Sonderabdruck aus
ARCHIV DER MATHEMATIK
Vol. X, 1959 BIRKHÄUSER VERLAG, BASEL UND STUTTGART Fasc. 1

Matrix Algebras and Groups Relatively Bounded in Norm

By J. L. MOTT in Edinburgh and HANS SCHNEIDER in Belfast

1. It is well-known that if ν is a norm on the algebra C_n of all n-th order complex matrices then $\nu(A) \geqq \varrho(A)$, where $\varrho(A)$ is the spectral radius of A, that is, the greatest modulus of the characteristic roots of A. As a preliminary, we complete a theorem of HOUSEHOLDER [6], [7] by giving sufficient, and now also necessary, conditions that there exists a norm ν such that $\nu(A) = \varrho(A)$ for a given matrix A of C_n. This leads naturally to the wider problem: to find necessary and sufficient conditions for the existence of a norm satisfying $\nu(A) = \varrho(A)$ for all A of a subset $\mathfrak{A}$ of C_n. If this is so we say that the norm ν is *minimal* on the set $\mathfrak{A}$.

We find that the conditions for the existence or otherwise of such a minimal norm are related to the behaviour of the ratio $\nu(A)/\varrho(A)$. We say that $\mathfrak{A}$ is *relatively bounded* if, for some norm ν, the ratio $\nu(A)/\varrho(A)$ is bounded on the subset $\mathfrak{A}$ of C_n. This condition is to be understood as implying that $\nu(A) = 0$ when $\varrho(A) = 0$. The norm ν here is arbitrary, but a result of OSTROWSKI [11] shows that if ν' is a second norm on C_n then $\nu'(A)/\nu(A)$ is bounded on C_n so that, if $\nu(A)/\varrho(A)$ is bounded for some particular norm, then it is in fact bounded for all norms. We show that the relatively bounded subalgebras of C_n are precisely the algebras of simultaneously diagonable, and therefore, commutative matrices. We shall demonstrate the existence of minimal norms on all relatively bounded matrix algebras and also on certain relatively bounded matrix groups.

2. We give the usual definition of a norm (see FADDEEVA [4], HOUSEHOLDER [6] or OSTROWSKI [11]). A (multiplicative) norm on the algebra C_n is a mapping ν of C_n into the non-negative real numbers satisfying

(i) $\nu(A) > 0$ if $A \neq O$;

(ii) $\nu(\lambda A) = |\lambda| \, \nu(A)$ for any scalar λ;

(iii) $\nu(A + B) \leqq \nu(A) + \nu(B)$;

(iv) $\nu(AB) \leqq \nu(A)\, \nu(B)$.

A consequence of these axioms is that $\nu(O) = 0$.

We first show that for any norm $\nu(A) \geqq \varrho(A)$. Various proofs of this result are known and we include a proof partly for completeness but also to emphasise the methods used in the rest of this note. We shall make use of the classical canonical form of a matrix (see, for example, TURNBULL and AITKEN [13] or WEDDERBURN [14]). Thus the classical canonical form Λ of A is a direct sum of certain upper triangular matrices:

(1) $$\Lambda = X^{-1} A X = \sum \oplus (\lambda_i I_i + U_i),$$

where the I_i are unit matrices, the U_i are auxiliary unit matrices with elements unity in the superdiagonal and zeros elsewhere, so that U_i is empty if I_i is of order 1, and the λ_i are the characteristic roots of A.

We shall use in our proof of this theorem and the next the two particular norms that follow. If $B = [b_{ij}]$ is a matrix of C_n we define the norm ν_1 by $\nu_1(B) = \max. \sum_j |b_{ij}|$, and then define a second norm ν_2 by $\nu_2(B) = \nu_1(X^{-1}BX)$, so that $\nu_2(A) = \nu_1(\Lambda)$.

Theorem 1. *Let A be an n-th order complex matrix. The greatest lower bound of $\nu(A)$ over all norms ν is the spectral radius $\varrho(A)$.*

Proof. By Ostrowski's theorem mentioned above we have, for any norm ν, a positive constant K such that $\nu(A^r) \geqq K\nu_2(A^r)$ for all positive integers r. It then follows, using also the definition of a norm, that

$$\nu(A)^r \geqq \nu(A^r) \geqq K\nu_2(A^r) = K\nu_1(\Lambda^r)\,.$$

But Λ^r has an element of absolute value $\varrho(A)^r$ in the leading diagonal so that

$$\nu_1(\Lambda^r) \geqq \varrho(A)^r$$

and hence

$$\nu(A)^r \geqq K\varrho(A)^r\,. \tag{2}$$

If $\varrho(A) = 0$ the theorem is clearly true, so we can suppose that $\varrho(A) > 0$. It then follows immediately from (2), which is true for all positive integers r, that $\nu(A) \geqq \varrho(A)$, for otherwise $\nu(A)^r/\varrho(A)^r < K$, which is positive, for sufficiently large values of r.

