SIAM J. NUMER. ANAL.
Vol. 15, No. 3, June 1978

0036-1429/78/1503-0013$1.00/0

NUMERICAL DETERMINATION OF THE FIELD OF VALUES OF A GENERAL COMPLEX MATRIX*

CHARLES R. JOHNSON†

Abstract. For an $n \times n$ complex matrix A, the convexity of $F(A) \equiv \{x^*Ax : x^*x = 1, x \in C^n\}$ and some simple observations are exploited to determine certain boundary points and tangents of $F(A)$. The result is a convergent computation scheme and an error measure for each approximation. Given the curvature of its boundary, the computational effort to determine $F(A)$ to a prespecified level of accuracy is $O(n^3)$.

1. Introduction. We consider the following question: how may the field of values (numerical range) of an arbitrary $n \times n$ complex matrix be numerically determined? The field of values arises in several theoretical as well as applicable contexts [1], [2], [5], [7], [8], [11], [15], [16], [17] and is, in many ways, as fundamental an object to associate with a matrix (operator) as the spectrum. A Gersgorin-type approach to estimation of the field of values has evolved in [4], [6], [10], and an exact representation of the field of values of a 2×2 matrix is given in [9]. However, no viable method for computing the field of values in the general case seems to have materialized. An approach is presented here using some known properties of the field of values and some additional observations regarding the boundary points.

Suppose $A = (a_{ij}) \in M_n(C)$, the $n \times n$ complex matrices; then the *field of values* of A is defined by

$$F(A) \equiv \{x^*Ax : x^*x = 1, x \in C^n\}. \tag{1}$$

Thus F may be thought of as a mapping from $M_n(C)$ into the subsets of the complex plane. We denote the spectrum, set of eigenvalues, of A by $\sigma(A)$. The following properties of F are well known (see e.g. [12], [13]).

(2) $F(A)$ is convex, closed and bounded.

(3) $\sigma(A) \subseteq F(A)$.

(4) For any unitary matrix U, $F(U^*AU) = F(A)$.

(5) $F(A + zI) = F(A) + z$ and $F(zA) = zF(A)$ for any complex number z.

(6) The boundary of $F(A)$, $\partial F(A)$, is a piecewise algebraic curve, and each point at which $\partial F(A)$ is not differentiable is an eigenvalue of A.

(7) If A is normal, then $F(A) = \mathrm{Co}(\sigma(A))$, where "Co" denotes the closed convex hull of a set.

(8) $F(A)$ is a segment of the real line if and only if A is Hermitian.

2. The theory. For an arbitrary element $B \in M_n(C)$, for convenience we define $H(B) \equiv \frac{1}{2}(B + B^*)$, the "Hermitian part" of B. Our approach rests primarily on (2) above and the following straightforward facts. Let $\lambda_M(B)$ denote the largest eigenvalue of B if $B \in M_n(C)$ is Hermitian.

* Received by the editors November 15, 1976.

† Institute for Physical Science and Technology, Department of Economics, University of Maryland, College Park, Maryland 20742 and Applied Mathematics Division, National Bureau of Standards, Washington, D.C. 20234.

OBSERVATIONS. *For any* $A \in M_n(C)$, *we have*

(9) $$e^{-i\theta}F(e^{i\theta}A) = F(A);$$

and

(10) $$\max_{z \in F(A)} \operatorname{Re}(z) = \max_{r \in F(H(A))} r = \lambda_M(H(A)).$$

Furthermore, the following three conditions on $x \in C^n$, *satisfying* $x^*x = 1$, *are equivalent*:

(10i) $$\operatorname{Re}(x^*Ax) = \max_{z \in F(A)} \operatorname{Re}(z);$$

(10ii) $$x^*H(A)x = \max_{r \in F(H(A))} r;$$

and

(10iii) $$H(A)x = \lambda_M(H(A))x.$$

Proof. Statement (9) follows from (5). The equivalence of (10i) and (10ii) follows from the calculation $\operatorname{Re}(x^*Ax) = \frac{1}{2}(x^*Ax + x^*A^*x) = x^*H(A)x$. If $\{x_1, \cdots, x_n\}$ is an orthonormal basis of eigenvectors of the Hermitian matrix $H(A)$, x_j associated with the eigenvalue λ_j, then x may be written $x = \sum_{n=1}^{n} \alpha_j x_j$ with $\sum_{j=1}^{n} \bar{\alpha}_j \alpha_j = 1$ (since $x^*x = 1$). Thus $x^*H(A)x = \sum_{j=1}^{n} \bar{\alpha}_j \alpha_j \lambda_j$, from which the equivalence of (10ii) and (10iii) is immediately deduced. The equivalence of (10i), (10ii) and (10iii), of course, implies statement (10). □

