QR 分解与非线性特征值问题*

李 仁 仓

（中国科学院计算中心）

QR DECOMPOSITION AND NONLINEAR EIGENVALUE PROBLEMS

Li Ren-cang

(Computing Center, Academia Sinica)

Abstract

For a given differentiable functional λ-matrix A(λ) of order n, some differentiability theorems for elements arising in its QR decompositions are proved, and several explicit expressions for some derivatives are given. Based on these results, an algorithm for solving nonlinear eigenvalue problems is proposed. By the way, it is pointed out that a counterexample given in [3] contains a fatal error. Also, extensions of these results to cases where A(λ) is not square or λ is not a single variable are made. Sevaral numerical tests are presented.

§1. 引 言

考察 $m\times n$ 矩阵 $A(\lambda)$，其中元素 $a_{ij}(\lambda)$ 均为复(实)变量 λ 的解析（至少有一阶导数)函数．称此类矩阵为泛函 λ-矩阵．特别，当 $a_{ij}(\lambda)$ 是 λ 的多项式时，$A(\lambda)$ 就是熟知的 λ-矩阵[4]．给定 $A(\lambda)\in C^{n\times n}(m=n)$，有时需确定其非线性特征值及其相应的特征向量，即求满足

$$A(\lambda)x=0 \tag{1.1}$$

和

$$y^H A(\lambda)=0 \tag{1.2}$$

的复数 λ 和 n 维向量 x，y（x，y 分别称为右、左特征向量)．今后使用的符号在本节末加以说明．显然所有特征值是函数 $\det A(\lambda)$ 的零点．

有些迭代法目前用于对此问题进行求解(见[1]—[4]，[7])，本文主要是研究 Kublanovskaya[1] 提出的一种迭代法：对某点 λ，对 $A(\lambda)$ 进行列选主元 QR 分解[9]，得

* 1987 年 7 月 23 日收到．

$$A(\lambda)\pi(\lambda) = Q(\lambda)R(\lambda), \tag{1.3}$$

其中 $\pi(\lambda)$ 是在进行列选主元中由列互换确定的置换阵，$Q(\lambda)$ 和 $R(\lambda)$ 分别是 $n\times n$ 阶酉矩阵和上三角阵．由列选主元可知，$R(\lambda)$ 的对角元满足

$$|r_{ii}(\lambda)| \geqslant |r_{ij}(\lambda)|,\ n \geqslant j > i \geqslant 1. \tag{1.4}$$

有了分解式(1.3)，然后假定 $\pi(\lambda)$ 在 λ 的一个充分小邻域内保持不变，确定 $r_{nn}(\lambda)$ 的“导数值”，最后用 Newton 法求 $r_{nn}(\lambda)$ 的零点及左右特征向量[1]

Jain-Singhal[3] 发现 Kublanovskaya 方法并非 Newton 法．事实上，[1]中给出的 $r_{nn}(\lambda)$ 的“导数值”在复数域上一般是不正确的． Kublanovskaya 方法的最终二次收敛性是由于它是 Rayleigh 商迭代的一种特殊情形[2]，而后者是最终二次收敛的[11]

[3]认为 $R(\lambda)$ 的元素(特别是对角元)通常是不可微的，因此，[3]提出了另外一种方法．这种方法在求 $\det A(\lambda)$ 的单重零点时有效．§2 证明，(1.3)中 $R(\lambda)$ 的对角元在某种条件下是可微的，并指出[3]中的反例有错误．

§3 是利用§2 的结论，构造一求解 (1.1)—(1.2) 的算法．此外，§§2 和 3 还讨论了 $m\times n$ $(m\geqslant n)$ 阶泛函 λ-矩阵．§4 是注记，特别考虑了

$$\lambda = (\lambda_1,\ \cdots,\ \lambda_l) \in \mathbf{C}\times\mathbf{C}\times\cdots\times\mathbf{C}$$

并非单变量的情形．§5 是数值例子．

符号 $\mathbf{C}^{m\times n}$表示 $m\times n$ 阶复数域上的矩阵全体；$\mathbf{C}^m = \mathbf{C}^{m\times 1}$；$\mathbf{C} = \mathbf{C}^1$；$\mathscr{U}_n$ 表示 $\mathbf{C}^{n\times n}$ 中酉矩阵的全体；$I^{(n)} \in \mathbf{C}^{n\times n}$ 表示单位阵，在阶数 n 不至于引起混淆的情形下常不标出；$I_j^{(n)} \in \mathbf{C}^{n\times j}$ $(1\leqslant j\leqslant n)$ 表示由 $I^{(n)}$ 的前 j 列组成的矩阵，$e_j^{(n)} \in \mathbf{C}^{n\times 1}$ 表示 $I^{(n)}$ 的第 j 列，有时会略去上标 (n)；上标 H, T 和$+$分别表示矩阵的共轭转置、转置和 Moore-Penrose 广义逆；$\equiv$表示恒等或定义．

§2. $R(\lambda)$ 对角元的可微性

[3]中曾以如下 2×2 阶泛函 λ-矩阵：

$$A(\lambda) = \begin{pmatrix} \lambda & 1 \\ 1 & \lambda \end{pmatrix} \tag{2.1}$$

为例，说明 $R(\lambda)$ 中的对角元是不可微的．[3]得此结论是由于其中对(2.1)的 QR 分解本身是错误的．事实上

$$A(\lambda) \neq \alpha\begin{pmatrix} \lambda\bar\lambda & \bar\lambda \\ \lambda & -\lambda\bar\lambda \end{pmatrix}\cdot\alpha\begin{pmatrix} \lambda(1+\lambda\bar\lambda) & \lambda(\lambda+\bar\lambda) \\ 0 & \bar\lambda(1-\lambda^2) \end{pmatrix}, \tag{2.2}$$

其中 $\alpha = [\lambda\bar\lambda(1+\lambda\bar\lambda)]^{-\frac{1}{2}}$，因为(2.2)右边两个矩阵的积位于(1,2)的元素为

$$\alpha^2[\lambda^2\bar\lambda(\lambda+\bar\lambda) + \bar\lambda^2(1-\lambda^2)] = \alpha^2[\bar\lambda^2 + \lambda^3\bar\lambda] \neq 1,\ \lambda \neq 0.$$

不难验证

$$A(\lambda) = (1+\lambda\bar\lambda)^{-\frac{1}{2}}\begin{pmatrix} \lambda & 1 \\ 1 & -\lambda \end{pmatrix}\cdot(1+\lambda\bar\lambda)^{-\frac{1}{2}}\begin{pmatrix} 1+\bar\lambda\lambda & \lambda+\bar\lambda \\ 0 & 1-\lambda^2 \end{pmatrix} \tag{2.3}$$

是 $A(\lambda)$ 的一个(列选主元的) QR 分解．显然，上式右边的上三角阵的对角元素，满足

$$(1+\lambda\bar\lambda)^{-\frac{1}{2}}(1+\lambda\bar\lambda) = 1 + O(|\lambda|^2),$$

$$(1+\bar{\lambda}\lambda)^{-\frac{1}{2}}(1-\lambda^2)=1+O(|\lambda|^2),$$

因而在 $\lambda=0$ 点是可微的.

