

## NORMS AND THE LOCALIZATION OF ROOTS OF MATRICES<sup>1</sup>

A. S. HOUSEHOLDER

Before I embark upon my topic, perhaps a bit of orientation might be in order. The topic lies in a borderland between numerical analysis and pure mathematics. The problems arise out of very important practical computational difficulties; the results are due almost exclusively to those who have rather more than a passing interest in computation; they have been published largely in books and journals devoted to numerical analysis and applied mathematics; but the more elaborate developments go rather beyond anything that can promise any direct and immediate practical application. Here we enter what I think must be classified as pure mathematics.

The phrase "numerical analysis" itself is a fairly recent coinage, dating back about twenty years to the advent of the electronic computer. What it refers to is neither pure mathematics nor classical applied mathematics. I would say that classical applied mathematics is primarily descriptive, at least in intent, whereas numerical analysis is primarily prescriptive in intent. To the extent that it is descriptive, it describes, or attempts to describe, processes and their outcome, but in terms that permit a rational choice among possibly competing alternatives. I leave you to characterize pure mathematics as you like, but I think it differs, at least in intent, from either. I make these points only for purposes of orientation and not argumentation.

My title is in two parts and is intended to indicate an intersection and not a union. A localization theorem for roots of a matrix is one that gives information concerning the location of the roots. There are two main classes, exclusion and inclusion. An inclusion theorem designates a set in the complex plane containing at least one root of a given matrix. An exclusion theorem designates a set containing none, so that its complement contains all. It is easy to see why such theorems would be of interest to numerical analysts. Any computation that does not employ exact arithmetic, and this means most, can produce only an approximation, at best, to the mathematically defined result. Naturally one is therefore interested in knowing the

---

An address delivered before the Knoxville, Tennessee meeting of the Society on November 11, 1967 by invitation of the Committee to Select Hour Speakers for Southeastern Sectional Meetings; received by the editors April 22, 1968.

<sup>1</sup> Research sponsored by the U. S. Atomic Energy Commission under contract with the Union Carbide Corporation.

limits of possible error, hence of knowing an inclusion set that contains the mathematically defined result. It is not farfetched to imagine that norms could be used to identify such inclusion sets. This is in fact the case, and my purpose is to show how norms have been used in particular to identify such sets for the roots of matrices. Not all localization theorems can be so obtained, or, at least, not all have been so obtained. But many have been, and I, with perhaps some bias, feel that these are the more interesting and more important ones.

It is easy to see why norms should be useful to the numerical analyst. They provide the obvious tools for measuring rates of convergence of sequences in  $n$ -space, and in the measurement of error. The rather surprising fact is that they seem not to have come into general use until the late 1950's, although they appear prominently in papers by Kantorovitch [24], Hotelling [14], and Lonseth [29] before 1950, and were reintroduced by Faddeeva [7] in 1950. But convergence questions are closely related to tests for the nonsingularity of a matrix, and again with questions concerning the location of the roots of a matrix. While I want to concentrate mainly on the third area, what I have to say has relevance also to each of the others.

The connection between problems of the three classes is easy to see. One special convergence problem concerns the vanishing of a sequence of powers  $B^n$  of a given fixed matrix. It is well known that this occurs if and only if all its roots lie within the unit disk. Hence any condition that assures that the roots lie within the unit disk is sufficient to assure the vanishing of the sequence  $B^n$ . This also assures the nonsingularity of the matrix  $I - B$ . But a nonsingularity criterion applied to the matrix  $\lambda I - A$  immediately gives information about where the roots of  $A$  cannot lie.

