ARKIV FÖR MATEMATIK Band 3 nr 40

Communicated 24 April 1957 by LENNART CARLESON

Metric criteria of normality for complex matrices of order less than 5

EDGAR ASPLUND[1]

I. Introduction

We denote a (finite-dimensional) complex Hilbert space by F. Its elements (vectors) are denoted f, g and the scalar product of $f, g \in F$ is written (f, g). The norm of $f \in F$ is $(f, f)^{\frac{1}{2}} = \|f\|$. Elements (matrices) of the algebra $B(F)$ of endomorphisms on F are denoted by capital letters other than B and F. The norm of $A \in B(F)$ is defined by $\|A\| = \sup_{f \in F} \|Af\| \cdot \|f\|^{-1}$. The adjoint A^* of A is defined by $(Af, g) = (f, A^* g)$ for all $f, g \in F$.

An element A of $B(F)$ is called normal if it commutes with its adjoint: $A^* A = A A^*$.

As is well known, $A \in B(F)$ is normal if and only if it can be written as a sum

$$A = \sum_1^m \lambda_k E_k, \tag{I.1}$$

where λ_k are complex scalars and $E_k \in B(F)$ satisfy the conditions

$$\sum_1^m E_k = I; \qquad E_j E_k = 0, \quad j \neq k; \qquad E_k = E_k^* = E_k^2. \tag{I.2}$$

The set $\operatorname{sp} A = \{\lambda_k \mid E_k \neq 0\}$ is called the spectrum of A. From eqs. (I.1) and (I.2) it is easy to conclude that for all polynomials $p(t)$ in one variable t with complex coefficients one has

$$\|p(A)\| = \max_{\lambda \in \operatorname{sp} A} |p(\lambda)|. \tag{I.3}$$

According to a theorem of v. Neumann [1], the following converse of (I.3) holds true. If Γ is a finite subset of the complex plane and

[1] This article was written at the Department of Mathematics, Royal Institute of Technology, Stockholm, while the author was holding a scholarship from the State Council for Technical Research.

$$\|p(A)\| = \max_{\lambda \in \Gamma} |p(\lambda)| \tag{I.4}$$

for all polynomials $p(t)$ in one variable, then A is normal and $\operatorname{sp} A \subseteq \Gamma$.

We call eq. (I.4) a metric criterion of normality. The following criterion

$$\|(p(A))^2\| = \|p(A)\|^2 \quad \text{for all polynomials } p(t) \tag{I.5}$$

is equivalent[1] with (I.4), because (I.5) implies

$$\|p(A)\| = \lim_{n\to\infty} \|(p(A))^n\|^{1/n} = \max_{\lambda \in \operatorname{sp}(p(A))} |\lambda| = \max_{\lambda \in \operatorname{sp} A} |p(\lambda)|$$

by theorems of Gelfand and Dunford. Actually, as every polynomial in A can be replaced by its residue modulo the minimal polynomial of A, a sufficient condition that A shall be normal is that (I.5) shall hold for every polynomial $p(t)$ of degree less than the minimal polynomial of A.

We will be mainly concerned in this article with a weakened form of (I.5), namely

$$\|A_\lambda^2\| = \|A_\lambda\|^2, \qquad A_\lambda = A - \lambda I \tag{1.6}$$

for all complex λ. It turns out that (I.6) implies normality only if $\dim F \le 4$.

Moyls and Marcus [2] have given another criterium of normality, whose domain of applicability coincides with that of eq. (I.6). If

$$W(A) = \{\lambda \mid \lambda = (Af, f)(f, f)^{-1},\ f \in F\}$$

is the range of values of A, the condition of Moyls and Marcus reads: $W(A)$ is equal to the convex hull of $\operatorname{sp} A$. They prove that this condition implies that A is normal if $\dim F \le 4$ by representing A as a triangular matrix (Schur's lemma). We give here in the last section another proof of their theorem which uses no special representation for A.

II. A characterization of normal matrices for dim $F \le 4$

II.1. *Introductory remarks*

According to the provious section, the condition

$$\|A_\lambda^2\| = \|A_\lambda\|^2,\ A_\lambda = A - \lambda I \quad \text{for all complex } \lambda \tag{II.1.1}$$

would imply that A is normal if $\dim F = 2$. Actually, the validity of eq. (II.1.1) as a criterion of normality for A reaches further. It is valid for $\dim F = 4$, and if $\dim F$ equals 2 or 3, it is possible to restrict the variation of λ and still have a necessary condition that A shall be normal. Thus, if $\dim F = 2$ and eq. (II.1.1) holds for one complex value λ only, then A is normal and the same

[1] This equivalency was pointed out to us by Vidar Thomée.

conclusion holds if $\dim F = 3$ and eq. (II.1.1) is valid for all values λ on some straight line in the complex plane.

