Circular Discs Containing Eigenvalues of Normal Matrices

Earl R. Barnes
Department of Mathematical Sciences
IBM Research Division
T. J. Watson Research Center
P.O. Box 218
Yorktown Heights, New York 10598

Dedicated to Alan J. Hoffman with great admiration on the occasion of his 65th birthday.

Submitted by Richard A. Brualdi

ABSTRACT

We give an inclusion theorem for the eigenvalues of a normal matrix using discs similar to the Gerschgorin discs. This theorem is used to obtain bounds on the eigenvalues of Jordan products, and a variant of Kantorovich's inequality for normal matrices.

1. INTRODUCTION

Let $A=(a_{ij})$ and $B=(b_{ij})$ be $n\times n$ normal matrices having eigenvalues $\lambda_1,\dots,\lambda_n$ and $\mu_1,\dots,\mu_n$ respectively. The Hoffman-Wielandt inequality [1] asserts the existence of a labeling of the eigenvalues such that

$$\sum_{i=1}^{n}|\lambda_i-\mu_i|^2\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}-b_{ij}|^2.$$

Shortly after I arrived at IBM in 1968 Alan Hoffman gave a very inspiring lecture there showing how this inequality can be used to obtain bounds for a graph partitioning problem that arises in laying out circuits on computer chips. He started by stating that he had discovered this inequality, with

655 Avenue of the Americas, New York, NY 10010 0024-3795/89/$3.50

Helmut Wielandt, in 1952, but had no application for it at the time. Now after many years he was very happy to have found an application for the inequality. I felt happy for the opportunity to learn about these new problems and happy to be arriving at a place where one is given a long time to find some use for his work. In the years that followed Hoffman and I coauthored several papers on subjects related to graph partitioning and the eigenvalues of real symmetric matrices. The present paper is an extension of our work in [2].

Let $A=(a_{ij})$ be an $n\times n$ normal matrix. Denote the eigenvalues of A by $\lambda_1,\dots,\lambda_n$. According to the Hoffman-Wielandt inequality applied to the matrices A and $B=\operatorname{diag}(a_{11},\dots,a_{nn})$ there is a labeling of the eigenvalues such that

$$\sum_{i=1}^{n}|\lambda_i-a_{ii}|^2\leqslant\sum_{i=1}^{n}\sum_{j\neq i}|a_{ij}|^2. \tag{1.1}$$

This inequality resembles Schur's inequality $\sum_{i=1}^{n}|\lambda_i|^2\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2$, which holds for any complex matrix A. However (1.1) need not hold for nonnormal matrices. For example, consider

$$A=\begin{pmatrix} 3 & -1 & -1 & -1\\ 1 & -3 & 1 & 1\\ -i & -i & 3i & -i\\ i & i & i & -3i\end{pmatrix}.$$

This matrix is not normal and satisfies $A^4=0$. It therefore has each of its eigenvalues equal to 0. For this matrix the inequality (1.1) reads $36\leqslant 12$, which is false.

Equation (1.1) says that the average of the numbers $|\lambda_i-a_{ii}|^2$ is less than or equal to the average of the numbers $\sum_{j\neq i}|a_{ij}|^2$. The work in this paper began with an attempt to find a better localization result for the eigenvalues of A in terms of the numbers a_{ii}, $\sum_{j\neq i}|a_{ij}|^2$, $i=1,\dots,n$.

Some initial results were obtained in [2]. The main result of this paper is a generalization to normal matrices of a theorem proved in [2] for real symmetric matrices. We also obtain bounds on the eigenvalues of the Jordan product of two normal matrices in terms of bounds on the eigenvalues of the individual matrices, and a generalization of Kantorovich's inequality for positive definite matrices to normal matrices.

Most of our results follow from simple geometric properties of circles. Consequently it is convenient to describe them in terms of circles. We begin by stating an elementary theorem about circles which will be useful in many of our proofs.

THEOREM 1.1. *If two lines through a point P meet a circle at points Q, Q' (possibly coincident) and R, R' (possibly coincident), respectively, then $PQ \times PQ' = PR \times PR'$. Note that in the case where P lies outside the circle and $Q = Q'$, the tangent PQ satisfies*

$$(PQ)^2 = PR \times PR'.$$

2. AN INCLUSION THEOREM FOR EIGENVALUES OF A

The main theorem of this paper is a generalization to normal matrices of the following theorem which Hoffman and I proved in [2] for real symmetric matrices.

THEOREM 2.1. *Let $A = (a_{ij})$ be a real symmetric matrix of order n. For a fixed index i let α and β be positive numbers satisfying*

$$\alpha\beta \geqslant \sum_{j \neq i} |a_{ij}|^2. \tag{2.1}$$

Then the interval $[a_{ii} - \alpha,\ a_{ii} + \beta]$ contains at least one eigenvalue of A.

Shortly after we published this result, we learned of the following more general result due to de Bruijn [3].