定理 2.1. 设 $A(\lambda)\in\mathbb{C}^{n\times n}$ 是可微的泛函 λ-矩阵. 集合 $\rho(A)$ 表示其正则点集, 即

$$\rho(A)=\{\lambda\in\mathbb{C}:\ \det A(\lambda)\neq 0\},\tag{2.4}$$

则对任意 $\lambda_0\in\rho(A)$, 存在 λ_0 的一个邻域 $\mathscr{B}(\lambda_0)\subset\rho(A)$ ($\rho(A)$ 是 $\mathbb{C}$ 上开集), 使得 $A(\lambda)$ 有一 QR 分解

$$A(\lambda)=Q(\lambda)R(\lambda),\ \lambda\in\mathscr{B}(\lambda_0),\tag{2.5}$$

满足 $R(\lambda)$ 的对角元素在 $\lambda=\lambda_0$ 点是可微的.

证明. 对 $A(\lambda_0)$ 进行 QR 分解:

$$A(\lambda_0)=Q_0R_0,\ R_0=(r_{ij}^{(0)})_{n\times n},\tag{2.6}$$

于是, 当 $|\lambda-\lambda_0|$ 充分小时,

$$Q_0^HA(\lambda)=Q_0^HA(\lambda_0)+Q_0^H\frac{d}{d\lambda}A(\lambda)|_{\lambda=\lambda_0}(\lambda-\lambda_0)+o(|\lambda-\lambda_0|),\tag{2.7}$$

其中 $\frac{d}{d\lambda}A(\lambda)$ 表示对 $A(\lambda)$ 的每个元素分别求导所得的 $n\times n$ 矩阵. 记

$$Q_0^HA(\lambda)\equiv R_0+E,\tag{2.8}$$

其中

$$E=Q_0^H\frac{d}{d\lambda}A(\lambda)\Big|_{\lambda=\lambda_0}(\lambda-\lambda_0)+o(|\lambda-\lambda_0|).\tag{2.9}$$

现在分块如下:

$$R_0=\begin{array}{cc} \begin{array}{cc} n-1 & 1 \end{array} & \\ \left(\begin{array}{c|c} R_{11} & R_{12}\\ \hline 0 & r_{nn}^{(0)}\end{array}\right) & \begin{array}{c} n-1\\ 1\end{array}\end{array},\ E=\begin{array}{cc} \begin{array}{cc} n-1 & 1 \end{array} & \\ \left(\begin{array}{c|c} E_{11} & E_{12}\\ \hline E_{21} & e_{nn}\end{array}\right) & \begin{array}{c} n-1\\ 1\end{array}\end{array}.\tag{2.10}$$

由(2.9)可知, 当 $|\lambda-\lambda_0|$ 充分小时, $(R_{11}+E_{11})^{-1}$ 存在, 故可令

$$P=E_{21}(R_{11}+E_{11})^{-1}\in\mathbb{C}^{1\times(n-1)}.\tag{2.11}$$

不难验证

$$Q_1^HQ_0^HA(\lambda)=\begin{pmatrix}(I+P^HP)^{-\frac{1}{2}} & 0\\ 0 & (1+PP^H)^{-\frac{1}{2}}\end{pmatrix}\begin{pmatrix}R_{11}+E_{11}+P^HE_{21} & R_{12}+E_{12}+P^H(r_{nn}^{(0)}+e_{nn})\\ 0 & r_{nn}^{(0)}+e_{nn}-P(R_{12}+E_{12})\end{pmatrix}$$

$$\equiv R_1(\lambda)\equiv\begin{pmatrix}R_{11}(\lambda) & *\\ 0 & r_{nn}(\lambda)\end{pmatrix},\tag{2.12}$$

其中

$$Q_1^H\equiv\begin{pmatrix}(I+P^HP)^{-\frac{1}{2}} & 0\\ 0 & (1+PP^H)^{-\frac{1}{2}}\end{pmatrix}\begin{pmatrix}I & P^H\\ -P & 1\end{pmatrix},\tag{2.13}$$

$$\begin{aligned}R_{11}(\lambda)&\equiv(I+P^HP)^{-\frac{1}{2}}(R_{11}+E_{11}+P^HE_{21})\\ &=R_{11}+I_{n-1}^tQ_0^H\frac{d}{d\lambda}A(\lambda)|_{\lambda=\lambda_0}I_{n-1}(\lambda-\lambda_0)+o(|\lambda-\lambda_0|),\end{aligned}\tag{2.14}$$

和

$$r_{nn}(\lambda)\equiv(1+PP^H)^{-\frac{1}{2}}[r_{nn}^{(0)}+e_{nn}-P(R_{12}+E_{12})]$$

$$-r_{nn}^{(0)}+\left(e_n^T Q_0^H \frac{d}{d\lambda} A(\lambda)\Big|_{\lambda=\lambda_0} e_n - e_n^T Q_0^H \frac{d}{d\lambda} A(\lambda)\Big|_{\lambda=\lambda_0} I_{n-1} R_{11}^{-1} R_{12}\right)(\lambda-\lambda_0)$$
$$+o(|\lambda-\lambda_0|). \tag{2.15}$$