To prove this we need two auxiliary theorems.

Theorem 1. *The following two statements are equivalent.*

1. $\|A^2\| = \|A\|^2$.
2. *There is a vector* $f \in F$ *such that* $A^* A f = A A^* f = \|A\|^2 f$.

Proof. Suppose that statement 1 is true. As $\dim F < \infty$, there is at least one vector $g \in F$ that satisfies

$$\|A^2 g\| = \|A\|^2 \|g\| \tag{II.1.2}$$

when statement 1 is true. From (II.1.2) one obtains

$$\|A\|^2 \|g\| = \|A^2 g\| \leq \|A\| \|A g\| \leq \|A\|^2 \|g\|. \tag{II.1.3}$$

Obviously, equality must hold throughout in eq. (II.1.3). Thus the equation

$$\|A f\| = \|A\| \|f\| \tag{II.1.4}$$

is satisfied by both $f = g$ and $f = A g$. However, a necessary (and also sufficient) condition for $f \in F$ to satisfy eq. (II.1.4) is $A^* A f = \|A\|^2 f$. Using this fact, we verify statement 2 with $f = A g$.

Conversely, 2 implies 1. For let $f \in F$ satisfy $A^* A f = A A^* f = \|A\|^2 f$. Then, if one puts $A^* f = g$,

$$\|A^2 g\| = (A^2 A^* f,\ A^2 A^* f)^{\frac{1}{2}} = (A f,\ A f)^{\frac{1}{2}} \|A\|^2 = \|A\|^2 (A^* A f,\ f)^{\frac{1}{2}} =$$
$$\|A\|^2 (A A^* f,\ f)^{\frac{1}{2}} = \|A\|^2 (A^* f,\ A^* f)^{\frac{1}{2}} = \|A\|^2 \|g\|.$$

This proves theorem 1.

Theorem 2. *Any vector* f_λ *which satisfies* $A_\lambda^* A_\lambda f_\lambda = A_\lambda A_\lambda^* f_\lambda = \|A_\lambda\|^2 f$ *is in the null space of* $A^* A - A A^*$. *To prove* A *normal, one need only exhibit* ($\dim F - 1$) *linearly independent vectors lying in the null space of* $A^* A - A A^*$.

Proof. If $A_\lambda^* A_\lambda f_\lambda = A_\lambda A_\lambda^* f_\lambda = \|A_\lambda\|^2 f_\lambda$, a simple computation shows that $(A^* A - A A^*) f_\lambda = 0$. The trace of $A^* A - A A^*$ is, however, zero. Thus if $A^* A - A A^*$ has zero as a ($\dim F - 1$)-tuple eigenvalue, the remaining eigenvalue must be zero too. This concludes the proof of theorem 2.

II.2. *The main theorem*

We are now ready to prove our main theorem.

Theorem 3. *If* Λ *is a subset of the complex plane and*

$$\|A_\lambda^2\| = \|A_\lambda\|^2, \qquad A_\lambda = A - \lambda I$$

for all $\lambda \in \Lambda$, *then* $A \in B(F)$ *is normal*

1. *trivially if* dim $F=1$.
2. *if* dim $F=2$ *and* Λ *is any point.*
3. *if* dim $F=3$ *and* Λ *is any straight line.*
4. *if* dim $F=4$ *and* Λ *is the whole complex plane.*

Proof. Statement 2 is proved directly by using theorem 1 and theorem 2. To prove statement 3 we have to exhibit two linearly independent vectors f_λ. It turns out that this may be accomplished by using vectors f_λ belonging to infinite values of λ. These are defined in the following way. Suppose f_λ, $\|f_\lambda\|=1$, satisfies $A_\lambda^* A_\lambda f_\lambda = A_\lambda A_\lambda^* f_\lambda = \|A_\lambda\|^2 f_\lambda$, i.e.

$$\left.\begin{aligned} A^* A f_\lambda - (\bar{\lambda} A f_\lambda + \lambda A^* f_\lambda) + |\lambda|^2 f_\lambda = \|A_\lambda\|^2 f_\lambda, \\ A A^* f_\lambda - (\bar{\lambda} A f_\lambda + \lambda A^* f_\lambda) + |\lambda|^2 f_\lambda = \|A_\lambda\|^2 f_\lambda. \end{aligned}\right\} \tag{II.2.1}$$

We rewrite eqs. (II.2.1) in the following way, using the abbreviations $\lambda/|\lambda|=\omega$, $\bar{\omega} A + \omega A^* = A_\omega$.