It will be convenient to establish certain notational conventions once for all. Lower case Greek letters will generally represent scalars real or complex, and lower case Latin letters will represent column vectors, real or complex, barring obvious exceptions such as the use of  $i, j, k$  for indices, and of  $n$  for the dimensionality of the space. Capitals, whether Latin or Greek, will generally represent matrices, ordinarily square and of order  $n$  unless otherwise indicated. A row vector will be indicated by, say,  $a^T$  (for simple transpose), or  $a^H$  (for conjugate transpose). So far as possible corresponding letters will be used to represent elements of matrices or vectors. Thus

$$A = (a_1, a_2, \dots, a_n) = (\alpha_{ij}),$$

but  $I$  will be used for the identity,  $e_i$  for its  $i$ th column, and  $e = \sum e_i$ . The notation  $A \leq B$  signifies that the matrices  $A$  and  $B$  are both real

and that  $\alpha_{ij} \leq \beta_{ij}$  for every  $i$  and  $j$ , and similarly for  $a \leq b$ . When vertical bars enclose a matrix or a vector, the elements are replaced by their moduli. Usually  $\lambda_j(A)$  will represent a root of  $A$ , and

$$\rho(A) = \max |\lambda_j(A)|.$$

I will often speak of vectors and points interchangeably.

A seminorm is a function on  $n$ -space to the nonnegative reals satisfying

$$(i) \quad \|\lambda a\| = |\lambda| \|a\|,$$

$$(ii) \quad \|a+b\| \leq \|a\| + \|b\|.$$

A seminorm is a norm if also

$$(iii) \quad \|a\| = 0 \rightarrow a = 0.$$

The reverse implication follows obviously from (i). Associated with any norm is a set [25], [17], [20]

$$K = \{x \mid \|x\| \leq 1\},$$

which has the following properties

$$(i') \quad a \in K, \quad |\omega| \leq 1 \rightarrow \omega a \in K.$$

This property is often called *equilibration*.

$$(ii') \quad K \text{ is convex.}$$

$$(iii') \quad \text{The origin is interior to } K.$$

$$(iv') \quad K \text{ is bounded and closed.}$$

Conversely, given any set  $K$  possessing these four properties, a norm can be defined by

$$\|x\| = \min [\nu \mid \nu \geq 0, x \in \nu K].$$

It is sometimes convenient to designate the norm by  $\|x\|_K$  and to say that it belongs to  $K$ . Whenever reference is made to a convex body it will be assumed that it possesses the four properties listed above.

Given any two convex bodies of this class,  $K_1$  and  $K_2$ , there exist  $\kappa_1 > 0$  and  $\kappa_2 > 0$  such that

$$K_1 \subset \kappa_2 K_2, \quad K_2 \subset \kappa_1 K_1.$$

In this sense all norms are equivalent. But in other respects they are by no means equivalent. The most commonly used norm is the ordinary Euclidean, or  $l_2$ , norm, which belongs to the unit sphere  $S$ :

$$\|x\|_S^2 = x^H x.$$

The  $l_\infty$  norm, defined by

$$\|x\|_E = \max |\xi_i|,$$

belongs to a kind of generalized hypercube  $E$ . Let

$$[x] = [\omega x \mid |\omega| \leq 1].$$

Then

$$E = \sum [e_i],$$

where the sum is the Minkowski sum, i.e., the set of all possible vectors obtained by selecting a vector from each of the sets  $[e_i]$  and adding. The  $l_1$  norm, defined by

$$\|x\|_{E'} = \sum |\xi_i| = e^T |x|$$

belongs to the convex hull  $E'$  of the union of the sets  $[e_i]$ .

Given any set  $K$  and any matrix  $P$ , define

$$PK = [Pk \mid k \in K].$$

If  $K$  is a convex body and  $P$  is nonsingular, then  $PK$  is a convex body. Also, clearly

$$\|x\|_{PK} = \|P^{-1}x\|_K.$$

Evidently  $PS$  is an ellipsoid, for any nonsingular  $P$ . If

$$G = \text{diag} (\gamma_1, \gamma_2, \dots, \gamma_n), \quad \gamma_i > 0,$$

and  $g = Ge$ , then

$$\|x\|_{GE} = \inf [\nu \mid |x| \leq \nu g],$$

while

$$\|x\|_{G^{-1}E'} = g^T |x|.$$

Since a matrix is a vector in  $n^2$ -space, matrix norms can be characterized by the same set of properties, and associated with convex bodies in  $n^2$ -space. But an additional property is desirable:

$$(iv) \quad \|AB\| \leq \|A\| \|B\|.$$

Any matrix norm will be assumed to possess this property. A function defined on  $n^2$ -space and possessing only the first three properties will be called a generalized matrix norm, and designated by the letter  $\nu$ . Thus