下面用归纳法证明定理 2.1：当 $n=1$ 时，结论是显然的．设定理对 $n-1$ 成立，继而证明结论对 n 也成立．由上面的叙述可知，存在 λ_0 的一个邻域 $\mathscr{B}_1(\lambda_0)\subset\rho(A)$，使(2.11)—(2.15)对 $\lambda\in\mathscr{B}_1(\lambda_0)$ 时均成立．由 (2.14) 知，$R_{11}(\lambda)\in \mathbf{C}^{(n-1)\times(n-1)}$ 为可微的泛函 λ-矩阵，于是，根据归纳假设，存在 λ_0 的一个邻域 $\mathscr{B}_2(\lambda_0)$，$Q_1(\lambda)\in\mathscr{U}_{n-1}$ 和 $\tilde{R}_1(\lambda)\in\mathbf{C}^{(n-1)\times(n-1)}$，使

$$R_{11}(\lambda)=Q_1(\lambda)\tilde{R}_1(\lambda),\ \lambda\in\mathscr{B}_2(\lambda_0), \tag{2.16}$$

并且 $\tilde{R}_1(\lambda)$ 的对角元在 $\lambda=\lambda_0$ 点可微．令

$$Q(\lambda)=Q_0'Q_1\begin{pmatrix} Q_1(\lambda) & 0\\ 0 & 1\end{pmatrix}, \tag{2.17}$$

$$R(\lambda)=\begin{pmatrix} \tilde{R}_1(\lambda) & *\\ 0 & r_{nn}(\lambda)\end{pmatrix}, \tag{2.18}$$

则对 $\mathscr{B}(\lambda_0)=\mathscr{B}_1(\lambda_0)\cap\mathscr{B}_2(\lambda_0)$，(2.5) 式成立，并且 $R(\lambda)$ 的对角元在 $\lambda=\lambda_0$ 点可微．证毕．

推论 2.1．设 $A(\lambda)$ 和 $\rho(A)$ 如定理 2.1 所示，则对任意 $\lambda_0\in\rho(A)$ 及置换阵 $\pi\in\mathbf{C}^{n\times n}$，存在 λ_0 的一个邻域 $\mathscr{B}(\lambda_0)\subset\rho(A)$，使

$$A(\lambda)\pi=Q(\lambda)R(\lambda),\ Q(\lambda)\in\mathscr{U}_n,\ \lambda\in\mathscr{B}(\lambda_0). \tag{2.19}$$

满足上三角阵 $R(\lambda)$ 的对角元在 $\lambda=\lambda_0$ 点可微，且

$$e_n^T R(\lambda)e_n=e_n^T R_0 e_n+\Big[e_n^T Q_0^H \frac{d}{d\lambda}A(\lambda)\Big|_{\lambda=\lambda_0}\pi e_n$$
$$-e_n^T Q_0^H \frac{d}{d\lambda}A(\lambda)\Big|_{\lambda=\lambda_0}\pi I_{n-1}(I_{n-1}^T R_0 I_{n-1})^{-1} I_{n-1}^T R_0 e_n\Big](\lambda-\lambda_0)$$
$$+o(|\lambda-\lambda_0|), \tag{2.20}$$

其中 $Q_0=Q(\lambda_0)$ 和 $R_0=R(\lambda_0)$．

注 2.1．在分解式(1.3)中，若对某 $\lambda_0\in\rho(\lambda_0)$，存在一个充分小的邻域，使 $\pi(\lambda)$ 保持不变，则 $A(\lambda)$ 有一个形如 (1.3) 并满足 $R(\lambda)$ 的对角元于 $\lambda=\lambda_0$ 点可微的列选主元 QR 分解．

(2.15)和(2.20)给出了 $R(\lambda)$ 右下角元素的导数，是§ 3 构造算法的基础．此外，认真考察定理 2.1 的证明，可以看出其中最关键的一步是(2.11)式有意义，即要求当 $|\lambda-\lambda_0|$ 充分小时，$(R_{11}+E_{11})^{-1}$ 存在，而在定理 2.1 中这一点是由 $\lambda_0\in\rho(A)$ 来保证的，因为 $\lambda_0\in\rho(A)\Rightarrow R_{11}$ 可逆．定理 2.1 的证明是通过逐步约化来实现的，因此，结合 (2.14) 可知，定理 2.1 中 $\lambda_0\in\rho(A)$ 可减弱为 $\lambda_0\in\mathbf{C}$ 使 $A(\lambda_0)$ 的前 $n-1$ 列是线性无关的（如此即可保证(2.6)，(2.10)中的 R_{11}^{-1} 存在）．

推论 2.2．设 $\lambda_0\in\mathbf{C}$ 使得 $A(\lambda_0)$ 的前 $n-1$ 列线性无关，其余条件与定理 2.1 同，则定理 2.1 的所有结论均成立．对推论 2.1，有如下结果：若 $A(\lambda_0)\pi$ 的前 $n-1$ 列线性无关，其余条件不变，则所有结论仍成立．

引理 2.1. 设 $C \in \mathbb{C}^{n\times n}$ 的前 $n-1$ 列线性无关，

$$C = Q_1R_1 = Q_2R_2$$

是 C 的两个 QR 分解，则存在对角阵 $D \in \mathscr{U}_n$，使得

$$Q_1 = Q_2D, \quad R_1 = D^H R_2.$$

证明. 先设 $\operatorname{rank} C = n-1$，则

$$R_i = \begin{array}{cc} \begin{array}{cc} \scriptstyle n-1 & \scriptstyle 1 \end{array} & \\ \left(\begin{array}{c|c} R_{11}^{(i)} & * \\ \hline 0 & 0 \end{array}\right) & \begin{array}{c} \scriptstyle n-1 \\ \scriptstyle 1 \end{array} \end{array}, \quad i = 1, 2,$$

其中 $R_{11}^{(i)} \in \mathbb{C}^{(n-1)\times(n-1)}$ 为上三角阵且 $\operatorname{rank} R_{11}^{(i)} = n-1$. 令

$$\widetilde{R}_1 = \begin{array}{cc} \begin{array}{cc} \scriptstyle n-1 & \scriptstyle 1 \end{array} & \\ \left(\begin{array}{c|c} R_{11}^{(1)} & * \\ \hline 0 & 1 \end{array}\right) & \begin{array}{c} \scriptstyle n-1 \\ \scriptstyle 1 \end{array} \end{array},$$

则易知存在

$$\widetilde{R}_2 = \begin{array}{cc} \begin{array}{cc} \scriptstyle n-1 & \scriptstyle 1 \end{array} & \\ \left(\begin{array}{c|c} R_{11}^{(2)} & * \\ \hline 0 & * \end{array}\right) & \begin{array}{c} \scriptstyle n-1 \\ \scriptstyle 1 \end{array} \end{array},$$

使得(注意到 $Q_2^HQ_1R_1 = R_2$)

$$Q_2^HQ_1\widetilde{R}_1 = \widetilde{R}_2 \Rightarrow Q_2^HQ_1 = \widetilde{R}_2\widetilde{R}_1^{-1} \equiv D.$$

故 $D \in \mathscr{U}_n$ 为上三角阵，于是 D 必是对角酉矩阵. 由此即得结论. 至于 $\operatorname{rank} C = n$ 时的证明，可仿此导出. 证毕.

