Journal of Computational Mathematics, Vol.10, No.1, 1992, 1–20.

COMPUTE MULTIPLY NONLINEAR EIGENVALUES*

Li Ren-cang

(Institute of Applied Physics and Computational Mathematics, Beijing, China)

Abstract

An incomplete QR decomposition called QR-like decomposition is proposed and studied. The developed theory enables us to construct two new algorithms for computing multiply nonlinear eigenvalues.

Several numerical tests are presented to illustrate their behavior in comparison with Kublanovskaya's approach.

§1. Introduction

Consider an n by n matrix $A(\lambda)$ whose entries $a_{ij}(\lambda)$ are analytic functions of a complex scalar λ (cf. $a_{ij}(\lambda)$ are functions which have at least first order derivatives of real scalar λ). We shall call such a matrix a functional λ-matrix[5]. However, if $a_{ij}(\lambda)$ are polynomials in λ, then $A(\lambda)$ is commonly known as a λ-matrix[8]. Values λ and the corresponding nonzero vectors x and y which satisfy

$$A(\lambda)x = 0, \quad y^H A(\lambda) = 0 \tag{1.1}$$

are the solutions to the nonlinear eigenvalue problem associated with $A(\lambda)$. In the above equations λ is known as a nonlinear eigenvalue, x a right and y a left nonlinear eigenvector (for the meaning of superscript H, see Notation below). Obviously, all the nonlinear eigenvalues of $A(\lambda)$ are also the roots of its characteristic equation

$$\det A(\lambda) = 0, \tag{1.2}$$

and vice versa.

By now, several iterative methods (see [5]-[9] and [12]) have been proposed for solving equations (1.1). Roughly speaking, they are just effective for computing simple nonlinear eigenvalues, and often converge slowly when a multiply nonlinear eigenvalue arises. In §2 we shall propose a kind of incomplete QR decomposition called QR-like decomposition, and then extend the differentiability theory for QR

* Received July 11, 1988.

decomposition (QRD) (see [6], [9] and [11]) to such decomposition. The extended theory is the basis of our algorithms for computing multiply nonlinear eigenvalue of $A(\lambda)$ in §3. In §4 and 5 we shall analyse the convergence theory of our algorithms. Numerical tests were made and are listed in §6 to illustrate their behavior.

Notation. We shall use $\mathbb{C}^{m\times n}$ ($\mathbb{R}^{m\times n}$) for the m by n complex (real) matrix set, $\mathbb{C}^m = \mathbb{C}^{m\times 1}$ ($\mathbb{R}^m = \mathbb{R}^{m\times 1}$), $\mathbb{C} = \mathbb{C}^1$ ($\mathbb{R} = \mathbb{R}^1$); $\mathcal{U}_n \subset \mathbb{C}^{n\times n}$ denotes the n by n unitary matrix set. $I^{(n)}$ is the n by n unit matrix, $e_j^{(n)}$ the jth column of $I^{(n)}$ and $I_j^{(n)} \equiv (e_1^{(n)}, \cdots, e_j^{(n)})$, $K_j^{(n)} \equiv (e_{n-j+1}^{(n)}, \cdots, e_n^{(n)})$. When no confusion arises, these superscripts (n) are usually omitted. A^H, A^T denote the conjugate transpose and transpose of A respectively, and $\|A\|_F$, $\|A\|_2$ the Frobenius norm and the spectral norm of A, respectively. For a matrix $A = (a_1, \cdots, a_n)$ where a_i are n column vectors, we define a column vector col A by col $A \equiv (a_1^T, \cdots, a_n^T)^T$. An adhoc notation is that for a given integer t and a matrix $A \in \mathbb{K}^{n\times n}$ we always partition it as

$$A = \begin{matrix} & \begin{matrix} n-t & t \end{matrix} \\ \begin{matrix} n-t \\ t \end{matrix} & \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \end{matrix}, \tag{1.3}$$

where $\mathbb{K} = \mathbb{C}$ or $\mathbb{R}$. Symbol $\otimes$ denotes the Kronecker product of matrices.

§2. QR-Like Decompositions

In this section, we first propose a kind of incomplete QR decompositions (QRDs), and then study its differentiability properties, which are the bases of our algorithms for computing multiply nonlinear eigenvalues in the forthcoming sections.

Definition 2.1. *Let $B \in \mathbb{C}^{n\times n}$ and $1 \le t \le n$ be an integer. A decomposition*

$$B = QR, \quad Q \in \mathcal{U}_n \tag{2.1}$$

is called a QR*-like decomposition with the index t* (QRD(t)), *if*

$$R = \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}, \tag{2.22}$$

where $R_{11} \in \mathbb{C}^{(n-t)\times(n-t)}$ is an upper triangular matrix.

From this definition, it follows that a QRD is always a QRD (t), but generally the converse is not true if $t \neq 1$, and if $t = 1$, QRD and QRD (1) are equivalent.

The following theorem is an extension of Theorem 2.1 in [9].

Theorem 2.1. *Let $A(\lambda) \in \mathbb{C}^{n\times n}$ be analytie and $\lambda_0 \in \mathbb{C}$. Assume that the first $n-t$ column vectors of $A(\lambda_0)$ are linearly independent.*

$$A(\lambda_0) = Q_0 R_0, \quad R_0 = \begin{pmatrix} R_{11}^{(0)} & R_{12}^{(0)} \\ 0 & R_{22}^{(0)} \end{pmatrix}$$

is its any QRD(t). *Then there exists a neighborhood $\mathcal{B}(\lambda_0) \subset \mathbb{C}$ of λ_0 such that we have* QRD(t) *of $A(\lambda)$,*

$$A(\lambda) = Q(\lambda)R(\lambda), \quad R(\lambda) = \begin{pmatrix} R_{11}(\lambda) & R_{12}(\lambda) \\ 0 & R_{22}(\lambda) \end{pmatrix} \quad \text{for} \quad \lambda \in \mathcal{B}(\lambda_0), \tag{2.3}$$

satisfying that $Q(\lambda_0) = Q_0$ and $R(\lambda_0) = R_0$, and that $R_{22}(\lambda)$ and the diagonal elements of $R_{11}(\lambda)$ are differentiable at λ_0, and all entries of $R(\lambda)$ and $Q(\lambda)$ are continuous with respect to $\lambda \in \mathcal{B}(\lambda_0)$.

Proof. $A(\lambda)$ may be read as

$$A(\lambda) = A(\lambda_0) + \frac{dA(\lambda)}{d\lambda}\Big|_{\lambda=\lambda_0} + O(|\lambda-\lambda_0|^2) \equiv A(\lambda_0) + E. \tag{2.4}$$

Therefore

$$Q_0^H A(\lambda) = Q_0^H A(\lambda_0) + Q_0^H E = R_0 + \widetilde{E}, \tag{2.5}$$

where $\widetilde{E} \equiv Q_0^H E$. Partition $\widetilde{E}$ conformally as $\widetilde{E} = \begin{pmatrix} \widetilde{E}_{11} & \widetilde{E}_{12} \\ \widetilde{E}_{21} & \widetilde{E}_{22} \end{pmatrix}$. By hypothesis, we know that $R_{11}^{(0)}$ is invertible. Together with (2.4) we deduce that, if $|\lambda - \lambda_0|$ is sufficiently small, then $R_{11}^{(0)} + \widetilde{E}_{11}$ is invertible. Let $P = \widetilde{E}_{21}(R_{11}^{(0)} + \widetilde{E}_{11})^{-1} \in \mathbb{C}^{t\times(n-t)}$. It is easy to verify that (refer to (2.4) and (2.5))

$$\begin{aligned} Q_1(\lambda)^H\, Q_0^H A(\lambda) &= \begin{pmatrix} (I+P^HP)^{-\frac{1}{2}} & 0 \\ 0 & (I+PP^H)^{-\frac{1}{2}} \end{pmatrix} \\ &\times \begin{pmatrix} R_{11}^{(0)} + \widetilde{E}_{11} + P^H\widetilde{E}_{21} & R_{12}^{(0)} + \widetilde{E}_{12} + P^H(R_{22}^{(0)} + \widetilde{E}_{22}) \\ 0 & R_{22}^{(0)} + \widetilde{E}_{22} - P(R_{12}^{(0)} + \widetilde{E}_{12}) \end{pmatrix} \\ &\equiv \begin{pmatrix} R_{11}(\lambda) & R_{12}(\lambda) \\ 0 & R_{22}(\lambda) \end{pmatrix} \end{aligned} \tag{2.6}$$

where

$$Q_1(\lambda)^H \equiv \begin{pmatrix} (I+P^HP)^{-\frac{1}{2}} & 0 \\ 0 & (I+PP^H)^{-\frac{1}{2}} \end{pmatrix} \begin{pmatrix} I & P^H \\ -P & I \end{pmatrix}, \tag{2.7}$$

$$R_{11}(\lambda) \equiv (I + P^H P)^{-\frac{1}{2}}(R_{11}^{(0)} + \widetilde{E}_{11} + P^H \widetilde{E}_{21}), \tag{2.8a}$$

$$R_{12}(\lambda) \equiv (I + P^H P)^{-\frac{1}{2}}[R_{12}^{(0)} + \widetilde{E}_{12} + P^H (R_{22}^{(0)} + \widetilde{E}_{22})], \tag{2.9}$$

$$R_{22}(\lambda) \equiv (I + PP^H)^{-\frac{1}{2}}[R_{22}^{(0)} + \widetilde{E}_{22} - P(R_{12}^{(0)} + \widetilde{E}_{12})]. \tag{2.10a}$$