- (i)  $\nu(\lambda A) = |\lambda| \nu(A),$
- (ii)  $\nu(A+B) = \nu(A) + \nu(B),$
- (iii)  $\nu(A) = 0 \rightarrow A = 0.$

But it can be shown [12] that for any generalized matrix norm  $\nu$ , there exists a scalar  $\kappa > 0$  such that the function defined by  $\|A\| = \kappa \nu(A)$  for all  $A$  satisfies (iv). The ordinary Euclidean matrix norm possesses all four properties.

Given any matrix norm, and an arbitrary vector  $a \neq 0$ ,

$$\|x\| = \|xa^H\|$$

defines a vector norm which, by virtue of (iv) is related to the matrix norm by

$$(C) \quad \|Ax\| \leq \|A\| \|x\|.$$

When a matrix norm satisfies (C) with any vector norm, it is said to be *consistent* with the vector norm. The Euclidean matrix norm is consistent with the Euclidean vector norm. Given any generalized matrix norm and any vector norm, there exists a scalar  $\kappa$  such that the function

$$\|A\| = \kappa \nu(A)$$

is a matrix norm, and it is consistent with the given vector norm.

All machinery has now been assembled for proving a quite general localization theorem. This is that

$$\rho(A) \leq \|A\|,$$

whatever the matrix  $A$  and whatever the matrix norm. The proof is quite trivial. In terms of any vector norm to which the matrix norm is consistent, if  $Ax = \lambda x$ , then

$$\|\lambda x\| = |\lambda| \|x\| = \|Ax\| \leq \|A\| \|x\|.$$

Hence every root  $\lambda_i(A)$  must satisfy

$$|\lambda_i(A)| \leq \|A\|.$$

Q.E.D.

This theorem has a kind of converse as follows [17]: For any matrix  $A$  and for any  $\epsilon > 0$ , there exists a matrix norm such that

$$\|A\| \leq \epsilon + \rho(A).$$

Thus while the norm  $\|A\|$  is never less than the spectral radius  $\rho(A)$  of a given matrix, special norms exist that come arbitrarily close to  $\rho(A)$  for that particular matrix.

Now let

$$A = B + C,$$

$$\lambda I - A = \lambda I - C - B = (\lambda I - C)[I - (\lambda I - C)^{-1}B].$$

Assuming  $\lambda I - C$  to be nonsingular, the condition  $\|(\lambda I - C)^{-1}B\| < 1$  is certainly sufficient to insure the nonsingularity of  $I - (\lambda I - C)^{-1}B$ ; hence of  $\lambda I - C - B$ . Hence the set

$$[\lambda \mid \|(\lambda I - C)^{-1}B\| < 1]$$

is certainly an exclusion set, which is to say that if  $\lambda$  satisfies the inequality it cannot be a root of  $A$ .

Two special interpretations of this will be made, but this necessitates the introduction of an important special class of matrix norms. Given any convex body  $K$  let

$$\text{lub}_K(A) = \sup_{x \neq 0} \|Ax\|_K / \|x\|_K.$$

It is easy to see that this defines a matrix norm, and that the matrix norm is consistent with the vector norm. More than that, there exists a vector  $x \neq 0$  such that

$$(S) \quad \|Ax\|_K = \|x\|_K \text{lub}_K(A).$$

This particular matrix norm is said to be subordinate to the vector norm, and it is unique. Among all matrix norms that are consistent with a given vector norm, the subordinate norm is minimal, in the sense that

$$\text{lub}_K(A) \leq \|A\|$$

for every  $A$ . On the other hand, given two convex bodies  $H$  and  $K$ , if for every matrix  $ab^H$  of rank 1 it is true that

$$\text{lub}_H(ab^H) \leq \text{lub}_K(ab^H),$$

then  $K = \kappa H$  for some  $\kappa$ , and therefore

$$\text{lub}_H(A) = \text{lub}_K(A)$$

for every matrix  $A$  [28], [37].

