$，即存在 Householder 变换 $\mathscr{H}$ 使 $\boldsymbol{y}=\mathscr{H}\boldsymbol{x}$，其变换矩阵是 $\boldsymbol{H}$. 事实上，若 $\boldsymbol{y}=\boldsymbol{x}$，则取 $\boldsymbol{H}=\boldsymbol{I}$ 即可. 若 $\boldsymbol{y}\neq\boldsymbol{x}$，那么由 $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$ 得

$$(\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x}+\boldsymbol{y})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}+\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}=0,$$

因此 $\boldsymbol{x}-\boldsymbol{y}$ 就是 Householder 变换所需的指向 $\boldsymbol{x}$ 一侧的法向量，将其单位化便得 $\boldsymbol{\omega}=\dfrac{\boldsymbol{x}-\boldsymbol{y}}{|\boldsymbol{x}-\boldsymbol{y}|}$. 将它代入(2.4-3)式，得到 Householder 矩阵

$$\boldsymbol{H}=\boldsymbol{I}-\frac{2(\boldsymbol{x}-\boldsymbol{y})(\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}}{|\boldsymbol{x}-\boldsymbol{y}|^2}=\boldsymbol{I}-2\,\frac{(\boldsymbol{x}-\boldsymbol{y})(\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}}{(\boldsymbol{x}-\boldsymbol{y})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y})}. \tag{2.4-4}$$

2.4 实方阵的正交相似化简

例 2 用 Householder 变换将向量 $\boldsymbol{x}=[-2,2,1]^{\mathrm{T}}$ 变换为与 $\boldsymbol{\varepsilon}_1=[1,0,0]^{\mathrm{T}}$ 同方向的向量，求其变换矩阵 $\boldsymbol{H}$.

解 由于 $|\boldsymbol{x}|=3$，故由题意知，要求 Householder 矩阵 $\boldsymbol{H}$ 使 $\boldsymbol{Hx}=3[1,0,0]^{\mathrm{T}}$.

令 $\boldsymbol{y}=[3,0,0]^{\mathrm{T}}$，则 $\boldsymbol{x}-\boldsymbol{y}=[-5,2,1]^{\mathrm{T}}$. 将它代入(2.4-4)式得

$$\boldsymbol{H}=\boldsymbol{I}-\frac{2}{30}\begin{bmatrix}-5\\2\\1\end{bmatrix}\begin{bmatrix}-5 & 2 & 1\end{bmatrix}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}-\frac{1}{15}\begin{bmatrix}25&-10&-5\\-10&4&2\\-5&2&1\end{bmatrix}$$

$$=\begin{bmatrix}-\frac{2}{3}&\frac{2}{3}&\frac{1}{3}\\\frac{2}{3}&\frac{11}{15}&-\frac{2}{15}\\\frac{1}{3}&-\frac{2}{15}&\frac{14}{15}\end{bmatrix}. \qquad \blacksquare$$

定理 2.4-2 任一 n 阶实方阵都正交相似于一个 n 阶上 Hessenberg 实矩阵.

证 由于 Householder 矩阵是对称正交矩阵，在数值计算中具有重要作用，所以我们用 Householder 矩阵作为变换矩阵.

2.4　实方阵的正交相似化简

设 A 的分块形式是

$$A=\left[\begin{array}{c:c} a_{11} & \boldsymbol{\gamma}^{\mathrm{T}} \\ \hdashline \boldsymbol{\beta} & \boldsymbol{A}_1 \end{array}\right].$$

其中 $n-1$ 维向量 $\boldsymbol{\beta}=[a_{21}\quad a_{31}\quad\cdots\quad a_{n1}]^{\mathrm{T}}$，$\boldsymbol{\gamma}^{\mathrm{T}}=[a_{12}\quad a_{13}\quad\cdots\quad a_{1n}]$，而 $\boldsymbol{A}_1$ 是 $n-1$ 阶方阵. 不妨设 $\boldsymbol{\beta}\neq\mathbf{0}$，否则不必作变换. 利用 Householder 变换可以把向量 $\boldsymbol{\beta}$ 变换为向量 $\boldsymbol{\alpha}=[\sigma|\boldsymbol{\beta}|\quad 0\quad\cdots\quad 0]^{\mathrm{T}}$，亦即存在 $n-1$ 阶 Householder 矩阵 $\boldsymbol{Q}_2$ 使

$$\boldsymbol{Q}_2\boldsymbol{\beta}=[\sigma\,|\boldsymbol{\beta}|\quad 0\quad\cdots\quad 0]^{\mathrm{T}},$$

其中 σ 是符号函数，取 $+1$ 或 -1（这取决于 a_{21} 的正负，主要是为了数值计算中避免损失有效数字）.

令
$$\boldsymbol{Q}_1=\left[\begin{array}{c:c} 1 & \\ \hdashline & \boldsymbol{Q}_2 \end{array}\right],$$

则 $\boldsymbol{Q}_1$ 是 n 阶对称正交矩阵（即 Householder 矩阵），且有

2.4 实方阵的正交相似化简

$$Q_1AQ_1^T=\left[\begin{array}{c:c}1 & \\ \hdashline & Q_2\end{array}\right]\left[\begin{array}{c:c}a_{11} & \gamma^T \\ \hdashline \beta & A_1\end{array}\right]\left[\begin{array}{c:c}1 & \\ \hdashline & Q_2^T\end{array}\right]=\left[\begin{array}{c:c}a_{11} & \gamma^TQ_2^T \\ \hdashline Q_2\beta & Q_2A_1Q_2^T\end{array}\right]=\left[\begin{array}{c:c}a_{11} & \gamma^TQ_2^T \\ \hdashline \sigma\ |\beta| & \\ 0 & Q_2A_1Q_2^T \\ \vdots & \\ 0 & \end{array}\right].$$

再对 $n-1$ 阶方阵 $A_2=Q_2A_1Q_2^T$ 继续上述过程，至多进行 $n-2$ 次 Householder 变换，便可将 A 变换为上 Hessenberg 矩阵. 这就是说，存在正交矩阵 Q（它可取为一系列 Householder 矩阵的乘积）使 $QAQ^{-1}=QAQ^T$ 为上 Hessenberg 矩阵，即 A 正交相似于一个上 Hessenberg 矩阵. ∎

从上述证明过程可以看出，若 A 是对称矩阵，则 $\gamma=\beta$，从而 $\gamma^TQ_2^T=(Q_2\gamma)^T=(Q_2\beta)^T=[\sigma|\beta|\quad 0\quad \cdots\quad 0]$，且 $Q_2A_1Q_2^T$ 也是对称矩阵，因此实对称矩阵正交相似于三对角线矩阵.

2.4 实方阵的正交相似化简

例 3 求正交矩阵 Q 使 QAQ^{-1} 为三对角线矩阵，其中

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

解 由于 A 是实对称矩阵，故存在正交矩阵 Q 使 $QAQ^{-1}=QAQ^{\mathrm{T}}$ 为三对角线矩阵.

将向量 $x=[1,0,1]^{\mathrm{T}}$ 变换为 $y=[1,1,0]^{\mathrm{T}}$ 的变换矩阵(它是 Householder 矩阵)，按公式(2.4-4)计算是

$$H=I-\frac{2(x-y)(x-y)^{\mathrm{T}}}{(x-y)^{\mathrm{T}}(x-y)}=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}-\begin{bmatrix}0\\-1\\1\end{bmatrix}\begin{bmatrix}0&-1&1\end{bmatrix}$$

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

于是，令 $Q=H$，则有

$$QAQ^{-1}=QAQ^{\mathrm{H}}=HAH^{\mathrm{T}}$$

$$=\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}\begin{bmatrix}1&0&1\\0&2&-1\\1&-1&1\end{bmatrix}\begin{bmatrix}1&0&0\\0&0&1\\0&1&0\end{bmatrix}=\begin{bmatrix}1&1&0\\1&1&-1\\0&-1&2\end{bmatrix}.$$

∎

2.4 实方阵的正交相似化简

值得注意的是，两个同阶的 Householder 矩阵的乘积一般不是 Householder 矩阵. 事实上，设 $\boldsymbol{H}_1$、$\boldsymbol{H}_2$ 是两个 n 阶 Householder 矩阵，$\boldsymbol{H}_1\boldsymbol{H}_2$ 当然是正交矩阵，但由于

$$(\boldsymbol{H}_1\boldsymbol{H}_2)^{\mathrm{T}} = \boldsymbol{H}_2^{\mathrm{T}}\boldsymbol{H}_1^{\mathrm{T}} = \boldsymbol{H}_2\boldsymbol{H}_1 \neq \boldsymbol{H}_1\boldsymbol{H}_2$$

所以它不是 Householder 矩阵. 仅当 $\boldsymbol{H}_1$ 与 $\boldsymbol{H}_2$ 可交换，即 $\boldsymbol{H}_1\boldsymbol{H}_2=\boldsymbol{H}_2\boldsymbol{H}_1$ 时，$\boldsymbol{H}_1\boldsymbol{H}_2$ 才是 Householder 矩阵.

