Canad. Math. Bull. Vol. 17 (2), 1974

THE GENERALIZED RAYLEIGH QUOTIENT

BY
M. V. PATTABHIRAMAN

1. **Introduction.** In this paper we generalize the concept of the Rayleigh quotient to a complex Banach space. Lord Rayleigh investigated the quotient

$$R(\mathbf{q}) = -\frac{\mathbf{q}'C\mathbf{q}}{\mathbf{q}'A\mathbf{q}} \tag{1}$$

considered as a function of the components of $\mathbf{q}$, in the case of a symmetric matrix pencil $A\lambda+C$ with A positive definite. It is known that $R(\mathbf{q})$ has a stationary value when $\mathbf{q}$ is a characteristic vector of $A\lambda+C$ and that

$$\lambda_i = -\frac{\mathbf{q}_i'C\mathbf{q}_i}{\mathbf{q}_i'A\mathbf{q}_i} \tag{2}$$

where $\mathbf{q}_i$ is a characteristic vector corresponding to the characteristic value λ_i. The case of unsymmetric pencils has been considered by Ostrowski [4] who has suggested the expression

$$R(\mathbf{q}, \mathbf{r}) = -\frac{\mathbf{r}'C\mathbf{q}}{\mathbf{r}'A\mathbf{q}} \tag{3}$$

and it has been proved (see Lancaster [3]) that "if $A\lambda+C$ is a simple matrix pencil and the characteristic values are defined in such a way that

$$R'AQ = I \quad \text{and} \quad R'CQ = -\Lambda \tag{4}$$

(where Q and R are the matrices of linearly independent right and left characteristic vectors, respectively and Λ is the diagonal matrix of characteristic values), then $R(\mathbf{q}, \mathbf{r})$ has a stationary value at $\mathbf{q}=\mathbf{q}_i$, $\mathbf{r}=\mathbf{r}_i$ (a pair of characteristic vectors associated with the characteristic value λ_i) and

$$R(\mathbf{q}_i, \mathbf{r}_i) = \lambda_i. \tag{5}$$

Lancaster [3] has further generalized this quotient to the case of a lambda matrix $D(\lambda)=\lambda^l A_0+\lambda^{l-1}A_1+\cdots+A_l$ where the A_i's are matrices, under the assumption that $D(\lambda)$ is a simple lambda matrix. Kummer [2] extends this idea to an infinite dimensional space. But he operates on a Hilbert space where one can take advantage of an inner product. In this paper we further generalize the Rayleigh quotient to an arbitrary Banach space and prove that the usual properties still obtain.

2. **Generalization of the Rayleigh quotient.** We consider an operator of the form

$$D(\lambda) = \lambda^l A_0 + \lambda^{l-1} A_1 + \cdots + A_l \tag{6}$$

where the A_i's are bounded linear operators from a complex Banach space B into itself and A_0 has a bounded inverse. We introduce the product space

$$\bar{B} = B_1 \times B_2 \times \cdots \times B_l (B_i = B,\ i = 1, 2, \ldots, l)$$

where any element $\bar{x} \in \bar{B}$ is given by

$$\bar{x} = (x^{(1)}, x^{(2)}, \ldots, x^{(l)}) x^{(i)} \in B (i = 1, 2, \ldots, l).$$

We also define operators A and C from $\bar{B} \to \bar{B}$ given by $A\bar{x} = \bar{y}$ where

$$y^{(k)} = A_0 x^{(l-k+1)} + A_1 x^{(l-k+2)} + \cdots + A_{k-1} x^{(l)} \qquad (k = 1, 2, \ldots, l) \tag{7}$$

and $C\bar{x} = \bar{y}$ where

$$y^{(k)} = -A_0 x^{(l-k)} - A_1 x^{(l-k+1)} - \cdots - A_{k-1} x^{(l-1)} \qquad k = 1, 2, \ldots, (l-1)$$

(8) and

$$y^{(l)} = A_l x^{(l)}.$$

(See [5] for more information on $\bar{B}$, A, and C.) Let $\bar{B}^*$ be the adjoint space of $\bar{B}$. It can easily be proved that any element $\bar{r}^*$ belonging to $\bar{B}^*$ can be written as

$$\bar{r}^* = (r^{(1)*}, r^{(2)*}, \ldots, r^{(l)*}) \qquad r^{(i)*} \in B^*.$$

Let

$$D^*(\lambda) = \lambda^l A_0^* + \lambda^{l-1} A_1^* + \cdots + A_l^* \tag{9}$$

be the adjoint operator from B^* into B^*, each A_i^* being the adjoint of A_i. We shall write $D^{1*}(\lambda)$ for $dD^*(\lambda)/d\lambda$. The following three results can easily be verified (cf. §2.2 of [2])

$$D(\lambda+h) = D(\lambda) + D^1(\lambda)h + 0(h^2) \qquad \text{as} \quad h \to 0 \tag{10}$$

$$D^*(\lambda+h) = D^*(\lambda) + D^{1*}(\lambda)h + 0(h^2) \qquad \text{as} \quad h \to 0 \tag{11}$$

$$D^{1*}(\lambda) = -D^*(\lambda) \frac{d}{d\lambda} [D^{*-1}(\lambda)] D^*(\lambda) \tag{12}$$