Subordinate to the Euclidean (or spherical) norm is the so-called spectral norm, whose value is the largest singular value:

$$\text{lub}_S(A) = \rho^{1/2}(A^H A).$$

Subordinate to the  $E$  norm is that whose value is the largest sum of the moduli of the elements in any row. Subordinate to the  $E'$  norm is the largest sum of the moduli of the elements in any column.

To make the first interpretation, let

$$\|X\| = \text{lub}_E(X),$$

and let  $C=D$ , the diagonal of  $A$ . Evidently

$$\text{lub}_E(X) = \max [\nu \mid |X| e \leq \nu e].$$

Hence

$$\text{lub}_E[(\lambda I - D)^{-1}B] < 1$$

is equivalent to

$$|(\lambda I - D)^{-1}B| e < e,$$

or to

$$|B| e < |\lambda I - D| e.$$

This is to say that if  $|B|e = p$ , then for  $\lambda$  to be root of  $A$  it must satisfy at least one of the inequalities

$$|\lambda - \delta_i| \leq \pi_i.$$

This is the so-called Gershgorin theorem [13] published in 1931, although it had been published fifty years earlier by Levy [25] for the real case in a different but equivalent form and has been rediscovered many times since [40]. It states that all roots of a given matrix lie in the union of a set of  $n$  disks whose centers are the diagonal elements of the matrix, and whose radii are the sums of the moduli of the off diagonal elements in the rows.

For another interpretation [4], let

$$\|X\| = \text{lub}_S(X)$$

and let  $B$  and  $C$  be normal. Evidently

$$\text{lub}_S[(\lambda I - C)^{-1}B] < 1$$

is implied by

$$\text{lub}_S[(\lambda I - C)^{-1}] \text{lub}_S(B) < 1.$$

It is easily verified that if  $N$  is normal, then

$$\text{lub}_S(N) = \rho(N).$$

Hence if

$$\gamma_i = \lambda_i(C),$$

any root  $\lambda$  of  $A$  must satisfy at least one of the inequalities

$$|\lambda - \gamma_i| \leq \rho(B).$$

But if  $A = B + C$ , then also  $A = (B - \mu I) + (C + \mu I)$  where  $\mu$  is any scalar. Hence any root  $\lambda$  must satisfy at least one of the inequalities

$$|\lambda - \mu - \gamma_i| \leq \rho(B - \mu I).$$

The interpretation is as follows: With any center  $\mu$ , draw a circle enclosing all roots of  $B$ . From this form  $n$  circles of equal radius by translating by  $\gamma_i$  for each  $i$ . The union of all disks enclosed by these circles contains all roots of  $A$ . This theorem is due to Wielandt [41]. A corollary is the theorem of Bendixson which states that if  $B$  is Hermitian and  $C$  is skew Hermitian all roots of  $A$  are contained in the rectangle whose vertical sides pass through the extreme roots of  $B$  and whose horizontal sides pass through the extreme roots of  $C$ .

When the matrix  $A$  is Hermitian a natural application of Wielandt's theorem is to the case in which  $B$  and  $C$  are also Hermitian and  $B$  is small. Thus bounds are obtained on the variations in the roots that result from small deviations in the elements. A generalization can be stated for the case that  $B$  and  $C$  are normalizable [4] but not necessarily normal, but the resulting theorem is much less simple and involves quantities not ordinarily available in computational practice.

An ordinary norm is a mapping of  $n$  space on the ordered set consisting of all nonnegative real numbers. A vectorial norm [24], [34] is a mapping of  $n$  space on the partially ordered set consisting of all points in the positive orthant of some  $k$  space. The partial ordering is the obvious one. A particular example would be the one for  $k = n$ , where the vectorial norm of  $x$  is defined to be  $|x|$ . If  $p(x)$  represents a vectorial norm with elements  $\pi_i(x)$ , then each  $\pi_i(x)$  is at least a seminorm. For example, consider partitioning the entire space  $C_n$  into subspaces  $C_{n_1}, \dots, C_{n_m}$ , say. Any norm in  $C_n$  induces a certain norm in each subspace, and  $\pi_i(x)$  can be taken equal to the norm induced in  $C_{n_i}$ .