作业（第1部分）

习题 2.3

1. 验证 $\boldsymbol{A}=\begin{bmatrix} -\sqrt{2}i & -4 \\ 4 & \sqrt{2}i \end{bmatrix}$ 是正规矩阵，并求酉矩阵 $\boldsymbol{U}$ 使 $\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}$ 为对角矩阵.

向量范数和矩阵范数

矩阵论（9）

厦門大學
XIAMEN UNIVERSITY
信息学院
（国家示范性软件学院）
School of Informatics
黄 煒
博士·副教授
Dr. Wei Huang

3 向量范数和矩阵范数

范数在数学上是一个重要概念，它从整体上刻划一个量的大小. 例如，在$[a,b]$上的实函数$f(x)$，如何表示这个函数的大小呢？由于$f(x)$在$[a,b]$上的不同点有不同的值，所以要求从整体上来表示其大小，而不是用局部的点来说这个函数的大小. 类似地，对于向量和矩数，要说明它们的大小，也要从整体着眼，而不是用局部的分量和元素. 简单地说，向量范数、矩阵范数是描述向量、矩阵"大小"的一种度量，正如内积空间中向量的长度可以整体表示向量的大小一样. 但范数是比长度更广泛的概念.

向量范数和矩阵范数在数值代数和矩阵分析中起着重要作用.

3.1 向量范数

定义 3.1-1 设 V 是数域 F 上的向量空间，对 V 中任一向量 $\boldsymbol{\alpha}$，都有一个实数 $\|\boldsymbol{\alpha}\|$ 与之对应，且满足下列三个条件：

(1) 正定性：$\|\boldsymbol{\alpha}\| \geqslant 0$，当且仅当 $\boldsymbol{\alpha}=\mathbf{0}$ 时才有 $\|\boldsymbol{\alpha}\|=0$；

(2) 齐次性：$\|k\boldsymbol{\alpha}\|=|k|\cdot\|\boldsymbol{\alpha}\|$，$k\in F$；

(3) 三角不等式：$\|\boldsymbol{\alpha}+\boldsymbol{\beta}\| \leqslant \|\boldsymbol{\alpha}\|+\|\boldsymbol{\beta}\|$；

则称 $\|\boldsymbol{\alpha}\|$ 为 $\boldsymbol{\alpha}$ 的**范数**. 定义了范数的向量空间称为**赋范向量空间**.

例如在 $\mathbf{C}^n$ 上，对于任一向量 $\boldsymbol{x}=[x_1, x_2, \cdots, x_n]^{\mathrm{T}}$，$\boldsymbol{x}$ 的长度 $|\boldsymbol{x}|=\sqrt{\sum_{i=1}^{n}|x_i|^2}$ 就是 $\boldsymbol{x}$ 的一种范数，称为**欧氏范数**，今后记为 $\|\boldsymbol{x}\|_2$，常称为 **2-范数**.

在 $\mathbf{C}^n$ 上的赋范，实质上就是在 $\mathbf{C}^n$ 上定义满足上述三个条件的一个实函数，因此 $\mathbf{C}^n$ 上可以找到无穷多种范数. 下面举出几种常用的范数.

$$\|\boldsymbol{x}\|_1=\sum_{i=1}^{n}|x_i|; \tag{3.1-1}$$

$$\|\boldsymbol{x}\|_\infty=\max_{1\leqslant i\leqslant n}|x_i|; \tag{3.1-2}$$

$$\|\boldsymbol{x}\|_p=\left(\sum_{i=1}^{n}|x_i|^p\right)^{\frac{1}{p}}①, 1\leqslant p\leqslant+\infty;$$

它们分别称为 $\boldsymbol{x}$ 的 1-范数，∞-范数和 p-范数. 由于

$$\|\boldsymbol{x}\|=\|\boldsymbol{x}-\boldsymbol{y}+\boldsymbol{y}\| \leqslant \|\boldsymbol{x}-\boldsymbol{y}\|+\|\boldsymbol{y}\|,$$

$$\|\boldsymbol{y}\|=\|\boldsymbol{y}-\boldsymbol{x}+\boldsymbol{x}\| \leqslant \|\boldsymbol{y}-\boldsymbol{x}\|+\|\boldsymbol{x}\|=\|\boldsymbol{x}-\boldsymbol{y}\|+\|\boldsymbol{x}\|,$$

3.1 向量范数

故有
$$-\|x-y\| \leqslant \|x\|-\|y\| \leqslant \|x-y\|,$$
即
$$\big|\|x\|-\|y\|\big| \leqslant \|x-y\|. \tag{3.1-3}$$
(3.1-3)式表明，当 y 趋向于 x 时，$\|y\|$ 趋向于 $\|x\|$. 因此，$\|x\|$ 是 x 的连续函数.

定理 设 $\|x\|_a$，$\|x\|_b$ 是 $\mathbf{C}^n$ 上 x 的任意两种范数，则存在两个与 x 无关的正常数 $K_2 \geqslant K_1 > 0$，使
$$K_1\|x\|_b \leqslant \|x\|_a \leqslant K_2\|x\|_b. \tag{3.1-4}$$

证 当 $x=0$时，(3.1-3)式显然成立. 下设 $x \neq 0$，记 φ 是 $\mathbf{C}^n$ 上的有界闭集：
$$\varphi=\left\{\xi=[\xi_1,\xi_2,\cdots,\xi_n]^{\mathrm{T}} \in \mathbf{C}^n \,\middle|\, \sum_{i=1}^{n}|\xi_i|^2=1\right\}.$$
由于 $\|\xi\|_a$、$\|\xi\|_b$ 是 ξ 的连续函数，所以在 $\xi \neq 0$ 时，$\dfrac{\|\xi\|_a}{\|\xi\|_b}$是 ξ 的连续函数，故 $\dfrac{\|\xi\|_a}{\|\xi\|_b}$ 在 φ 上取得最大值 K_2 和最小值 $K_1>0$，即当 $\xi \in \varphi$ 时，
$$K_1 \leqslant \frac{\|\xi\|_a}{\|\xi\|_b} \leqslant K_2. \tag{3.1-5}$$

对于任一 $x \neq 0$，令 $\xi=\dfrac{x}{\|x\|_2}$，则由范数的齐次性得 $\|\xi\|_2=\left\|\dfrac{x}{\|x\|_2}\right\|_2=\dfrac{1}{\|x\|_2}$ $\|x\|_2=1$，从而 $\xi \in \varphi$. 于是，由(3.1-5)式知，
$$K_1 \leqslant \frac{\|\xi\|_a}{\|\xi\|_b}=\frac{\|x\|_a/\|x\|_2}{\|x\|_b/\|x\|_2}=\frac{\|x\|_a}{\|x\|_b} \leqslant K_2,$$
即
$$K_1\|x\|_b \leqslant \|x\|_a \leqslant K_2\|x\|_b. \quad \blacksquare$$

3.1 向量范数

称满足(3.1-4)式的范数 $\|x\|_a$ 与 $\|x\|_b$ 是等价的. 因此这个定理可以叙述为，$\mathbf{C}^n$ 上任何两种不同的范数都是等价的.

例 1 设 A 是一个正定矩阵，对任意实向量 $x\in\mathbf{R}^n$，定义函数

$$\|x\|_A=\sqrt{x^{\mathrm{T}}Ax},\tag{3.1-6}$$

证明它是一种向量范数(称为加权范数).

证 因为 A 正定，所以只要 $x\neq 0$，就有 $x^{\mathrm{T}}Ax>0$，从而 $\|x\|_A>0$，又若 $\|x\|_A=0$，即 $x^{\mathrm{T}}Ax=0$，则 $x=0$. 因此 $\|\cdot\|_A$ 满足正定性条件.

对任何实数 $k\in\mathbf{R}$，有

$$\begin{aligned}\|kx\|_A&=\sqrt{(kx)^{\mathrm{T}}A(kx)}=\sqrt{k^2x^{\mathrm{T}}Ax}\\&=|k|\sqrt{x^{\mathrm{T}}Ax}=|k|\,\|x\|_A.\end{aligned}$$

因此 $\|\cdot\|$ 满足齐次性条件.

由于 A 正定，故存在可逆矩阵 P 使 $A=P^{\mathrm{T}}P$. 于是，

$$\|x\|_A=\sqrt{x^{\mathrm{T}}Ax}=\sqrt{x^{\mathrm{T}}P^{\mathrm{T}}Px}=\sqrt{(Px)^{\mathrm{T}}(Px)}=\|Px\|_2.$$

因此

$$\begin{aligned}\|x+y\|_A&=\|P(x+y)\|_2=\|Px+Py\|_2\\&\leqslant\|Px\|_2+\|Py\|_2=\|x\|_A+\|y\|_A,\end{aligned}$$

即 $\|\cdot\|_A$ 满足三角不等式.