THEOREM 2.2. *Let A be an $n \times n$ Hermitian matrix with eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ on the real line. Through each pair of neighboring eigenvalues draw a semicircle as in Figure 1. Denote by $\mathscr{R}$ the region bounded by these semicircles and the semicircle through λ_1 and λ_n. Let x be an arbitrary unit*

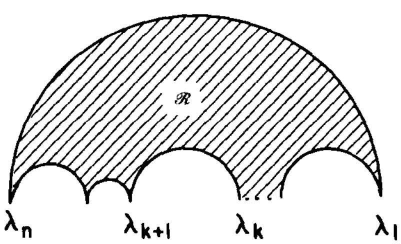

FIG. 1.

vector, and let

$$\mu = x^*Ax, \qquad \sigma = \|Ax - \mu x\|. \tag{2.2}$$

Then the point (μ, σ) lies in $\mathscr{R}$.

To see that Theorem 2.1 follows from Theorem 2.2 take x to be the ith unit coordinate vector. Then

$$\mu = a_{ii} \quad \text{and} \quad \sigma = \sqrt{\sum_{j \neq i} |a_{ij}|^2}\,.$$

Choose k such that $\lambda_{k+1} \leqslant a_{ii} \leqslant \lambda_k$. This is always possible, since the diagonal terms of A lie in the convex hull of the eigenvalues of A. Let h denote the vertical distance from a_{ii} to the set $\mathscr{R}$. Since $(a_{ii}, \sqrt{\sum_{j \neq i}|a_{ij}|^2}\,) \in \mathscr{R}$, we have $h \leqslant \sqrt{\sum_{j \neq i}|a_{ij}|^2}$. It follows from Theorem 1.1 that

$$(a_{ii} - \lambda_{k+1})(\lambda_k - a_{ii}) = h^2 \leqslant \sum_{j \neq i} |a_{ij}|^2.$$

Thus if (2.1) holds, either $\alpha \geqslant a_{ii} - \lambda_{k+1}$ or $\beta \geqslant \lambda_k - a_{ii}$, so that the interval $[a_{ii} - \alpha, a_{ii} + \beta]$ contains at least one of the eigenvalues λ_k, λ_{k+1}.

We will generalize Theorems 2.1 and 2.2 to normal matrices. First we establish some conventions. In general the entries and eigenvalues of a normal matrix are complex numbers. So in what follows, the entries in all matrices and vectors encountered will be treated as complex numbers. We will identify a complex number $\lambda = \xi + i\eta$ with the point (ξ, η) in the Cartesian plane. The inner product of two points λ and μ will be denoted by $\lambda \cdot \mu$.

THEOREM 2.3. *Let $A = (a_{ij})$ be an $n \times n$ normal matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ in the complex plane. Let x be a unit vector, and set $\mu = x^*Ax$. Let α and β be any positive numbers satisfying*

$$\alpha\beta = \|Ax - \mu x\|^2,$$

and let l be a line segment through μ, drawn such that μ divides l into parts

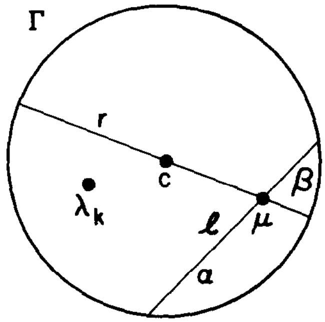

FIG. 2.

of length α and β respectively. Then any disc having l as a chord contains an eigenvalue of A.

See Figure 2.

Proof. Let Γ denote a circle drawn through the endpoints of l. Denote the center and radius of Γ by c and r respectively. Consider a diameter of Γ drawn through c and μ. It follows from Theorem 1.1 that

$$(r-|c-\mu|)(r+|c-\mu|)=\alpha\beta=\|Ax-\mu x\|^2.$$

This implies that

$$r^2=\|Ax-\mu x\|^2+|\mu-c|^2.$$

To complete the proof of the theorem we must show that A has an eigenvalue λ satisfying

$$|\lambda-c|^2\leqslant\|Ax-\mu x\|^2+|\mu-c|^2.$$

Let $\lambda_1,\dots,\lambda_n$ denote the eigenvalues of A, and let $u_1,\dots,u_n$ be a corresponding set of orthonormal eigenvectors. We then have $A=\sum_{i=i}^n\lambda_i u_i u_i^*$, so that

$$\mu=x^*Ax=\sum_{i=1}^n\lambda_i|u_i^*x|^2. \tag{2.3}$$

Let $p_i=|u_i^*x|^2$, $i=1,\dots,n$. Then each $p_i\geqslant 0$ and $\sum_{i=1}^n p_i=\|x\|^2=1$.