Definition. In the product space $\bar{B}$ let $\bar{q}$ denote the vector whose components are $(\lambda^{l-1}q, \lambda^{l-2}q, \ldots, \lambda q, q)$, $q \in B$ and in the adjoint space $\bar{B}^*$ let the functional $\bar{r}^*$ have the representation

$$\bar{r}^* = (\lambda^{l-1} r^*, \lambda^{l-2} r^*, \ldots, \lambda r^*, r^*)$$

where $r^* \in B^*$. We then define the expression

$$R(q, r^*, \lambda) = -\frac{\bar{r}^* C \bar{q}}{\bar{r}^* A \bar{q}} \tag{13}$$

to be the generalized Rayleigh quotient, where A and C are as given in (7) and (8).

https://doi.org/10.4153/CMB-1974-049-4 Published online by Cambridge University Press

The natural extension of the technique used by Lancaster [3] and Kummer [2] yields the following result:

LEMMA 2.1. *The expression for* $R(q, r^*, \lambda)$ *given in* (13) *can be written in the form*

$$R(q, r^*, \lambda) = \lambda - \frac{r^* D(\lambda) q}{r^* D^1(\lambda) q} \qquad q \in B, \quad r^* \in B^*. \tag{14}$$

We now prove the stationary property of the generalized Rayleigh quotient along the same lines as suggested by Lancaster [3]. Note that R is a mapping from $\bar{B} \times \bar{B}^* \times C$ into C. The weak or Gateaux differential of R at (q_i, r_i^*, λ_i) is (see [1], p. 110).

$$R^1(q_i, r_i^*, \lambda_i) = \lim_{\varepsilon \to 0} \frac{R(q_i + \varepsilon q, r_i^* + \varepsilon r, \lambda_i + \varepsilon\lambda) - R(q_i, r_i^*, \lambda_i)}{\varepsilon}$$

We show that $R^1(q_i, r_i^*, \lambda_i) = 0$.

THEOREM 2.1. *If* $r_i^* D^1(\lambda_i) q_i \neq 0$ *and* $D(\lambda_i) q_i = 0$, $D^*(\lambda_i) r_i^* = 0$, *then the expression* $R(q, r^*, \lambda)$ *defined by* (13) *or* (14) *has a stationary value at* (q_i, r_i^*, λ_i).

Proof. Let us make arbitrarily small variations in q, r^*, and λ from a set of characteristic values and characteristic vectors of $D(\lambda)$ and its adjoint operator $D^*(\lambda)$ and show that the resulting change in R is zero to the first order.

Replacing q_i, r_i^*, and λ_i by $q_i + \varepsilon q$, $r_i^* + \varepsilon r^*$, and $\lambda_i + \varepsilon\lambda$ where $\varepsilon > 0$ and q, r^*, λ are arbitrary and using (10) along with the fact that $D(\lambda_i) q_i = 0$, $D^*(\lambda_i) r_i^* = 0$, we find that as $\varepsilon \to 0$

$$(r_i^* + \varepsilon r^*) D(\lambda_i + \varepsilon\lambda)(q_i + \varepsilon q) = r_i^* D^1(\lambda_i) q_i (\varepsilon\lambda) + 0(\varepsilon^2)$$

$$(r_i^* + \varepsilon r^*) D^1(\lambda_i + \varepsilon\lambda)(q_i + \varepsilon q) = r_i^* D^1(\lambda_i) q_i + 0(\varepsilon)$$

Thus

$$\begin{aligned} R(q_i + \varepsilon q, r_i^* + \varepsilon r^*, \lambda_i + \varepsilon\lambda) &= R(q_i, r_i^*, \lambda_i) + \delta R \\ &= \lambda_i + \varepsilon\lambda - \frac{(r_i^* + \varepsilon r^*) D(\lambda_i + \varepsilon\lambda)(q_i + \varepsilon q)}{(r_i^* + \varepsilon r^*) D^1(\lambda_i + \varepsilon\lambda)(q_i + \varepsilon q)} \\ &= \lambda_i + \varepsilon\lambda - \frac{r_i^* D^1(\lambda_i) q_i (\varepsilon\lambda) + 0(\varepsilon^2)}{r_i^* D^1(\lambda_i) q_i + 0(\varepsilon)} \end{aligned}$$

Now, assuming $r_i^* D^1(\lambda_i) q_i \neq 0$, the above gives

$$R(q_i, r_i^*, \lambda_i) + \delta R = \lambda_i + \varepsilon\lambda - \varepsilon\lambda + 0(\varepsilon^2)$$

But $R(q_i, r_i^*, \lambda_i) = \lambda_i$ using (14) and the fact that $D(\lambda_i) q_i = 0$. We thus have $\delta R = 0$ to the first order in variation.