综上所述，$\|\cdot\|_A$ 是一种向量范数. ∎

3.1 向量范数

例 2 举例说明，当 $0<p<1$ 时，

$$\|\boldsymbol{x}\|_p=\left(\sum_{i=1}^n|x_i|^p\right)^{\frac{1}{p}}$$

不是向量范数.

解 取 $\boldsymbol{x}=[1,0,0,\cdots,0]^{\mathrm{T}}$，$\boldsymbol{y}=[0,1,0,\cdots,0]^{\mathrm{T}}$，则 $\|\boldsymbol{x}\|_p=\|\boldsymbol{y}\|_p=1$. 但 $\boldsymbol{x}+\boldsymbol{y}=[1,1,0,\cdots,0]^{\mathrm{T}}$，

$$\|\boldsymbol{x}+\boldsymbol{y}\|_p=2^{\frac{1}{p}}>2=\|\boldsymbol{x}\|_p+\|\boldsymbol{y}\|_p,$$

因此它不是向量范数. ∎

例 3 由于一个 $m\times n$ 矩阵 $\mathbf{A}=[a_{ij}]$ 可以把它拉直看成一个 mn 维向量，所以 $\mathbf{C}^{mn}$ 上的任何一种向量范数都可以作为 $m\times n$ 矩阵的范数. 例如，$\mathbf{C}^{mn}$ 上的 2-范数

$$\|\mathbf{A}\|_F=\sqrt{\sum_{i=1}^m\sum_{j=1}^n|a_{ij}|^2}=\sqrt{\operatorname{tr}(\mathbf{A}^{\mathrm{H}}\mathbf{A})}, \tag{3.1-7}$$

是 $\mathbf{A}$ 的范数，称为 Frobenius 范数，简称 F-范数. ∎

3.2 矩阵范数

在 3.1 节中，我们把矩阵看成是拉直的向量，而利用向量范数来规定矩阵的范数. 但若把 $m\times n$ 矩阵 $\boldsymbol{A}$ 看成是矩阵集合 $\mathbf{C}^{m\times n}$ 中的元素，由于这个集合 $\mathbf{C}^{m\times n}$ 对矩阵加法和数乘运算是封闭的，并满足 1.1 节所说的八条运算规则，所以 $\mathbf{C}^{m\times n}$ 也是“向量”空间. 为了避免混淆，一般称 $\mathbf{C}^{m\times n}$ 是线性空间，我们将在第六章论述线性空间的一般概念.

但对于矩阵，还必须考虑矩阵的乘法运算，或者从矩阵分解(即将一个矩阵看成若干个有特殊性质的矩阵的乘积)来考虑，都需要在一个矩阵的范数与其各因子的范数之间有一种确定的关系. 例如，为了用矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$ 的范数 $\|\boldsymbol{A}\|$、$\|\boldsymbol{B}\|$ 来控制 $\boldsymbol{AB}$ 的范数 $\|\boldsymbol{AB}\|$，就要求有

$$\|\boldsymbol{AB}\| \leqslant \|\boldsymbol{A}\| \, \|\boldsymbol{B}\|. \tag{3.2-1}$$

如果直接把向量范数移植过来定义矩阵范数就未必能保证上述不等式成立. 事实上，若按 $\mathbf{C}^{mn}$ 中的∞-范数来定义 $m\times n$ 矩阵 $\boldsymbol{A}=[a_{ij}]$的范数，则 $\|\boldsymbol{A}\| = \max\limits_{\substack{1\leqslant i\leqslant m\\1\leqslant j\leqslant n}} |a_{ij}|$，那么对于方阵 $\boldsymbol{A}=\boldsymbol{B}=\begin{bmatrix}1 & 1\\1 & 1\end{bmatrix}$，$\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的乘积 $\boldsymbol{AB}=\begin{bmatrix}2 & 2\\2 & 2\end{bmatrix}$，从而 $\|\boldsymbol{AB}\|=2$，但 $\|\boldsymbol{A}\|=\|\boldsymbol{B}\|=1$，显然 $\|\boldsymbol{AB}\| > \|\boldsymbol{A}\| \, \|\boldsymbol{B}\|$.

因此，我们定义矩阵范数如下.

3.2 矩阵范数

定义 3.2-1 对于任一 $m\times n$ 复矩阵 $\boldsymbol{A}$，都有一个实数 $\|\boldsymbol{A}\|$ 与之对应，且满足

(1) 正定性 $\|\boldsymbol{A}\|\geqslant 0$，当且仅当 $\boldsymbol{A}=\boldsymbol{O}$ 时才有 $\|\boldsymbol{A}\|=0$；

(2) 齐次性 $\|k\boldsymbol{A}\|=|k|\,\|\boldsymbol{A}\|$，$\forall k\in\mathbf{C}$；

(3) 三角不等式 $\|\boldsymbol{A}+\boldsymbol{B}\|\leqslant\|\boldsymbol{A}\|+\|\boldsymbol{B}\|$；

(4) 相容性 当 $\boldsymbol{A}=\boldsymbol{B}\boldsymbol{G}$ 时有 $\|\boldsymbol{A}\|\leqslant\|\boldsymbol{B}\|\,\|\boldsymbol{G}\|$，则称 $\|\boldsymbol{A}\|$ 为 $\boldsymbol{A}$ 的范数.

与向量的情况一样，矩阵范数也是多种多样的.

例 1 设 $\boldsymbol{A}=[a_{ij}]$ 是 $m\times n$ 复矩阵，证明

(1) $$\|\boldsymbol{A}\|_F=\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}=\sqrt{\operatorname{tr}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})},$$

(2) $$\|\boldsymbol{A}\|_{M_1}=\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|$$

是矩阵范数.

证 它们满足矩阵范数定义的前三条是显然的，现只证它们满足矩阵范数定义的第四条.

令 $\boldsymbol{B}=[b_{jk}]$ 是 $n\times s$ 复矩阵，则 $\boldsymbol{AB}$ 是 $m\times s$ 矩阵，它的第 i 行第 k 列的元素是

$$(\boldsymbol{AB})_{ik}=\sum_{j=1}^{n}a_{ij}b_{jk},\ i=1,\cdots,m;k=1,\cdots,s.$$

3.2 矩阵范数

(1) $$\|AB\|_F=\left(\sum_{i=1}^{m}\sum_{k=1}^{s}\left|\sum_{j=1}^{n}a_{ij}b_{jk}\right|^2\right)^{\frac{1}{2}}\leqslant\left[\sum_{i=1}^{m}\sum_{k=1}^{s}\left(\sum_{j=1}^{n}|a_{ij}|\cdot|b_{jk}|\right)^2\right]^{\frac{1}{2}}$$

$$\leqslant\left[\sum_{i=1}^{m}\sum_{k=1}^{s}\left(\sum_{j=1}^{n}|a_{ij}|^2\right)\left(\sum_{j=1}^{n}|b_{jk}|^2\right)\right]^{\frac{1}{2}}$$

$$=\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2\right)^{\frac{1}{2}}\cdot\left(\sum_{j=1}^{n}\sum_{k=1}^{s}|b_{jk}|^2\right)^{\frac{1}{2}}=\|A\|_F\ \|B\|_F.$$

(2) $$\|AB\|_{M_1}=\sum_{i=1}^{m}\sum_{k=1}^{s}\left|\sum_{j=1}^{n}a_{ij}b_{jk}\right|\leqslant\sum_{i=1}^{m}\sum_{k=1}^{s}\sum_{j=1}^{n}|a_{ij}|\cdot|b_{jk}|$$

$$\leqslant\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|\right)\cdot\left(\sum_{j=1}^{n}\sum_{k=1}^{s}|b_{jk}|\right)=\|A\|_{M_1}\ \|B\|_{M_1}.$$ ■

在矩阵的理论和运算中，常常要作矩阵 $\boldsymbol{A}$ 与向量 $\boldsymbol{x}$ 的乘法，且乘积 $\boldsymbol{Ax}$ 是向量，这就涉及矩阵范数 $\|\boldsymbol{A}\|$ 与向量范数 $\|\boldsymbol{x}\|$、$\|\boldsymbol{Ax}\|$ 之间的关系。若将向量看成矩阵，由矩阵范数的相容性应有 $\|\boldsymbol{Ax}\|\leqslant\|\boldsymbol{A}\|\ \|\boldsymbol{x}\|$。但矩阵范数与向量范数又是有差异的，如果它们各自取范数，则这个不等式未必成立。例如，向量范数取 1-范数，矩阵范数取 F-范数，那么对于向量 $\boldsymbol{x}=[0,2]^{\mathrm{T}}$、矩阵 $\boldsymbol{A}=\begin{bmatrix}1&4\\1&4\end{bmatrix}$ 来说，$\boldsymbol{Ax}=[8,8]^{\mathrm{T}}$，于是 $\|\boldsymbol{Ax}\|_1=8+8=16$，$\|\boldsymbol{x}\|_1=0+2=2$，$\|\boldsymbol{A}\|_F=\sqrt{2(1^2+4^2)}=\sqrt{34}$。显然 $\|\boldsymbol{Ax}\|_1>\|\boldsymbol{A}\|_F\ \|\boldsymbol{x}\|_1$。

因此，希望能有这样的矩阵范数，使它能与向量范数协调。

3.2 矩阵范数

定义 3.2-2 设 $\|\cdot\|$ 是向量范数，$\|\cdot\|_M$ 是矩阵范数，如果对任意 $m\times n$ 矩阵 $\boldsymbol{A}$ 和任意 n 维向量 $\boldsymbol{x}$，均有

$$\|\boldsymbol{A}\boldsymbol{x}\| \leqslant \|\boldsymbol{A}\|_M \|\boldsymbol{x}\|, \tag{3.2-2}$$

则称 $\|\cdot\|_M$ 为与向量范数 $\|\cdot\|$ 协调的矩阵范数.