Let  $P_1, P_2, \dots, P_m$  be orthogonal projectors such that  $\sum P_i = I$ . Then any vector  $x$  can be resolved into mutually orthogonal components

$$x_i = P_i x, \quad x = \sum x_i, \quad x_i^H x_j = 0 \quad \text{for } i \neq j,$$



and any matrix  $A$  can be expressed as

$$A = \sum \sum P_i A P_j.$$

Let

$$A_{ij} = P_i A P_j,$$

and given any norm over the entire space  $C_n$ , define

$$\nu(A_{ij}) = \sup_{P_j x \neq 0} \|P_i A P_j x\| / \|P_j x\|.$$

Also let

$$\nu(A_{ii}) = \inf_{P_i x \neq 0} \|P_i A P_i x\| / \|P_i x\|.$$

These definitions are general, but for a fixed  $A$  let

$$\nu_{ij} = \nu(A_{ji}), \quad \nu_{ii} = \nu(A_{ii}),$$

and define the matrix

$$M(A) = \begin{pmatrix} \nu_{11} & -\nu_{12} & \cdots \\ -\nu_{21} & \nu_{22} & \cdots \\ \cdots & \cdots & \cdots \end{pmatrix}.$$

This is of the form  $D - B$ , where  $D$  is a nonnegative diagonal and  $B$  is nonnegative. It can also be written in the form  $\mu I - C$ , where  $C$  is nonnegative and  $\mu > 0$ . Such matrices have been much studied, and if, in particular,  $\mu > \rho(C)$ , then the matrix is called an  $M$  matrix (or, by some, a  $K$  matrix, after Koteljanski) and it is nonsingular. The theorem of interest here is that if  $M(A)$  is an  $M$  matrix, then  $A$  is nonsingular. When this result is applied to the matrix  $\lambda I - A$ , there is obtained a natural generalization of the Gershgorin theorem, and, indeed, it is true generalization. The theorem itself comes from taking the projections to be on the coordinate axes and choosing the norm as before [31], [32], [10], [11], [8].

It seems that vectorial norms were first introduced by Kantorovich [24] in 1939 in a different connection. More recently they have been studied in some depth by Robert [34], [35], [36].

The theorems just discussed are based on a partitioning of the matrix  $A$  in some way. This amounts to forming  $A = D + B$  where  $D$  is block diagonal. I will only mention some recent and still unpublished results of Joel Brenner [5] who forms  $D^{-1}B$ , or, equivalently,

$(\lambda I - D)^{-1}B$ . On applying the  $E$ -norm the result is a quite different generalization of the Gershgorin theorem.

So far only exclusion theorems have been considered. To pass on to inclusion theorems and the associated inclusion sets, it turns out that for an important class the complement of any inclusion set is also an inclusion set. Hence the common boundary separates the complex plane into two sets that are disjoint except for the boundary, each set being an inclusion set. Often the common boundary can be excluded, sometimes not. In any event the common boundary is called a separation locus, and the associated theorem a separation theorem.

The best known example of a separation theorem is that which states that for any Hermitian matrix, any Rayleigh quotient  $x^H A x / x^H x$  lies between the greatest and least of the roots. Hence it divides the real axis into two rays, each of which is an inclusion set. Moreover, the Rayleigh quotient itself can be removed from both rays unless it is one of the extreme roots. If  $A$  is normal the Rayleigh quotient lies in the convex hull of the roots, and this statement can be rephrased by saying that any line through a Rayleigh quotient is a separation locus. If the Rayleigh quotient is not itself a root, then the line itself can be removed from both half-planes. The corresponding theorem when  $A$  is normalizable but not normal is rather more complicated and involves a quotient that is slightly different. This will not be given here.