3. **An iterative procedure.** Let us assume that we have an approximation λ_0 to a characteristic value λ of $D(\lambda)$. Let ω be an arbitrary element of B and z^* an arbitrary element of the adjoint space B^*. We construct two sequences $\{q_v\}$ and $\{r_v^*\}$ according to the following rule:

https://doi.org/10.4153/CMB-1974-049-4 Published online by Cambridge University Press

$$q_\nu = D^{-1}(\lambda_\nu)w, \qquad r_\nu^* = D^{*-1}(\lambda_\nu)z^* \tag{15}$$

and define

$$\lambda_{\nu+1} = \lambda_\nu - \frac{r_\nu^*\{D(\lambda_\nu)q_\nu\}}{r_\nu^*\{D^1(\lambda_\nu)q_\nu\}} \tag{16}$$

assuming that the denominator is non-zero.

LEMMA 3.2. *The sequence $\{\lambda_\nu\}$ defined by the recurrence formula* (16) *is equivalent to*

$$\lambda_{\nu+1} = \lambda_\nu + \frac{z^*D^{-1}(\lambda_\nu)w}{z^*\left\{\frac{d}{d\lambda}D^{-1}(\lambda_\nu)w\right\}} \tag{17}$$

Proof. The expression obtained by substituting (15) in (16) can be simplified to (17) if one takes advantage of (12) and the property of the adjoint operator, viz $g^*(Tf)=(T^*g^*)f$.

Let λ_j be a characteristic value of $D(\lambda)$. We can expand $D^{-1}(\lambda)$ in a Laurent Series around λ_j. We write

$$D^{-1}(\lambda) = P^-(\lambda)+P^+(\lambda)$$

where $P^-(\lambda)$ contains all the negative powers of $(\lambda-\lambda_j)$ and $P^+(\lambda)$ all the positive powers of $(\lambda-\lambda_j)$. In particular if λ_j is a simple pole of $D^{-1}(\lambda)$ we write

$$D^{-1}(\lambda) = \frac{P_j}{\lambda-\lambda_j}+P^+(\lambda). \tag{18}$$

Typical of our further generalizations of Kummer's results is the following theorem:

THEOREM 3.2. *Let λ_j be a simple pole of $D^{-1}(\lambda)$, let the sequence $\{\lambda_\nu\}$ given by* (16) *tend to λ_j and suppose that $z^*P_jw\neq 0$. Then $\{\lambda_\nu\}$ converges at least quadratically to λ_j.*

Proof. Using (17) we have

$$\lambda_{\nu+1}-\lambda_j = (\lambda_\nu-\lambda_j)+\frac{z^*D^{-1}(\lambda_\nu)w}{z^*\left\{\left[\frac{d}{d\lambda}D^{-1}(\lambda_\nu)\right]w\right\}}$$

$$= \frac{z^*\left\{\frac{d}{d\lambda}[(\lambda-\lambda_j)D^{-1}(\lambda)]w_{\lambda=\lambda_\nu}\right\}}{z^*\left\{\left[\frac{d}{d\lambda}D^{-1}(\lambda)\right]w_{\lambda=\lambda_\nu}\right\}}$$

https://doi.org/10.4153/CMB-1974-049-4 Published online by Cambridge University Press

using (18) we have

$$(\lambda_{v+1}-\lambda_j) = \frac{z^*\left\{\left[\frac{d}{d\lambda}(\lambda-\lambda_j)P^+(\lambda)\right]w_{\lambda=\lambda_v}\right\}}{z^*\left\{\left(-\frac{P_j}{(\lambda_v-\lambda_j)^2}+\frac{d}{d\lambda}[P^+(\lambda)]_{\lambda=\lambda_v}\right)w\right\}}$$

$$\frac{\lambda_{v+1}-\lambda_j}{(\lambda_v-\lambda_j)^2} = \frac{z^*P^+(\lambda_v)w+z^*\left\{(\lambda_v-\lambda_j)\frac{d}{d\lambda}P^+(\lambda_v)\right\}w}{-z^*P_jw+(\lambda_v-\lambda_j)^2z^*\left\{\frac{d}{d\lambda}[P^+(\lambda)]w_{\lambda=\lambda_v}\right\}}$$

If $z^*P_jw\neq 0$, then the denominator is non-zero as λ_v tends to λ_j and we have quadratic convergence.