例 2 设 $\boldsymbol{x}=[x_1,x_2,\cdots,x_n]^{\mathrm{T}}$，$\boldsymbol{A}=[a_{ij}]$ 是 $m\times n$ 矩阵，证明 $\|\boldsymbol{A}\|_{M_1}$ 是与向量的 1-范数协调的.

证 由于 $\|\boldsymbol{A}\boldsymbol{x}\|_1=\sum_{i=1}^{m}\left|\sum_{j=1}^{n}a_{ij}x_j\right|\leqslant\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|\cdot|x_j|$

$$\leqslant\left(\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|\right)\left(\sum_{j=1}^{n}|x_j|\right)=\|\boldsymbol{A}\|_{M_1}\|x\|_1,$$

故 $\|\boldsymbol{A}\|_{M_1}$ 是与向量的 1-范数协调的. ∎

给出一种与向量范数协调的矩阵范数，这就是**诱导范数**（也称**算子范数**）：

$$\|\boldsymbol{A}\| = \max_{\|\boldsymbol{x}\|=1}\|\boldsymbol{A}\boldsymbol{x}\|. \tag{3.2-3}$$

由于 $\|\boldsymbol{A}\boldsymbol{x}\|$ 是 $\boldsymbol{x}$ 的连续函数，所以对给定的 $\boldsymbol{A}$ 来说，$\|\boldsymbol{A}\boldsymbol{x}\|$ 在有界闭集 $\|\boldsymbol{x}\|=1$ 上是可以取得最大值的，即存在这样的向量 $\boldsymbol{x}^*$，$\|\boldsymbol{x}^*\|=1$ 且使 $\|\boldsymbol{A}\boldsymbol{x}^*\|=\|\boldsymbol{A}\|$.

下面证明，诱导范数满足范数定义中的四个条件.

(1) 设 $\boldsymbol{A}\neq\boldsymbol{O}$，则存在 $\boldsymbol{x}^*$，使 $\|\boldsymbol{x}^*\|=1$ 及 $\boldsymbol{A}\boldsymbol{x}^*\neq\boldsymbol{0}$. 从而 $\|\boldsymbol{A}\boldsymbol{x}^*\|>0$，故有 $\|\boldsymbol{A}\|=\max_{\|\boldsymbol{x}\|=1}\|\boldsymbol{A}\boldsymbol{x}\|\geqslant\|\boldsymbol{A}\boldsymbol{x}^*\|>0$.

3.2 矩阵范数

如果 $\boldsymbol{A}=\boldsymbol{O}$，则对任意的 $\boldsymbol{x}$ 有 $\boldsymbol{Ax}=\boldsymbol{0}$，故 $\|\boldsymbol{Ax}\|=0$. 因此 $\|\boldsymbol{A}\|=\max\limits_{\|\boldsymbol{x}\|=1}\|\boldsymbol{Ax}\|=0$.

(2) $\|k\boldsymbol{A}\|=\max\limits_{\|\boldsymbol{x}\|=1}\|k\boldsymbol{Ax}\|=|k|\max\limits_{\|\boldsymbol{x}\|=1}\|\boldsymbol{Ax}\|=|k|\,\|\boldsymbol{A}\|$.

(3) 对于 $\boldsymbol{A}+\boldsymbol{B}$，可以找到 $\boldsymbol{x}^*$，使 $\|\boldsymbol{x}^*\|=1$ 及 $\|\boldsymbol{A}+\boldsymbol{B}\|=\|(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}^*\|$，因而

$$\|\boldsymbol{A}+\boldsymbol{B}\|=\|(\boldsymbol{A}+\boldsymbol{B})\boldsymbol{x}^*\|=\|\boldsymbol{Ax}^*+\boldsymbol{Bx}^*\|\leqslant\|\boldsymbol{Ax}^*\|+\|\boldsymbol{Bx}^*\|$$
$$\leqslant\max_{\|\boldsymbol{x}\|=1}\|\boldsymbol{Ax}\|+\max_{\|\boldsymbol{x}\|=1}\|\boldsymbol{Bx}\|=\|\boldsymbol{A}\|+\|\boldsymbol{B}\|.$$

(4) 对于 $\boldsymbol{AB}$，可以找到 $\boldsymbol{x}^*$ 使 $\|\boldsymbol{x}^*\|=1$ 及 $\|\boldsymbol{AB}\|=\|\boldsymbol{ABx}^*\|$，因而

$$\|\boldsymbol{AB}\|=\|\boldsymbol{ABx}^*\|=\|\boldsymbol{A}\cdot\boldsymbol{Bx}^*\|\leqslant\|\boldsymbol{A}\|\,\|\boldsymbol{Bx}^*\|\leqslant\|\boldsymbol{A}\|\cdot\max_{\|\boldsymbol{x}\|=1}\|\boldsymbol{Bx}\|$$
$$=\|\boldsymbol{A}\|\,\|\boldsymbol{B}\|.$$

对于矩阵的诱导范数来说，必有 $\|\boldsymbol{I}\|=1$. 事实上，$\|\boldsymbol{I}\|=\max\limits_{\|\boldsymbol{x}\|=1}\|\boldsymbol{Ix}\|=\max\limits_{\|\boldsymbol{x}\|=1}\|\boldsymbol{x}\|=1$.

在(3.2-3)式中，当向量范数取为 1-范数，∞-范数和 2-范数时，对应的矩阵诱导范数分别称为 $\boldsymbol{A}$ 的 1-范数，∞-范数和 2-范数，依次记为 $\|\boldsymbol{A}\|_1$，$\|\boldsymbol{A}\|_\infty$ 和 $\|\boldsymbol{A}\|_2$.

定理 3.2-1 设 $\boldsymbol{A}=[a_{ij}]$ 是 $m\times n$ 矩阵，则

(1) $$\|\boldsymbol{A}\|_1=\max_{1\leqslant j\leqslant n}\sum_{i=1}^{m}|a_{ij}|;\tag{3.2-4}$$

(2) $$\|\boldsymbol{A}\|_\infty=\max_{1\leqslant i\leqslant m}\sum_{j=1}^{n}|a_{ij}|;\tag{3.2-5}$$

(3) $\|\boldsymbol{A}\|_2=\sqrt{\lambda_1}$，$\lambda_1$ 是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的最大特征值. (3.2-6)

3.2　矩阵范数

证　(1) 设 $\|\boldsymbol{x}\|_1=1$，即 $\sum\limits_{j=1}^{n}|x_j|=1$，则

$$\|\boldsymbol{A}\boldsymbol{x}\|_1=\sum_{i=1}^{m}\left|\sum_{j=1}^{n}a_{ij}x_j\right|\leqslant\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|\,|x_j|=\sum_{j=1}^{n}\Big(\sum_{i=1}^{m}|a_{ij}|\Big)|x_j|$$

$$\leqslant\Big(\max_{1\leqslant j\leqslant n}\sum_{i=1}^{m}|a_{ij}|\Big)\sum_{j=1}^{n}|x_j|=\max_{1\leqslant j\leqslant n}\sum_{i=1}^{m}|a_{ij}|.$$

再证
$$\max_{\|\boldsymbol{x}\|_1=1}\|\boldsymbol{A}\boldsymbol{x}\|_1=\max_{1\leqslant j\leqslant n}\sum_{i=1}^{m}|a_{ij}|.$$

假设当 $j=k$ 时 $\sum\limits_{i=1}^{m}|a_{ij}|$ 取得最大值，即 $\max\limits_{1\leqslant j\leqslant n}\sum\limits_{i=1}^{m}|a_{ij}|=\sum\limits_{i=1}^{m}|a_{ik}|$. 取 $\boldsymbol{x}^*=[0\ \cdots\ 0\ \overset{k}{1}\ 0\ \cdots\ 0]^{\mathrm{T}}$，则显然有 $\|\boldsymbol{x}^*\|_1=1$，且

$$\|\boldsymbol{A}\boldsymbol{x}^*\|_1=\sum_{i=1}^{m}\left|\sum_{j=1}^{n}a_{ij}x_j^*\right|=\sum_{i=1}^{n}|a_{ik}|=\max_{1\leqslant j\leqslant n}\sum_{i=1}^{m}|a_{ij}|.$$

因此，(3.2-4)式成立.