We have

$$\sum_{i=1}^{n} p_i|\lambda_i - c|^2 = \sum_{i=1} p_i|\lambda_i - \mu + \mu - c|^2$$

$$= \sum_{i=1}^{n} p_i|\lambda_i - \mu|^2 + |\mu - c|^2$$

$$= \sum_{i=1}^{n} p_i\left\{|\lambda_i|^2 - \lambda_i\bar{\mu} - \bar{\lambda}_i\mu + |\mu|^2\right\} + |\mu - c|^2$$

$$= x^*A^*Ax - |\mu|^2 + |\mu - c|^2$$

$$= \|Ax - \mu x\|^2 + |\mu - c|^2 = r^2.$$

Let λ_k denote the eigenvalue of A nearest c. Then

$$|\lambda_k - c|^2 \leqslant \sum_{i=1}^{n} p_i|\lambda_i - c|^2 = r^2.$$

This completes the proof of the theorem. ■

COROLLARY. *The disc described in Theorem 2.3 has at least one eigenvalue of A in its exterior, or on its boundary.*

Proof. $\max_j |\lambda_j - c|^2 \geqslant \sum_{i=1}^{n} p_i|\lambda_i - c|^2 = r^2.$ ■

Suppose we take x to be the ith unit coordinate vector and apply Theorem 2.3 to a real symmetric matrix. In this case the disc bounded by Γ intersects the real axis in a line segment l' which is divided by μ into parts of lengths α' and β'. By Theorem 1.1 we have $\alpha'\beta' = \alpha\beta = \sum_{j \neq i}|a_{ij}|^2$. Thus in this case we can infer that the disc bounded by Γ contains an eigenvalue of A from Theorem 2.1. Since Theorem 2.3 clearly implies Theorem 2.1, we see that in this case the two theorems are equivalent.

Theorem 2.2 can also be inferred from Theorem 2.3. To see this let μ and σ be given by (2.2). Assume that $(\mu, \sigma) \notin \mathscr{R}$. To be specific, assume that (μ, σ) lies below $\mathscr{R}$. Select consecutive eigenvalues λ_k, λ_{k+1} of A such that $\lambda_{k+1} \leqslant \mu \leqslant \lambda_k$. Draw a circle through the points (μ, σ) and $(\mu, -\sigma)$ having center $\frac{1}{2}(\lambda_k + \lambda_{k+1})$. Clearly this circle encloses no eigenvalue of A. Since

$\sigma^2 = \|Ax - \mu x\|^2$, this contradicts the conclusion of Theorem 2.3. Therefore (μ, σ) does not lie below $\mathscr{R}$.

Assume that (μ, σ) lies above $\mathscr{R}$. In this case we can draw a circle through (μ, σ) and $(\mu, -\sigma)$ having center $\frac{1}{2}(\lambda_1 + \lambda_n)$. This circle encloses all eigenvalues of A. But this contradicts the corollary to Theorem 2.3, so (μ, σ) cannot lie above $\mathscr{R}$. It follows that $(\mu, \sigma) \in \mathscr{R}$ as claimed.

In order to obtain the analog of Theorem 2.2 for normal matrices we need some properties of the Voronoi diagram of the eigenvalues of A. Voronoi diagrams are discussed in [4, Chapter 5]. Here we will describe just the properties of these diagrams needed for our generalization of Theorem 2.2.

When the eigenvalues of A are collinear, Theorem 2.2 remains valid as stated as long as the semicircles are drawn normal to the complex plane. So in what follows, we assume the eigenvalues of A are not collinear.

The Voronoi diagram of the eigenvalues $\lambda_1, \dots, \lambda_n$ of A is a partitioning of the Cartesian plane into regions $V(i)$, $i = 1, \dots, n$, such that $\lambda_i \in V(i)$ and $V(i)$ contains all points in the plane that are closer to λ_i than to any eigenvalue $\lambda_j \neq \lambda_i$. Mathematically $V(i)$ is defined as follows: For $\lambda_j \neq \lambda_i$ the points closer to λ_i than to λ_j lie on the λ_i side of the half plane determined by the perpendicular bisector of λ_i and λ_j. We denote this half plane by $H(\lambda_i, \lambda_j)$. Then

$$V(i) = \bigcap_{j \neq i} H(\lambda_i, \lambda_j).$$

$V(i)$ is called the Voronoi polygon associated with λ_i. The Voronoi diagram for the five λ's in Figure 3 is represented by the broken-line graph. Consider the straight-line dual of the Voronoi diagram. This is the graph on the points

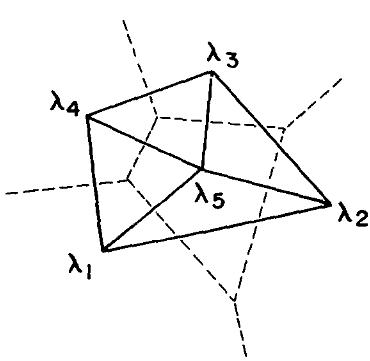

FIG. 3.

$\lambda_1,\ldots,\lambda_n$ obtained by connecting two points λ_i and λ_j by a straight line if $V(i)$ and $V(j)$ have a common edge in the Voronoi diagram. In Figure 3 the dual graph is represented by solid lines. The dual graph defines a triangularization of the convex hull of the points $\lambda_1,\ldots,\lambda_n$. The triangles are called Delaunay triangles. They are unique and have the interesting property that their circumscribing circles contain none of the points $\lambda_1,\ldots,\lambda_n$ in their interiors. It is this property that allows us to prove the following generalization of Theorem 2.2.