A quite general separation theorem will now be obtained for a normal matrix [3], [19], [20]. For a matrix that is only normalizable a slightly different norm is required but the proof is much the same [21]. Suppose

$$A = P\Lambda P^H, \quad P^H = P^{-1}, \\ \Lambda = \text{diag} (\lambda_1, \lambda_2, \dots, \lambda_n).$$

Let  $\alpha(\lambda)$  and  $\beta(\lambda)$  be any polynomials in  $\lambda$ , and

$$\gamma(\lambda) = \alpha(\lambda)/\beta(\lambda).$$

Then

$$\gamma(A) = P\gamma(\Lambda)P^H.$$

With the Euclidian norm, for any  $z$ ,

$$\|\gamma(A)z\| = \|P\gamma(\Lambda)P^H z\| = \|\gamma(\Lambda)y\|,$$

where  $y = P^H z$  and  $\|y\| = \|z\|$ . Hence

$$\|\gamma(A)z\| \geq \|y\| \min |\gamma(\lambda_i)| = \|z\| \min |\gamma(\lambda_i)|.$$

This is true for any  $z$ ; hence, in particular, it holds if

$$z = \beta(A)v,$$

with  $v$  arbitrary. But then

$$\gamma(A)z = \alpha(A)v;$$

hence, for any  $v$

$$\|\alpha(A)v\| \geq \|\beta(A)v\| \min |\gamma(\lambda_i)|.$$

Similarly,

$$\|\alpha(A)v\| \leq \|\beta(A)v\| \max |\gamma(\lambda_i)|.$$

The two inequalities together imply that the locus

$$|\alpha(\lambda)/\beta(\lambda)| = \|\alpha(A)v\|/\|\beta(A)v\|$$

is a separation locus provided the right-hand side of the equation is defined.

A mild amount of manipulation can be applied now to show from this that if, in particular,

$$[\beta(A)v]^H[\alpha(A)v] = 0$$

then for an arbitrary scalar  $\eta$ ,

$$\operatorname{Re} [\eta\alpha(\lambda)\overline{\beta(\lambda)}] = 0$$

defines a separation locus, and the Rayleigh quotient theorem comes from taking  $\beta$  to be a constant and  $\alpha(\lambda)$  linear. In fact, given any polynomial in  $\lambda$  and  $\bar{\lambda}$ :

$$\theta(\lambda, \bar{\lambda}) = \theta_0 + \theta_{01}\bar{\lambda} + \theta_{01}\lambda + \cdots,$$

this will be called a separation polynomial in case

$$v^H\theta(A, A^H)v = 0,$$

since for any such polynomial

$$\operatorname{Re} \eta\theta(\lambda, \bar{\lambda}) = 0$$

defines a separation locus, where again  $\eta$  is arbitrary.

Returning to the general separation theorem, if  $\beta(\lambda)$  is constant and  $\alpha(\lambda)$  linear, the locus is a circle, and it can be shown that the circle of smallest diameter is the one whose center is the Rayleigh quotient. If

$$v^Hv = 1, \quad v^HA v = \mu,$$

then there is at least one root in the circular disk

$$|\lambda - \mu| \leq \|p\|, \quad p = Av - \mu v.$$

Now let  $P$  be any matrix of  $r \leq n$  orthonormal columns, let  $H$  be any matrix of order  $r$ , and define  $AP - PH = R$ . Let

$$Hv = \eta v, \quad v^H v = 1.$$

Then

$$(A - \eta I)Pv = Rv.$$

Since  $P^H P = I$ , it follows that

$$\|Pv\| = \|v\|;$$

hence

$$\text{glb } (A - \eta I) \leq \text{lub } (R).$$

But  $\eta$  is any root of  $H$ . Hence the disk of radius  $\text{lub}(R)$  centered at any root of  $H$  is an inclusion set. This simple proof is due to Stewart [unpublished]. If the disks are all disjoint, then their union contains at least  $r$  roots of  $A$ . Unfortunately, the union of two overlapping disks could contain only a single root, which would then lie in their intersection. A rather more complicated proof, however, due to Kahan [23], shows that this cannot be so when  $A$  is Hermitian. In that case each root of  $H$  can be paired off with a root of  $A$  in such a way that the distance between pairs is at most  $\text{lub}(R)$ . The question remains open when  $A$  is normal but not Hermitian. For obvious reasons theorems such as this are called multiple inclusion theorems.