(2) 请读者自证(3.2-5)式.

(3) 由于
$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2=\langle\boldsymbol{A}\boldsymbol{x},\boldsymbol{A}\boldsymbol{x}\rangle=\overline{\langle\boldsymbol{A}\boldsymbol{x},\boldsymbol{A}\boldsymbol{x}\rangle}=\overline{(\boldsymbol{A}\boldsymbol{x})^{\mathrm{T}}\ \overline{(\boldsymbol{A}\boldsymbol{x})}}=(\boldsymbol{A}\boldsymbol{x})^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}$$
$$=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}\geqslant 0,$$

所以 n 阶 Hermite 矩阵 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的特征值必是非负数，不妨设为 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_n\geqslant 0$，对应的两两正交的单位特征向量依次为 $\boldsymbol{v}_1,\boldsymbol{v}_2,\cdots,\boldsymbol{v}_n$.

3.2 矩阵范数

对任意一个 2- 范数是 1 的向量 $\boldsymbol{x}$，它可表示为

$$\boldsymbol{x} = c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + \cdots + c_n\boldsymbol{v}_n,$$

且有

$$|c_1|^2 + |c_2|^2 + \cdots + |c_n|^2 = 1.$$

由于

$$\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^H\boldsymbol{A}\Big(\sum_{i=1}^n c_i\boldsymbol{v}_i\Big) = \sum_{i=1}^n c_i(\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{v}_i) = \sum_{i=1}^n \lambda_i c_i \boldsymbol{v}_i,$$

所以有

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2 = \boldsymbol{x}^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{x} = \Big(\sum_{j=1}^n c_j\boldsymbol{v}_j\Big)^H\Big(\sum_{i=1}^n \lambda_i c_i\boldsymbol{v}_i\Big) = \sum_{i=1}^n\sum_{j=1}^n \lambda_i c_i \bar{c}_j \boldsymbol{v}_j^H\boldsymbol{v}_i. \qquad (3.2\text{-}7)$$

因为 $\boldsymbol{v}_j^H\boldsymbol{v}_i = \overline{\langle \boldsymbol{v}_j^T\bar{\boldsymbol{v}}_i\rangle} = \overline{\langle \boldsymbol{v}_j, \boldsymbol{v}_i\rangle} = \langle \boldsymbol{v}_i, \boldsymbol{v}_j\rangle = \delta_{ij}$，故由(3.2-7)式得

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2 = \sum_{i=1}^n \lambda_i |c_i|^2 \leqslant \lambda_1 \sum_{i=1}^n |c_i|^2 = \lambda_1,$$

两边开方便有 $\|\boldsymbol{A}\boldsymbol{x}\|_2 \leqslant \sqrt{\lambda_1}$.

另一方面，$\|\boldsymbol{v}_1\|_2 = 1$，且 $\|\boldsymbol{A}\boldsymbol{v}_1\|_2^2 = \boldsymbol{v}_1^H\boldsymbol{A}^H\boldsymbol{A}\boldsymbol{v}_1 = \boldsymbol{v}_1^H \cdot \lambda_1\boldsymbol{v}_1 = \lambda_1\boldsymbol{v}_1^H\boldsymbol{v}_1 = \lambda_1$，故 $\|\boldsymbol{A}\boldsymbol{v}_1\|_2 = \sqrt{\lambda_1}$. 因此(3.2-6)式成立. ∎

由于 $\|\boldsymbol{A}\|_1$ 是每一列所有的元素的模之和构成的列和集合中的最大值，所以也称它为**列和范数**.

类似地，由于 $\|\boldsymbol{A}\|_\infty$ 是每一行所有的元素的模之和构成的行和集合中的最大值，所以也称它为**行和范数**.

3.2 矩阵范数

例 3 设 A 是 n 阶方阵，$\|A\|$ 是诱导范数，证明当 $\|A\|<1$ 时，$I-A$ 可逆，且有

$$\|(I-A)^{-1}\|\leqslant\frac{1}{1-\|A\|}.$$

证 若 $I-A$ 不可逆，则齐次线性方程组 $(I-A)x=O$ 有非零解，即存在 $x\neq 0$ 使 $Ax=x$，从而

$$\|x\|=\|Ax\|\leqslant\|A\|\ \|x\|<\|x\|.$$

但这是不可能的，故 $I-A$ 可逆.

令 $B=(I-A)^{-1}$，则有 $B(I-A)=I$，即

$$B=I+BA.$$

于是

$$\|B\|=\|I+BA\|\leqslant\|I\|+\|BA\|\leqslant 1+\|B\|\ \|A\|,$$

即

$$(1-\|A\|)\|B\|\leqslant 1.$$

因此

$$\|(I-A)^{-1}\|=\|B\|\leqslant\frac{1}{1-\|A\|}.\quad\blacksquare$$

矩阵范数都是指诱导范数.

3.2 矩阵范数

例 4 设 n 阶方阵 $\boldsymbol{A}$ 是可逆的，则线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 有唯一解 $\boldsymbol{x}=\boldsymbol{A}^{-1}\boldsymbol{b}$. 讨论当 $\boldsymbol{b}$ 有小的误差 $\delta\boldsymbol{b}$ 时，解的误差 $\delta\boldsymbol{x}$ 是多大的问题.

由方程 $\boldsymbol{A}(\boldsymbol{x}+\delta\boldsymbol{x})=\boldsymbol{b}+\delta\boldsymbol{b}$ 得，

$$\boldsymbol{A}\delta\boldsymbol{x}=\boldsymbol{b}-\boldsymbol{Ax}+\delta\boldsymbol{b}=\delta\boldsymbol{b},$$

故 $\delta\boldsymbol{x}=\boldsymbol{A}^{-1}\cdot\delta\boldsymbol{b}$，因此有 $\|\delta\boldsymbol{x}\|=\|\boldsymbol{A}^{-1}\delta\boldsymbol{b}\|\leqslant\|\boldsymbol{A}^{-1}\|\ \|\delta\boldsymbol{b}\|$.

另一方面，由 $\boldsymbol{Ax}=\boldsymbol{b}$ 推出 $\|\boldsymbol{b}\|\leqslant\|\boldsymbol{A}\|\ \|\boldsymbol{x}\|$，故有 $\dfrac{1}{\|\boldsymbol{x}\|}\leqslant\dfrac{\|\boldsymbol{A}\|}{\|\boldsymbol{b}\|}$. 于是

$$\frac{\|\delta\boldsymbol{x}\|}{\|\boldsymbol{x}\|}\leqslant\|\boldsymbol{A}\|\ \|\boldsymbol{A}^{-1}\|\ \frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}=\mathrm{Cond}(\boldsymbol{A})\ \frac{\|\delta\boldsymbol{b}\|}{\|\boldsymbol{b}\|}, \tag{3.2-8}$$

其中

$$\mathrm{Cond}(\boldsymbol{A})=\|\boldsymbol{A}\|\ \|\boldsymbol{A}^{-1}\|, \tag{3.2-9}$$

称为方阵 $\boldsymbol{A}$ 的**条件数**. ■

(3.2-8)式的左端是 $\boldsymbol{x}$ 的相对误差，而右端是 $\boldsymbol{b}$ 的相对误差的 Cond($\boldsymbol{A}$)倍. 当 $\boldsymbol{A}$ 的条件数很大时，$\boldsymbol{b}$ 的微小相对误差可能产生 $\boldsymbol{x}$ 的很大的相对误差，因而解的可靠性低.

条件数对研究问题的性态起重要的作用，在数值计算方法部分将再做一般的讨论.

值得注意的是，方阵 $\boldsymbol{A}$ 的条件数的值取决于所取的矩阵范数，由于常用的矩阵范数是∞-范数、1-范数及 2-范数，所以把(3.2-9)式改写为

$$\mathrm{Cond}_p(\boldsymbol{A})=\|\boldsymbol{A}\|_p\|\boldsymbol{A}^{-1}\|_p,\quad p=\infty,1,2. \tag{3.2-9$'$}$$

再则，任一方阵的条件数必大于或等于 1，因为 $\|\boldsymbol{A}\|\ \|\boldsymbol{A}^{-1}\|\geqslant\|\boldsymbol{AA}^{-1}\|=\|\boldsymbol{I}\|=1$.

3.2 矩阵范数

例 5 设 $A=\begin{bmatrix} 1 & 5 & -2 \\ -2 & 1 & 0 \\ 3 & -8 & 2 \end{bmatrix}$，求 $\mathrm{Cond}_\infty(A)$ 和 $\mathrm{Cond}_1(A)$.