THEOREM 2.4. *Let A be an $n\times n$ normal matrix with noncollinear eigenvalues $\lambda_1,\ldots,\lambda_n$ in the complex plane. Let $T_1,\ldots,T_{n-1}$ denote the Delaunay triangles defined by these eigenvalues. Let S_i denote the solid sphere of least volume containing T_i, $i=1,\ldots,n-1$, and let S denote the solid sphere of least volume containing all eigenvalues of A. Note that each of the spheres $S_1,\ldots,S_{n-1}$, S has its center in the complex plane, and no S_i has an eigenvalue of A in its interior. Let $\mathcal{R}$ consist of the points above the complex plane which are in S but not in the interior of any of the spheres S_i, $i=1,\ldots,n-1$. For any unit vector x, the point (μ,σ), defined by*

$$\mu = x^*Ax, \qquad \sigma = \|Ax-\mu x\|$$

lies in $\mathcal{R}$.

An example of $\mathcal{R}$ for four points is shown in Figure 4. When the eigenvalues of A are collinear, the theorem remains valid with $\mathcal{R}$ replaced by a region similar to Figure 1. This result can be easily obtained as the limiting case of the situation when the eigenvalues are noncollinear.

Proof of Theorem 2.4. It follows from (2.3) that μ lies in the convex hull of the eigenvalues of A. Therefore μ lies in one of the Delaunay triangles.

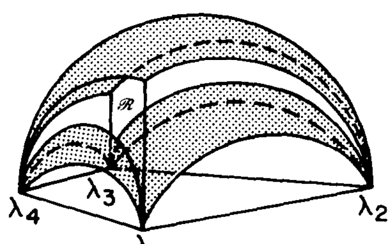

FIG. 4.

Suppose $\mu \in T_k$. To show that $(\mu, \sigma) \in \mathscr{R}$ we will first show that (μ, σ) is not in the interior of S_k. The proof is by contradiction. Assume that (μ, σ) is in the interior of S_k. We can then construct a solid sphere $S_k' \subset S_k$, concentric with S_k and having (μ, σ) on its surface. Consider the disc Γ where S_k' intersects the complex plane. Since (μ, σ) is on the surface of S_k', Γ has a chord of length 2σ, centered at μ. It follows from Theorem 2.3, by taking $\alpha = \beta = \sigma$, that Γ contains an eigenvalue of A. But Γ is contained in the interior of S_k, and S_k has no eigenvalues of A in its interior. Thus we have a contradiction. It follows that (μ, σ) is not in the interior of S_k.

To complete the proof we must show that (μ, σ) is inside or on the boundary of S. Assume that (μ, σ) is in the exterior of S. We can then construct a sphere $S' \supset S$, concentric with S, having (μ, σ) on its surface. The intersection of S' with the complex plane is a disc enclosing all eigenvalues of A. But this disc has a chord of length 2σ centered at μ. This contradicts the corollary to Theorem 2.3. We must therefore have (μ, σ) inside or on the boundary of S. ∎

3. APPLICATIONS

3.1. Bounds for the Eigenvalues of Jordan Products

Our first application is an extension of de Bruijn's bounds [3] for the eigenvalues of the Jordan product of two Hermitian matrices to normal matrices.

Let A and B be $n \times n$ normal matrices. We wish to find bounds on $\operatorname{Re} x^*A^*Bx$ as x ranges over the set of n-dimensional unit complex vectors. It is assumed that the smallest discs containing the eigenvalues of A and B are known. The reason for interest in these bounds is that

$$\operatorname{Re} x^*A^*Bx = \tfrac{1}{2}x^*(A^*B + B^*A)x. \tag{3.1}$$

Clearly $A^*B + B^*A$ is Hermitian. Thus as x varies over the set of unit vectors, $2\operatorname{Re} x^*A^*Bx$ varies between the smallest and largest eigenvalues of $A^*B + B^*A$. This matrix is the so-called Jordan product of A and B. The problem of determining bounds on the eigenvalues of $A^*B + B^*A$ in terms of bounds in the eigenvalues of A and B was first posed by Olga Taussky Todd in [5]. In the case of Hermitian matrices, bounds have been obtained in [3], [6], and [7].

Let x be a unit vector, and define

$$\mu = x^*Ax, \qquad \nu = x^*Bx.$$

We use the notation $\mu = \alpha + i\beta$, $\nu = \xi + i\eta$ to express μ and ν as complex numbers. Define

$$\sigma = \|Ax - \mu x\| \qquad \text{and} \qquad \delta = \|Bx - \nu x\|.$$

We can then write

$$Ax = \mu x + \sigma u, \qquad Bx = \nu x + \delta w$$

where u and w are unit vectors and

$$x^* u = x^* w = 0.$$

We have

$$x^* A^* Bx = \bar{\mu}\nu + \sigma\delta u^* w.$$

Let $\operatorname{Re} u^* w = \varepsilon$. Then $|\varepsilon| \leqslant 1$ and

$$\operatorname{Re} x^* A^* Bx = \alpha\xi + \beta\eta + \sigma\delta\varepsilon. \tag{3.2}$$