Kummer's theorems 3.1.2 and 3.2.1 of [2] concerning the rate of convergence for isolated poles of general order and the existence of convergence neighbourhoods also generalize immediately.

4. **Example for the Rayleigh quotient technique.** We consider an operator in l^1 (which we know is not a Hilbert space) and for simplicity we shall choose the polynomial operator $D(\lambda)$ given by $D(\lambda)=A\lambda-I$, where A is an operator from l^1 to l^1 given by

$$A(x_1, x_2, \ldots, x_n, \ldots) = \left(x_1, \frac{x_2}{2}, \frac{x_3}{3}, \ldots\right)$$

We know that $\lambda=1$ is a characteristic value of $D(\lambda)$ with characteristic vector $(1, 0, 0, \ldots, 0, \ldots)$. We shall obtain a sequence of iterates converging to this characteristic value starting from an initial value $\lambda_0=\frac{3}{4}$. In equation (15) w and z^* are now vectors of the space l^1 and its adjoint space m, the space of bounded sequences. It is easily seen that the adjoint operator $A^*:m\to m$ is given by

$$A^*x^* = \left\{\frac{x_n}{n}\right\} \tag{19}$$

where $x^*=\{x_n\}\in m$. We then have $D^*(\lambda)=A^*\lambda-I^*$. We choose $\lambda_0=\frac{3}{4}$, $w=(1, \frac{1}{2}, 0, \ldots)\in l$ and $z^*=(1, 1, \ldots, 1, \ldots,)\in m$. Then we have $q_v=D^{-1}(\lambda_v)w$. If $q_v=(x_1, x_2, \ldots,)$ it is found that $x_1=1/(\lambda_v-1)$, $x_2=1/(\lambda_v-2)$, $x_3=0$, $x_4=0, \ldots$. We find similarly that $r_v^*=(1/(\lambda_v-1), 2/(\lambda_v-2), 3/(\lambda_v-3), \ldots)$ and since $D^1(\lambda_v)=A$ we obtain

$$r_v^*D^1(\lambda_v)q_v = \frac{2\lambda_v^2+6\lambda_v+5}{(\lambda_v-1)^2(\lambda_v-2)^2}.$$

The iteration formula now gives

$$\lambda_{v+1} = 1+\frac{(\lambda_v-1)^2}{(\lambda_v-1)^2+(\lambda_v-2)^2}$$

https://doi.org/10.4153/CMB-1974-049-4 Published online by Cambridge University Press

starting with $\lambda_0=\frac{3}{4}$, we get

$$\lambda_1 = 1+\frac{1}{1+5^2}, \qquad \lambda_2 = 1+\frac{1}{1+5^4}, \qquad \lambda_3 = 1+\frac{1}{1+5^8}, \ldots$$

and in general $\lambda_n=1+1/(1+25^n)$ so that $\lambda_n \to 1$. Note also that

$$\frac{(\lambda_{v+1}-1)}{(\lambda_v-1)^2} = \frac{1}{(\lambda_v-1)^2+(\lambda_v-2)^2} \to 1 \quad \text{as} \quad \lambda_v \to 1$$

so that we do have quadratic convergence.

This paper forms part of the doctoral dissertation submitted to the University of Calgary by the author under the supervision of Professor P. Lancaster, University of Calgary. This research has been supported by the University of Calgary and the National Research Council of Canada.

References

1. E. Hille and R. S. Philips, *Functional analysis and semigroups*, American Mathematical Society Colloquium Publications, 1965.

2. H. Kummer, *Zur Praktischen Behandlung nichtlinearer Eigenwertaufgaben abgeschlossener linearer operatoren*, Giessen, 1964, Mitteilunger aus dem Mathem Seminar Giessen.

3. P. Lancaster, *Lambda matrices and vibrating systems*, Pergamon Press, 1966.

4. A. M. Ostrowski, *On the convergence of the Rayleigh quotient iteration for the computation of the characteristic vectors*, Archive for Rational Mechanics and Analysis, Vol. 1, No. 3, 1958, pp. 233–241.

5. M. V. Pattabhiraman and P. Lancaster, *Spectral properties of a polynomial operator*, Research Paper #45, Dept. of Math., Univ. of Calgary, 1968, Numenshe Mathematik, Vol. 13, 1969, pp. 247–259.

6. A. Wilansky, *Functional Analysis*, Blaisdell Publishing Company, New York, 1964.

University of Calgary,
Alberta, Canada