解 $\|A\|_\infty=\max\{8,3,13\}=13$,

$$\|A\|_1=\max\{6,14,4\}=14.$$

又
$$A^{-1}=-\frac{1}{4}\begin{bmatrix} 2 & 6 & 2 \\ 4 & 8 & 4 \\ 13 & 23 & 11 \end{bmatrix},$$

$$\|A^{-1}\|_\infty=\max\left\{\frac{10}{4}, 4, \frac{47}{4}\right\}=\frac{47}{4},$$

$$\|A^{-1}\|_1=\max\left\{\frac{19}{4}, \frac{37}{4}, \frac{17}{4}\right\}=\frac{37}{4}.$$

因此，由(3.2-9′)式得

$$\mathrm{Cond}_\infty(A)=13\times\frac{47}{4}=\frac{611}{4},\quad \mathrm{Cond}_1(A)=14\times\frac{37}{4}=\frac{259}{2}.$$ ∎

上面讨论了给定向量范数，确定由此向量范数诱导的矩阵范数. 反之，给定了矩阵范数，也可确定由此矩阵范数诱导的向量范数.

3.2 矩阵范数

定理 2.3-2 设 $\|A\|_*$ 是矩阵范数，则存在向量范数 $\|x\|$ 使

$$\|Ax\| \leqslant \|A\|_* \|x\|.$$

证 任取一非零向量 β，由于 $x\beta^H$ 是矩阵，所以可以定义 x 的函数为 $\|x\|$：

$$\|x\| = \|x\beta^H\|_*.$$

下证 $\|x\|$ 是 x 的范数，且 $\|x\|$ 与 $\|A\|_*$ 协调.

由于 $\beta\neq 0$，所以只要 $x\neq 0$ 就有 $x\beta^H$ 不是零矩阵，故 $\|x\| = \|x\beta^H\|_* > 0$. 因此 $\|\cdot\|$ 满足正定性.

显然，$\|\cdot\|$ 满足齐次性.

又由于 $$\begin{aligned}\|x+y\| &= \|(x+y)\beta^H\|_* = \|x\beta^H + y\beta^H\|_* \\ &\leqslant \|x\beta^H\|_* + \|y\beta^H\|_* = \|x\| + \|y\|.\end{aligned}$$

故 $\|\cdot\|$ 满足三角不等式.

最后，因 $$\|Ax\| = \|Ax\beta^H\|_* \leqslant \|A\|_* \|x\beta^T\|_* = \|A\|_* \|x\|,$$

故 $\|x\|$ 与 $\|A\|_*$ 协调. ∎

3.3 矩阵范数

对于方阵来说，其“大小”自然可以用 3.2 节所述的矩阵范数来描述，但从前两章可知，方阵的特征值及特征向量是方阵的本质特性，从而希望能用与特征值有关的数来刻划矩阵的“大小”，这个数就是方阵的谱半径. 此外，还要讨论谱半径与矩阵范数的关系.

定义 3.3-1 方阵 $\boldsymbol{A}$ 的所有不同特征值组成的集合称为 $\boldsymbol{A}$ 的**谱**，记为 $Sp(\boldsymbol{A})$，并称特征值的模的最大值为 $\boldsymbol{A}$ 的**谱半径**，记为 $\rho(\boldsymbol{A})$.

例如，对于方阵

$$\boldsymbol{A}=\begin{bmatrix}1-\mathrm{i} & 3\\ -1 & 1+\mathrm{i}\end{bmatrix},$$

由于 $\det(\boldsymbol{A}-\lambda\boldsymbol{I})=\begin{vmatrix}(1-\lambda)-\mathrm{i} & 3\\ -1 & (1-\lambda)+\mathrm{i}\end{vmatrix}=(1-\lambda)^2+4$，故 $\boldsymbol{A}$ 有两个不同的特征值：$\lambda_1=1+2\mathrm{i}$，$\lambda_2=1-2\mathrm{i}$，所以

$$Sp(\boldsymbol{A})=\{1+2\mathrm{i},1-2\mathrm{i}\},\ \rho(\boldsymbol{A})=\sqrt{5}.$$

由谱半径的定义可知，方阵 $\boldsymbol{A}$ 的所有特征值分布在复平面上以原点为中心、$\rho(\boldsymbol{A})$ 为半径的圆上.

方阵的谱半径与矩阵范数不同，它没有正定性和三角不等式，也没有相容性. 例如，$\boldsymbol{A}=\begin{bmatrix}0 & 0\\ 3 & 0\end{bmatrix}\neq\boldsymbol{O}$，但 $\rho(\boldsymbol{A})=0$，再取 $\boldsymbol{B}=\begin{bmatrix}2 & 3\\ 0 & 2\end{bmatrix}$，则 $\boldsymbol{A}+\boldsymbol{B}=\begin{bmatrix}2 & 3\\ 3 & 2\end{bmatrix}$，$\boldsymbol{AB}=\begin{bmatrix}0 & 0\\ 6 & 9\end{bmatrix}$，从而

$$\rho(\boldsymbol{A}+\boldsymbol{B})=5,\quad \rho(\boldsymbol{AB})=9.$$

3.3 矩阵范数

但由于 $\rho(\boldsymbol{B})=2$，所以 $\rho(\boldsymbol{A})+\rho(\boldsymbol{B})=2$，$\rho(\boldsymbol{A})\rho(\boldsymbol{B})=0$。因此

$$\rho(\boldsymbol{A}+\boldsymbol{B})>\rho(\boldsymbol{A})+\rho(\boldsymbol{B}),\quad \rho(\boldsymbol{AB})>\rho(\boldsymbol{A})\rho(\boldsymbol{B}).$$

不过，谱半径与矩阵范数还是有一定关系的，正如下述定理所表明的那样。

定理 3.3-1[①] 方阵 $\boldsymbol{A}$ 的谱半径不大于 $\boldsymbol{A}$ 的任何一种诱导范数，即

$$\rho(\boldsymbol{A})\leqslant\|\boldsymbol{A}\|. \tag{3.3-1}$$

证 设 λ 是 $\boldsymbol{A}$ 的任一特征值，$\boldsymbol{x}\neq\boldsymbol{0}$ 是属于 λ 的特征向量，则由 $\boldsymbol{Ax}=\lambda\boldsymbol{x}$ 得

$$|\lambda|\ \|\boldsymbol{x}\|=\|\lambda\boldsymbol{x}\|=\|\boldsymbol{Ax}\|\leqslant\|\boldsymbol{A}\|\ \|\boldsymbol{x}\|.$$

因 $\boldsymbol{x}\neq\boldsymbol{0}$，故 $\|\boldsymbol{x}\|\neq 0$，所以 $|\lambda|\leqslant\|\boldsymbol{A}\|$。

由于 λ 是 $\boldsymbol{A}$ 的任一特征值，从而(3.3-1)式成立. ∎

例 1 证明对任意可逆矩阵 $\boldsymbol{A}$，有

$$\|\boldsymbol{A}\|_2=\sqrt{\rho(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})}=\sqrt{\rho(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})}, \tag{3.3-2}$$

且当 $\boldsymbol{A}$ 是 Hermite 矩阵时，$\|\boldsymbol{A}\|_2=\rho(\boldsymbol{A})$。

证 按(3.2-6)式知，$\|\boldsymbol{A}\|_2=\sqrt{\lambda_1}$，其中 λ_1 是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的最大特征值，故 $\lambda_1=\rho(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})$，因而(3.3-2)式的第一个等式成立.

又因 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})\boldsymbol{A}^{-1}$，故 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 与 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 相似，从而它们有相同的特征值。因此它们的最大特征值也相同，于是(3.3-2)式的第二个等式成立.

3.3 矩阵范数

当 $\boldsymbol{A}$ 是 Hermite 矩阵时，$\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$，故

$$\|\boldsymbol{A}\|_2=\sqrt{\rho(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})}=\sqrt{\rho(\boldsymbol{A}^2)}=\sqrt{[\rho(\boldsymbol{A})]^2}=\rho(\boldsymbol{A}). \quad \blacksquare$$

定理 3.3-2 设 $\boldsymbol{A}$ 是 n 阶方阵，则对任意 $\varepsilon>0$，存在某种矩阵范数 $\|\cdot\|_*$，使

$$\|\boldsymbol{A}\|_* \leqslant \rho(\boldsymbol{A})+\varepsilon. \tag{3.3-3}$$

证 由本章 2.1 节知，存在可逆矩阵 $\boldsymbol{P}$ 使 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}=\mathrm{diag}(\boldsymbol{J}_1,\cdots,\boldsymbol{J}_s)$，其中

$$\boldsymbol{J}_i=\begin{bmatrix}\lambda_i & 1 & & \\ & \lambda_i & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i\end{bmatrix}_{r_i\times r_i}, \quad i=1,2,\cdots,s.$$

令 $r_i\times r_i$ 矩阵 $\boldsymbol{D}_i$ 为如下对角矩阵：

$$\boldsymbol{D}_i=\mathrm{diag}(1,\varepsilon,\cdots,\varepsilon^{r_i-1}) \quad (i=1,2,\cdots,s),$$