The above observations may be exploited to give computable representations of boundary points of and tangents to $F(A)$. (See Fig. 1.) Note that (10) implies that the line parallel to the imaginary axis at a distance $\lambda_M(H(A))$ is a tangent (not necessarily unique) to $F(A)$. (Note that this tangency does not necessarily occur on the real axis.) Statement (9) may be used via rotation to determine other tangents.

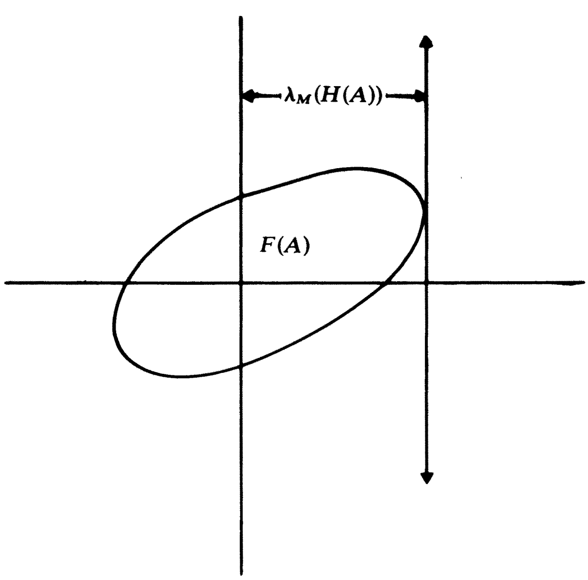

FIG. 1

For an angle θ, $0 \leqq \theta < 2\pi$, we define

$$\lambda_\theta \equiv \lambda_M(H(e^{i\theta}A)) \tag{11}$$

and let $x_\theta \in C^n$ be an arbitrary (there may be many) associated unit eigenvector:

$$H(e^{i\theta}A)x_\theta = \lambda_\theta x_\theta, \qquad x_\theta^* x_\theta = 1. \tag{12}$$

We denote by L_θ the line $\{e^{-i\theta}(\lambda_\theta + ti): t \in R\}$, and we denote by $H_\theta = e^{-i\theta}\{c : \mathrm{Re}\,(c) \leqq \lambda_\theta\}$, the half-plane determined by L_θ. Statement (9) together with the above comments reveal

THEOREM 1. *The line L_θ is a supporting hyperplane for the convex set $F(A)$, and $F(A) \subseteq H_\theta$ for all $0 \leqq \theta < 2\pi$.*

We may also deduce the point at which L_θ is tangent to $F(A)$.

THEOREM 2. *The complex number*

$$p_\theta \equiv x_\theta^* A x_\theta \tag{13}$$

is a boundary point of $F(A)$, and, furthermore, $p_\theta \in L_\theta \cap F(A)$. ($L_\theta$ is a tangent at p_θ.)

Given any complex number $z \notin F(A)$, we may choose θ so that L_θ is a separating hyperplane and $z \notin H_\theta$ (this utilizes the convexity of $F(A)$). This allows the following representations of the field of values.

THEOREM 3. *For $A \in M_n(C)$ we have* $\mathrm{Co}\,\{p_\theta : 0 \leqq \theta < 2\pi\} = F(A) = \bigcap_{0 \leqq \theta < 2\pi} H_\theta$.

We note that because, for example, of the possibility of "flat" portions of $\partial F(A)$ and the nonuniqueness of x_θ, the set $\{p_\theta : 0 \leqq \theta < 2\pi\}$ is not uniquely determined by A. Its closed convex hull, however, is. Along flat portions of $F(A)$, for example, any boundary point may be attained as p_θ by a particular choice of x_θ.

From Theorems 1 and 2 a discrete analogue of Theorem 3 follows with equalities replaced by set containments. Let θ now denote a set of "mesh" points $\theta = \{\theta_1, \cdots, \theta_k\}$ where $0 \leqq \theta_1 < \theta_2 < \cdots < \theta_k < 2\pi$. We define $F_{\mathrm{In}}(A, \theta) \equiv \mathrm{Co}\,\{p_{\theta_1}, \cdots, p_{\theta_k}\}$ and $F_{\mathrm{Out}}(A, \theta) = \bigcap_{1 \leqq j \leqq k} H_{\theta_j}$ and we may now give our main result which provides constructive inner and outer approximating sets for $F(A)$.