THEOREM 3.1. *Let S and R denote the solid spheres of least volume containing the eigenvalues of A and B respectively. For any unit vector x we can choose points $(\alpha', \beta', \sigma')$ and (ξ', η', δ') on the surfaces of S and R, respectively, such that*

$$\operatorname{Re} x^* A^* Bx = \alpha'\xi' + \beta'\eta' + \sigma'\delta'. \tag{3.3}$$

Proof. The proof is similar to the proof of Theorem 3.2 in [3]. We begin with Equation (3.2). For the moment assume that $\varepsilon \geqslant 0$. Let Γ_S and Γ_R denote the intersections of the solid spheres S and R with the complex plane. The eigenvalues of A and B lie in the discs Γ_S and Γ_R, respectively. For any point (α, β) in Γ_S let $\varphi(\alpha, \beta)$ be defined so that $\varphi(\alpha, \beta) \geqslant 0$ and the point $(\alpha, \beta, \varphi(\alpha, \beta))$ is on the surface of the sphere S. For (ξ, η) in Γ_R let $\psi(\xi, \eta)$ be defined in a similar manner for R. When (α, β, σ) and (ξ, η, δ) are defined as in (3.2), it follows from Theorem 2.4 that $\varphi(\alpha, \beta) \geqslant \sigma$ and $\psi(\xi, \eta) \geqslant \delta$.

Let t be a real parameter, and for $|t|$ sufficiently small let $(\alpha(t), \beta(t))$ and $(\xi(t), \eta(t))$ be parametric curves lying in Γ_S and Γ_R, respectively, and defined by

$$\alpha(t) = \alpha e^{-t}, \qquad \beta(t) = \beta e^{t}, \qquad \xi(t) = \xi e^{t}, \qquad \eta(t) = \eta e^{-t}.$$

Here e denotes the base of the natural logarithm, but any positive number $\neq 1$ could be used. Note that

$$\alpha(t)\xi(t)+\beta(t)\eta(t)=\alpha\xi+\beta\eta. \tag{3.4}$$

Also, as t increases from 0, both curves $(\alpha(t),\beta(t))$, $(\xi(t),\eta(t))$ become unbounded. As $(\alpha(t),\beta(t))$ approaches the boundary of Γ_S, $\varphi(\alpha(t),\beta(t))\to 0$. Similarly, as $(\xi(t),\eta(t))$ approaches the boundary of Γ_R, $\psi(\xi(t),\eta(t))\to 0$.

Since $\varphi(\alpha(0),\beta(0))=\varphi(\alpha,\beta)\geqslant\sigma$, $\psi(\xi(0),\eta(0))=\psi(\xi,\eta)\geqslant\delta$, and $\varepsilon\geqslant 0$, it follows from (3.4) that

$$\alpha(t')\xi(t')+\beta(t')\eta(t')+\varphi(\alpha(t'),\beta(t'))\psi(\xi(t'),\eta(t'))=\alpha\beta+\beta\eta+\sigma\delta\varepsilon$$

for some $t'\geqslant 0$. If we take $\alpha'=\alpha(t')$, $\beta'=\beta(t')$, $\xi'=\xi(t')$, $\eta'=\eta(t')$, $\sigma'=\varphi(\alpha(t'),\beta(t'))$, $\delta'=\psi(\xi(t'),\eta(t'))$ we obtain the conclusion of the theorem for $\varepsilon\geqslant 0$.

If $\varepsilon<0$ the theorem can be proved in a similar manner. In this case define ψ as before and define $\varphi(\alpha,\beta)$ to be $\leqslant 0$ and such that the point $(\alpha,\beta,\varphi(\alpha,\beta))$ is on the surface of the sphere S for (α,β) in Γ_S. ■

Remark. It is clear that Theorem 3.1 remains valid if S and R are replaced by any solid spheres centered in the complex plane and containing S and R. In what follows we assume that such spheres are known. We will obtain bounds on (3.3) in terms of the centers and radii of these spheres.

Let S_A and S_B be the solid spheres of least volume containing the eigenvalues of A and B, respectively. Identify the complex plane with the Cartesian plane, and denote the centers and radii of S_A and S_B by $(h_1,h_2,0)$, r and $(k_1,k_2,0)$, ρ, respectively. It is clear from (3.3) that we can obtain an upper bound for $\operatorname{Re} x^*A^*Bx$ by solving the maximization problem

$$\begin{aligned}
&\text{maximize} && \alpha\xi+\beta\eta+\sigma\delta\\
&\text{subject to} && (\alpha-h_1)^2+(\beta-h_2)^2+\sigma^2=r^2\\
& && (\xi-k_1)^2+(\eta-k_2)^2+\delta^2=\rho^2.
\end{aligned} \tag{3.5}$$

Similarly, we can find a lower bound on $\operatorname{Re} x^*A^*Bx$ by minimizing the objective in (3.5).

