On the Convergence of the Rayleigh Quotient Iteration for the Computation of the Characteristic Roots and Vectors. IV

(Generalized Rayleigh Quotient for Nonlinear Elementary Divisors)

A. M. OSTROWSKI

48. In this part* we shall give a quadratically convergent iteration rule for computing an eigenvalue of a matrix to which there corresponds a non-linear elementary divisor. Our method generalizes the rule discussed in Sections 1—8, Part I, to non-symmetric matrices and combines it with rules for hastening the convergence of an iteration of the first order.

Let A be an $n \times n$ matrix with an eigenvalue σ, and assume that the maximal exponent L of an elementary divisor of A corresponding to σ, is >1. Choose two (row) vectors, α and β, and consider for a λ_0 which is different from all eigenvalues of A the linear equations

$$(A-\lambda_0 I)\,\xi' = \alpha', \qquad \eta(A-\lambda_0 I) = \beta, \tag{183}$$

defining a couple of vectors ξ, η.

Now form the generalized Rayleigh quotient of ξ and η as defined by (72), and put

$$\lambda_1 = R(\xi,\eta) = \frac{\eta A \xi'}{\eta \xi'}, \tag{184}$$

assuming that

$$\eta \xi' \neq 0. \tag{184°}$$

Then from (184) and (183) we obtain λ_1 as a rational function of λ_0,

$$\lambda_1 = \varphi(\lambda_0). \tag{185}$$

We shall have to prove first that, under suitable hypotheses, the iteration by the function $\varphi(\lambda)$ defined by (185) has σ as a point of attraction. This is a consequence of the following

Theorem. *In the notation and under the assumptions of Section* 48, *there exists a matrix H depending only on A and on the choice of σ, such that if α and β are*

* This paper was prepared in part at Numerical Analysis Research, University of California at Los Angeles, and in part under a National Bureau of Standards contract with The American University, Washington, D.C., with the sponsorship of the Office of Naval Research. I am grateful to Mr. G. CULLER and Dr. E. V. HAYNSWORTH for discussions.

chosen to satisfy

(186) $$\beta H \alpha' \neq 0,$$

then we have

(187) $$\varphi'(\sigma) = 1 - \frac{1}{L}.$$

In particular, if (186) *is satisfied, then the condition* (184°) *is also satisfied when* λ_0 *is sufficiently near to* σ.

49. (187) shows that the iteration by the function $\varphi(\lambda)$ converges rather slowly, though linearly; even for $L=2$ the derivative is $\frac{1}{2}$. However, since $\varphi(\lambda)$ is a rational function, it follows from (187) that, as soon as (186) is satisfied, we have a development

(188) $$\varphi(\lambda) = \sigma + \left(1 - \frac{1}{L}\right)(\lambda - \sigma) + \sum_{\nu=2}^{\infty} a_\nu (\lambda - \sigma)^\nu,$$

and therefore different methods of acceleration are applicable.

Consider for an $\alpha \neq 1$ the function

(189) $$\varphi^*(\lambda) = \frac{1}{1-\alpha}\left(\varphi(\lambda) - \alpha\lambda\right).$$

We verify at once that $\varphi^*(\sigma) = \sigma$ and obtain for the value of the derivative of φ^* in σ:

(190) $$\varphi^{*\prime}(\sigma) = \frac{1 - \frac{1}{L} - \alpha}{1 - \alpha}.$$

If now the value of L is known, we can choose $\alpha = 1 - \frac{1}{L}$ and obtain for the corresponding function φ^*:

(191) $$\varphi_L(\lambda) = L\varphi(\lambda) - (L-1)\lambda$$

with vanishing derivative at the point σ. *The iteration by means of this function then converges quadratically to the value* σ.