It follows that, if $|\lambda - \lambda_0|$ is sufficiently small such that $\|P\|_2 < 1$, then (from (2.4), (2.5), (2.7)-(2.10a))

$$R_{11}(\lambda) = R_{11}^{(0)} + I_{n-t}^T Q_0^H \frac{dA(\lambda)}{d\lambda}\Big|_{\lambda=\lambda_0} I_{n-t}(\lambda - \lambda_0) + O(|\lambda - \lambda_0|^2), \tag{2.8b}$$

$$R_{22}(\lambda) = R_{22}^{(0)} + \Big[K_t^T Q_0^H \frac{dA(\lambda)}{d\lambda}\Big|_{\lambda=\lambda_0} K_t - K_t^T Q_0^H \frac{dA(\lambda)}{d\lambda}\Big|_{\lambda=\lambda_0} I_{n-t} R_{11}^{(0)^{-1}} R_{12}^{(0)}\Big](\lambda - \lambda_0)$$
$$+ O(|\lambda - \lambda_0|^2), \tag{2.10b}$$

and $\lim_{\lambda\to\lambda_0} \|Q_1(\lambda) - I^{(n)}\|_2 = \lim_{\lambda\to\lambda_0} \|R_{12}(\lambda) - R_{12}^{(0)}\|_2 = 0$. From (2.4)-(2.6), we know that there exists a neighborhood $\mathcal{B}_1(\lambda_0)$ of λ_0 such that equations (2.7)-(2.10) remain true for all $\lambda \in \mathcal{B}_1(\lambda_0)$. On the other hand, by Theorem 2.1 in [9], it follows by induction that there exists a neighborhood $\mathcal{B}_2(\lambda_0)$ of λ_0 such that we have QRDs of $R_{11}(\lambda) = Q_2(\lambda)R_2(\lambda)$ for $\lambda \in \mathcal{B}_2(\lambda_0)$ which satisfy that the diagonal entries of $R_2(\lambda)$ are differentiable at λ_0 and all entries of $Q_2(\lambda)$ and $R_2(\lambda)$ are continuous with respect to $\lambda \in \mathcal{B}_2(\lambda_0)$, and $Q_2(\lambda_0) = I^{(n-t)}$, $R_2(\lambda_0) = R_{11}^{(0)}$. Let

$$Q(\lambda) = Q_0 Q_1(\lambda) \begin{pmatrix} Q_2(\lambda) & 0 \\ 0 & I^{(t)} \end{pmatrix} \quad \text{and} \quad R(\lambda) = \begin{pmatrix} R_2(\lambda) & * \\ 0 & R_{22}(\lambda) \end{pmatrix}.$$

Then QRD(t) $A(\lambda) = Q(\lambda)R(\lambda)$ for $\lambda \in \mathcal{B}(\lambda_0) \equiv \mathcal{B}_1(\lambda_0) \cap \mathcal{B}_2(\lambda_0)$ meet the needs of Theorem 2.1.

From Theorem 2.1 and its proof, we can deduce straightforwardly

Corollary 2.1. Let $A(\lambda) \in \mathbb{C}^{n\times n}$ be analytic, $\pi \in \mathbb{C}^{n\times n}$ a permutation matrix, $1 \le t < n$ and $\lambda_0 \in \mathbb{C}$. Assume that the first $n - t$ column vectors of $A(\lambda_0)\pi$ are linearly independent, and $A(\lambda_0)\pi = Q_0 R_0$, with $R_0 = \begin{pmatrix} R_{11}^{(0)} & R_{12}^{(0)} \\ 0 & R_{22}^{(0)} \end{pmatrix}$, is its any QRD($t$). Then there exists a neighborhood $\mathcal{B}(\lambda_0) \subset \mathbb{C}$ of λ_0 such that we have QRD(t) of $A(\lambda)\pi$,

$$A(\lambda)\pi = Q(\lambda)R(\lambda), \quad R(\lambda) = \begin{pmatrix} R_{11}(\lambda) & R_{12}(\lambda) \\ 0 & R_{22}(\lambda) \end{pmatrix} \quad \text{for} \quad \lambda \in \mathcal{B}(\lambda_0),$$

satisfying that $Q(\lambda_0) = Q_0$ and $R(\lambda_0) = R_0$, and that $R_{22}(\lambda)$ and the diagonal elements of $R_{11}(\lambda)$ are differentiable at λ_0,

$$R_{22}(\lambda) \equiv K_t^T R(\lambda) K_t = K_t^T R_0 K_t + \Big[K_t^T Q_0^H \frac{dA(\lambda)}{d\lambda}\Big|_{\lambda=\lambda_0} \pi K_t$$
$$- K_t^T Q_0^H \frac{dA(\lambda)}{d\lambda}\Big|_{\lambda=\lambda_0} \pi I_{n-t} (I_{n-t}^T R_0 I_{n-t})^{-1} I_{n-t}^T R_0 K_t \Big] (\lambda - \lambda_0)$$
$$+ O(|\lambda - \lambda_0|^2), \qquad (2.11)$$

and all entries of $R(\lambda)$ and $Q(\lambda)$ are continuous with respect to $\lambda \in \mathcal{B}(\lambda_0)$.

Remark 2.1. From the proof of Theorem 2.1, especially from equation (2.10), it follows that, under the conditions of Theorem 2.1, there exists a positive Hermitian matrix $B(\lambda)$, which is continuous with respect to $\lambda \in \mathcal{B}(\lambda_0)$ and satisfies $x^H B(\lambda) x \geq x^H x$ (for all $x \in \mathbb{C}^t$ and $\lambda \in \mathcal{B}(\lambda_0)$), such that $B(\lambda)^{-1} R_{22}(\lambda) \equiv B(\lambda)^{-1} K_t^T R(\lambda) K_t$ is analytic in $\mathcal{B}(\lambda_0)$. In fact, we can set $B(\lambda) = (I + PP^H)^{\frac{1}{2}}$.

We also have a theorem similar to the uniqueness theorem of QRD in [9].

Theorem 2.2. *Suppose the first $n - t (1 \leq t < n)$ column vectors of matrix $C \in \mathbb{C}^{n \times n}$ are linearly independent, and $C = Q_1 R_1 = Q_2 R_2$ are its two QRD(t)s. Then*

$$Q_1 = Q_2 D \quad \text{and} \quad R_1 = D^H R_2 \qquad (2.12)$$

where matrix D is of form

$$D = \begin{pmatrix} D_{11} & \\ & D_{22} \end{pmatrix} \in \mathcal{U}_n \quad \text{where} \quad D_{11} = \text{diag}\,(d_1, \cdots, d_{n-t}) \in \mathcal{U}_{n-t},\ D_{22} \in \mathcal{U}_t. \qquad (2.13)$$

Proof. Partition $R_i (i = 1, 2)$ as $R_i = \begin{pmatrix} R_{11}^{(i)} & R_{12}^{(i)} \\ 0 & R_{22}^{(i)} \end{pmatrix}$ for $i = 1, 2$. By the hypotheses, it follows that $R_{11}^{(i)} \in \mathbb{C}^{(n-t)\times(n-t)} (i = 1, 2)$ are nonsingularly upper triangular matrices. Choosing $\varepsilon \in \mathbb{C}$ such that $R_{22}^{(1)} + \varepsilon I^{(t)}$ is invertible. Let $\widetilde{R}_1 = \begin{pmatrix} R_{11}^{(1)} & R_{12}^{(1)} \\ & R_{22}^{(1)} + \varepsilon I \end{pmatrix}$. Then by $Q_2^H Q_1 R_1 = R_2$ we know that there exists $\widetilde{R}_2 \in \mathbb{C}^{n \times n}$ of form $\widetilde{R}_2 = \begin{pmatrix} R_{11}^{(2)} & * \\ 0 & * \end{pmatrix}$, such that

$$Q_2^H Q_1 \widetilde{R}_1 = \widetilde{R}_2 \Rightarrow Q_2^H Q_1 = \widetilde{R}_2 \widetilde{R}_1^{-1} \equiv D \in \mathcal{U}_n. \qquad (2.14)$$

Because of the form of $\widetilde{R}_1$ and $\widetilde{R}_2$, we deduce that $D \in \mathcal{U}_n$ must be of form $D = \begin{pmatrix} D_{11} & D_{12} \\ 0 & D_{22} \end{pmatrix}$, where $D_{11} \in \mathcal{U}_{n-t} \subset \mathbb{C}^{(n-t)\times(n-t)}$ is an upper triangular matrix, and $D_{22} \in U_t$. Therefore D_{11} must be diagonal and $D_{12} = 0$, i.e. D is of form (2.13). Thus (2.12) is proved.

§3. Algorithms

In constructing our algorithms we were motivated by optimization methods for solving problems of sums of squares (see e.g. [2], Chapter 6).