THEOREM 4. *For $A \in M_n(C)$ and θ as above, we have*

$$F_{\mathrm{In}}(A, \theta) \subseteq F(A) \subseteq F_{\mathrm{Out}}(A, \theta).$$

(See Fig. 2.)

The second containment in Theorem 4 is valid in any event, but, in case the θ_j are sufficiently numerous and well spaced that $\bigcap_{1 \leqq j \leqq k} H_{\theta_j}$ is bounded, then $F_{\mathrm{Out}}(A, \theta)$ is

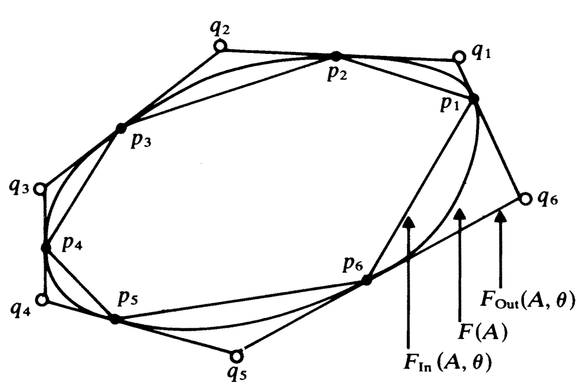

FIG. 2

more useful as an outer estimate. It is also simply determined in this event. In case they intersect and do not coincide, we denote by q_{θ_j} the intersection point of L_{θ_j} and $L_{\theta_{j+1}}$, where $j=1,\cdots,k$ and $k+1$ is identified with 1. If it is bounded, we then have the following simple alternate representation of $F_{\text{Out}}(A,\theta)$:

$$\bigcap_{1\leqq j\leqq k} H_{\theta_j}=\operatorname{Co}\{q_{\theta_1},\cdots,q_{\theta_k}\}. \tag{14}$$

The sets $F_{\text{In}}(A,\theta)$ and $F_{\text{Out}}(A,\theta)$ are then in a sense dual to each other and their boundaries are similarly determined. The boundary of $F_{\text{In}}(A,\theta)$ is simply the union of the k line segments $[p_{\theta_1},p_{\theta_2}],\cdots,[p_{\theta_{k-1}},p_{\theta_k}],[p_{\theta_k},p_{\theta_1}]$ while $\partial F_{\text{Out}}(A,\theta)$ is the union of the k line segments $[q_{\theta_1},q_{\theta_2}],\cdots,[q_{\theta_k},q_{\theta_1}]$. The points q_{θ_j} may be explicitly determined in a straightforward, though somewhat tedious, way. We abbreviate $\delta_j=\theta_{j+1}-\theta_j$, and, omitting the calculations, we find the result is

$$q_{\theta_j}=e^{-i\theta_j}\left(\lambda_{\theta_j}+\frac{\lambda_{\theta_j}\cos\delta_j-\lambda_{\theta_{j+1}}}{\sin\delta_j}i\right). \tag{15}$$

As a measure of the accuracy of the approximations in Theorem 4 we adopt a normalized difference between the areas of $F_{\text{Out}}(A,\theta)$ and $F_{\text{In}}(A,\theta)$. We denote the former area by $\text{Area}_{\text{Out}}(A,\theta)$ and the latter by $\text{Area}_{\text{In}}(A,\theta)$, and our error measure is

$$\Delta(A,\theta)\equiv(\text{Area}_{\text{Out}}(A,\theta)-\text{Area}_{\text{In}}(A,\theta))/\text{Area}_{\text{Out}}(A,\theta) \tag{16}$$

so that the approximation is good when Δ is near 0. Each of $F_{\text{In}}(A,\theta)$ and $F_{\text{Out}}(A,\theta)$ is a polygon determined by its k extreme points. A compact formula for the area of such a polygon may be derived in a straightforward way by division into triangles. If P is a k-sided polygon in the complex plane whose vertices are $c_1,\cdots,c_k$, given in counterclockwise order, then the area of P may be written as

$$\text{Area}\,(P)=\frac{1}{2}\operatorname{Im}\,(\bar{c}_1c_2+\bar{c}_2c_3+\cdots+\bar{c}_{k-1}c_k+\bar{c}_kc_1), \tag{17}$$

where "Im" denotes the imaginary part of a complex number.