If the value of L is not known, then it is still best to use the iterating function

(192) $$\varphi_2(\lambda) = 2\varphi(\lambda) - \lambda.$$

The iteration by this function converges quadratically for $L=2$, while in the case of a general L we obtain a derivative

(193) $$\varphi_2'(\sigma) = 1 - \frac{2}{L}$$

which is less than $1 - \frac{1}{L} = \varphi'(\sigma)$. On the other hand, if $L > 2$, the value of L will usually be recognized easily after a certain number of steps, and then the corresponding function $\varphi_L(\lambda)$ can be formed.

The use of $\varphi_2(\lambda)$ is best under the assumption that the value $L=2$ is the most probable of all $L>1$. On the other hand, if all values of L between 2 and n are more or less equally probable, we shall take

(194) $$\alpha = \frac{1 - \frac{1}{n} + \frac{1}{2}}{2},$$

and obtain readily

$$|\varphi^{*\prime}(\sigma)| \leqq \frac{1-\frac{2}{n}}{1+\frac{2}{n}} . \tag{195}$$

50. An iteration with quadratic convergence can be obtained in our case, if L is not known, by the Steffensen method, which, however, implies a double amount of computational work, for it uses a combination of λ, $\varphi(\lambda)$ and $\varphi(\varphi(\lambda))$. In this method we form the function

$$\Phi(\lambda) = \frac{\varphi(\varphi(\lambda)) - \varphi(\lambda)^2}{\lambda - 2\varphi(\lambda) + \varphi(\varphi(\lambda))}, \tag{196}$$

which for $\lambda = \lambda_0$ usually gives a better approximation than

$$\lambda_2 = \varphi(\lambda_1) = \varphi(\varphi(\lambda_0)).$$

In practice the value of $\varphi(\lambda_0)$ is of course computed by obtaining successively the values of λ_1 and λ_2 and by taking

$$\Phi(\lambda_0) = \frac{\lambda_0 \lambda_2 - \lambda_1^2}{\lambda_0 - 2\lambda_1 + \lambda_2}. \tag{197}$$

If then we put

$$\lambda_1' = \Phi(\lambda_0'), \quad \lambda_0' = \lambda_0, \tag{198}$$

we can consider the passage from λ_0' to λ_1' as a direct iteration by the iterating function $\Phi(\lambda)$.

However, this iteration converges quadratically. We prove this by using the result of our paper*. If we replace there λ_1 and λ_2 by 2, y by λ, ζ by σ and $\alpha_1 = \alpha_2$ by $1 - \frac{1}{L}$, we obtain from formula (21) *l.c.*

$$\frac{\Phi(\lambda) - \sigma}{(\lambda - \sigma)^2} = L^2 T_z + O(\lambda - \sigma)$$

where T_z is obtained from the formulae (15) and (9) *l.c.*:

$$T_z = \left(1 - \frac{1}{L}\right)^2 E(\varphi(\lambda)) - \left(1 - \frac{1}{L}\right) E(\lambda),$$

$$E(\lambda) = a_2 + a_3(\lambda - \sigma) + \cdots,$$

$$T_z = \left(1 - \frac{1}{L}\right)^2 a_2 - \left(1 - \frac{1}{L}\right) a_2 + O(\lambda - \sigma)$$

$$= -\frac{1}{L}\left(1 - \frac{1}{L}\right) a_2 + O(\lambda - \sigma),$$

$$\frac{\Phi(\lambda) - \sigma}{(\lambda - \sigma)^2} = -(L - 1)\, a_2 + O(\lambda - \sigma). \tag{199}$$

The one theoretically unsurmountable difficulty in the practical application of this method appears to be the fact that the bilinear form in (186) is unknown

* OSTROWSKI, A.: Über Verfahren von STEFFENSEN und HOUSEHOLDER zur Konvergenzverbesserung von Iterationen [MAURO PICONE zum 70. Geburtstag, ZAMP Vol. VII, 218–229 (1956)].

so long as the transformation of A to the Jordan canonical form has not been carried out, and this presupposes the knowledge of the eigenvalues. However, in practice this is hardly a difficulty at all, since it is infinitely improbable that for α and β taken at random, (186) would not be satisfied.