Let $\lambda_* \in \mathbb{C}$ be a multiply nonlinear eigenvalue, and rank $A(\lambda_*) = n-t(1 \le t < n)$ (we shall deal with the case $t = n$ at the end of this section). Now suppose μ is sufficiently close to λ_*, and

$$A(\mu)\pi(\mu) = Q(\mu)R(\mu), \quad R(\mu) = \begin{pmatrix} R_{11}(\mu) & R_{12}(\mu) \\ 0 & R_{22}(\mu) \end{pmatrix} \tag{3.1}$$

is a QRD(t) with column pivoting. Then $\|R_{22}(\mu)\|_F \ll \|(R_{11}(\mu), R_{12}(\mu))\|_F$; moreover

$$\lim_{\mu \to \lambda_*} |\det R_{11}(\mu)| = \text{ a positive number}.$$

Here we need to explain what QRD(t) with column pivoting is: it is commonly known that computing a QRD with column pivoting of a matrix $B \in \mathbb{C}^{n\times n}$ by using Housholder transformations is generally implemented through $n-1$ reductions and before each reduction a permutation of some two columns is necessary. Different from these, a QRD(t) with column pivoting of a matrix is computed without the last $t-1$ reductions in computing a QRD with column pivoting of the matrix.

From Theorem 2.1, it follows that there exists a QRD(t)

$$A(\lambda)\pi(\mu) = Q(\lambda)R(\lambda), \quad R(\lambda) = \begin{pmatrix} R_{11}(\lambda) & R_{12}(\lambda) \\ 0 & R_{22}(\lambda) \end{pmatrix} \quad \text{for} \quad \lambda \in \mathcal{B}(\mu) \tag{3.2}$$

such that $R_{22}(\lambda)$ is differentiable at μ. Now we face a problem: how to improve μ as an approximation to λ_* to get a new and more accurate (expected) one. Since $R_{22}(\lambda)$ can be read as $R_{22}(\lambda) = R_{22}(\mu) + R'_{22}(\mu)(\lambda - \mu) + O(|\lambda - \mu|^2)$, ignoring the second and higher order terms, we approximately let $R_{22}(\lambda) \approx R_{22}(\mu) + R'_{22}(\mu)(\lambda - \mu)$ and the improved approximation $\widetilde{\mu}$ to λ_* is naturally chosen such that

$$\|R_{22}(\mu) + R'_{22}(\mu)(\widetilde{\mu} - \mu)\|_F^2 = \min_{\lambda \in \mathbb{C}} \|R_{22}(\mu) + R'_{22}(\mu)(\lambda - \mu)\|_F^2.$$

Therefore (assume $\|R'_{22}(\mu)\|_F^2 \neq 0$)

$$\widetilde{\mu} = \mu - \frac{[\,\text{col } R'_{22}(\mu)\,]^H \cdot \text{col } R_{22}(\mu)}{\|R'_{22}(\mu)\|_F^2}. \tag{3.3}$$

Formally we have the following algorithm.

Algorithm 3.1. Compute a multiply nonlinear eigenvalue λ_* of $A(\lambda) \in \mathbb{C}^{n\times n}$ with $\operatorname{rank} A(\lambda_*) = n - t$ known in advance by some means.

a) Give an initial approximation μ_0 to λ^*.

b) Compute

$$A(\mu_i) \quad \text{and} \quad \left.\frac{dA(\lambda)}{d\lambda}\right|_{\lambda=\mu_i} \equiv A_1^{(i)}, i = 0, 1, \cdots.$$

c) Compute QRD (or QRD(t)) with column pivoting of $A(\mu_i)$

$$A(\mu_i)\pi_i = Q_i R_i, \quad R_i = \begin{pmatrix} R_{11}^{(i)} & R_{12}^{(i)} \\ 0 & R_{22}^{(i)} \end{pmatrix}.$$

d) Compute $R_{22}'^{(i)} = K_t^T Q_i^H A_1^{(i)} \pi_i K_t - K_t^T Q_i^H A_1^{(i)} \pi_i I_{n-t} {R_{11}^{(i)}}^{-1} R_{12}^{(i)} \in \mathbb{C}^{t\times t}$.

e) Compute $\mu_{i+1} = \mu_i - \dfrac{[\operatorname{col} R_{22}'^{(i)}]^H \cdot \operatorname{col} R_{22}^{(i)}}{\|R_{22}'^{(i)}\|_F^2}$.

f) If the needed accuracy is attained, stop; otherwise go to b).

A practical and very difficult problem is how to get $\operatorname{rank} A(\lambda_*)$ at the beginning of our computing, which is crucial. If $\mu = \mu_0$ is a good approximation of λ_* and $A(\mu)\pi = QR$ is a QRD with column pivoting (denote $R = (r_{ij})$), then from the observations made at the beginning of this section it follows that the diagonal entries r_{ii} $(i = 1, \cdots, n)$ of upper triangular matrix R have the property

$$|r_{n-t\,n-t}| = \min_{1\le i\le n-t} |r_{ii}| \gg |r_{n-t+1\,n-t+1}| = \max_{n-t+1\le i\le n} |r_{ii}|. \tag{3.4}$$

Property (3.4) is very important and is often used in the determination of the numerical rank of a matrix (see e.g. [1], [14]). In our concerns, we choose a threshold $\varepsilon > 0$, and then find the smallest integer t such that

$$|r_{n-t+1\,n-t+1}| < \varepsilon |r_{11}| \le |r_{n-t\,n-t}|, \tag{3.5}$$

and approximately let $\operatorname{rank} A(\lambda_*) = r \approx n - t$.

These observations show that if we truly use (3.5) to estimate $\operatorname{rank} A(\lambda_*)$, then it is necessary to have a very good initial approximation μ_0 to λ_*. But in some cases good initial approximations are not available. To deal with such cases, we propose the following algorithm which is a modified (maybe improved) version of Algorithm 3.1.

Algorithm 3.2. Compute a multiply nonlinear eigenvalue λ_* of $A(\lambda) \in \mathbb{C}^{n\times n}$.

a) Give an initial approximation μ_0 to λ^*.

b) Compute

$$A(\mu_i) \quad \text{and} \quad \frac{dA(\lambda)}{d\lambda}\Big|_{\lambda=\mu_i} \equiv A_1^{(i)}, i = 0, 1, \cdots.$$

c) Compute QRD (or QRD(t)) with column pivoting of $A(\mu_i)$ $A(\mu_i)\pi_i = Q_i R_i$, where $R_i = (\, r_{kj}^{(i)} \,)$. Find the smallest integer $p \quad (1 \le p < n)$ such that $|r_{n-p+1\, n-p+1}| < \varepsilon |r_{11}| \le |r_{n-p\, n-p}|$, and let $t = p$; if such p does not exist, let $t = 1$. Partition R_i comformally as before: $R_i = \begin{pmatrix} R_{11}^{(i)} & R_{12}^{(i)} \\ 0 & R_{22}^{(i)} \end{pmatrix}$.

d) Compute $R_{22}^{\prime(i)} = K_t^T Q_i^H A_1^{(i)} \pi_i K_t - K_t^T Q_i^H A_1^{(i)} \pi_i I_{n-t} R_{11}^{(i)^{-1}} R_{12}^{(i)} \in \mathbb{C}^{t\times t}$.

e) Compute $\mu_{i+1} = \mu_i - \dfrac{[\text{ col } R_{22}^{\prime(i)}\,]^H \cdot \text{col } R_{22}^{(i)}}{\|R_{22}^{\prime(i)}\|_F^2}$.

f) If the needed accuracy is attained, **stop**; otherwise go to b).

In Algorithm 3.2, t is modified for every iteration. But we think it is not necessary to do so in practical computations. In fact, the behavior of t in Algorithm 3.2 must be like this: in the first few iterations $t = 1$, and then t will increase for another few iterations, and finally t becomes a constant integer. Therefore we can improve the estimations of t in Algorithm 3.2 as: first let $t = 1$ and do several iterations, and then use (3.5) to modify t. With this modified t, do several other iterations, and then again use (3.5) to modify this modified t, and so on. The threshold ε can also be replaced by a sequence of thresholds $\varepsilon_i (i = 0, 1, \cdots)$, where ε_i is used in ith iteration.

We recall that at the beginning of this section we noted the trivial case rank A (λ_*) $= 0$, in which $\lim\limits_{\mu\to\lambda_*} A(\mu) = A(\lambda_*) = 0$. By the same ideas as in Algorithm 3.1 and 3.2, if μ is an approximation to λ_*, then the improved approximation $\tilde{\mu}$ to λ_* is

$$\tilde{\mu} = \mu - \frac{[\text{ col } A'(\mu)\,]^H \cdot \text{col } A(\mu)}{\|A'(\mu)\|_F^2} \tag{3.6}$$

where $A'(\mu) = \dfrac{d}{d\lambda} A(\lambda)\Big|_{\lambda=\mu}$.

Remark 3.1. In Algorithm 3.1, suppose we do not know rank$A(\lambda_*)$, but simply let it equal $n - 1$, i.e. $t = 1$. In Algorithm 3.2 we remove the procedure for determining t but simply let $t = 1$. Then Algorithms 3.1 and 3.2 both degenerate to Kublanovskaya's approach (see [11] for proofs).