3. An algorithm. The above ideas may be used to specify an algorithm for the determination of $F(A)$. We focus on $F_{\text{In}}(A,\theta)$ as our approximation to $F(A)$. Suppose that $A\in M_n(C)$ and a prespecified level of tolerance $\Delta>0$ are given.

I. Check to see if A is Hermitian. If so, compute the largest and smallest eigenvalues of A and the line segment joining them is $F(A)$. If not, proceed to the next step.

II. Check A for normality ($A^*A=AA^*$). If A is normal, compute all eigenvalues of A; their convex hull is $F(A)$. If A is not normal continue.

III. Choose a k-mesh $\theta_j=(j-1)(2\pi/k)$, $j=1,2,\cdots,k$, with $k\geqq 3$. For each θ_j, construct $H(e^{i\theta_j}A)$ and compute λ_{θ_j} and x_{θ_j} satisfying (11) and (12). For each j this is a well-defined Hermitian eigenvalue-eigenvector problem.

IV. Compute p_{θ_j} via (13) and q_{θ_j} via the formula (15) for each $j=1,2,\cdots,k$.

V. Using, for example, the formula (17), compute $\Delta(A,\{\theta_1,\cdots,\theta_k\})$ defined by (16). If $\Delta(A,\{\theta_1,\cdots,\theta_k\})\leqq\Delta$, continue to the next step; and, if not, replace k by $2k$ and return to step III. In that step only k new computations need be made.

VI. Plot $F_k(A)\equiv F_{\text{In}}(A,\{\theta_1,\cdots,\theta_k\})$ as an approximation to $F(A)$. Note that $\partial F_k(A)$ is the union of the line segments joining p_{θ_j} to $p_{\theta_{j+1}}$, $j=1,\cdots,k$, so that $F_k(A)$ is easily plotted.

Remark 1. The above method considers only $\lambda_M(H(e^{i\theta}A))$. An alternate approach could be taken in which two diametrically opposite boundary points of $F(A)$ are determined from each $e^{i\theta}A$ by considering both the largest and smallest eigenvalues (and associated eigenvectors) of $H(e^{i\theta}A)$. This is the case since, for any $B \in M_n(C)$, the smallest eigenvalue of $H(B)$ is just $-\lambda_M(H(-B))$. Thus sufficient knowledge about the smallest eigenvalue of $H(e^{i\theta}A)$ renders unnecessary the consideration of $\lambda_{\theta+\pi}$. This suggests an alternate algorithm in which a mesh of θ's only over the interval $[0, \pi)$ is used, but two eigenvalue-eigenvector determinations are made for each.

Remark 2. In producing $F_k(A)$ by the given algorithm, a separate eigenvalue-eigenvector computation is required for each θ_j. It is not explored here, but information gained from the solution to the eigenproblem for θ_j should assist in the analysis of the eigenproblem associated with θ_{j+1}. Since λ_θ and x_θ are continuous functions of θ, this should be especially true when δ_j is "small".

Remark 3. The numerical radius of $A \in M_n(C)$ is defined by

$$r(A) \equiv \max_{z \in F(A)} |z|. \tag{18}$$

the greatest distance from the origin achieved on the compact set $F(A)$. Since it is clear that the maximum is achieved on $\partial F(A)$ (in fact, at an extreme point), it follows that

$$r(A) = \max_{0 \leqq \theta \leqq 2\pi} |\lambda_\theta|. \tag{19}$$

(If $0 \in F(A)$) the absolute value bars may be omitted.) Numerically $r(A)$ could be determined by essentially the same algorithm as above (with comparison of the absolute values of the finite number of boundary points). However, since only one element of $\partial F(A)$ is desired, a modification of the strategy might be more economical. First, a comparatively coarse mesh might be chosen in an attempt to localize a smaller region for the optimizing θ. On this smaller region a finer mesh could be used to approximate that θ which generates $r(A)$. Of course the q_{θ_j}'s still may play a role by helping determine which regions to omit from consideration.