51. Lemma 1. *Consider the matrix*

$$A_0 = \sigma I + U, \tag{200}$$

I being the unity matrix of order l and U the corresponding auxiliary unity matrix of order l, which has 1's in the first superdiagonal and zeros elsewhere. Then we have, for $\lambda \neq \sigma$, $\lambda \to \sigma$:

$$(A_0 - \lambda I)^{-2} = l \frac{U^{l-1}}{(\lambda-\sigma)^{l+1}} + O\left(\frac{1}{(\lambda-\sigma)^l}\right), \tag{201}$$

$$U(A_0 - \lambda I)^{-2} = (l-1) \frac{U^{l-1}}{(\lambda-\sigma)^{l}} + O\left(\frac{1}{(\lambda-\sigma)^{l-1}}\right). \tag{202}$$

Proof. Putting $\varkappa = \frac{1}{\lambda-\sigma}$ we have, since $U^l = 0$,

$$\begin{aligned}((\sigma-\lambda) I + U)^{-2} &= \varkappa^2 (I - \varkappa U)^{-2} \\ &= \sum_{\nu=0}^{l-1} (\nu+1) \varkappa^{\nu+2} U^\nu,\end{aligned}$$

and, multiplying this by U,

$$U((\sigma-\lambda) I + U)^{-2} = \sum_{\nu=0}^{l-2} (\nu+1) \varkappa^{\nu+2} U^{\nu+1},$$

as $U^l = 0$. Taking out the highest terms on the right, we obtain (201) and (202).

52. Lemma 2. *Let σ be an eigenvalue of the matrix A to which correspond elementary divisors with the maximal exponent $L > 1$. Then there exists a matrix $H \neq 0$ such that we have for $\lambda \to \sigma$:*

$$(A - \lambda I)^{-2} = L \frac{H}{(\lambda-\sigma)^{L+1}} + O\left(\frac{1}{(\lambda-\sigma)^L}\right), \tag{203}$$

$$A(A - \lambda I)^{-2} - \sigma (A - \lambda I)^{-2} = (L-1) \frac{H}{(\lambda-\sigma)^{L}} + O\left(\frac{1}{(\lambda-\sigma)^{L-1}}\right). \tag{204}$$

53. Proof. Since the assertions (203) and (204) are invariant with respect to a similarity transformation, though H does change, we can assume from the beginning that A is given in the Jordan canonical form. We can therefore write A as a direct sum,

$$A = \dot{\sum_i} A_i \dotplus B, \tag{205}$$

where B is a matrix with eigenvalues $\neq \sigma$, while each A_i is a matrix of order m_i given by

$$A_i = \sigma I_{m_i} + U_{m_i} \qquad (i = 1, \ldots, k), \; L = \max_i m_i, \tag{206}$$

I_{m_i} and U_{m_i} having meanings analogous to those of I and U in Lemma 1.

54. Then for $\lambda \to \sigma$ we obviously have

$$(B - \lambda I)^{-2} = O(1),$$

and therefore

$$(A - \lambda I)^{-2} = \sum_{i=1}^{k}{}^{\boldsymbol{\cdot}} (A_i - \lambda I_{m_i})^{-2} \dotplus O(1).$$

Applying the result of Lemma 1 to each of the first k terms in the direct sum on the right, we obtain

$$(A - \lambda I)^{-2} = L \frac{\sum_i{}^{\boldsymbol{\cdot}} U_{m_i}^{L-1}}{(\lambda - \sigma)^{L+1}} + O\left(\frac{1}{(\lambda-\sigma)^L}\right).$$

Here in the direct sum

$$H = \sum_i{}^{\boldsymbol{\cdot}} U_{m_i}^{L-1} \tag{207}$$

we must sum over all i with $m_i = L$. However, for $m_i < L$ we have $U_{m_i}^{L-1} = 0$, and therefore we can write

$$H = \sum_{i=1}^{k}{}^{\boldsymbol{\cdot}} U_{m_i}^{L-1},$$

where $H \neq 0$, since there exist $m_i = L$.