§4. Convergence Analysis

Before giving our convergence analysis, we first derive some properties of Algorithms 3.1 and 3.2 and some preliminaries.

For a fixed integer t and $\lambda = \mu$, suppose the first $n-t$ column vectors are linearly independent and the following are two QRD(t)s of $A(\mu)\pi$ (refer to Theorem 2.2):

$$A(\mu)\pi = Q(\mu)R(\mu) = Q(\mu)D \cdot (D^H R(\mu)), \quad R(\mu) = \begin{pmatrix} R_{11}(\mu) & R_{12}(\mu) \\ 0 & R_{22}(\mu) \end{pmatrix}, \tag{4.1a}$$

where D is of form (2.13). Denote $\widetilde{R}(\mu) = D^H R(\mu)$ and conformally partition it as

$$\widetilde{R}(\mu) = \begin{pmatrix} \widetilde{R}_{11}(\mu) & \widetilde{R}_{12}(\mu) \\ 0 & \widetilde{R}_{22}(\mu) \end{pmatrix}. \tag{4.1b}$$

Then from Theorem 2.1, it follows that there exist two QRD(t)s in a neighborhood $\mathcal{B}(\mu)$ of μ, say $A(\lambda)\pi = Q(\lambda)R(\lambda) = \widetilde{Q}(\lambda)\widetilde{R}(\lambda)$, which have properties stated in Theorem 2.1. Partition $R(\lambda)$ and $\widetilde{R}(\lambda)$ as

$$R(\lambda) = \begin{pmatrix} R_{11}(\lambda) & R_{12}(\lambda) \\ 0 & R_{22}(\lambda) \end{pmatrix}, \quad \widetilde{R}(\lambda) = \begin{pmatrix} \widetilde{R}_{11}(\lambda) & \widetilde{R}_{12}(\lambda) \\ 0 & \widetilde{R}_{22}(\lambda) \end{pmatrix}. \tag{4.2}$$

Then from (2.11) and (2.13) we have

$$\begin{cases} R_{22}(\lambda) = R_{22}(\mu) + R'_{22}(\mu)(\lambda-\mu) + O(|\lambda-\mu|^2), \\ \widetilde{R}_{22}(\lambda) = \widetilde{R}_{22}(\mu) + \widetilde{R}'_{22}(\mu)(\lambda-\mu) + O(|\lambda-\mu|^2), \\ \widetilde{R}_{22}(\mu) = D_{22}^H R_{22}(\mu),\ \widetilde{R}'_{22}(\mu) = D_{22}^H R'_{22}(\mu). \end{cases} \tag{4.3}$$

Hence

$$\|\widetilde{R}'_{22}(\mu)\|_F^2 = \|R'_{22}(\mu)\|_F^2. \tag{4.4}$$

On the other hand, from (4.3) we also have (refer to [13, pp.25-28])

$$\text{col } \widetilde{R}_{22}(\mu) = (I^{(t)} \otimes D_{22}^H) \text{ col } R_{22}(\mu), \quad \text{col } \widetilde{R}'_{22}(\mu) = (I^{(t)} \otimes D_{22}^H) \text{ col } R'_{22}(\mu). \tag{4.5}$$

Therefore

$$\begin{aligned} [\text{ col } \widetilde{R}'_{22}(\mu)]^H \text{ col } \widetilde{R}_{22}(\mu) &= [\text{ col } R'_{22}(\mu)]^H (I^{(t)} \otimes D_{22}) \cdot (I^{(t)} \otimes D_{22}^H) \text{ col } R_{22}(\mu) \\ &= [\text{ col } R'_{22}(\mu)]^H \text{ col } R_{22}(\mu). \end{aligned} \tag{4.6}$$

(4.4) and (4.6) show that

$$f(\mu,t,\pi) \equiv \|R'_{22}(\mu)\|_F^2 \quad \text{and} \quad g(\mu,t,\pi) \equiv [\text{ col } R'_{22}(\mu)]^H \text{ col } R_{22}(\mu) \tag{4.7}$$

are two functions depending only on μ, t and π and for fixed μ, t and π, different QRD(t)s of $A(\mu)\pi$ do not change their values. So step $g(\mu,t,\pi)/f(\mu,t,\pi)$ in Algorithms 3.1 and 3.2 do not vary with different QRD(t)s of $A(\mu)\pi$.

Property 4.1.

$$f(\mu,t,\pi\widetilde{\pi}) = f(\mu,t,\pi), \quad g(\mu,t,\pi\widetilde{\pi}) = g(\mu,t,\pi), \tag{4.8}$$

where $\widetilde{\pi} = \operatorname{diag}(\pi_1,\pi_2)$; $\pi_1 \in \mathbb{C}^{(n-t)\times(n-t)}$ and $\pi_2 \in \mathbb{C}^{t\times t}$ are both permutation matrices.

Proof. Let $A(\mu)\pi = Q(\mu)R(\mu)$ be a QRD(t) and $R(\mu)$ be of form (4.1). Then

$$Q(\mu)^H A(\mu)\pi\widetilde{\pi} = R(\mu)\widetilde{\pi} = \begin{pmatrix} R_{11}(\mu)\pi_1 & R_{12}(\mu)\pi_1 \\ 0 & R_{22}(\mu)\pi_2 \end{pmatrix}.$$

Compute a QRD $R_{11}(\mu)\pi_1 = Q_1(\mu)\widetilde{R}_{11}(\mu)$ and let $\widetilde{Q}(\mu) = Q(\mu)\begin{pmatrix} Q_1(\mu) & \\ & I^{(t)} \end{pmatrix}$. Then the decomposition $A(\mu)\pi\widetilde{\pi} = \widetilde{Q}(\mu)\widetilde{R}(\mu)$ is a QRD(t). Partition $\widetilde{R}(\mu)$ as (4.1b). It is easy to verify that

$$\widetilde{R}_{11}(\mu) = Q_1^H(\mu)R_{11}(\mu)\pi_1, \; \widetilde{R}_{12}(\mu) = Q_1^H(\mu)R_{12}(\mu)\pi_1, \; \widetilde{R}_{22}(\mu) = R_{22}(\mu)\pi_2. \tag{4.9}$$

On the other hand, similarly to (4.2), (4.3) and (4.5), by (2.11) we have (with the same notation)

$$\widetilde{R}'_{22}(\mu) = R'_{22}(\mu)\pi_2. \tag{4.10}$$

Combining (4.9), (4.10) with (4.7) will lead to (4.8).

Generally, permutation matrices are not uniquely determined by column pivoting in the processes of decompositions. But we have

Property 4.2. Let π, $\widetilde{\pi}$ be two permutation matrices which result from QRD(t)s with column pivoting of $A(\lambda_*)$ ($\operatorname{rank} A(\lambda_*) = n-t$). Then

$$f(\lambda_*,t,\pi) > 0 \quad \Leftrightarrow \quad f(\lambda_*,t,\widetilde{\pi}) > 0. \tag{4.11}$$

Proof. Assume on the contrary that (4.11) is not true. Without loss of generality, we assume that $f(\lambda_*,t,\pi) > 0$ but $f(\lambda_*,t,\widetilde{\pi}) = 0$.

Note that $A(\lambda_*)\pi = Q^*R^*$, $A(\lambda_*)\widetilde{\pi} = \widetilde{Q}^*\widetilde{R}^*$, $R^* = \begin{pmatrix} R_{11}^* & R_{12}^* \\ 0 & 0 \end{pmatrix}$, $\widetilde{R}^* = \begin{pmatrix} \widetilde{R}_{11}^* & \widetilde{R}_{12}^* \\ 0 & 0 \end{pmatrix}$. Therefore (by Theorem 2.1) there are QRD(t)s of $A(\lambda)\pi$ and $A(\lambda)\widetilde{\pi}$ in a neighborhood of λ_*

$$A(\lambda)\pi = Q(\lambda)R(\lambda), \quad A(\lambda)\widetilde{\pi} = \widetilde{Q}(\lambda)\widetilde{R}(\lambda), \tag{4.12}$$

such that (refer to (4.1))

$$\begin{cases} R_{11}(\lambda), \widetilde{R}_{11}(\lambda) \quad \text{invertible}, \\ R_{22}(\lambda) = R'_{22}(\lambda_*)(\lambda - \lambda_*) + O(|\lambda - \lambda_*|^2), \quad \widetilde{R}_{22}(\lambda) = O(|\lambda - \lambda_*|^2), \\ R'_{22}(\lambda_*) \neq 0, \end{cases} \tag{4.13}$$

if $|\lambda - \lambda_*|$ is sufficiently small. Let

$$V(\lambda) = \begin{pmatrix} R_{11}(\lambda) & R_{12}(\lambda) \\ 0 & I^{(t)} \end{pmatrix}, \ \widetilde{V}(\lambda) = \begin{pmatrix} \widetilde{R}_{11}(\lambda) & \widetilde{R}_{12}(\lambda) \\ 0 & I^{(t)} \end{pmatrix},$$

$$D(\lambda) = \begin{pmatrix} I^{(n-t)} & \\ & R_{22}(\lambda) \end{pmatrix}, \ \widetilde{D}(\lambda) = \begin{pmatrix} I^{(n-t)} & \\ & \widetilde{R}_{22}(\lambda) \end{pmatrix}.$$