It is readily verified that  $\text{lub}(R)$  is minimized for given  $P$  by taking  $H = P^H A P$ . This generalizes in a natural way the theorem that for a given vector  $v$  the minimal inclusion disk has its center at the Rayleigh quotient.

Multiple inclusion theorems of a somewhat different sort can be obtained directly from the general separation theorem stated above. Two polynomials  $\alpha(\lambda)$  and  $\beta(\lambda)$  are said to be orthogonal with respect to the vector  $v$  in case  $[\alpha(A)v]^H [\beta(A)v] = 0$ . For any vector  $v \neq 0$  it is possible to define a set of mutually orthogonal polynomials  $\phi_i(\lambda)$ , unique up to a normalizing factor. If  $A$  is Hermitian it can be shown that the zeros are all real, that the zeros of consecutive  $\phi$ 's separate each other, and that the segment joining any two zeros of a given  $\phi$ , contains at least one root of  $A$  [6], [3], [20].

The general separation theorem for a normal matrix  $A$  can be extended by use of a different norm to a normalizable matrix  $M$  [21],

but application requires that  $M$  be expressed in the form  $B^{-1}A$ , where  $B$  is positive Hermitian and  $A$  is normal with respect to  $B$ . By this is meant that

$$A^H B^{-1} A = A B^{-1} A^H.$$

Such a decomposition is not easily made, but in important cases it is given in advance. That is to say, the problem often arises in the form  $Ax = \lambda Bx$ , where most commonly  $A$  is even Hermitian.

Many techniques have been used to derive inclusion theorems. For Hermitian matrices it is possible to base them on the Rayleigh-quotient theorem [19], but this has been seen to be a consequence of the general separation theorem. Again for Hermitian matrices, Lehmann [26] uses the minimax theorem to obtain multiple inclusion theorems. It turns out that at least some of those can also be derived from the general separation theorem [22].

Among the exclusion theorems, however, while many of them can be derived from the theory of norms, there are a number of others that cannot, or at least not easily. But about these I will let others speak.

#### REFERENCES

1. F. L. Bauer, *Theory of norms*, Computer Sci. Dept. Stanford Univ. Tech Report No. CS 75, 1967.
2. F. L. Bauer and C. T. Fike, *Norms and exclusion theorems*, Numer. Math. 2 (1960), 137-41.
3. F. L. Bauer and A. S. Householder, *Moments and characteristic roots*, Numer. Math. 2 (1960), 42-53.
4. ———, *Absolute norms and characteristic roots*, Numer. Math. 3 (1961), 241-246.
5. J. L. Brenner, *New root-location theorems for partitioned matrices*, 1967, Unpublished.
6. Hans Bückner, *Die praktische Behandlung von Integralgleichungen*, Springer-Verlag, Berlin, 1952.
7. V. N. Faddeeva, *Computational methods of linear algebra*, GITTL, Moscow 1950; English transl., Dover, New York, 1959.
8. D. G. Feingold and R. S. Varga, *Block diagonally dominant matrices and generalizations of the Gerschgorin circle theorem*, Pacific J. Math. 12 (1963), 1241-1250.
9. Miroslav Fiedler, *Some estimates of spectra of matrices*. Symposium on the Numerical Treatment of Ordinary Differential Equations, Integral and Integro-differential Equations (Rome, 1960), Birkhauser, Basel, 1960, pp. 33-36. (Russian)
10. Miroslav Fiedler and Vlastimil Pták, *Some inequalities for the spectrum of a matrix*, Mat.-Fys.. Časopis. Slovensk. Akad. Vied. 10 (1960), 148-166. (Russian)
11. ———, *Generalized norms of matrices and the location of the spectrum*, Czechoslovak Math. J. 12 (1962), 558-571. (Russian)
12. Noël Gastinel, *Matrices du second degré et normes générales en analyse numérique linéaire*, Publ. Sci. Tech. Ministère de l'Air Notes Tech. No. 110, Paris, 1962.