Problem (3.5) can be solved as follows. For any feasible values of the variables α, β, and σ we have

$$\begin{aligned}\alpha\xi+\beta\eta+\sigma\delta &= \alpha(\xi-k_1)+\beta(\eta-k_2)+\sigma\delta+k_1\alpha+k_2\beta\\ &\leqslant \rho\sqrt{\alpha^2+\beta^2+\sigma^2}+k_1\alpha+k_2\beta \qquad (3.6)\end{aligned}$$

by the Cauchy-Schwartz inequality. Equality holds if

$$\xi-k_1=\frac{\rho\alpha}{\sqrt{\alpha^2+\beta^2+\sigma^2}},\quad \eta-k_2=\frac{\rho\beta}{\sqrt{\alpha^2+\beta^2+\sigma^2}},\quad \delta=\frac{\rho\sigma}{\sqrt{\sigma^2+\beta^2+\sigma^2}}. \qquad (3.7)$$

Suppose we choose α, β, and σ to solve the problem

$$\begin{aligned}&\text{maximize} && \rho\sqrt{\alpha^2+\beta^2+\sigma^2}+k_1\alpha+k_2\beta\\ &\text{subject to} && (\alpha-h_1)^2+(\beta-h_2)^2+\sigma^2=r^2. \qquad (3.8)\end{aligned}$$

Then by choosing ξ, η, and δ as in (3.7) we clearly have a solution of (3.5). Since (3.5) and (3.8) have the same maximum value, we can obtain the bound we need by solving (3.8).

LEMMA. *The problem* (3.8) *always has a solution in which* $\sigma=0$.

Proof. It is clear that (3.8) always has a solution. It is also clear that if $h_1=h_2=0$ there is a solution with $\sigma=0$. So assume $h_1^2+h_2^2\neq 0$. Let $(\alpha_1,\beta_1,\sigma_1)$ be a solution of (3.8). If $\sigma_1=0$ there is nothing to prove. So assume $\sigma_1\neq 0$. We will show that there is a point on the surface of S_A with $\sigma=0$ and for which the objective value in (3.8) is at least as large as it is at $(\alpha_1,\beta_1,\sigma_1)$.

Let $r_1=\sqrt{\alpha_1^2+\beta_1^2+\sigma_1^2}$. Since $h_1^2+h_2^2\neq 0$, the sphere $\alpha^2+\beta^2+\sigma^2=r_1^2$ intersects S_A in a circle Γ normal to the complex plane and centered in the complex plane. Clearly $(\alpha_1,\beta_1,\sigma_1)\in\Gamma$. For points (α,β,σ) on Γ we have

$$\begin{aligned}&\rho\sqrt{\alpha^2+\beta^2+\sigma^2}+k_1\alpha+k_2\beta\\ &\quad=\rho r_1-\tfrac{1}{2}\left\{(\alpha-k_1)^2+(\beta-k_2)^2+\sigma^2\right\}+\tfrac{1}{2}\left\{r_1^2+k_1^2+k_2^2\right\}. \qquad (3.9)\end{aligned}$$

This shows that (3.8) achieves its maximum value on Γ at the point nearest $(k_1, k_2, 0)$. An easy application of Theorem 1.1 shows that this point can be chosen to have $\sigma = 0$. Moreover, this maximum value is at least as large as the value of (3.8) at $(\alpha_1, \beta_1, \sigma_1)$. This completes the proof of the lemma. ∎

We can now replace the problem (3.8) by the following problem in two variables:

$$\begin{aligned} &\text{maximize} \quad \rho\sqrt{\alpha^2+\beta^2} + k_1\alpha + k_2\beta \\ &\text{subject to} \quad (\alpha - h_1)^2 + (\beta - h_2)^2 = r^2. \end{aligned} \tag{3.10}$$

This problem has a trivial solution if $h_1 = h_2 = 0$ or if $k_1 = k_2 = 0$. In the first case the maximum value in (3.10) is given by $r(\rho + \sqrt{k_1^2 + k_2^2})$. In the second case the maximum value is given by $\rho(r + \sqrt{h_1^2 + h_2^2})$. So in what follows we assume that $h_1^2 + h_2^2 \neq 0$ and $k_1^2 + k_2^2 \neq 0$.

To simplify the discussion we introduce the vector notation $h = (h_1, h_2)$ and $k = (k_1, k_2)$. The 2×2 rotation matrix

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

will be denoted by J. There are several ways to reduce (3.10) to a problem involving a single variable. We have chosen the following.