Thus

$$\begin{aligned} &A(\lambda)\pi = Q(\lambda)D(\lambda)V(\lambda), \\ &A(\lambda)\widetilde{\pi} = \widetilde{Q}(\lambda)\widetilde{D}(\lambda)\widetilde{V}(\lambda) \Rightarrow A(\lambda) = Q(\lambda)D(\lambda)M(\lambda) = \widetilde{Q}(\lambda)\widetilde{D}(\lambda)\widetilde{M}(\lambda), \end{aligned}$$

where $M(\lambda) = V(\lambda)\pi^T$ and $\widetilde{M}(\lambda) = \widetilde{V}(\lambda)\widetilde{\pi}^T$. From (4.13) we know if $|\lambda - \lambda_*|$ is sufficiently small, then $M(\lambda)$ and $\widetilde{M}(\lambda)$ are both invertible. Hence

$$D(\lambda)M(\lambda)\widetilde{M}(\lambda)^{-1} = Q(\lambda)^H\widetilde{Q}(\lambda)\widetilde{D}(\lambda) \text{ if } |\lambda - \lambda_*| \text{ is sufficiently small}. \tag{4.14}$$

Let $N(\lambda) = M(\lambda)\widetilde{M}(\lambda)^{-1}$ and $\widetilde{N}(\lambda) = Q(\lambda)^H\widetilde{Q}(\lambda)$. They are both invertible if $|\lambda - \lambda_*|$ is sufficiently small. Partition $N(\lambda)$ and $\widetilde{N}(\lambda)$ conformally as

$$N(\lambda) = \begin{pmatrix} N_{11}(\lambda) & N_{12}(\lambda) \\ N_{21}(\lambda) & N_{22}(\lambda) \end{pmatrix}, \ \widetilde{N}(\lambda) = \begin{pmatrix} \widetilde{N}_{11}(\lambda) & \widetilde{N}_{12}(\lambda) \\ \widetilde{N}_{21}(\lambda) & \widetilde{N}_{22}(\lambda) \end{pmatrix}.$$

Then (4.14) becomes

$$\begin{pmatrix} N_{11}(\lambda) & N_{12}(\lambda) \\ R_{22}(\lambda)N_{21}(\lambda) & R_{22}(\lambda)N_{22}(\lambda) \end{pmatrix} = \begin{pmatrix} \widetilde{N}_{11}(\lambda) & \widetilde{N}_{12}(\lambda)\widetilde{R}_{22}(\lambda) \\ \widetilde{N}_{21}(\lambda) & \widetilde{N}_{22}(\lambda)\widetilde{R}_{22}(\lambda) \end{pmatrix}. \tag{4.15}$$

On the other hand, (4.13) implies $\lim\limits_{\lambda\to\lambda_*} R_{22}(\lambda) = \lim\limits_{\lambda\to\lambda_*} \widetilde{R}_{22}(\lambda) = 0$, combining which with (4.15) gives $\lim\limits_{\lambda\to\lambda_*} \widetilde{N}_{21}(\lambda) = R_{22}(\lambda)N_{21}(\lambda) = 0$ and $\lim\limits_{\lambda\to\lambda_*} N_{12}(\lambda) = \widetilde{N}_{12}(\lambda)\widetilde{R}_{22}(\lambda) = 0$. Therefore from the continuity of $N(\lambda)$ and $\widetilde{N}(\lambda)$ follows that $N_{22}(\lambda)$ and $\widetilde{N}_{22}(\lambda)$ are both invertible if $|\lambda - \lambda_*|$ is sufficiently small.

From (4.15),

$$\begin{aligned} R_{22}(\lambda)N_{22}(\lambda) = \widetilde{N}_{22}(\lambda)\widetilde{R}_{22}(\lambda) &\Rightarrow R_{22}(\lambda) = \widetilde{N}_{22}(\lambda)\widetilde{R}_{22}(\lambda)N_{22}(\lambda)^{-1} \\ &\Rightarrow \frac{R_{22}(\lambda)}{\lambda - \lambda_*} = \widetilde{N}_{22}(\lambda)\frac{\widetilde{R}_{22}(\lambda)}{\lambda - \lambda_*}N_{22}(\lambda)^{-1}. \end{aligned}$$

Letting $\lambda \to \lambda_*$ gives (refer to (4.13)) $0 \neq R'_{22}(\lambda_*) = \lim_{\lambda\to\lambda_*} \frac{R_{22}(\lambda)}{\lambda - \lambda_*} = \lim_{\lambda\to\lambda_*} \widetilde{N}_{22}(\lambda)\frac{\widetilde{R}_{22}(\lambda)}{\lambda - \lambda_*}$ $N_{22}(\lambda)^{-1} = 0$, a contradiction, which completes our proof.

Based on Property 4.2, we define

Definition 4.1. *Suppose λ_* is a nonlinear eigenvalue of $A(\lambda) \in \mathbb{C}^{n\times n}$ and $\mathrm{rank}A(\lambda_*) = n - t\,(1 \le t \le n)$. For $t = n$ define*

$$\triangle(\lambda_*) = \begin{cases} 1, & \text{if} \quad \left\| \frac{dA(\lambda)}{d\lambda}\Big|_{\lambda=\lambda_*} \right\| > 0, \\ 0, & \text{otherwise} . \end{cases}$$

For $1 \le t < n$, define

$$\triangle(\lambda_*) = \begin{cases} 1, & \text{if} \quad f(\lambda_*, t, \pi_*) > 0\,, \\ 0, & \text{otherwise}\,, \end{cases}$$

where π_* is a permutation matrix which results from a QRD(t) with column pivoting of $A(\lambda_*)$.

From the proof of Theorem 2.1, we can easily obtain (refer to [9])

Lemma 4.1. *Suppose the first $n - t$ column vectors of matrix $C_1 \in \mathbb{C}^{n\times n}$ are linearly independent, $C_1 = Q_1R_1$ is a QRD(t) of C_1. Let $C_2 \in \mathbb{C}^{n\times n}$. Then for any $\varepsilon > 0$, there exists a QRD(t) $C_2 = Q_2R_2$ such that $\|Q_1 - Q_2\|_2 < \varepsilon$, and $\|R_1 - R_2\|_2 < \varepsilon$, if $\|C_1 - C_2\|_2$ is sufficiently small.*

Combining this lemma, Theorem 2.2 and (2.11), we will have

Lemma 4.2. *Let $A(\lambda) \in \mathbb{C}^{n\times n}$ be analytic with respect to $\lambda \in \mathcal{D} \subset \mathbb{C}$, where $\mathcal{D}$ is an open connected set of $\mathbb{C}$, and π a permutation matrix. If the first $n-t$ column vectors of $A(\lambda)\pi$ are linearly independent for $\lambda \in \mathcal{D}$, then $f(\lambda, t, \pi)$ is continuous with respect to $\lambda \in \mathcal{D}$.*

Theorem 4.1. *Let $A(\lambda) \in \mathbb{C}^{n\times n}$ be analytic, and $\pi_i = \pi_*$ independent of i for μ_i sufficiently close to λ_*. If $\triangle(\lambda_*) = 1$, then Algorithm 3.1 is locally quadratically convergent.*

Proof. Suppose we have decompositions (4.1a) with $\pi = \pi_*$. From Theorem 2.1, it is reasonable to let (refer to (4.3)) $R_{22}(\lambda) = R_{22}(\mu_i) + R'_{22}(\mu_i)(\lambda - \mu_i) + O(|\lambda - \mu_i|^2)$. Let $\lambda = \lambda_*$. Then

$$\begin{aligned} 0 &= R_{22}(\mu_i) + R'_{22}(\mu_i)(\lambda_* - \mu_i) + O(|\lambda_* - \mu_i|^2) \\ \Rightarrow 0 &= \mathrm{col}R_{22}(\mu_i) + \mathrm{col}\ R'_{22}(\mu_i)(\lambda_* - \mu_i) + O(|\lambda_* - \mu_i|^2). \end{aligned}$$

Multiplying the two sides of the last equation by $[\mathrm{col}\ R'_{22}(\mu_i)]^H$, we get

$$0 = [\mathrm{col}\ R'_{22}(\mu_i)]^H \mathrm{col}\ R_{22}(\mu_i) + \|R'_{22}(\mu_i)\|_F^2(\lambda_* - \mu_i) + O(|\lambda_* - \mu_i|^2). \tag{4.16}$$

Since $\triangle(\lambda_*) = 1$, by Lemma 4.2 and Property 4.2 we know that $\|R'_{22}(\mu_i)\|_F^2 > \frac{1}{2}f(\lambda_*, t, \pi_*) > 0$ for μ_i sufficiently close to λ_*. Therefore by (4.16) we have

$$\begin{aligned}0 &= \|R'_{22}(\mu_i)\|_F^{-2}[\text{col } R'_{22}(\mu_i)]^H \text{col } R_{22}(\mu_i) + (\lambda_* - \mu_i) + O(|\lambda_* - \mu_i|^2)\\ &= -(\mu_{i+1} - \mu_i) + (\lambda_* - \mu_i) + O(|\lambda_* - \mu_i|^2) = \lambda_* - \mu_{i+1} + O(|\lambda_* - \mu_i|^2)\\ &\Rightarrow \lambda_* - \mu_{i+1} = O(|\lambda_* - \mu_i|^2).\end{aligned}$$

Remark 4.1. For the trivial case $t = n$, i.e. $A(\lambda_*) = 0$, similarly to the above proof, we can prove that, if $\triangle(\lambda_*) = 1$, then the iteration algorithm (3.6) is also locally quadratically convergent.