4. Convergence and complexity. Note that $F_k(A) \subseteq F(A)$ for all k, but $F_k(A) \to F(A)$ as $k \to \infty$ or as $\Delta \to 0$. It would, however, be useful to understand the rate of improvement in the approximation and to have an idea of the number of mesh points necessary to assure a given level of accuracy. It should be noted that the accuracy depends rather intimately on the underlying shape (primarily the uniformity of the curvature of its boundary) of $F(A)$ and only on the shape. The accuracy does not depend on n, the dimension of A, except, perhaps through the underlying eigenvalue-eigenvector computations. To the end of understanding the accuracy as a function of k, the number of mesh points, we study two cases, in a sense a "best" case and a case which cannot actually occur but which illustrates the nature of the error when the boundary points of the inner estimate are especially poorly placed relative to those of the outer estimate.

First, suppose that

$$F_{\text{Out}}\left(A, \left\{0, \frac{2\pi}{k}, \frac{4\pi}{k}, \cdots, (k-1)\frac{2\pi}{k}\right\}\right)$$

is a regular polygon for all k and that $F_k(A)$, the inner approximation, is the inscribed polygon whose vertices are the midpoints of the outer approximation's edges. This is equivalent to assuming that $F(A)$ is a circular disc. In this event a straightforward application of geometry yields that the ratio of the inner and outer areas is $\sin^2(90(k-2)/k)$, measuring angles in degrees. Thus

$$\Delta(A, \theta)=\cos^2\left(90\frac{k-2}{k}\right)=\sin^2\frac{180}{k}, \tag{20}$$

as a function of k. Therefore $\Delta(A, \theta)$ approaches 0 like $1/k^2$ and the convergence is quadratic in the sense that doubling the number of mesh points divides the error by four. For example, a 5 degree mesh (72 mesh points) gives an error of at most .0019 while a 10 degree mesh (36 mesh points) has an error of at most .0076, les than 1%.

Alternatively, suppose that the outer estimate is still a regular polygon for all k but that the inner estimate is the worst it can be, the limiting case of a polygon whose vertices are every other vertex of the outer estimate. For convenience, we suppose $k=2l$ so that the inner polygon has l vertices and the outer has $2l$. The ratio of the inner to the outer area then turns out to be

$$2\tan\left(90\frac{l-2}{l}\right)\sin\left(90\frac{l-1}{l}\right)\cos 90\left(\frac{l-1}{l}\right),$$

which may be simplified to

$$\cos\frac{180}{l} \quad \text{or} \quad \cos\frac{360}{k}.$$

Therefore,

$$\Delta(A, \theta)=1-\cos\frac{360}{k}, \tag{21}$$

which still approaches 0 like $1/k^2$. The convergence remains quadratic in the sense mentioned previously, but, of course, the level of error measure for a given number of mesh points is higher. For example, a 5 degree mesh yields an error of at most .0038 while a 10 degree mesh has an error of at most .0152. These are twice as much as in the previous case, and, in fact, for k large this error function is twice the previous one.

In the event $\partial F(A)$ has "flat" portions, the extreme case is a polygonal $F(A)$. In this case A may be normal (which is picked up in step II), or, if not, two facts should be noted. First of all, $F(A)=\mathrm{Co}\,(\sigma(A))$ and each of the eigenvalues which is a vertex of $F(A)$ has an eigenspace which is orthogonal to the eigenspaces of all other eigenvalues [8]. Secondly, $F_k(A)$ will converge to $F(A)$ in a finite number of steps although the number is not bounded independent of A. In any event, since it is contained in $F(A)$ and not necessarily in $F_k(A)$ for "small" k, $\mathrm{Co}\,(\sigma(A))$ might be used where feasible to assist the given algorithm in determining $F(A)$. The case in which the boundary of $F(A)$ has "flat" portions but is not polygonal has not been considered here and may lead to slow convergence. It should also be noted that coincidence of boundary points of approximations to $F(A)$ may occur only when $F(A)$ has vertices.

Since the amount of effort necessary to compute $F(A)$ by the above algorithm to a desired degree of accuracy depends only on the shape and not the dimension n, the amount of effort is, for a given fifficulty in shape, on the same order as that of solving the Hermitian eigenvalue-eigenvector problem necessary to obtain a boundary point. We thus conclude the necessary effort is $O(n^3)$, where, of course, the constant is a sizeable multiple (determined by the number of mesh points necessary to obtain the desired accuracy for this shape of $F(A)$) of the constant for a single eigenvalue problem. Keep in mind that our error measure (16) does not take into account inaccuracy in the eigenvalue-eigenvector computations but that it is generous in that the true boundary must lie strictly in between the inner and outer estimates.