引理 2.2. 设 $C_1 \in \mathbb{C}^{n\times m}$ 的前 $n-1$ 列线性无关，$C_1 = Q_1R_1$ 是 C_1 的 QR 分解. 又 $C_2 \in \mathbb{C}^{n\times m}$，则对任意 $\varepsilon > 0$，只要 $\|C_1 - C_2\|$ 充分小，就存在 C_2 的某 QR 分解 $C_2 = Q_2R_2$，使得 $\|Q_1 - Q_2\| < \varepsilon$，$\|R_1 - R_2\| < \varepsilon$，其中 $\|\cdot\|$ 为某矩阵范数.

定理 2.2. 设 $A(\lambda) \in \mathbb{C}^{n\times m}$ 是可微的泛函 λ-矩阵，$\pi \in \mathbb{C}^{n\times m}$ 为置换阵，$A(\lambda)\pi = Q(\lambda)R(\lambda)$ 是 $A(\lambda)\pi$ 的 QR 分解. 记 $R(\lambda)$ 的主对角元为 $r_{ii}(\lambda)$，则对任一使 $A(\lambda_0)\pi$ 的前 $n-1$ 列线性无关的 λ_0，$|r_{ii}(\lambda_0)|$ 是唯一确定的，并且存在 λ_0 的某邻域 $\mathscr{B}(\lambda_0)$，使得 $\lambda \in \mathscr{B}(\lambda_0)$ 时

$$|r_{ii}(\lambda)| = |a_0^{(i)}(\lambda_0) + a_1^{(i)}(\lambda_0)(\lambda - \lambda_0) + o(|\lambda - \lambda_0|)|, \quad i = 1, \cdots, n,$$

其中函数 $|a_j^{(i)}(\lambda_0)|$ 也是由 λ_0 唯一确定且关于 λ_0 是连续的. 此外，若有

$$|r_{ii}(\lambda)| = |b_0^{(i)}(\lambda_0) + b_1^{(i)}(\lambda_0)(\lambda - \lambda_0) + o(|\lambda - \lambda_0|)|, \quad i = 1, \cdots, n,$$

则存在 $\eta_i(\lambda_0) \in \mathbb{C}$，$|\eta_i(\lambda_0)| = 1$，使

$$b_j^{(i)}(\lambda_0) = \eta_i(\lambda_0)a_j^{(i)}(\lambda_0), \quad i = 1, \cdots, n; \ j = 0, 1.$$

注 2.2. 当 $A(\lambda) \in \mathbb{C}^{n\times m}$ 是二阶可微泛函 λ-矩阵时，上述各式中出现的 $o(|\lambda - \lambda_0|)$ 均可用 $O(|\lambda - \lambda_0|^2)$ 代替.

引理 2.3. 设 $f(\lambda)$ 是 $\lambda \in \mathbb{C}$ 的函数，且有渐近式

$$f(\lambda) = |f_0(\lambda_0) + f_1(\lambda_0)(\lambda - \lambda_0) + O(|\lambda - \lambda_0|^2)|, \quad \lambda_0 \in \mathbb{C}, \tag{2.21}$$

满足 $|f_1(\lambda_0)|$ 关于 λ_0 连续. 若迭代 $\lambda_{s+1} = \lambda_s - f_0(\lambda_s)/f_1(\lambda_s)$ $(s = 0, 1, \cdots)$ 收敛于 λ_*，其中 $f(\lambda_*) = f_0(\lambda_*) = 0$，$|f_1(\lambda_*)| \neq 0$，则收敛最终是二次的.

证明. 首先注意 $f_1(\lambda_*) \neq 0$，$|f_1(\lambda)|$ 是连续的，故当 s 充分大时，有

$$f(\lambda_{s+1}) = O([f(\lambda_s)]^2).$$

由此结合(2.21)式即得所论．证毕．

现在考察非方阵的情形．设 $B(\lambda)\in \mathbf{C}^{m\times n}$ $(m\geqslant n)$．$\rho(B)$ 表示 $B(\lambda)$ 的正则点集，即

$$\rho(B)=\{\lambda\in \mathbf{C}:\ \operatorname{rank}B(\lambda)=n\},$$

这里约定

$$\sup_{\lambda\in \mathbf{C}}\operatorname{rank}B(\lambda)=n. \tag{2.22}$$

对任意 $\lambda\in\rho(B)$，$B(\lambda)$ 有分解式([9])

$$B(\lambda)=Q(\lambda)R(\lambda), \tag{2.23}$$

其中 $Q(\lambda)\in\mathscr{U}_m$，

$$R(\lambda)=\begin{array}{l}\left(\begin{array}{ccc} r_{11}(\lambda) & \cdots & r_{1n}(\lambda)\\ & \ddots & \vdots\\ 0 & & r_{nn}(\lambda)\\ \hline & 0 & \end{array}\right)\begin{array}{l} n\\ \\ \\ m-n\end{array}\end{array}, \tag{2.24}$$

或列选主元分解

$$B(\lambda)\pi(\lambda)=Q(\lambda)R(\lambda). \tag{2.25}$$

为方便起见，也称 $r_{ii}(\lambda)$ $(1\leqslant i\leqslant n)$ 为形如(2.24)的矩阵 $R(\lambda)$ 的对角元．

定理 2.3. 设 $B(\lambda)\in\mathbf{C}^{m\times n}$ 是可微的，$\rho(B)$ 非空，则对任意 $\lambda_0\in\rho(B)$，存在某邻域 $\mathscr{B}(\lambda_0)\subset\rho(B)$，使 $B(\lambda)$ 有如下 QR 分解：

$$B(\lambda)=Q(\lambda)R(\lambda),\quad \lambda\in\mathscr{B}(\lambda_0). \tag{2.26}$$