Remark 4.2. Choosing ε appropriately in Algorithm 3.2 can also guarantee its local quadratic convergence if $\triangle(\lambda_*) = 1$.

§5. The Condition $\triangle(\lambda_*) = 1$

$A(\lambda) = (a_{ij}(\lambda))$ is analytic; so is $\det A(\lambda)$. Let λ_* be a nonlinear eigenvalue of $A(\lambda)$. Then the algebraic multiplicity of λ_* is finite, i.e. there exist an analytic function $h(\lambda)$ and a positive integer, say $m(\lambda_*)$, such that

$$\det A(\lambda) = (\lambda - \lambda_*)^{m(\lambda_*)} h(\lambda),\ h(\lambda_*) \neq 0. \tag{5.1}$$

For convenience use, we define the algebraic multiplicity $m(\lambda) = 0$ for $\lambda \in \mathbb{C}$ which is not a nonlinear eigenvalue.

The main results of this section are two theorems below.

Theorem 5.1. *Let $A(\lambda)$ be analytic and $\text{rank}A(\lambda_0) = n - t (0 \le t \le n)$. Then there exist matrices $M(\lambda)$, $N(\lambda) \in \mathbb{C}^{n\times n}$ which are continuously invertible in a neighborhood of λ_0, such that*

$$A(\lambda) = M(\lambda)\begin{pmatrix} I^{(n-t)} & & & \\ & (\lambda-\lambda_0)^{\kappa_1} & & \\ & & \ddots & \\ & & & (\lambda-\lambda_0)^{\kappa_t}\end{pmatrix} N(\lambda), \tag{5.2}$$

where integers κ_i ($1 \le \kappa_1 \le \cdots \le \kappa_t < +\infty$), called partial multiplicities of λ_0, are unique.

If $A(\lambda)$ is a λ-matrix, i.e. $a_{ij}(\lambda)$ are all polynomials in λ, Theorem 5.1 is just the well-known representation theorem of local Smith form (refer to [3, pp.330-333]). To the best of our knowledge, there are no similar results for cases when $a_{ij}(\lambda)$ are analytic functions, so we will give a proof of Theorem 5.1. We see that, if λ_0 is a nonlinear eigenvalue, then $\kappa_1 + \cdots + \kappa_t = m(\lambda_0)$ (refer to (5.1)).

Theorem 5.2. *Let $A(\lambda)$ be analytic, λ_* a nonlinear eigenvalue of it and rank$A(\lambda_*) = n - t(1 \le t \le n)$. Assume also that the partial multiplicities are $\kappa_i, i = 1, \cdots, t$. Then $\triangle(\lambda_*) = 1 \Longleftrightarrow \kappa_1 = 1$.*

Proof of Theorem 5.1. We shall first prove the existence of (5.2) by induction in $m(\lambda_0)$. To this end, we see that if $m(\lambda_0) = 0$, then choosing $M(\lambda) = A(\lambda)$ and $N(\lambda) = I^{(n)}$ completes the proof.

Suppose for all $\lambda_0 \in \mathbb{C}$ which satisfy $1 \le m(\lambda_0) < m$, an integer, we have representation (5.2). Then it suffices to prove that, when $m(\lambda_0) = m$, we also have representation (5.2). To this end, by Theorem 2.1 and Remark 2.1, we deduce that there is a QRD(t) ($n - t \equiv$rank$A(\lambda_0)$)

$$A(\lambda)\pi = Q(\lambda)R(\lambda),\ R(\lambda) = \begin{pmatrix} R_{11}(\lambda) & R_{12}(\lambda) \\ 0 & R_{22}(\lambda) \end{pmatrix} \quad \text{for} \quad \lambda \in \mathcal{B}(\lambda_0) \tag{5.4}$$

satisfying that $Q(\lambda)$ and $R(\lambda)$ are continuous, $R_{11}(\lambda)$ is continuously invertible, and moreover there is a positively continuous Hermitian matrix $B(\lambda)$ such that $\widetilde{R}_{22}(\lambda) \equiv B(\lambda)^{-1}R_{22}(\lambda)$ is analytic. Since $\widetilde{R}_{22}(\lambda_0) = 0$, we can find an integer, say $\kappa_1 \ge 1$, such that $\widehat{R}_{22}(\lambda) \equiv \widetilde{R}_{22}(\lambda)\,/(\lambda - \lambda_0)^{\kappa_1}$ is analytic and $\widehat{R}_{22}(\lambda_0) \ne 0$. Thus the multiplicity $\widehat{m}(\lambda_0)$ of λ_0, as a nonlinear eigenvalue of $\widehat{R}_{22}(\lambda)$, is $\widehat{m}(\lambda_0) = m(\lambda_0) - \kappa_1 < m$. Hence by the hypotheses of induction, there are matrices $\widehat{M}(\lambda)$, $\widehat{N}(\lambda) \in \mathbb{C}^{t\times t}$ which are continuously invertible in a neighborhood of λ_0 such that

$$\widehat{R}_{22}(\lambda) = \widehat{M}(\lambda) \begin{pmatrix} (\lambda-\lambda_0)^{\tau_1} & & \\ & \ddots & \\ & & (\lambda-\lambda_0)^{\tau_t} \end{pmatrix} \widehat{N}(\lambda), \tag{5.5}$$

where $0 = \tau_1 = \cdots = \tau_k < 1 \le \tau_{k+1} \le \cdots \le \tau_t$. Let

$$M(\lambda) = Q(\lambda) \begin{pmatrix} I^{(n-t)} & \\ & B(\lambda)\widehat{M}(\lambda) \end{pmatrix}, \quad N(\lambda) = \begin{pmatrix} R_{11}(\lambda) & R_{12}(\lambda) \\ 0 & \widehat{N}(\lambda) \end{pmatrix} \pi^T$$

and $\kappa_i = \kappa_1 + \tau_i$, $i = 1, \cdots, t$. Then together with (5.4), we get a representation (5.2) of $A(\lambda)$.

Now we are in a position to prove the uniqueness of $\kappa_1, \cdots, \kappa_t$. The case $n - t \equiv$ rank $A(\lambda_0) = n$ (i.e., $t = 0$) is trivial. In the following $t > 0$ is always assumed.

Suppose there are two representations

$$A(\lambda) = M(\lambda) \begin{pmatrix} I^{(n-t)} & & & \\ & (\lambda-\lambda_0)^{\kappa_1} & & \\ & & \ddots & \\ & & & (\lambda-\lambda_0)^{\kappa_t} \end{pmatrix} N(\lambda)$$

$$= \widetilde{M}(\lambda) \begin{pmatrix} I^{(n-t)} & & & \\ & (\lambda-\lambda_0)^{\tau_1} & & \\ & & \ddots & \\ & & & (\lambda-\lambda_0)^{\tau_t} \end{pmatrix} \widetilde{N}(\lambda), \tag{5.6}$$

where $1 \le \kappa_1 \le \cdots \le \kappa_t < +\infty$, $1 \le \tau_1 \le \cdots \le \tau_t < +\infty$. Let

$$W(\lambda) = \widetilde{M}(\lambda)^{-1} M(\lambda) = \begin{pmatrix} W_{11}(\lambda) & W_{12}(\lambda) \\ W_{21}(\lambda) W_{22}(\lambda) & \end{pmatrix},$$

$$\widetilde{W}(\lambda) = \widetilde{N}(\lambda) N(\lambda)^{-1} = \begin{pmatrix} \widetilde{W}_{11}(\lambda) & \widetilde{W}_{12}(\lambda) \\ \widetilde{W}_{21}(\lambda) \widetilde{W}_{22}(\lambda) & \end{pmatrix}.$$

Then by (5.6),

$$\begin{pmatrix} W_{11}(\lambda) & W_{12}(\lambda)A_2(\lambda) \\ W_{21}(\lambda) & W_{22}(\lambda)A_2(\lambda) \end{pmatrix} = \begin{pmatrix} \widetilde{W}_{11}(\lambda) & \widetilde{W}_{12}(\lambda) \\ \widetilde{A}_2(\lambda)\widetilde{W}_{21}(\lambda) & \widetilde{A}_2(\lambda)\widetilde{W}_{22}(\lambda) \end{pmatrix}, \tag{5.7}$$

where

$$A_2(\lambda) = \begin{pmatrix} (\lambda-\lambda_0)^{\kappa_1} & & \\ & \ddots & \\ & & (\lambda-\lambda_0)^{\kappa_t} \end{pmatrix}, \quad \widetilde{A}_2(\lambda) = \begin{pmatrix} (\lambda-\lambda_0)^{\tau_1} & & \\ & \ddots & \\ & & (\lambda-\lambda_0)^{\tau_t} \end{pmatrix}.$$

Let $\lambda = \lambda_0$ in (5.7). We see $W_{21}(\lambda_0) = \widetilde{W}_{12}(\lambda_0) = 0 \Rightarrow W_{22}(\lambda_0)$ and $\widetilde{W}_{22}(\lambda_0)$ are both invertible. Thus by the continuity of $M(\lambda)$, $N(\lambda)$ and $\widetilde{M}(\lambda)$, $\widetilde{N}(\lambda)$, we deduce that $W_{22}(\lambda)$ and $\widetilde{W}_{22}(\lambda)$ are invertible if $|\lambda-\lambda_0|$ is sufficiently small.