13. S. Gerschgorin, *Über die Abgrenzung der Eigenwerte einer Matrix*, Izv. Akad. Nauk SSSR. Otd. Mat. Estestv. Nauk, (1931), 749–754.
14. Harold Hotelling, *Some new methods in matrix calculation*, Ann. Math. Statistics **14** (1943) 1–34.
15. A. S. Householder, *On norms of vectors and matrices*, Rep. ORNL 1756, 1954.
16. ———, *On the convergence of matrix iterations*, J. Assoc. Comput. Mach. **3** (1956), 314–324.
17. ———, *The approximate solution of matrix problems*, J. Assoc. Comput. Mach. **5** (1958), 205–234.
18. ———, *Minimal matrix norms*, Monatsh. Math. **63**, (1959), 344–350.
19. ———, *Localization of the characteristic roots of matrices. Recent advances in matrix theory*, Univ. of Wisconsin Press, Madison, Wis., 1964, pp. 39–60.
20. ———, *The theory of matrices in numerical analysis*, Blaisdell, New York, 1964.
21. ———, *Separation theorems for normalizable matrices*, Numer. Math. **9** (1966) 46–50.
22. ———, *Moments and characteristic roots. II*, Numer. Math. **11** (1968), 126–128.
23. W. Kahan, *Inclusion theorems for clusters of eigenvalues of Hermitian matrices*, Computer Science, Univ. of Toronto Press, Toronto, 1967.
24. L. V. Kantorovitch, *The method of successive approximations for functional equations*, Acta Math. **71** (1939), 63–97.
25. A. Kolmogoroff, *Zur Normierbarkeit eines allgemeinen topologischen linearen Raumes*, Studia Math. **5** (1934), 29–33.
26. N. J. Lehmann, *Optimale Eigenwerteinschliessungen*, Numer. Math. **5** (1963), 246–272.
27. Lucien Levy, *Sur la possibilité de l'équilibre électrique*, C. R. Acad. Sci. Paris **93** (1881), 706–708.
28. Ju. I. Ljubič, *Operator norms of matrices*, Uspehi Mat. Nauk **18**, no. 4 (112) (1963), 161–164. (Russian)
29. A. T. Lonseth, *The propagation of error in linear problems*, Trans. Amer. Math. Soc. **62** (1947), 193–212.
30. A. M. Ostrowski, *A regularity condition for a class of partitioned matrices*, Univ. of Wisconsin MRC Technical Summary Report No. 132, Madison, Wis., 1960.
31. ———, *On some metrical properties of operator matrices and matrices partitioned into blocks*, Univ. of Wisconsin MRC Technical Summary Report No. 138, Madison, Wis., 1960.
32. ———, *On some metrical properties of operator matrices and matrices partitioned into blocks*, J. Math. Anal. Appl. **2** (1960), 161–209.
33. ———, *A regularity condition for a class of partitioned matrices*, Compositio Math. **15** (1961), 23–27.
34. F. Robert, *Normes vectorielles de vecteurs et de matrices*, Rev. Française Traitement Information (Chiffres) **7** (1964), 261–299.
35. ———, *Sur les normes vectorielles régulières sur un espace vectoriel de dimension fini*, C. R. Acad. Sci. Paris **261** (1965), 5173–5176.
36. ———, *Recherche d'une M-matrice parmi les minorantes d'un opérateur linéaire*, Numer. Math. **9** (1966), 189–199.
37. Hans Schneider and W. G. Strang, *Comparison theorems for supremum norms*, Numer. Math. **4** (1962), 15–20.

38. Josef Stoer, *A characterization of Hölder norms*, J. Soc. Indust. Appl. Math. 12 (1964), 634–648.
39. ———, *On the characterization of least upper bound norms in matrix space*, Numer. Math. 6 (1964), 302–314.
40. Olga Taussky, *A recurring theorem on determinants*, Amer. Math. Monthly 56 (1949), 672–676.
41. Helmut Wielandt, *On eigenvalues of sums of normal matrices*, Pacific J. Math. 5 (1955), 633–638.

OAK RIDGE NATIONAL LABORATORY AND  
UNIVERSITY OF TENNESSEE