Thus $\varrho(A)$ is a lower bound for $\nu(A)$. To show that it is the greatest lower bound let D be a diagonal matrix given by $D = \text{diag.}\{1, \varepsilon, \ldots, \varepsilon^{n-1}\}$ for some $\varepsilon > 0$, and write $\Lambda_1 = D^{-1}\Lambda D$. Define a third norm ν_3 by $\nu_3(B) = \nu_1(D^{-1}X^{-1}BXD)$ for all B of C_n. Then consideration of Λ_1 shows immediately that

$$\nu_3(A) = \nu_1(\Lambda_1) \leqq \varrho(A) + \varepsilon\,,$$

so that $\varrho(A)$ is the greatest lower bound for $\nu(A)$, and the theorem is proved.

3. We now consider the class of matrices for each of which a norm exists such that the greatest lower bound of the preceding theorem is attained, that is, the nature of a matrix A for which a norm exists that is minimal on A alone.

We shall call a characteristic root of A *maximal* if it is not exceeded in modulus by any other root.

Theorem 2. *Let A be an n-th order complex matrix. Then the following four conditions are equivalent:*

(i) *Each maximal characteristic root of A is a simple zero of the minimum polynomial (that is, the elementary divisors belonging to each maximal root are simple).*
(ii) *The semi-group $(A)^+$ of positive powers of A is relatively bounded.*
(iii) *There exists a norm minimal on $(A)^+$.*
(iv) *There exists a norm ν for which $\nu(A) = \varrho(A)$.*

Proof. It is immediate that (iii) implies (ii) and almost immediate that (iv) implies (iii), for with (iv) we have

$$\nu(A^r) \leqq \nu(A)^r = \varrho(A)^r = \varrho(A^r)\,,$$

so that (iii) holds (and in fact $\nu(A^r) = \varrho(A^r)$, from Theorem 1).

That (i) implies (iv) (see HOUSEHOLDER [6]) can be proved as follows. If every characteristic root of A is maximal the result is immediate with norm ν_3 as in the proof of Theorem 1. Otherwise, define the diagonal matrix D as before and suppose that $\varrho'(A)$ is the greatest modulus of the characteristic roots of A other than the maximal roots. Then if (i) holds we have by the same argument as before that $\nu_3(A) \leqq \max\{\varrho(A), \varrho'(A) + \varepsilon\}$, so that, since $\varrho'(A) < \varrho(A)$, it follows on taking ε sufficiently small that $\nu_3(A) = \varrho(A)$.

To prove the main part of the theorem, that (ii) implies (i), we shall show that the negation of (i) implies the negation of (ii). Thus suppose that some characteristic root of A is not a simple zero of the minimum polynomial. Then $A \neq 0$, so that if $\varrho(A) = 0$ then $\nu(A) > 0 = \varrho(A)$, and hence $(A)^+$ is not relatively bounded. If, on the other hand, $\varrho(A) > 0$ then the matrix Λ^r of the proof of Theorem 1 has an element of absolute value $r\varrho(A)^{r-1}$ in the superdiagonal, whence $\nu_1(\Lambda^r) > r\varrho(A)^{r-1}$. Thus for all positive r

$$\nu_2(A^r)/\varrho(A^r) \geqq \nu_1(\Lambda^r)/\varrho(A)^r > r/\varrho(A)\,,$$

and in this case also $(A)^+$ is not relatively bounded. Thus (ii) implies (i), and the proof of the theorem is complete.

4. We now consider the existence of a minimal norm, not on a single matrix A, but on a subalgebra $\mathfrak{A}$ of C_n. We consider first the case when $\mathfrak{A}$ is an algebra of polynomials in A. This widening of the domain on which the norm is minimal naturally leads to a strengthening of the conditions for its existence. We give such necessary and sufficient conditions in the following theorem.