Suppose $\kappa_1 = \cdots = \kappa_p < \kappa_{p+1} \le \cdots \le \kappa_t$. From (5.7) we have

$$W_{22}(\lambda)A_2(\lambda) = \widetilde{A}_2(\lambda)\widetilde{W}_{22}(\lambda) \Rightarrow W_{22}(\lambda)\frac{A_2(\lambda)}{(\lambda-\lambda_0)^{\kappa_1}}\widetilde{W}_{22}(\lambda)^{-1} = \frac{\widetilde{A}_2(\lambda)}{(\lambda-\lambda_0)^{\kappa_1}}. \tag{5.8}$$

Since $\lim\limits_{\lambda\to\lambda_0} W_{22}(\lambda)\frac{A_2(\lambda)}{(\lambda-\lambda_0)^{\kappa_1}}\widetilde{W}_{22}(\lambda)^{-1}$ = a constant matrix with rank p, together with (5.8), we get

$$\tau_1 = \cdots = \tau_p = \kappa_1 < \tau_{p+1} \le \cdots \le \tau_t. \tag{5.9}$$

Hence (5.8) becomes

$$W_{22}(\lambda)\begin{pmatrix} I^{(p)} & & & \\ & (\lambda-\lambda_0)^{\kappa_{p+1}-\kappa_1} & & \\ & & \ddots & \\ & & & (\lambda-\lambda_0)^{\kappa_t-\kappa_1}\end{pmatrix}$$

$$=\begin{pmatrix} I^{(p)} & & & \\ & (\lambda-\lambda_0)^{\tau_{p+1}-\kappa_1} & & \\ & & \ddots & \\ & & & (\lambda-\lambda_0)^{\tau_t-\kappa_1}\end{pmatrix}\widetilde{W}_{22}(\lambda).$$

If $p=t$, then the uniqueness of κ_i is proved; if $p<t$, then by similar arguments, we can also prove that, if $\kappa_{p+1}-\kappa_1=\cdots=\kappa_q-\kappa_1<\kappa_{q+1}-\kappa_1\le\cdots$ $(q>p)$, then $\tau_{p+1}-\kappa_1=\cdots=\tau_q-\kappa_1<\tau_{q+1}-\kappa_1\le\cdots$, from which and (5.9) follows $\tau_{p+1}=\ldots=\tau_q=\kappa_{p+1}=\cdots=\kappa_q<\tau_{q+1}\le\cdots$. After several repeated procedures, we finally obtain $\kappa_i=\tau_i\ (1\le i\le t)$.

A proof of Theorem 5.2 may be given by combining the first part of the above proof with Property 4.2. We omit the details here. Theorems 5.1 and 5.2 enable us to deduce the following interesting corollary.

Corollary 5.1. Let $A(\lambda)\in\mathbb{C}^{n\times n}$ be analytic, λ_* a nonlinear eigenvalue with the multiplicity $m(\lambda_*)=3$. If $\operatorname{rank}A(\lambda_*)\le n-2$, then Algorithm 3.1 (3.2) is locally quadratically convergent.

Proof. Since $m(\lambda_*)=3$ and $\operatorname{rank}A(\lambda_*)\le n-2$, the partial multiplicities of λ_* must be $\{1,2\}$ or $\{1,1,1\}$. Thus by Theorem 5.2, $\triangle(\lambda_*)=1$ together with Theorem 4.1, we know that Algorithm 3.1 (3.2) is locally quadratically convergent.

§6. Numerical Tests

To demonstrate the behavior of our algorithms in §4, numerical tests were made on IBM-PC/XT. Double precision arithmatic was used throughout.

Consider the following problem: $A(\lambda)=A_0+A_1\lambda+A_2\lambda^2, n=4$, with

$$A_0=\begin{pmatrix} -16. & 16. & 0. & 32. \\ -32. & 34. & 4. & 66. \\ 16. & -18. & 8. & -34. \\ -48. & 52. & -4. & 101.\end{pmatrix},\quad A_1=\begin{pmatrix} 12. & -12. & 0. & -24. \\ 24. & -26. & -4. & -50. \\ -12. & 14. & -5. & 26. \\ 36. & -40. & 1. & -78.\end{pmatrix},$$

$$A_2 = \begin{pmatrix} -4. & 4. & 0. & 8. \\ -8. & 8. & 0. & 16. \\ 4. & -4. & 3. & -8. \\ -12. & 12. & -3. & 25. \end{pmatrix}.$$

Its three multiply nonlinear eigenvalues are

Nonlinear eigenvalue λ	Multiplicity $m(\lambda)$	Partial multiplicities
$1.$	3	$\kappa_1 = 1, \kappa_2 = 2;$
$\frac{3+\sqrt{7}}{2}$	2	$\kappa_1 = 1, \kappa_2 = 1;$
$\frac{3-\sqrt{7}}{2}$	2	$\kappa_1 = 1, \kappa_2 = 1.$

Moreover, rank $A(\lambda_*) = 2$ for any of the above three multiply nonlinear eigenvalues.

In our tests, QRDs with column pivoting were performed throughout. And in tables below numbers $|r_{33}^{(i)}|$ and $|r_{44}^{(i)}|$ are listed as the absolute values of the last two diagonal elements of upper triangular matrix R_i in each iteration.

Table 6.1. $\mu_0 = 1.5 - 0.5i$

	Algorithm 3.1			Kublanovskaya's Approach		
No.(j)	$\lvert\mu_j - 1\rvert$	$\lvert r_{33}^{(j)}\rvert$	$\lvert r_{44}^{(j)}\rvert$	$\lvert\mu_j - 1\rvert$	$\lvert r_{33}^{(j)}\rvert$	$\lvert r_{44}^{(j)}\rvert$
0	7.1E–01	3.0E–01	2.4E–01	7.1E–01	3.0E–01	2.4E–01
1	1.9E–01	8.4E–02	1.9E–02	2.8E–01	1.2E–01	4.2E–02
2	1.6E–02	7.5E–03	1.5E–04	1.3E–01	6.1E–02	1.0E–02
3	2.2E–04	1.0E–04	2.7E–08	6.6E–02	3.0E–02	2.5E–03
4	2.5E–08	1.2E–08	1.1E–11	3.3E–02	1.5E–02	6.2E–04
5	1.2E–15	7.7E–15	1.8E–32	1.6E–02	7.6E–03	1.6E–04
6				8.2E–03	3.8E–03	3.9E–05
⋮				⋮	⋮	⋮
10				5.1E–04	2.4E–04	1.5E–07
⋮				⋮	⋮	⋮
15				1.7E–05	8.0E–06	1.7E–10
⋮				⋮	⋮	⋮
19				2.1E–06	9.9E–07	2.7E–12

Table 6.1 displays the computational results of Algorithm 3.1 with $t = 2$ and Kublanovskaya's approach, both with starting point $\mu_0 = 1.5 - 0.5i$ and converging to $\lambda_* = 1$. From this table, we see Algorithm 3.1 is indeed quadratically convergent while Kublanovskaya's approach converges linearly (approximately $|\mu_{j+1} - 1| \approx 0.5|\mu_j - 1|$) and to reach less than half the accuracy secured by using Algorithm 3.1, it needs, however, roughly 4 times iterations needed by Algorithm 3.1.

Tables 6.2 and 6.3 also display the computational results of Algorithm 3.1 with $t = 2$ and of Kublanovskaya's approach. With starting point $\mu_0 = 1.5 + 1.5i$ and converging to $\lambda_* = \dfrac{3+\sqrt{7}}{2}$ in Table 6.2, the two algorithms need approximately the same number of iterations to reach approximately the same accuracy. The author has also constructed several other tests, and got the idea that, when the two sequences obtained by Algorithm 3.1 and Kublanovskaya's approach respectively converge to the same nonlinear eigenvalue with all partial multiplicities as 1, then the two algorithms behave similarly.