Let $(\alpha(\zeta), \beta(\zeta))$ denote the point of intersection of the line $h_1\alpha + h_2\beta = \zeta$ and the circle $(\alpha - h_1)^2 + (\beta - h_2)^2 = r^2$. These curves only intersect for $|h|^2 - r|h| \leqslant \zeta \leqslant |h|^2 + r|h|$. When they intersect in two points, we choose the one for which the objective in (3.10) has the larger value. A brief calculation shows that

$$\beta(\zeta) = \frac{h_2}{|h|^2}\zeta \pm \frac{h_1}{|h|^2}\sqrt{|h|^2r^2 - (\zeta - |h|^2)^2},$$

$$\alpha(\zeta) = \frac{h_1}{|h|^2}\zeta \mp \frac{h_2}{|h|^2}\sqrt{|h|^2r^2 - (\zeta - |h|^2)^2}.$$

If we make the change of variable $\zeta = |h|^2 + |h|rz, \ -1 \leqslant z \leqslant 1$, the objective in (3.10) can be written in terms of z as

$$f(z)=\rho\sqrt{r^2+|h|^2+2|h|rz}+\frac{r}{|h|}(h\cdot k)z\pm\frac{r}{|h|}(h\cdot Jk)\sqrt{1-z^2}+h\cdot k, \tag{3.11}$$

where $\pm$ is defined so that $\pm h\cdot Jk=|h\cdot Jk|$. The maximum in (3.10) is just the maximum of (3.11) as z ranges over the interval $-1\leqslant z\leqslant 1$. If $h\cdot Jk\neq 0$, that is, if h and k are linearly independent, we have $f'(-1)=+\infty$ and $f'(1)=-\infty$. There is therefore a value of z in the open interval $(-1,1)$ where $f'(z)=0$. Moreover, $f''(z)<0$ on this interval, so there is precisely one value of z such that $f'(z)=0$. In general one may need a computer to determine this value of z. However, in special cases, $f(z)$ can be maximized explicitly. From now on we restrict our attention to one of these cases.

Assume h and k are linearly dependent, so that $h\cdot Jk=0$ and $h\cdot k=\pm|h||k|$. This includes the case studied in [3], where A and B are Hermitian matrices. In this case we can compute the maximum value of f explicitly. For example, if $h\cdot k\geqslant 0$, then $f(z)$ is monotone increasing in z and attains the maximum value

$$(\rho+|k|)(r+|h|)$$

at $z=1$.

If $h\cdot k<0$, maximizing f on the interval $[-1,1]$ is a little less straightforward. f is strictly concave, but not necessarily monotone on this interval. If f is monotone, its maximum value is attained at an endpoint of the interval and is given by

$$\max\{f(-1),f(1)\}=\max\{(\rho-|k|)(r+|h|),\rho|r-|h||+|k|(r-|h|)\}$$

If f is not monotone on the interval $-1\leqslant z\leqslant 1$, we have

$$f'(-1)f'(1)<0.$$

This follows because f is strictly concave. The condition $f'(-1)f'(1)<0$ can be written as

$$|k||r-|h||<\rho|h|<|k|(r+|h|). \tag{3.12}$$

If this condition holds, we have

$$f'(z) = |k|r - \frac{r\rho|h|}{\sqrt{r^2 + |h| - 2r|h|z}} = 0$$

for some $z \in (-1, 1)$. In this case the maximum value of f on this interval is given by

$$\frac{|k|^2 r^2 - |k|^2 |h|^2 + |h|^2 \rho^2}{2|h|\,|k|}.$$

This completes the solution of (3.10) for the case where h and k are linearly dependent. In a similar way we can obtain a lower bound on $\operatorname{Re} x^* A^* Bx$ by minimizing $\alpha\xi + \beta\eta + \sigma\delta$ subject to the constraints in (3.5). In this case we have

$$\begin{aligned}\alpha\xi + \beta\eta + \sigma\delta &= \alpha(\xi - k_1) + \beta(\eta - k_2) + \sigma\delta + k_1\alpha + k_2\beta \\ &\geqslant -\rho\sqrt{\alpha^2 + \beta^2 + \sigma^2} + k_1\alpha + k_2\beta\end{aligned}$$

with equality if

$$\xi - k_1 = -\frac{\rho\alpha}{\sqrt{\alpha^2 + \beta^2 + \sigma^2}}, \qquad \eta - k_2 = -\frac{\rho\beta}{\sqrt{\alpha^2 + \beta^2 + \sigma^2}},$$

$$\sigma = -\frac{\rho\sigma}{\sqrt{\alpha^2 + \beta^2 + \sigma^2}}.$$

It follows that

$$\min(\alpha\xi + \beta\eta + \sigma\delta) = -\max\Big(\rho\sqrt{\alpha^2 + \beta^2 + \sigma^2} - k_1\alpha - k_2\beta\Big),$$

where the min and max are taken subject to (α, β, σ) and (ξ, η, δ) in the boundaries of S_A and S_B respectively. This maximization problem is the same as (3.8) with k replaced by $-k$. It can therefore be solved by the technique we have developed for (3.8). We give the solution in the next theorem without repeating the details of the proof.