满足 $R(\lambda)$ 的主对角元素在 $\lambda=\lambda_0$ 点可微．

证明．因为 $\operatorname{rank}B(\lambda_0)=n$，故存在 $B_{\lambda_0}\in\mathbf{C}^{m\times(m-n)}$ 使 $\det\,(B(\lambda_0),\ B_{\lambda_0})\neq 0$．因此，存在 λ_0 的充分小邻域 $\mathscr{B}_1(\lambda_0)$，使

$$\det\,(B(\lambda),\ B_{\lambda_0})\neq 0,\quad \lambda\in\mathscr{B}_1(\lambda_0). \tag{2.27}$$

由定理 2.1 知，存在 λ_0 的某邻域 $\mathscr{B}(\lambda_0)\subset\mathscr{B}_1(\lambda_0)$，使得如下 QR 分解

$$(B(\lambda),\ B_{\lambda_0})=Q(\lambda)\tilde{R}(\lambda),\quad \lambda\in\mathscr{B}(\lambda_0) \tag{2.28}$$

满足 $\tilde{R}(\lambda)$ 的对角元在 $\lambda=\lambda_0$ 点可微．现将 $\tilde{R}(\lambda)$ 分块

$$\tilde{R}(\lambda)=(\overset{n}{R(\lambda)}\,|\,\overset{m-n}{R_1(\lambda)}),$$

则比较(2.28)式两端，即得(2.26)，并且满足定理的要求．证毕．

推论 2.3. 设 $B(\lambda)$ 和 $\rho(B)$ 同定理 2.2，则对任意 $\lambda_0\in\rho(B)$ 及置换 π，存在 λ_0 的某邻域 $\mathscr{B}(\lambda_0)$，使得 $B(\lambda)\pi$ 有如下 QR 分解：

$$B(\lambda)\pi=Q(\lambda)R(\lambda),\quad \lambda\in\mathscr{B}(\lambda_0); \tag{2.29}$$

满足 $R(\lambda)$ 的主对角元在 $\lambda=\lambda_0$ 点可微．

3. 算　　法

利用定理 2.1 及证明，可以得出计算 $n\times n$ 阶可微泛函 λ-矩阵的特征值的算法．

算法 1. 计算 $A(\lambda)\in\mathbf{C}^{n\times n}$ 的非线性特征值．

a) 给定可微的 $A(\lambda)\in\mathbf{C}^{n\times n}$ 及其非线性特征值的初始近似值 μ_0．

b) 计算 $A(\mu_i)$ 和 $\frac{d}{d\lambda}A(\lambda)|_{\lambda=\mu_i}\equiv A_1^{(i)}$, $i=0, 1, 2, \cdots$.

c) 对 $A(\mu_i)$ 进行列选主元 QR 分解:

$$A(\mu_i)\pi_i = Q_i R_i,$$

其中 $Q_i\in\mathscr{U}_n$,

$$R_i=\begin{array}{cc} & \begin{array}{cc} n-1 & 1 \end{array} \\ & \left(\begin{array}{c|c} R_{11}^{(i)} & R_{12}^{(i)} \\ \hline 0 & r_{nn}^{(i)} \end{array}\right)\begin{array}{l} n-1 \\ 1 \end{array} \end{array}.$$

d) 计算

$$r_{nn}^{(i)\prime}=e_n^T Q_i^H A_1^{(i)}\pi_i e_n - e_n^T Q_i^H A_1^{(i)}\pi_i I_{n-1}R_{11}^{(i)^{-1}}R_{12}^{(i)}.$$

e) 计算

$$\mu_{i+1}=\mu_i - r_{nn}^{(i)}/r_{nn}^{(i)\prime}.$$

f) 如果已达所需精度,结束迭代;否则转 b).

若需要相应的左右特征向量,则只需在近似值 μ_k 满足精度要求后,利用 c) 中的分解,可以容易地得出(见[1]或[3]). 对于单重特征值 λ_*, 当 $A(\lambda)$ 二阶可微时,算法 1 是最终二阶收敛的. 这是因为若将 $r_{nn}(\lambda)$ 表示成(见推论 2.2, 定理 2.2)

$$|r_{nn}(\lambda)| = |a_1^{(m)}(\lambda_*)(\lambda-\lambda_*)+O(|\lambda-\lambda_*|^2)|,$$

则 $a_1^{(m)}(\lambda_*)\neq 0$. 再利用引理 2.3, 便得结论.

注 3.1. 在算法 1 开始几步迭代中, 可以使用修改 Newton 法, 以期得到较好的效果,如用 Newton 下山法[10]控制步长. 此外, 对重根也可作了适当地处理[10].

注 3.2. 由于 $r_{nn}^{(0)}+e_{nn}-P(R_{12}+E_{12})$ (见(2.12))的可微性与 $A(\lambda)$ 相同,因此,当 $A(\lambda)$ 具有较高阶可微性时,可以利用它来构造收敛阶较高的算法.

下面讨论如何计算 $B(\lambda)\in C^{m\times n}$ $(m\geqslant n)$ 的非正则点(若 $\rho(B)\neq C$).

算法 2. 计算 $B(\lambda)\in C^{m\times n}$ $(m\geqslant n)$ 的非正则点:

a) 给定可微的 $B(\lambda)\in C^{m\times n}$ 及 $\lambda_*\notin\rho(B)$ 的初始近似值 μ_0.

b) 计算 $B(\mu_i)$ 和 $\frac{d}{d\lambda}B(\lambda)\Big|_{\lambda=\mu_i}=B_1^{(i)}$, $i=0, 1, 2, \cdots$.

c) 对 $B(\mu_i)$ 进行列选主元 QR 分解(见(2.25))

$$B(\mu_i)\pi_i = Q_i R_i,$$

其中 $Q_i\in\mathscr{U}_m$,

$$R_i=\begin{array}{cc} & \begin{array}{cc} n-1 & 1 \end{array} \\ & \left(\begin{array}{c|c} R_{11}^{(i)} & R_{12}^{(i)} \\ \hline & r_{nn}^{(i)} \\ 0 & \\ \hline & 0 \end{array}\right)\begin{array}{l} n-1 \\ 1 \\ m-n \end{array} \end{array}.$$

d) 计算 $Q_i^H B_1^{(i)}\pi_i$, 并分块

$$Q_i^H B_1^{(i)}\pi_i=\begin{array}{cc} & \begin{array}{cc} n-1 & 1 \end{array} \\ & \left(\begin{array}{c|c} B_{11}^{(i)} & B_{12}^{(i)} \\ B_{21}^{(i)} & B_{22}^{(i)} \end{array}\right)\begin{array}{l} n-1 \\ m-n+1 \end{array} \end{array}.$$

e) 计算

$$r^{(i)}=B_{22}^{(i)}-B_{21}^{(i)}R_{11}^{(i)-1}R_{12}^{(i)}=(\xi_1^{(i)},\ \cdots,\ \xi_{m-n+1}^{(i)})^T\in\mathbf{C}^{m-n+1}.$$

f) 计算

$$\mu_{i+1}=\mu_i-\xi_1^{(i)}r_{nn}^{(i)}/r^{(i)H}r^{(i)}. \tag{3.1}$$

g) 若满足精度要求，结束迭代；否则转 b).