5. An example. Shown below in Fig. 3 is the field of values of the doubly stochastic matrix

$$A=\begin{pmatrix} .3 & .4 & .3 \\ 0 & .5. & 5 \\ .7 & .1 & .2 \end{pmatrix},$$

which was determined by the algorithm in section 3 with a 5° mesh. This led to a conjecture regarding an inclusion region for the field of values of a doubly stochastic matrix which has since been proved in [12a].

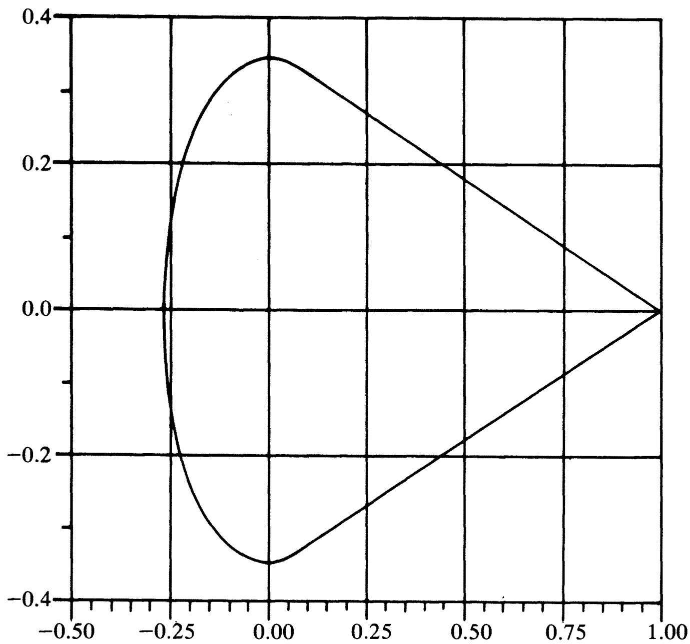

FIG. 3

REFERENCES

[1] K. ARROW AND M. MCMANUS, *A note on dynamic stability*, Econometrica, 26 (1958), pp. 448–454.

[2] C. S. BALLANTINE AND C. R. JOHNSON, *Accretive Matrix Products*, Linear and Multilinear Algebra, 3 (1975), pp. 169–185.

[3] C. S. BALLANTINE, *Numerical range of a matrix: Some effective criteria*, to appear.

[4] C. R. JOHNSON, *Gersgorin sets and the field of values*, J. Math. Anal. Appl., 45 (1974), pp. 416–419.

[5] ———, *A Lyapunov theorem for angular cones*, J. Res. Nat. Bur. Standards Sect. B, 78B (1974), pp. 7–10.

[6] ———, *A Gersgorin inclusion set for the field of values of a finite matrix*, Proc. Amer. Math. Soc., 41 (1973), pp. 57–60.

[7] ———, *D-stability and real and complex quadratic forms*, Linear Algebra and Appl., 9 (1974), pp. 89–94.

[8] ———, *Normality and the numerical range*, Ibid., 15 (1976), pp. 89–94.

[9] ———, *Computation of the field of values of a 2 by 2 matrix*, J. Res. Nat. Bur. Standards Sect. B, 78B (1974), pp. 105–107.

[10] ———, *Numerical location of the field of values*, Linear and Multilinear Algebra, 3 (1975), pp. 9–14.

[11] ———, *The field of values and spectra of positive definite multiples*, J. Res. Nat. Bur. Standards Sect. B, 78B (1974), pp. 197–198.

[12] ———, *Functional characterizations of the field of values and the convex hull of the spectrum*, Proc. Amer. Math. Soc., 61 (1977), pp. 201–204.

[12a] ———, *An inclusion region for the field of values of a doubly stochastic matrix based on its graph*, Aequationes Math., to appear.

[13] R. KIPPENHAHN, *Über den Wertevorrat einer Matrix*, Math. Nachr., 6 (1951), pp. 193–228.

[14] F. MURNAGHAN, *On the field of values of a square matrix*, Proc. Nat. Acad. Sci. U.S.A., 18 (1932), pp. 246–248.

[15] O. TAUSSKY, *A remark concerning the similarity of a finite matrix A and A^**, Math. Z., 117 (1970), pp. 189–190.

[16] ———, *The role of symmetric matrices in the study of general matrices*, Linear Algebra and Appl., to appear.

[17] H. WIELANDT, *On the Eigenvalues of $A+B$ and AB*, J. Res. Nat. Bur. Standards Sect. B, 77B (1973), pp. 61–63.