This proves (203).

55. Further, from (205) we have

$$A(A - \lambda I)^{-2} - \sigma(A - \lambda I)^{-2} = \sum_{i=1}^{k}{}^{\boldsymbol{\cdot}} (A_i - \sigma I_{m_i})(A - \lambda I_{m_i})^{-2} \dotplus O(1).$$

Applying (202) to the first k terms in the sum on the right, we have now

$$(A_i - \sigma I_{m_i})(A_i - \lambda I_{m_i})^{-2} = U_{m_i}(A_i - \lambda I_i)^{-2} = (m_i - 1)\frac{U_{m_i}^{m_i - 1}}{(\lambda-\sigma)^{m_i}} + O\left(\frac{1}{(\lambda-\sigma)^{m_i-1}}\right).$$

Therefore

$$A(A - \lambda I)^{-2} - \sigma(A - \lambda I)^{-2} = (L-1)\frac{\sum_i U_{m_i}^{L-1}}{(\lambda-\sigma)^L} + O\left(\frac{1}{(\lambda-\sigma)^{L-1}}\right),$$

where the numerator of the first fraction on the right is given by (207). This proves (204).

56. Proof of the theorem of Section 48. Take the matrix H occurring in the formulae (203), (204) of Lemma 2 and assume that the condition (186) is satisfied. Then from (184) and (185) for $\lambda_0 \to \sigma$ we have

$$\varphi(\lambda_0) - \sigma = \frac{\eta A \xi' - \sigma(\eta \xi')}{\eta \xi} = \frac{\beta A (A - \lambda I)^{-2} \alpha' - \sigma \beta (A - \lambda_0 I)^{-2} \alpha'}{\beta (A - \lambda_0 I)^{-2} \alpha'}.$$

From (203) and (204)

$$\varphi(\lambda_0) - \sigma = \frac{(L-1)\dfrac{\beta H \alpha'}{(\lambda_0-\sigma)^L} + O\left(\dfrac{1}{(\lambda_0-\sigma)^{L-1}}\right)}{L\dfrac{\beta H \alpha'}{(\lambda_0-\sigma)^{L+1}} + O\left(\dfrac{1}{(\lambda_0-\sigma)^{L}}\right)} = (\lambda_0 - \sigma)\frac{L-1+O(\lambda_0-\sigma)}{L+O(\lambda_0-\sigma)}.$$

This proves assertion (187) of the theorem. The last part of the theorem follows immediately by (203) from

$$\eta\,\xi' = \beta(A-\lambda_0 I)^{-2}\alpha' = L\,\frac{\beta H \alpha'}{(\lambda_0-\sigma)^{L+1}} + O\left(\frac{1}{(\lambda_0-\sigma)^{L}}\right).$$

57. We consider as an example the matrix

$$A = \begin{pmatrix} 2 & 0 & 3 \\ 1 & 1 & 3 \\ 1 & -1 & 1 \end{pmatrix}.$$

Here $(A-\lambda I)$ has the elementary divisors $(\lambda-1)^2$ and $(\lambda-2)$, and we have

$$D = |A-\lambda I| = (\lambda-1)^2(2-\lambda) = -(\lambda^3-4\lambda^2+5\lambda-2).$$

We easily obtain

$$D(A-\lambda I)^{-1} = \begin{pmatrix} \lambda^2-2\lambda+4, & -3, & 3\lambda-3 \\ \lambda+2, & \lambda^2-3\lambda-1, & 3\lambda-3 \\ \lambda-2, & 2-\lambda, & \lambda^2-3\lambda+2 \end{pmatrix},$$

$$X = \frac{D^2}{\lambda-1}(A-\lambda I)^{-2}$$

$$= \begin{pmatrix} \lambda^3-3\lambda^2+12\lambda-16, & -9\lambda+15, & 6\lambda^2-15\lambda+9 \\ 2\lambda^2+4\lambda-12, & \lambda^3-5\lambda^2-\lambda+11, & 6\lambda^2-15\lambda+9 \\ 2\lambda^2-8\lambda+8, & -2\lambda^2+8\lambda-8, & \lambda^3-5\lambda^2+8\lambda-4 \end{pmatrix}$$