$$\left.\begin{aligned} A_\omega f_\lambda - |\lambda|^{-1} A^* A f_\lambda = |\lambda|^{-1}(|\lambda|^2 - \|A_\lambda\|^2) f_\lambda, \\ A_\omega f_\lambda - |\lambda|^{-1} A A^* f_\lambda = |\lambda|^{-1}(|\lambda|^2 - \|A_\lambda\|^2) f_\lambda. \end{aligned}\right\} \tag{II.2.2}$$

Taking the difference of the two eqs. (II.2.1) we get

$$(A^* A - A A^*) f_\lambda = 0.$$

If now λ tends to infinity in such a way that ω approaches a limit, it is possible to pick out a convergent sequence f_{λ_n}, whose limit f_ω is an eigenvector of A_ω:

$$A_\omega f_\omega = m_\omega f_\omega, \tag{II.2.3}$$

and which also by continuity has the property

$$(A^* A - A A^*) f_\omega = 0. \tag{II.2.4}$$

Moreover, m_ω is the smallest eigenvalue of the self-adjoint matrix A_ω. We demonstrate this by proving that $A_\omega - (m_\omega - \varepsilon) I$ is a positive self-adjoint matrix for an arbitrary positive ε (the matrix A is said to be positive if it is self-adjoint and $(Af, f) \geq 0$ for every vector f). Namely, this matrix is this sum of the three matrices

$$A_\omega - |\lambda|^{-1} A^* A - |\lambda|^{-1}(|\lambda|^2 - \|A_\lambda\|^2) I, \qquad |\lambda|^{-1} A^* A + \frac{\varepsilon}{2} I$$

and

$$(-m_\omega + |\lambda|^{-1}(|\lambda|^2 - \|A_\lambda\|^2) I + \frac{\varepsilon}{2} I,$$

the first of which is positive by the definition of $\|A_\lambda\|$. The second and third will be positive for all sufficiently large $\lambda = \lambda_n$ corresponding to the convergent sequence f_{λ_n}.

We are thus able to obtain two eigenvectors f_ω and $f_{-\omega}$, corresponding to the eigenvalues m_ω and $m_{-\omega}$ of A_ω and $A_{-\omega} = -A_\omega$ respectively. But $m_{-\omega}$, the smallest eigenvalue of $-A_\omega$, is obviously the same as the largest eigenvalue M_ω of A_ω. If $m_\omega \neq M_\omega$, then for dim $F=3$ we have satisfied the requirements of theorem 2 and statement 3 is proved. If $m_\omega = M_\omega$, then $A_\omega = m_\omega I$ and we have

$$A^* = \bar{\omega}\, m_\omega I - \bar{\omega}^2 A,$$

which is enough for normality in any case.

When we start out to prove statement 4 we can thus assume the existence of f_ω and $f_{-\omega}$ satisfying

$$\left.\begin{aligned} A_\omega f_\omega &= m_\omega f_\omega \\ A_\omega f_{-\omega} &= M_\omega f_{-\omega} \end{aligned}\right\} \tag{II.2.5}$$

with $m_\omega < M_\omega$. As Λ is now the whole complex plane, we can construct in the same way for an $\omega' \neq \pm\omega$ eigenvectors $f_{\omega'}$, $f_{-\omega'}$ satisfying

$$\left.\begin{aligned} A_{\omega'} f_{\omega'} &= m_{\omega'} f_{\omega'} \\ A_{\omega'} f_{-\omega'} &= M_{\omega'} f_{-\omega'} \end{aligned}\right\} \tag{II.2.6}$$

with $m_{\omega'} \neq M_{\omega'}$. Now, either we have enough vectors for use in theorem 2 to prove A normal or else f_ω, $f_{-\omega}$ and $f_{\omega'}$, $f_{-\omega'}$ span the same two-dimensional subspace $F_1 \subset F$. As A_ω and $A_{\omega'}$ are two independent linear homogeneous functions of A and A^* we conclude from eqs. (II.2.5) and (II.2.6) that this subspace is reduced by both A and A^*. Thus F_1 is spanned by two eigenvectors of A corresponding to different eigenvalues λ_1 and λ_2 (A acts as a normal matrix on F_1 because $A^*A - AA^*$ annihilates F_1, therefore $\lambda_1 = \lambda_2$ would contradict $m_\omega < M_\omega$). If we put $\omega'' = i(\lambda_1 - \lambda_2)\,|\lambda_1 - \lambda_2|^{-1}$, it is easy to verify that every vector of F_1 is an eigenvector of $A_{\omega''}$ with the same eigenvalue, namely $2\,Im(\bar{\lambda}_1 \lambda_2)\,|\lambda_1 - \lambda_2|^{-1}$. When we then repeat the construction of eigenvectors $f_{\omega''}$ and $f_{-\omega''}$ corresponding to the smallest and largest eigenvalue of $A_{\omega''}$ respectively, it is clear that at least one of $f_{\omega''}$ and $f_{-\omega''}$ does not belong to F_1 (i.e. the subspace generated by f_ω and $f_{-\omega}$) or else $A_{\omega''} = m_{\omega''} I$, i.e. $A^* = \bar{\omega}'' m_{\omega''} I - \bar{\omega}''^2 A$, and in either either case A must be normal. Thus all statements of theorem 4 are proved.