Theorem 3. *Let $C[A]$ be the algebra of all polynomials in A, and $AC[A]$ the algebra of such polynomials without constant term. Then the following conditions are equivalent:*
(i) *The matrix A is diagonable.*
(ii) *The algebra $AC[A]$ is relatively bounded.*
(iii) *The algebra $C[A]$ is relatively bounded.*
(iv) *There exists a norm minimal on $C[A]$.*

Proof. Evidently (iv) implies (iii), and (iii) implies (ii) since the algebra $AC[A]$ is contained in the algebra $C[A]$. Also (i) implies (iv), since we can take for the norm here the norm ν_2 of Theorem 1. Thus we have only to prove that (ii) implies (i).

We do this as in the preceding theorem by showing that the negation of (i) implies the negation of (ii). Let $\lambda_1', \ldots, \lambda_r'$ denote the distinct non-zero characteristic roots of A, and define the polynomial $p(x)$ by $p(x) = x\prod_{i=1}^{r}(x - \lambda_i')$. Then for each characteristic root λ (possibly zero) of A we have $p(\lambda) = 0$ but $p'(\lambda) \neq 0$. The matrix polynomial $P = p(A)$ belongs to $AC[A]$, and if $\Lambda = X^{-1}AX = \sum\oplus(\lambda_i I_i + U_i)$ is the classical canonical form of A as in (1), then

$$X^{-1}PX = p(\Lambda) = \sum\oplus\{p'(\lambda_i)\,U_i + p''(\lambda_i)\,U_i^2/2! + \cdots + p^{(m-1)}(\lambda_i)\,U_i^{m-1}/(m-1)!\}\,,$$

where m is the order of I_i. Thus $p(\Lambda)$ has zeros throughout its leading diagonal and hence $\varrho(P) = \varrho(\Lambda) = 0$. On the other hand if (i) does not hold, so that A is not diagonable, then at least one U_i of the direct sum (1) is not empty whence, since $p'(\lambda_i) \neq 0$, the superdiagonal of $p(A)$ is not zero. Thus $\nu_1(A) > 0$ and hence $\nu_2(P) > 0$. We conclude P is not relatively bounded and (ii) does not hold. Thus (ii) implies (i), and the theorem is proved.

5. We now generalise Theorem 3 by considering the conditions under which a minimal norm can exist on a general subalgebra $\mathfrak{A}$ of C_n. In doing this we use a theorem essentially due to Motzkin and Taussky [9] (cf. also [3], [12]): *If $\mathfrak{A}$ is a linear space of matrices each of which is diagonable, then all matrices of $\mathfrak{A}$ are simultaneously diagonable and therefore commute.* This theorem was proved in [9] for the case of $\dim \mathfrak{A} = 2$. To pass to the general case, let $A_1, A_2, \ldots, A_s$ be a linear basis for $\mathfrak{A}$, and note that, since the space generated by each pair of matrices A_i and A_j consists of diagonable matrices, each A_i and A_j commute. Thus $A_1, \ldots, A_s$ is a set of commutative diagonable matrices and the matrices are therefore simultaneously diagonable. The general theorem follows.

We shall call $\mathfrak{A}$ *diagonable* if all matrices of $\mathfrak{A}$ can be diagonalized simultaneously.

Theorem 4. *Let $\mathfrak{A}$ be an algebra of n-th order complex matrices. Then the following conditions are equivalent:*

(i) *The algebra $\mathfrak{A}$ is diagonable.*

(ii) *The algebra $\mathfrak{A}$ is relatively bounded.*

(iii) *There exists a norm minimal on $\mathfrak{A}$.*

Proof. We have only to prove that (ii) implies (i). To do this let A be a matrix of $\mathfrak{A}$. The algebra $A\,C[A]$ is contained in $\mathfrak{A}$ and so, if (ii) holds, is relatively bounded. Then, by Theorem 3, A is diagonable, and then (i) follows from the theorem of Motzkin and Taussky quoted above.

6. We could have moved from section 3 in a different direction, by taking as a subset of C_n more general than a single matrix A not an algebra but a group; and this we now do. One particular case of a group on which a minimal norm evidently does exist is that of a group of simultaneously diagonable matrices, for we can consider the algebra generated by the group and so have the existence of a minimal norm assured by Theorem 4. We shall exhibit, however, certain relatively bounded groups which are not diagonable, or even commutative, and show that a minimal norm exists for them.