and

$$Y = \frac{D^2}{\lambda-1}A(A-\lambda I)^{-2}$$

$$= \begin{pmatrix} 2(\lambda^3-4), & -6(\lambda^2-\lambda-1), & 3(\lambda^3-\lambda^2-2\lambda+2) \\ \lambda^3+5\lambda^2-8\lambda-4, & \lambda^3-11\lambda^2+14\lambda+2, & 3(\lambda^3-\lambda^2-2\lambda+2) \\ \lambda^3-3\lambda^2+4, & -(\lambda^3-3\lambda^2+4), & \lambda^3-5\lambda^2+8\lambda-4) \end{pmatrix}.$$

From the definition in Section 48,

$$\varphi(\lambda) = \frac{\beta Y \alpha'}{\beta X \alpha'}. \tag{208}$$

Since we have

$$Y - X = (\lambda-1)\begin{pmatrix} \lambda^2+4\lambda-8 & -3(2\lambda-3) & 3(\lambda-1)^2 \\ \lambda^2+4\lambda-8 & -3(2\lambda-3) & 3(\lambda-1)^2 \\ (\lambda-2)^2 & -(\lambda-2)^2 & 0 \end{pmatrix} \equiv (\lambda-1)Z,$$

it follows that

$$\frac{\Phi(\lambda)-1}{\lambda-1} = \frac{\beta Z \alpha'}{\beta X \alpha'}. \tag{209}$$

For $\lambda = 1$, Z and X become respectively

$$Z_0 = \begin{pmatrix} -3 & 3 & 0 \\ -3 & 3 & 0 \\ 1 & -1 & 0 \end{pmatrix}, \qquad X_0 = \begin{pmatrix} -6 & 6 & 0 \\ -6 & 6 & 0 \\ 2 & -2 & 0 \end{pmatrix} = 2Z_0,$$

and we see that as $\lambda \to 1$ the ratio (209) tends to $\frac{1}{2}$ if $\beta Z_0 \alpha' \neq 0$. Z_0 is therefore the matrix H in the condition (186), and if we write

$$\alpha = (a_1, a_2, a_3), \quad \beta = (b_1, b_2, b_3),$$

this condition becomes

$$(a_1 - a_2)(-3b_1 - 3b_2 + b_3) \neq 0.$$

We see that α and β have to be chosen in such a way that $a_1 \neq a_2$, $b_3 \neq 3(b_1 + b_2)$.

If, for instance, we take

$$\alpha = (0, 1, 0), \quad \beta = (1, 0, 0),$$

we have from (208)

$$\varphi(\lambda) = 2\frac{\lambda^2 - \lambda - 1}{3\lambda - 5} \tag{210}$$

and from (192)

$$\varphi_2(\lambda) = \frac{\lambda^2 + \lambda - 4}{3\lambda - 5}. \tag{211}$$

If we start with $\lambda_0 = 0$ we obtain from (210) the following values of λ_ν and $\frac{1-\lambda_\nu}{1-\lambda_{\nu-1}}$:

ν	λ_ν	$\frac{1-\lambda_\nu}{1-\lambda_{\nu-1}}$
0	0.	
1	.4	.6
2	.6526	.579
3	.80648	.557
4	.89599	.5375
5	.94565	.5225
6	.97214	.5126

Starting again with $\lambda_0 = 0$ we obtain from (211)

ν	λ_ν	$\frac{(1-\lambda_{\nu-1})^2}{1-\lambda_\nu}$
0	0	
1	.8	5
2	.984616	2.046
3	.999884338	2.0^3347
4	$.9^8331229$	2.0^72006

Certenago-Montagnola
Lugano, Svizzera

(Received February 2, 1959)