II.3. *Counterexample for* dim $F=5$

We now construct a non normal matrix A of order 5 such that

$$\|A_\lambda^2\| = \|A_\lambda\|^2, \quad A_\lambda = A - \lambda I \tag{II.3.1}$$

for all complex λ. Let F be the direct sum of two mutually orthogonal subspaces F_1 and F_2, dim $F_1 = 3$ and dim $F_2 = 2$. Let A, A_λ be represented by the block matrices

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \qquad A_\lambda = \begin{pmatrix} A_1 - \lambda I_1 & 0 \\ 0 & A_2 - \lambda I_2 \end{pmatrix} = \begin{pmatrix} A_{1\lambda} & 0 \\ 0 & A_{2\lambda} \end{pmatrix}.$$

From the definition of the norm of A it is obvious that

$$\|A_\lambda\| = \max\{\|A_{1\lambda}\|,\ \|A_{2\lambda}\|\}.$$

We choose A_1 to be a normal matrix such that

$$\|A_{1\lambda}\| \geq 1 + |\lambda|.$$

This can be accomplished by choosing $2,\ -1 \pm i\sqrt{3}$ as eigenvalues of A_1. Then we take A_2 to be a non normal matrix of norm 1, e.g. $A_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, which thus satisfies

$$\|A_{2\lambda}\| \leq 1 + |\lambda|$$

Then $$\|A_\lambda^2\| = \max\{\|A_{1\lambda}^2\|,\ \|A_{2\lambda}^2\|\} = \|A_{1\lambda}^2\| = \|A_{1\lambda}\|^2 = \|A_\lambda\|^2$$

which proves the assertion of eq. (II.3.1).

III. Properties of eigenvalues which lie on the boundary of the range of values of a matrix

Moyls and Marcus (2) have proved that the eigenvectors corresponding to eigenvalues of $A \in B(F)$ lying on the boundary of the range of values $W(A) = \{\lambda \mid \lambda = (Af, f)(f, f)^{-1},\ f \in F\}$ are eigenvectors also of A^*. It then follows in the same way as in the proof of our theorem 3 that if only one (simple) eigenvalue of A lies in the interior of $W(A)$ (this must always be the case when $\dim F \leq 4$ and $W(A) =$ convex hull of $\operatorname{sp} A$), then A is normal. Moyls and Marcus prove their result representing A as a triangular matrix by means of Schur's lemma, but the theorem is really a consequence of a simple property of the boundary of $W(A)$ and can be proved without using any special representation.

Theorem 4. *If $f \in F$ is an eigenvector of $A \in B(F)$ corresponding to an eigenvalue λ which lies on the boundary of the range of values of A, then f is also an eigenvector of A^*.*

We prove theorem 5 by a variational method. Put $f_\mu = f + \mu g$ and determine

$$\begin{aligned} d\lambda &= \{d[(Af_\mu, f_\mu)(f_\mu, f_\mu)^{-1}]\}_{\mu=0} \\ &= \{d[(\lambda f + \mu A g,\ f + \mu g)(f + \mu g,\ f + \mu g)^{-1}]\}_{\mu=0} \\ &= [d\mu\,(Ag, f) + \lambda\, d\bar{\mu}\,(f, g)]\,(f, f)^{-1} - \lambda\,(f, f)^{-1}\,[d\mu\,(g, f) + d\bar{\mu}\,(f, g)] \\ &= d\mu\,[(Ag, f) - \lambda\,(g, f)]\,(f, f)^{-1} \\ &= d\mu\,(g,\ A^* f - \bar{\lambda} f)\,(f, f)^{-1}. \end{aligned}$$

Since, however, $d\lambda$ is restricted because λ lies on the boundary of $W(A)$ but $d\mu$ is not, we must have

$$(g, A^* f - \bar{\lambda} f) = 0$$

for all $g \in F$, which proves the theorem.

REFERENCES

1. VON NEUMANN, J., Eine Spektraltheorie für allgemeine Operatoren eines unitären Raumes, Math. Nachr. *4*, 258–281 (1951).
2. MOYLS, B. N., and MARCUS, M. D., Field convexity of a square matrix, Proc. Amer. Math. Soc. *6*, 981–983 (1955).

Tryckt den 25 juni 1957

Uppsala 1957. Almqvist & Wiksells Boktryckeri AB