We call a matrix of C_n a *group matrix* if it is a member of some group of n-th order matrices. A matrix A is a group matrix if and only if A is either non-singular or has zero as a simple root of its minimum equation (Farahat and Mirsky [5] or Barnes and Schneider [2].) Further, if Z is the group inverse of A in a particular group $\mathfrak{G}$ then it is the inverse of A in any other group that contains A. For there exists a matrix Y in C_n such that, for all B in $\mathfrak{G}$, $Y^{-1}BY$ is a direct sum given by $Y^{-1}BY = B_1 \oplus O$, where B_1 is non-singular and of the same order for all B of $\mathfrak{G}$ and O may be empty, so that the group inverse of A is the matrix Z for which $Y^{-1}ZY = A_1^{-1} \oplus O$. Thus the group (A) consisting of AZ and of $(A)^+$ and $(Z)^+$, the semi-groups of positive powers of A and Z respectively, is the intersection of all matrix groups containing A and is naturally called the cyclic group generated by A.

We now find conditions for the existence of a minimal norm on such a cyclic group.

Theorem 5. *Let A be a group matrix in C_n. Then the following conditions are equivalent:*

(i) *Each maximal characteristic root, and each root of least non-zero modulus, is a simple zero of the minimum polynomial.*

(ii) *The cyclic group (A) is relatively bounded.*

(iii) *There exists a norm minimal on (A).*

Proof. All three conditions hold trivially in the case of $A = O$. We can therefore suppose that $A \neq O$, and it then follows since A is a group matrix that $\varrho(A) > 0$.

It is obvious that (iii) implies (ii). To prove that (ii) implies (i) we suppose that λ_0 is a non-zero characteristic root of A of least modulus. If Z is the group inverse of A, then $\varrho(Z) = |\lambda_0|^{-1}$, and $(A)^+$ and $(Z)^+$ are relatively bounded by our assumption. Condition (i) then follows from Theorem 2 applied to the semi-groups $(A)^+$ and $(Z)^+$.

To prove that (i) implies (iii) let $X^{-1}AX$ be the canonical form of A as in (1), so that $X^{-1}AX = \sum \oplus (\lambda_i I_i + U_i)$. Also

$$X^{-1}ZX = \sum \oplus \{\mu_i I_i - \mu_i^2 U_i + \mu_i^3 U_i^2 - \cdots + (-1)^{m-1} \mu_i^m U_i^{m-1}\} \tag{3}$$

and $X^{-1}AZX = \sum \oplus \delta_i I_i$ where $\mu_i = \delta_i = 0$ if $\lambda_i = 0$ and $\mu_i = \lambda_i^{-1}$, $\delta_i = 1$ otherwise. If, as in Theorems 1 and 2, ν_3 is the norm given by $\nu_3(B) = \nu_1(D^{-1}X^{-1}BXD)$, where $D = (1, \varepsilon, \ldots, \varepsilon^{n-1})$ and $0 < \varepsilon < \varrho(A) - \varrho'(A)$, then we know already from Theorem 2 that $\nu_3(A) = \varrho(A)$ and ν_3 is minimal on A and hence on $(A)^+$. But this norm is also minimal on Z, and hence on $(Z)^+$, if ε is sufficiently small. This we show in a similar way, but have to consider now the non-zero characteristic roots of least modulus. Let λ_0 be asab ove and $|\lambda_1|$ be the least modulus of roots λ for which $|\lambda| > |\lambda_0|$. If there are no such roots of modulus $|\lambda_1|$ the result is trivial. Otherwise we have from the condition (i) that all roots of modulus $|\lambda_0|$ are simple zeros of the minimum polynomial and so we have from (3) that $\nu_3(Z) = |\lambda_0|^{-1} = \varrho(Z)$ provided that ε is such that

$$1 + |\varepsilon\mu_1| + |\varepsilon\mu_1|^2 + \cdots + |\varepsilon\mu_1|^{m-1} \leqq |\mu_0/\mu_1|,$$

which is certainly true if $\varepsilon < |\lambda_1| - |\lambda_0|$. Thus for ε sufficiently small ν_3 is a norm minimal on both $(A)^+$ and $(Z)^+$ and so, since $\nu_3(AZ) = 1 = \varrho(AZ)$, ν_3 is a norm minimal on (A).

7. The cyclic group (A) above is a particular commutative group, but a minimal norm can exist on a non-commutative group.