递推关系(3.1)是上述算法的关键，求得的步骤如下(为方便起见，略去上标 i)：

首先，注意到

$$\begin{aligned}Q_i^HB(\lambda)\pi_i&=R_i+Q_i^H\left.\frac{d}{d\lambda}B(\lambda)\right|_{\lambda=\mu_i}\pi_i(\lambda-\mu_i)+o(|\lambda-\mu_i|)\\&=R_i+Q_i^HB_1\pi_i(\lambda-\mu_i)+o(|\lambda-\mu_i|)\equiv R_i+F,\end{aligned}$$

其中 $F\to 0(|\lambda-\mu_i|\to 0)$. 现将 F 分块

$$F=\begin{array}{c}\begin{array}{cc}\scriptstyle n-1 & \scriptstyle 1\end{array}\\ \left(\begin{array}{c|c}F_{11}&F_{12}\\\hline F_{21}&F_{22}\end{array}\right)\begin{array}{l}\scriptstyle n-1\\ \scriptstyle m-n+1\end{array}\end{array},$$

则

$$R_i+F=\begin{array}{c}\begin{array}{cc}\scriptstyle n-1 & \scriptstyle 1\end{array}\\ \left(\begin{array}{c|c}R_{11}+F_{11}&R_{12}+F_{12}\\\hline F_{21}&r_{nn}e_1+F_{22}\end{array}\right)\begin{array}{l}\scriptstyle n-1\\ \scriptstyle m-n+1\end{array}\end{array}.$$

其次，令

$$P=F_{21}(R_{11}+F_{11})^{-1}\in\mathbf{C}^{(m-n+1)\times(n-1)},$$

$$\tilde{Q}^H=\begin{pmatrix}(I+P^HP)^{-\frac{1}{2}}&0\\0&(I+PP^H)^{-\frac{1}{2}}\end{pmatrix}\begin{pmatrix}I&P^H\\-P&I\end{pmatrix},$$

则，当 $|\lambda-\mu_i|$ 充分小时

$$\begin{aligned}\tilde{Q}^HQ_i^HB(\lambda)\pi_i&=\begin{pmatrix}(I+P^HP)^{-\frac{1}{2}}&0\\0&(I+PP^H)^{-\frac{1}{2}}\end{pmatrix}\begin{pmatrix}R_{11}+F_{11}+P^HF_{21}&F_{12}+R_{12}+P^H(r_{nn}e_1+F_{22})\\0&r_{nn}e_1+F_{22}-P(F_{12}+R_{12})\end{pmatrix}\\&\equiv\begin{array}{c}\begin{array}{cc}\scriptstyle n-1 & \scriptstyle 1\end{array}\\ \left(\begin{array}{c|c}\tilde{R}_{11}&*\\\hline 0&\tilde{r}\end{array}\right)\end{array}.\end{aligned}$$

易知

$$\tilde{r}=\begin{pmatrix}r_{nn}\\0\\\vdots\\0\end{pmatrix}+r(\lambda-\mu_i)+o(|\lambda-\mu_i|)\in\mathbf{C}^{m-n+1}.$$

为取 $\lambda\in\mathbf{C}$ 使 $\|\tilde{r}\|_2=\min$, 只考虑其线性项. 因此得到了最小二乘问题：求

$$\|r_{nn}e_1+r(\lambda-\mu_i)\|_2=\min \tag{3.2}$$

的极小解. 由此立得(3.1)[9].

注 3.3. 在算法 2 的 c) 步的分解中 R_i 不一定是“上三角阵”(见(2.24))，而只需分解成

$$R_i=\begin{array}{c}\begin{array}{cc}\scriptstyle n-1 & \scriptstyle 1\end{array}\\ \left(\begin{array}{c|c}R_{11}^{(i)}&R_{12}^{(i)}\\\hline 0&R_{22}^{(i)}\end{array}\right)\begin{array}{l}\scriptstyle n-1\\ \scriptstyle m-n+1\end{array}\end{array}.$$

最后，在 f) 步中，迭代式改为

$$\mu_{i+1} = \mu_i - r^{(i)H} R_{22}^{(i)} / r^{(i)H} r^{(i)}. \tag{3.3}$$

注 3.4. 当 $m > n$ 时,算法 2 的收敛比较复杂,并且即使在收敛的情形下,一般说收敛速度是不超过线性的.

§4. 当 λ 是多变量的情形

前面仅局限于单变量情形,即 $\lambda \in \mathbb{C}$. 现在设

$$\lambda = (\lambda_1, \cdots, \lambda_l) \in \overbrace{\mathbb{C} \times \mathbb{C} \times \cdots \times \mathbb{C}}^{l\text{个}}, \quad l > 1. \tag{4.1}$$

定理 4.1. 设 $A(\lambda) \in \mathbb{C}^{n \times m}$, λ 如(4.1)所示 $\lambda^{(0)} \in \mathbb{C} \times \cdots \times \mathbb{C}$ 使

$$\det A(\lambda^{(0)}) \neq 0. \tag{4.2}$$

若 $A(\lambda)$ 于 $\lambda^{(0)}$ 点可微,即

$$A(\lambda) = A(\lambda^{(0)}) + \sum_{i=1}^{l} \frac{\partial}{\partial \lambda_i} A(\lambda)\Big|_{\lambda=\lambda^{(0)}} (\lambda_i - \lambda_i^{(0)}) + o(\|\lambda - \lambda^{(0)}\|), \tag{4.3}$$