那么有

$$\boldsymbol{D}_i^{-1}\boldsymbol{J}_i\boldsymbol{D}_i=\begin{bmatrix}\lambda_i & \varepsilon & & \\ & \lambda_i & \ddots & \\ & & \ddots & \varepsilon \\ & & & \lambda_i\end{bmatrix}, \quad i=1,2,\cdots,s.$$

再令 $\boldsymbol{D}=\mathrm{diag}(\boldsymbol{D}_1,\cdots,\boldsymbol{D}_s)$，则

$$\|\boldsymbol{D}^{-1}\boldsymbol{J}\boldsymbol{D}\|_\infty=\|\boldsymbol{D}^{-1}\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\boldsymbol{D}\|_\infty=\max_{1\leqslant i\leqslant n}(|\lambda_i|+\varepsilon)=\rho(\boldsymbol{A})+\varepsilon.$$

3.3 矩阵范数

另一方面，由于

$$\|(D^{-1}JD)^H(D^{-1}JD)\|_\infty \leqslant \max_{1\leqslant j\leqslant n}(|\lambda_j|+\varepsilon)^2 = (\rho(A)+\varepsilon)^2,$$

所以有
$$\|D^{-1}P^{-1}APD\|_2^2 = \|D^{-1}JD\|_2^2 = \rho[(D^{-1}JD)^H(D^{-1}JD)]$$
$$\leqslant \|(D^{-1}JD)^H(D^{-1}JD)\|_\infty \leqslant (\rho(A)+\varepsilon)^2,$$

两端开方后得
$$\|D^{-1}P^{-1}APD\|_2 \leqslant \rho(A)+\varepsilon.$$

定义 $\|A\|_* = \|D^{-1}P^{-1}APD\|_2$，显然，$\|\cdot\|_*$ 是矩阵范数，因此(3.3-3)式成立. ∎

例 2 设 A 是可逆矩阵，证明

$$\mathrm{Cond}(A) \geqslant \rho(A)\rho(A^{-1}) = \frac{\max\limits_{\lambda\in S_p(A)}|\lambda|}{\min\limits_{\lambda\in S_p(A)}|\lambda|}.$$

证 由于 $\rho(A)\leqslant\|A\|$，$\rho(A^{-1})\leqslant\|A^{-1}\|$，所以

$$\mathrm{Cond}(A) = \|A\|\,\|A^{-1}\| \geqslant \rho(A)\rho(A^{-1}).$$

若 $\lambda_1,\cdots,\lambda_n$ 是 A 的所有特征值，则 A^{-1} 的所有特征值是 $\frac{1}{\lambda_1},\cdots,\frac{1}{\lambda_n}$. 因此

$$\rho(A) = \max_{\lambda\in Sp(A)}|\lambda|,$$

而
$$\rho(A^{-1}) = \max_{\lambda\in Sp(A)}\left|\frac{1}{\lambda}\right| = \frac{1}{\min\limits_{\lambda\in Sp(A)}|\lambda|},$$

作业（第2部分）

习题 3.3

1. 求 $\boldsymbol{A}=\begin{bmatrix}1+\mathrm{i} & 3\\ 2 & 1-\mathrm{i}\end{bmatrix}$ 的谱半径 $\rho(\boldsymbol{A})$.

线性方程组的数值解法

数值计算（9）

厦門大學
XIAMEN UNIVERSITY

信息学院
（国家示范性软件学院）
School of Informatics
黄 炜
博士·副教授
Dr. Wei Huang

2 线性方程组的数值解法

科学研究与工程技术中许多实际问题的数值求解，常常归结为求解线性方程组，如电路网络问题，船体放样中建立三次样条函数的问题，用最小二乘法求实验数据的拟合曲线问题，以及用差分法或有限元法求解微分方程等，都要求解线性方程组.

即使在社会科学，数量经济等领域，也会遇到求解线性方程组的问题，如投入产出分析等.

按系数矩阵阶数的高低和含零元素的多少，线性方程组可以分为两类：①低阶稠密方程组，即系数矩阵阶数不高，含零元素很少；②高阶稀疏方程组，即系数矩阵阶数高，零元素极多.

关于线性方程组的数值解法主要有两大类.

1. 直接法

这类方法是指在没有舍入误差影响的条件下，经过有限次四则运算可以求得方程组的准确解的一些方法. 但由于实际计算时舍入误差是不可避免的，所以它们也只能求得近似解.

2. 迭代法

这类方法是用某种极限过程去逐步逼近方程组的准确解的一些方法. 这类方法编程较容易，但要考虑迭代过程的收敛性、收敛速度等问题. 由于实际计算时，只能做有限步，从而得到的也是近似解，所以要估计截断误差的大小.

当系数矩阵是高阶稀疏矩阵时，一般优先考虑迭代法.

2.1 Gauss主元消去法

在矩阵论中已经讲过，可以通过消元或直接三角分解把线性方程组化为一个或两个三角形方程组，这就是直接法的基本原理.

设有线性方程组

$$\begin{cases} a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n=b_1, \\ a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n=b_2, \\ \qquad\qquad\vdots \\ a_{n1}x_1+a_{n2}x_2+\cdots+a_{nn}x_n=b_n, \end{cases} \tag{2.1-1}$$

其矩阵形式为

$$\boldsymbol{Ax}=\boldsymbol{b}. \tag{2.1-1'}$$

Gauss 消去法是对该线性方程组的增广矩阵$[\boldsymbol{A} \mid \boldsymbol{b}]$施以行初等变换，化 $\boldsymbol{A}$ 为上三角矩阵，从而得到三角形方程组，这是消去过程. 然后再回代求解. 其具体过程如下：

为了叙述清晰，将(2.1-1′)式改写为

$$\boldsymbol{A}^{(1)}\boldsymbol{x}=\boldsymbol{b}^{(1)}.$$

当 $a_{11}^{(1)}\neq 0$ 时，第一步消元后得

$$[\boldsymbol{A}^{(2)} \mid \boldsymbol{b}^{(2)}]=\left[\begin{array}{cccc|c} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} & b_1^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} & b_2^{(2)} \\ & \vdots & & \vdots & \vdots \\ & a_{n2}^{(2)} & \cdots & a_{nn}^{(2)} & b_n^{(2)} \end{array}\right],$$

2.1 Gauss主元消去法

所需计算的是：

(1) 行乘数 $l_{i1}=\dfrac{a_{i1}^{(1)}}{a_{11}^{(1)}}, i=2,3\cdots,n,$

(2) $\begin{cases} a_{ij}^{(2)}=a_{ij}^{(1)}-l_{i1}a_{1j}^{(1)}, i,j=2,3,\cdots,n, \\ b_i^{(2)}=b_i^{(1)}-l_{i1}b_1^{(1)}, i=2,3,\cdots,n. \end{cases}$

一般地，假定已完成了 $k-1$ 步消元，则得

$$[\boldsymbol{A}^{(k)} \mid \boldsymbol{b}^{(k)}]=\left[\begin{array}{cccccc:c} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & & \cdots & a_{1n}^{(1)} & b_1^{(1)} \\ & a_{22}^{(2)} & \cdots & & \cdots & a_{2n}^{(2)} & b_2^{(2)} \\ & & \ddots & & & \vdots & \vdots \\ & & & a_{kk}^{(k)} & \cdots & a_{kn}^{(k)} & b_k^{(k)} \\ & & & \vdots & & \vdots & \vdots \\ & & & a_{nk}^{(k)} & \cdots & a_{nn}^{(k)} & b_n^{(k)} \end{array}\right].$$

若 $a_{kk}^{(k)}\neq 0$，则可以进行第 k 步消元，从而得到$[\boldsymbol{A}^{(k+1)} \mid \boldsymbol{b}^{(k+1)}]$，需要计算的是：

$$\left.\begin{array}{l} (1)\ 行乘数\ l_{ik}=\dfrac{a_{ik}^{(k)}}{a_{kk}^{(k)}}, i=k+1,k+2,\cdots,n; \\ (2)\ \begin{cases} a_{ij}^{(k+1)}=a_{ij}^{(k)}-l_{ik}a_{kj}^{(k)}, i,j=k+1,k+2,\cdots,n, \\ b_i^{(k+1)}=b_i^{(k)}-l_{ik}b_k^{(k)}, i=k+1,k+2,\cdots,n. \end{cases} \end{array}\right\} \qquad (2.1\text{-}2)$$

2.1 Gauss主元消去法

这样完成 $n-1$ 步消元后，得到

$$[\boldsymbol{A}^{(n)} \mid \boldsymbol{b}^{(n)}] = \left[\begin{array}{cccc|c} a_{11}^{(1)} & a_{12}^{(1)} & \cdots & a_{1n}^{(1)} & b_1^{(1)} \\ & a_{22}^{(2)} & \cdots & a_{2n}^{(2)} & b_2^{(2)} \\ & & \ddots & \vdots & \vdots \\ & & & a_{nn}^{(n)} & b_n^{(n)} \end{array}\right].$$

当 $a_{nn}^{(n)} \neq 0$ 时，用回代过程求解上述三角形方程组，得

$$\begin{cases} x_n = \dfrac{b_n^{(n)}}{a_{nn}^{(n)}}, \\ x_i = \dfrac{b_i^{(i)} - \sum\limits_{j=i+1}^{n} a_{ij}^{(i)} x_j}{a_{ii}^{(i)}} \quad (i = n-1, n-2, \cdots, 1). \end{cases} \tag{2.1-3}$$

现在估计 Gauss 消去法的计算量.