Theorem 6. *Let $\mathfrak{G}$ be a group of n-th order unitary matrices. Then there exists a norm minimal on $\mathfrak{G}$.*

Proof. Define the norm ν_4 by $\nu_4(B) = \sqrt{\varrho(B^*B)}$ for all B of C_n. Then $\nu_4(A) = 1 = \varrho(A)$ for all A of $\mathfrak{G}$.

This theorem allows us to show the existence of a minimal norm on a wider class of groups. A group $\mathfrak{G}$ is *bounded* if for some norm ν, and hence for all norms ν by Ostrowski's theorem, $\nu(A)$ is bounded for all A of $\mathfrak{G}$. Thus, in particular, a finite group is bounded. By Auerbach's Theorem [1] (see also Littlewood [8] and Murnaghan [10]) any bounded group of non-singular matrices is similar to a unitary group. We now have, almost at once,

Theorem 7. *Let $\mathfrak{G}$ be a bounded group of n-th order complex matrices. Then $\mathfrak{G}$ is relatively bounded, and there exists a norm minimal on $\mathfrak{G}$.*

Proof. If $\mathfrak{G}$ consists of the zero matrix alone, the result is obvious. Otherwise there exists a matrix Y of C_n such that $Y^{-1}AY = A_1 \oplus O$ for each A of $\mathfrak{G}$, where A_1 is non-singular and of the same order for each A. By AUERBACH's Theorem we can choose Y so that each A_1 is unitary. We now define the norm ν_5 by $\nu_5(B) = \nu_4(Y^{-1}BY)$ for all B of C_n, so that $\nu_5(A) = \nu_4(Y^{-1}AY) = 1 = \varrho(A)$ for all A of $\mathfrak{G}$. This proves the theorem.

8. We could have formulated some of our theorems in a slightly different way, strengthening them a little in one direction but weakening them in another. For if ν and ν' are norms, not on the whole of C_n but merely on a subalgebra containing the subset $\mathfrak{A}$ of C_n, then $\nu'(A)/\nu(A)$ is bounded on $\mathfrak{A}$. Thus we could have called a subset relatively bounded if there exists a norm ν on an algebra containing $\mathfrak{A}$ such that $\nu(A)/\varrho(A)$ is bounded on $\mathfrak{A}$. With this definition the conditions of the theorems are changed correspondingly. Thus in Theorem 4, for example, (iii) is replaced by (iii)': *ϱ is a norm on $\mathfrak{A}$.*

References

[1] H. AUERBACH, Sur les groupes bornés de substitutions linéares. C. r. Acad. Sci., Paris **195**, 1367—1369 (1932).

[2] W. E. BARNES — H. SCHNEIDER, The group membership of a polynomial in an element algebraic over a field. Arch. Math. **8**, 166—168 (1957).

[3] M. P. DRAZIN, Algebraic and diagonable rings. Canadian J. Math. **8**, 341—354 (1956).

[4] V. N. FADDEEVA, Computational methods of linear algebra. National Bureau of Standards Report 1644 (1950).

[5] H. K. FARAHAT — L. MIRSKY, A condition for diagonability of matrices. Amer. Math. Monthly **63**, 410—412 (1956).

[6] A. S. HOUSEHOLDER, On the convergence of matrix iterations. Oak Ridge National Laboratory-1833, Physics (1955) and Journal Association Computing Machinery **3**, 314—324 (1956).

[7] A. S. HOUSEHOLDER, The approximate solution of matrix problems. Journal Association Computing Machinery **5**, 205—243 (1958).

[8] D. E. LITTLEWOOD, The Theory of Group Characters and Matrix Representations of Groups. Oxford 1940.

[9] T. S. MOTZKIN — O. TAUSSKY, Pairs of matrices with property *L*, II. Trans. Amer. Math. Soc. **80**, 387—401 (1955).

[10] F. D. MURNAGHAN, The Theory of Group Representations. Baltimore 1938.

[11] A. OSTROWSKI, Über Normen von Matrizen. Math. Z. **63**, 2—18 (1955).

[12] O. TAUSSKY, Commutativity in finite matrices. Amer. Math. Monthly **64**, 229—235 (1957).

[13] H. W. TURNBULL — A. C. AITKEN, An Introduction to the Theory of Canonical Matrices. London 1932.

[14] J. H. M. WEDDERBURN, Lectures on Matrices. New York 1934.

Eingegangen am 18. 6. 1958