其中

$$\|\lambda - \lambda^0\| = \left(\sum_{i=1}^{l} |\lambda_i - \lambda_i^{(0)}|^2 \right)^{\frac{1}{2}},$$

则存在 $\lambda^{(0)}$ 的一个邻域 $\mathscr{B}(\lambda^{(0)})$, 使 $A(\lambda)\pi$ 有一 QR 分解

$$A(\lambda)\pi = Q(\lambda)R(\lambda), \quad \lambda \in \mathscr{B}(\lambda^{(0)}) \tag{4.4}$$

满足 $R(\lambda)$ 的对角元在 $\lambda = \lambda^{(0)}$ 点可微,并且

$$\begin{aligned} e_n^T R(\lambda) e_n = e_n^T R_0 e_n &+ \sum_{i=1}^{l} \Big[e_n^T Q_0^H \frac{\partial}{\partial \lambda_i} A(\lambda)\Big|_{\lambda=\lambda^{(0)}} \pi e_n \\ &- e_n^T Q_0^H \frac{\sigma}{\sigma \lambda_i} A(\lambda)\Big|_{\lambda=\lambda^{(0)}} \pi I_{n-1} (I_{n-1}^T R_0 I_{n-1})^{-1} I_{n-1}^T R_0 e_n \Big] (\lambda_i - \lambda_i^{(0)}) \\ &+ o(\|\lambda - \lambda^{(0)}\|), \end{aligned} \tag{4.5}$$

其中 π 为置换阵, $Q_0 = Q(\lambda^{(0)})$ 和 $R_0 = R(\lambda^{(0)})$.

利用这个定理,也可以得出计算 $\det A(\lambda)$ 的零点 $\lambda^{(*)}$ 的算法. 主要关键是 (4.5) 式. 采用列选主元 QR 分解,只考虑到(4.5)右边的线性项为止. 求解

$$\begin{cases} e_n^T R(\lambda) e_n \approx 0, \\ \|\lambda - \lambda^{(0)}\| = \min. \end{cases} \tag{4.6}$$

其迭代公式为

$$\lambda_i = \lambda_i^{(0)} - \frac{\bar{\alpha}_i}{\sum_{i=1}^{l} |\alpha_i|^2} e_n^T R_0 e_n, \quad 1 \leqslant i \leqslant l, \tag{4.7}$$

其中,简记

$$e_n^T R(\lambda) e_n = e_n^T R_0 e_n + \sum_{i=1}^{l} \alpha_i (\lambda_i - \lambda_i^{(0)}) + o(\|\lambda - \lambda^{(0)}\|).$$

由(4.7)不难写出算法,从略.

§§ 2—3 中的关于 $B(\lambda)\in C^{m\times n}$ $(m\geq n)$ 的一些结果（如定理 2.2）可以推广到 λ 为多变量的情形．值得注意的是，在推广算法 2 时，需要解类似于 (3.2) 的一个系数矩阵为 $(m-n+1)\times l$ 阶的最小二乘问题．

§ 5. 数值例子

考虑如下二次特征值问题([2—4])：

$$A(\lambda)=A_0+\lambda A_1+\lambda^2 A_2, \tag{5.1}$$

其中

$$A_0=\begin{pmatrix}121 & 18.9 & 15.9\\ 0 & 2.7 & 0.145\\ 11.9 & 3.64 & 15.5\end{pmatrix},\quad A_1=\begin{pmatrix}7.66 & 2.45 & 2.1\\ 0.23 & 1.04 & 0.223\\ 0.6 & 0.756 & 0.658\end{pmatrix},$$

$$A_2=\begin{pmatrix}17.6 & 1.28 & 2.89\\ 1.28 & 0.824 & 0.413\\ 2.89 & 0.413 & 0.725\end{pmatrix}.$$

这个问题有三对分离很好的非线性特征值（10 位有效数字）：

$$\lambda_1=-0.9179981715+1.760584204i,\quad \lambda_4=\bar{\lambda}_1,$$
$$\lambda_2=0.09472172578+2.522876588i,\quad \lambda_5=\bar{\lambda}_2,$$
$$\lambda_3=-0.8848302463+8.44151219i,\quad \lambda_6=\bar{\lambda}_3.$$

表 1 列出了用算法 1 和 Jain-Singhal 方法[3]在 IBM-PC/XT (AT) 上的计算结果．

表 1

No.	初始值	Jain-Singhal 法		算法 1	
		迭代次数	计算值	迭代次数	计算值
1	0+0.0001i	20	−0.8848303 −8.441512i	14	−0.9179981 +1.760584i
2	0.1+0.1i	9	−0.8848302 +8.441512i	7	−0.8848303 +8.441513i
3	−0.9+1.7i	3	−0.9179982 +1.760584i	3	−0.9179982 +1.760584i
4	−1.0+1.5i	4	−0.9179981 +1.760584i	3	−0.9179983 +1.760584i
5	0+2i	10	0.09472174 +2.522877i	7	−0.9179982 +1.760584i
6	0+2.5i	5	0.09472177 +2.522877i	4	0.09472174 +2.522877i
7	0+3i	7	−0.8848304 +8.441512i	8	0.09472174 +2.522877i
8	0+10i	3	−0.8848301 +8.441512i	3	−0.8848303 +8.441512i
9	0+100i	7	−0.8848304 +8.441513i	7	−0.8848303 +8.441512i
10	100+100i	8	−0.8848304 +8.441513i	8	−0.8848304 +8.441513i

在所列计算特征值中有 6 位有效数字．可以看出，二者所需的迭代次数相当，尽管它们会收敛于不同的特征值．

其次，考察如下问题([4],[2])：

$$A_0=\begin{pmatrix}-1+2\alpha^2 & \alpha(1-\alpha^2-2\beta^2) & 2\alpha^2\beta^2 & -\alpha\beta^2(\alpha^2+\beta^2)\\ 2\alpha & -(\alpha^2+2\beta^2) & 2\alpha\beta^2 & -\beta^2(\alpha^2+\beta^2)\\ 1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\end{pmatrix},$$

$$A_1=\begin{pmatrix}3\alpha & -(1+\alpha^2+2\beta^2) & \alpha(1+2\beta^2) & -\beta^2(\alpha^2+\beta^2)\\ 2 & 0 & 0 & 0\\ 0 & 2 & 0 & 0\\ 0 & 0 & 2 & 0\end{pmatrix},$$

$$A_2=I^{(4)}.$$

同样 $A(\lambda)=A_0+A_1\lambda+A_2\lambda^2$，这里 $\alpha\geqslant 0$ 是参数和 $\beta=1+\alpha$． $\det A(\lambda)$ 的零点之间的分离程度随 α 的改变而改变[4]．特别，$\alpha=0$ 时，它有三重特征值 $\pm i$ 和二重特征值零．