由(2.1-2)式知，第 k 步消元的计算工作量为：

要做乘除的次数是

$$(n-k) + (n-k)^2 + (n-k) = (n-k)(n-k+2),$$

要做加减的次数是

$$(n-k)^2 + (n-k) = (n-k)(n-k+1).$$

2.1 Gauss主元消去法

因此，消元过程的总运算量是

$$\text{乘除次数} = \sum_{k=1}^{n-1}(n-k)(n-k+2) = \frac{n^3}{3}+\frac{n^2}{2}-\frac{5}{6}n,$$

$$\text{加减次数} = \sum_{k=1}^{n-1}(n-k)(n-k+1) = \frac{n^3}{3}-\frac{n}{3}.$$

由(2.1-3)式知，回代求解所需的总运算量是

$$\text{乘除次数} = 1+\sum_{i=1}^{n-1}(n-i+1) = \sum_{i=1}^{n}(n-i+1) = \frac{n^2}{2}+\frac{n}{2},$$

$$\text{加减次数} = \sum_{i=1}^{n-1}[1+(n-i-1)] = \sum_{i=1}^{n-1}(n-i) = \frac{n^2}{2}-\frac{n}{2}.$$

因此，当 n 很大时，求解线性方程组(2.1-1)的 Gauss 消去法所需的总的乘除次数和加减次数分别为

$$\text{乘除次数} = \frac{n^3}{3}+n^2-\frac{n}{3}\approx\frac{n^3}{3},$$

$$\text{加减次数} = \frac{n^3}{3}+\frac{n^2}{2}-\frac{5}{6}n\approx\frac{n^3}{3},$$

并且主要的计算工作量是在消元过程中，相对于消元过程的运算量，回代求解过程的运算量可忽略不计.

在 Gauss 消去法中，$a_{kk}^{(k)}(k=1,2,\cdots,n-1)$称为**主元**. 消元过程中若出现 $a_{kk}^{(k)}=0$，则消元法不能进行下去，即使 $a_{kk}^{(k)}\neq 0$，但绝对值很小，把它作为除数会使其他元素的数量级急剧增大，因而舍入误差随之增大，致使结果不可靠.

2.1 Gauss主元消去法

实用的 Gauss 消去法应该在消元过程的每一步，都在可能范围内选择绝对值较大的元素作为主元，以避免舍入误差的增长，这就是 Gauss 主元消去法.

常用的一种主元消去法是按列选主元，即对 $1\leqslant k\leqslant n-1$，在消去过程的第 k 步，选取 $c_k=\max\limits_{k\leqslant i\leqslant n}|a_{ik}^{(k)}|$ 作为主元，并将其所在的行与第 k 行对换后再进行第 k 步消元. 这样便保证每个行乘数的绝对值都不大于 1.

Gauss 列主元消去法的算法

1) 消元过程　消元结果冲掉 $\mathbf{A}$，行乘数冲掉 $a_{ik}(i>k)$，det 存放行列式.

1° $1\Rightarrow \det$.

2° 对于 $k=1,2,\cdots,n-1$，做①～⑤：

① 按列选主元 c_k，即确定 i_k 使

$$c_k=|a_{i_k k}|=\max_{k\leqslant i\leqslant n}|a_{ik}|.$$

② 若 $c_k=0$，则输出错误信息；停机.

③ 若 $i_k=k$，转(4)，否则执行换行：

$$a_{kj}\Leftrightarrow a_{i_k,j}\quad (j=k,k+1,\cdots,n),$$

$$b_k\Leftrightarrow b_{i_k},\qquad -\det\Rightarrow \det.$$

④ 消元计算：

2.1 Gauss主元消去法

$$l_{ik}=\frac{a_{ik}}{a_{kk}}\Rightarrow a_{ik}\qquad (i=k+1,\cdots,n),$$

$$a_{ij}-l_{ik}a_{kj}\Rightarrow a_{ij}\qquad (i,j=k+1,\cdots,n),$$

$$b_i-l_{ik}b_k\Rightarrow b_i\qquad (i=k+1,\cdots,n).$$

⑤ $a_{kk}\cdot\det\Rightarrow\det$.

3° 若 $a_{nn}=0$，则判断 b_n 是否等于零：

① $b_n\neq 0$，则输出“无解”信息；停机.

② $b_n=0$，则输出“非唯一解”，停机.

若 $a_{nn}\neq 0$，则 $a_{nn}\det\Rightarrow\det$.

2）回代过程　解冲掉 $\boldsymbol{b}$.

$$x_n=\frac{b_n}{a_{nn}}\Rightarrow b_n,$$

$$x_i=\frac{b_i-\sum\limits_{j=i+1}^{n}a_{ij}b_j}{a_{ii}}\Rightarrow b_i\quad (i=n-1,\cdots,2,1).$$

3）输出 $x_1,x_2,\cdots,x_n$ 及 det. det 值的大小可以作为方程组是否病态的参考. ▌

2.1 Gauss主元消去法

例 1 用舍入三位有效数字求解线性方程组

$$\begin{cases} 0.0300x_1 + 58.9x_2 = 59.2, \\ 5.31x_1 - 6.10x_2 = 47.0. \end{cases}$$

（准确解是 $x_1 = 10.0, x_2 = 1.00$）

解 1）不选主元的 Gauss 消去法.

$$\left[\begin{array}{cc:c} 0.0300 & 58.9 & 59.2 \\ 5.31 & -61.0 & 47.0 \end{array}\right] \to \left[\begin{array}{c:c:c} 0.0300 & 58.9 & 59.2 \\ \hdashline 177 & -1.04\times 10^4 & -1.05\times 10^4 \end{array}\right],$$

其中左下角虚线下的数是行乘数 $l_{21} = \dfrac{5.31}{0.0300} = 177$.

回代求解，得

$$x_2 = \frac{-1.05\times 10^4}{-1.04\times 10^4} = 1.01, \quad x_1 = \frac{(59.2 - 58.9x_2)}{0.0300} = -10.0.$$

我们看到，x_2 仅有 0.01 的误差，但在求 x_1 时，由于“小主元”作除数而产生 x_1 的计算结果失真，所以这个解是无效的.

2）按列选主元的 Gauss 主元消去法.

$$\left[\begin{array}{cc:c} 0.0300 & 58.9 & 59.2 \\ 5.31 & -6.10 & 47.0 \end{array}\right] \to \left[\begin{array}{cc:c} 5.31 & -6.10 & 47.0 \\ 0.0300 & 58.9 & 59.2 \end{array}\right] \to \left[\begin{array}{c:c:c} 5.31 & -6.10 & 47.0 \\ \hdashline 0.00565 & 58.9 & 58.9 \end{array}\right],$$

回代求解，得 $x_2 = \dfrac{58.9}{58.9} = 1.00, \quad x_1 = \dfrac{(47.0 + 6.10x_2)}{5.31} = 10.0.$

这个解中 x_1、x_2 的三位数字都是有效数字.

2.1 Gauss主元消去法

当线性方程组的系数矩阵的元素大小相差很大时，则可能引进因丢失有效数字而产生的误差，并且舍入误差的影响严重，即使用 Gauss 主元消去法得到的解也不可靠. 为避免这一问题，可将系数矩阵的行元素按比例增减以使元素的变化减小. 例如，用每行元素的最大模除该行各元素，使它们的模都不大于 1，这叫**行标度化**. 标度化的目的是要找到真正的主元. 消元过程仍是对原方程组进行的，只不过在 Gauss 列主元消去法的算法中，按列选主元 c_k 时，应修改为

$$c_k = \max_{k \leqslant i \leqslant n} \frac{|a_{ik}^{(k)}|}{s_i^{(k)}},$$

其中 $$s_i^{(k)} = \max_{k \leqslant j \leqslant n} |a_{ij}^{(k)}|, \quad i = k, k+1, \cdots, n.$$

作业（第2部分）

习题 2

1. 用 Gauss 消去法和列主元消去法求解

$$\begin{cases} 4x_1 - 1.24x_2 + 0.3x_3 = -11.04, \\ 2x_1 + 4.5x_2 + 0.36x_3 = 0.02, \\ 0.5x_1 + 1.1x_2 + 3.1x_3 = 6. \end{cases}$$

謝謝觀看！

厦門大學
XIAMEN UNIVERSITY
信息学院
（国家示范性软件学院）
School of Informatics
黃 煒
博士，副教授
Dr. Wei Huang