Table 6.2. $\mu_0 = 1.5 + 1.5i$

	Algorithm 3.1			Kublanovskaya's Approach		
No.(j)	$\lvert\mu_j - \frac{3+\sqrt{7}}{2}\rvert$	$\lvert r_{33}^{(j)}\rvert$	$\lvert r_{44}^{(j)}\rvert$	$\lvert\mu_j - \frac{3+\sqrt{7}}{2}\rvert$	$\lvert r_{33}^{(j)}\rvert$	$\lvert r_{44}^{(j)}\rvert$
0	1.8E-01	7.5E-01	4.3E-01	1.8E-01	7.5E-01	4.3E-01
1	1.1E-01	5.4E-01	2.7E-01	3.6E-02	3.8E-01	7.9E-02
2	5.2E-02	4.4E-01	1.3E-01	2.2E-03	2.9E-02	5.1E-03
3	1.1E-02	1.3E-01	2.5E-02	7.7E-06	1.0E-04	1.8E-05
4	3.9E-04	5.1E-03	9.0E-04	9.3E-11	1.2E-09	2.1E-10
5	7.8E-08	1.0E-06	1.8E-07	2.2E-16	2.9E-15	7.1E-16
6	3.8E-15	4.9E-14	9.6E-15			

Table 6.3. $\mu_0 = 10 - 10i$

	Algorithm 3.1			Kublanovskaya's Approach		
No.(j)	$\lvert\mu_j - 1\rvert$	$\lvert r_{33}^{(j)}\rvert$	$\lvert r_{44}^{(j)}\rvert$	$\lvert\mu_j - 1\rvert$	$\lvert r_{33}^{(j)}\rvert$	$\lvert r_{44}^{(j)}\rvert$
0	1.3E+01	5.4E+01	1.1E+01	1.3E+01	5.4E+01	1.1E+01
1	6.9E+00	1.4E+01	5.5E+01	1.0E+00	5.0E-01	4.3E-01
2	3.6E+00	4.2E+00	2.5E+00	3.4E-01	1.5E-01	6.2E-02
⋮	⋮	⋮	⋮	⋮	⋮	⋮
5	1.5E+00	7.3E-01	4.6E-01	3.9E-02	1.8E-02	8.8E-04
⋮	⋮	⋮	⋮	⋮	⋮	⋮
10	1.1E+00	4.3E-01	4.1E-01	1.2E-03	5.6E-04	8.5E-07
11	7.4E-01	3.5E-01	2.7E-01	6.1E-04	2.8E-04	2.1E-07
12	5.5E-01	2.4E-01	1.5E-01	3.0E-04	1.4E-04	5.4E-08
13	8.7E-02	4.0E-02	4.3E-03	1.5E-04	6.9E-05	1.3E-08
14	2.6E-03	1.2E-03	4.0E-06	7.5E-05	3.5E-05	3.2E-09
15	2.3E-06	1.1E-06	3.2E-12	3.7E-05	1.7E-05	8.1E-10
16	2.0E-12	9.2E-13	3.1E-15	1.9E-05	8.7E-06	2.0E-10

In Table 6.3 a bad initial approximation $\mu_0 = 10+10i$ was chosen. Algorithm 3.1 needs about 16 iterations to make $|\mu_j - 1| \sim 10^{-12}$, while Kublanovskaya's approach needs about as many to make $|\mu_j - 1| \sim 10^{-5}$. The reason for the large number of iterations is that, in quite a number of iterations at the beginning, Algorithm 3.1 converges very slowly, and only if μ_j comes sufficiently close to 1, does the quadratic convergence of Algorithm 3.1 begin to show; but on the other hand, Kublanovskaya's approach needs few iterations to get a μ_j close to 1, and then μ_j moves slowly to 1. As we noted in §3, Algorithm 3.1 needs good initial approximations, not just for the estimation of rank$A(\lambda_*)$. Tables 6.4 and 6.5 below display the computational results of an improved version of Algorithm 3.2 called Algorithm 3.2(r), which means that the first r iterations are Kublanovskaya's approach, i.e. Algorithm 3.2 with $t = 1$ and then back to Algorithm 3.2 with $t = 2(=\text{rank}A(\lambda_*))$.

Comparing Tables 6.4 and 6.5 with Table 6.3, we can find that, when no good initial approximation is available, we'd better use Algorithm 3.2. From Table 6.5 we can also see that choosing r appropriately often could reduce the number of iterations.

Table 6.4. $\mu_0 = 10 - 10i$

	Algorithm 3.2 (1)			Algorithm 3.2 (2)		
No.(j)	$\lvert\mu_j - 1\rvert$	$\lvert r_{33}^{(j)}\rvert$	$\lvert r_{44}^{(j)}\rvert$	$\lvert\mu_j - 1\rvert$	$\lvert r_{33}^{(j)}\rvert$	$\lvert r_{44}^{(j)}\rvert$
0	1.3E+01	5.4E+01	1.1E+01	1.3E+01	5.4E+01	1.1E+01
1	1.0E+00	5.0E-01	4.3E-01	1.0E+00	5.0E-01	4.3E-01
2	4.1E-01	1.8E-01	8.7E-01	3.4E-01	1.5E-01	6.2E-02
3	2.4E-02	1.1E-02	3.3E-04	1.0E-02	4.7E-03	6.0E-05
4	3.0E-04	1.4E-04	5.3E-08	8.8E-05	4.1E-05	4.5E-09
5	1.5E-08	7.1E-09	1.0E-15	3.2E-09	1.5E-09	4.1E-15
6	2.3E-15	5.2E-15	1.5E-15	4.3E-15	6.1E-15	4.6E-15

Table 6.5. Algorithm 3.2(r)

r	μ_0	ν	$\lvert\mu_\nu - 1\rvert$	$\lvert r_{33}^{(\nu)}\rvert$	$\lvert r_{44}^{(\nu)}\rvert$
3	10+10i	7	8.4E–15	7.7E–15	1.0E–15
4	10+10i	7	9.2E–13	4.3E–13	4.1E–15
5	10+10i	8	2.3E–15	5.2E–15	8.6E–17
1	10–10i	6	2.3E–15	5.2E–15	1.5E–15
2	10–10i	6	4.3E–15	6.1E–15	4.6E–16
3	10–10i	7	8.4E–15	7.7E–15	1.0E–15
1	100+100i	6	5.4E–15	6.1E–15	8.5E–16
2	100+100i	5	6.7E–14	3.1E–14	8.6E–16

§7. Remarks About Real Cases

It is commonly known that, if a complex variable function, is differentiable at every point in an open connected domain in the complex plane, then it must be analytic in that domain. But unfortunately, we have no counterpart for a real variable function. We recall that in the above sections $A(\lambda) \in \mathbb{C}^{n\times n}(\lambda \in \mathbb{C})$ is always assumed to be analytic; therefore one may ask how to modify the conditions of results in §§1-6 for real cases. When $A(\lambda) \in I\!R^{n\times n}(\lambda \in I\!R)$ is real analytic, all our conclusions remain true just by replacing symbol $\mathbb{C}$ by $I\!R$. That $A(\lambda) \in$

$\mathbb{R}^{n\times n}(\lambda \in \mathbb{R})$ is real analytic is an extremely strong assumption and generally it is not necessary. Slight modifications are

(i) In §2 and §3, that $A(\lambda) \in \mathbb{R}^{n\times n}(\lambda \in \mathbb{R})$ is differentiable is necessary. In such cases all terms $O(|\lambda - \lambda_*|^2)$ must be replaced by $o(|\lambda - \lambda_*|)$, but if in addition $A(\lambda)$ is second order differentiable, then all conclusions remain true just by replacing symbol $\mathbb{C}$ by $\mathbb{R}$.

(ii) In §4, if $A(\lambda) \in \mathbb{R}^{n\times n}(\lambda \in \mathbb{R})$ is differentiable, all conclusions remain true except Theorem 4.1 in which it suffices to assume that $A(\lambda)$ is second order differentiable.

(iii) In §5, it suffices to assume that $A(\lambda) \in \mathbb{R}^{n\times n}(\lambda \in \mathbb{R})$ is real analytic.

References

[1] P. Deuflhard and W. Sautter, On rank-deficient pseudoinverses, *Linear Algebra Appl.*, **29**(1980), 91-111.

[2] R. Fletcher, Practical Methods of Optimization, Vol. 1, John Wiley & Sons, 1980.

[3] I. Gohberg, P. Lancaster and L. Rodman, *Matrix Polynomials*, Academic Press, 1982.

[4] G.H. Golub and C.F. Van Loan, Matrix Computations, The John Hopkins Univ. Press, 1983.

[5] N.K. Jain and K. Singhal, On Kublanovskaya's approach to the solution of the generalized latent value problems for functional $\lambda-$matrices, *SIAM J. Numer. Anal.*, **20**(1983), 1062-1067.

[6] V.N. Kublanovskaya, On an approach to the solution of the generalized latent value problems for $\lambda-$matrices, *SIAM J. Numer. Anal.*, **7**(1970), 532-537.

[7] P. Lancaster, A generalized Rayleigh quotient iteration for Lambda-matrices, *Arch. Rat. Mech. Anal.*, **8**(1961), 309-322.

[8] P. Lancaster, Lambda-Matrices and Vibrating Systems, Pergamon Press, Oxford, 1966.

[9] Li, Ren-cang, QR decomposition and nonlinear eigenvalue problems, *Math. Numer. Sinica*, **11**:4 (1989), 374-385.

[10] Li, Ren-cang, Compute critical points of a stability problem, *Math. Numer. Sinica*, **12**:3, 250–258 (1990).

[11] Li, Ren-cang, QR decomposition and Kublanovskaya's approach, to appear.

[12] A. Ruhe, Algorithms for nonlinear eigenvalue problems, *SIAM J. Numer. Anal.*, **10** (1973), 672-687.

[13] Sun Ji-guang, Matrix Perturbation Analysis, Science Press, Beijing, 1987.

[14] G.W. Stewart, Rank degeneracy, *SIAM J. Sci. Stat. Comput.*, **5**(1984), 91-111.