下面是 α 取 0.0 和 0.1 时的计算结果．计算中的结束迭代的准则是

$$|r_{44}^{(s)}|<\varepsilon_1$$

和

$$|\mathrm{Re}(\lambda_s-\lambda_{s+1})|/|\mathrm{Re}(\lambda_s)|<\varepsilon_2,\ 若\ |\mathrm{Re}(\lambda_s)|>\varepsilon_2,$$

或

$$|\mathrm{Re}(\lambda_s-\lambda_{s+1})|<\varepsilon_2,\ 否则$$

和

$$|\mathrm{Im}(\lambda_s-\lambda_{s+1})|/|\mathrm{Im}(\lambda_s)|<\varepsilon_2,\ 若\ |\mathrm{Im}(\lambda_s)|>\varepsilon_2,$$

或

$$|\mathrm{Im}(\lambda_s-\lambda_{s+1})|<\varepsilon_2,\ 否则.$$

表2 $\alpha=0$ (算法 1)

No.	初始值	迭代次数 (s)	计算特征值 (λ_s)	ε_1	ε_2
1	$2+i$	22	$0.136081\times10^{-4}+0.4677132\times10^{-4}i$	5×10^{-9}	5×10^{-4}
2	$2+2i$	25	$0+1.000002i$	5×10^{-6}	5×10^{-6}
3	$10+10i$	9	$-0.5481967\times10^{-3}+1.004006i$	1×10^{-4}	5×10^{-3}

最后，举例说明非方阵的泛函 λ-矩阵的计算问题：

$$B(\lambda)=B_0+B_1\lambda+B_2\lambda^2\in \mathbf{C}^{4\times3},$$

$$B_0=\begin{pmatrix}-1&0&0\\0&-2&0\\1&0&0\\0&1&0\end{pmatrix},\quad B_1=\begin{pmatrix}2&-3&0\\2&0&0\\0&2&0\\0&0&2\end{pmatrix},\quad B_2=\begin{pmatrix}1&0&0\\0&1&0\\0&0&1\\0&0&0\end{pmatrix}.$$

它的非正则点为 $\pm i$ 和 0．

表 3　$\alpha=0.1$（算法 1）

No.	初 始 值	迭代次数 (s)	计算特征值 (λ_s)	ε_1	ε_2	被逼近特征值的类型
1	$-0.1+0.5i$	8	-0.7167993×10^{-6} $+1.000003i$	5×10^{-7}	5×10^{-6}	复(单)
2	$0.1+0.9i$	5	0.2701969×10^{-5} $+0.999997i$	1×10^{-7}	5×10^{-6}	复(单)
3	$-1.9+0.2i$	8	0.09999985 $+0.208481\times10^{-7}i$	5×10^{-9}	1×10^{-7}	实(单)
4	$1.9+0.2i$	8	0.7250966×10^{-6} $+0.3219782\times10^{-7}i$	5×10^{-8}	1×10^{-6}	实(单)
5	$-0.1+1.01i$	13	-0.099984 $+1.099999i$	5×10^{-7}	5×10^{-6}	复(单)
6	$0.5+10i$	11	-0.2155021×10^{-4} $+1.099993i$	5×10^{-6}	5×10^{-5}	复(单)

注．表 3 中最后一列是根据所算结果进行分析而得，其中(单)表示收敛于 $A(\lambda)$ 的单重非线性特征值．此外，当实、虚部很小时，所列结果不准．

由表 4 可以看出，算法 2 的收敛速度非常慢．但当 λ 是多变量且变量个数恰好为 $m-n+1$ 时，则相当于算法 2 的自然推广，亦即 Newton 法（这里假定最小二乘问题的系数矩阵可逆）．因此，可以期望有较好的收敛性质．

表 4（算法 2）

No.	初 始 值	迭代次数 (s)	计算值 (λ_s)	ε_1	ε_2
1	$0.9+1.1i$	30	$0.4385618\times10^{-4}+0.9999702i$	5×10^{-5}	5×10^{-5}
2	$0+0.8i$	22	$0+0.9998837i$	5×10^{-5}	5×10^{-5}
3	$0.1+10i$	22	$0.2946093\times10^{-5}+1.000140i$	1×10^{-4}	5×10^{-5}
4	$0.001+0.001i$	79	$0.4947933\times10^{-4}+0.4947885\times10^{-4}i$	5×10^{-5}	5×10^{-6}

感谢孙继广老师的热情指导和审稿人对原稿提出的许多有益的建议。

参　考　文　献

[1] V. N. Kublanovskaya, On an approach to the solution of the generalized latent value problem for λ-matrices, *SIAM J.Numer. Anal.*, 7(1970), 532—537.
[2] A. Ruhe, Algorithms for the nonlinear eigenvalue problem, *ibid.*, 10(1973), 672—689.
[3] N. K. Jain, K. Singhal, On Kublanovskaya's approach to the solution of the generalized latent value problem for functional λ-matrices, *ibid.*, 20(1983), 1062—1070.
[4] P. Lancaster, Lambda-Matrices and Vibrating Systems, Pergamon Press, Oxford, 1966.
[5] 孙继广，矩阵扰动分析，科学出版社，北京，1987.
[6] I., Gohberg, P. Lancaster, L. Rodman, Matrix Polynomials, Academic Press, 1982.
[7] 薛兆清，薛彦才，求解非线性矩阵特征值问题的一个 Kublanovskaya 型方法(未发表).
[8] J. H. Wilkinson, The Algebraic Eigenvalue problem, Claredon Press, Oxford, 1965.
[9] G. H. Golub, C. Van Loan, Matrix Computations, the Johns Hopkins University Press, 1983.
[10] 曹志浩，张玉德，李瑞遐，矩阵计算与方程求根，人民教育出版社，1979.
[11] P. Lancaster, A generalized Rayleigh-quotient iteration for lambda-matrices, *Arch. Rat. Mech. Anal.*, 8(1961), 309—322.