THEOREM 3.2. *Let A and B be normal matrices, and let the smallest discs containing their eigenvalues have centers $h=(h_1, h_2)$, $k=(k_1, k_2)$ and radii r, ρ, respectively. If h and k are linearly dependent, then all eigenvalues of the Jordan product A^*B+B^*A are contained in the interval $[m, M]$, where*

$$M=\begin{cases} 2(\rho+|k|)(r+|h|) & \text{if } \; h\cdot k\geqslant 0, \\ \dfrac{|k|^2r^2-|k|^2|h|^2+|h|^2\rho^2}{|h|\,|k|} & \text{if } \; h\cdot k<0 \text{ and } (3.12) \text{ holds}, \\ 2\max\{(\rho-|k|)(r+|h|),\, \rho|r-|h||+|k|(r-|h|)\} & \\ \quad \text{if } \; h\cdot k<0 \text{ and } (3.12) \text{ fails}, & \end{cases}$$

$$m=\begin{cases} -\dfrac{|k|^2r^2-|k|^2|h|^2+|h|^2\rho^2}{|h|\,|k|} & \text{if } \; h\cdot k>0 \text{ and } (3.12) \text{ holds}, \\ -2\max\{(\rho-|k|)(r+|h|),\, \rho|r-|h||+|k|(r-|h|)\} & \\ \quad \text{if } \; h\cdot k>0 \text{ and } (3.12) \text{ fails}, & \\ -2(\rho+|k|)(r+|h|) & \text{if } \; h\cdot k\leqslant 0. \end{cases}$$

This theorem remains valid if the smallest discs containing the eigenvalues of A and B are replaced by any discs containing these smallest discs. This follows from the remark following Theorem 3.1.

We close this section by constructing a class of normal matrices for which the bounds given in Theorem 3.2 are sharp.

Let $x, x_2, \dots, x_{n-1}, y$ and $x, y_2, \dots, y_{n-1}, y$ be two sets of orthonormal n-dimensional complex vectors, and let h and k be two nonzero complex numbers satisfying $h=tk$ for some $t>0$. Let r and ρ be two positive numbers satisfying

$$\frac{r}{|h|}+\frac{\rho}{|k|}<1.$$

Define complex numbers $\lambda_1, \dots, \lambda_n$ and $\mu_1, \dots, \mu_n$ as follows:

$$\lambda_1=\left(1-\frac{r}{|h|}\right)h, \qquad \lambda_n=\left(1+\frac{r}{|h|}\right)h,$$

$$\mu_1=\left(1-\frac{\rho}{|k|}\right)k, \qquad \mu_n=\left(1+\frac{\rho}{|k|}\right)k.$$

Choose $\lambda_2,\dots,\lambda_{n-1}$ and $\mu_2,\dots,\mu_{n-1}$ to be any complex numbers satisfying

$$|\lambda_j - h| \leqslant r \text{ and } |\mu_j - k| \leqslant \rho, \qquad j = 2,\dots,n-1.$$

Define

$$A = \lambda_1 xx^* + \lambda_n yy^* + \sum_{j=2}^{n-1} \lambda_j x_j x_j^*$$

and

$$B = \mu_1 xx^* + \mu_n yy^* + \sum_{j=2}^{n-1} \mu_j y_j y_j^*.$$

By construction, the eigenvalues of A are $\lambda_1,\dots,\lambda_n$ and satisfy $|\lambda_j - h| \leqslant r$ for each j. Similarly the eigenvalues of B are $\mu_1,\dots,\mu_n$, and $|\mu_j - k| \leqslant \rho$ for each j. We will show that the bounds given in Theorem 3.2 are sharp for matrices A and B constructed in this manner.

Since $h \cdot k = |h||k| > 0$, we have $M = 2(\rho + |k|)(r + |h|)$. Note that

$$(A^*B + B^*A)y = (\bar{\lambda}_n \mu_n + \bar{\mu}_n \lambda_n)y,$$

so that $\bar{\lambda}_n \mu_n + \bar{\mu}_n \lambda_n$ is an eigenvalue of $A^*B + B^*A$. But

$$\bar{\lambda}_n \mu_n + \bar{\mu}_n \lambda_n = 2\left(1 + \frac{r}{|h|}\right)\left(1 + \frac{\rho}{|k|}\right)|h||k| = M,$$

so the upper bound in Theorem 3.2 is sharp.

Consider the lower bound. Since $r/|h| + \rho/|k| < 1$, the inequality (3.12) does not hold and so

$$\begin{aligned} m &= -2\max\{(\rho - |k|)(r + |h|), \rho|r - |h|| + |k|(r - |h|)\} \\ &= -2\{\rho(|h| - r) + |k|(r - |h|)\} = 2(|h| - r)(|k| - \rho). \end{aligned}$$

Note that $(A^*B + B^*A)x = (\bar{\lambda}_1\mu_1 + \bar{\mu}_1\lambda_1)x$ and

$$\bar{\lambda}_1\mu_1 + \bar{\mu}_1\lambda_1 = 2\left(1 - \frac{r}{|h|}\right)\left(1 - \frac{\rho}{|k|}\right)|h||k| = m.$$

This shows that the lower bound is also sharp.

3.2. Kantorovich's Inequality

In [8] there appears the following generalization of an inequality due to Kantorovich [9, p. 410].

Let A be a positive definite Hermitian matrix with eigenvalues $0 < \lambda_1 \leqslant \cdots \leqslant \lambda_n$, and let

$$\chi^2 = \frac{\lambda_n}{\lambda_1}$$

be the spectral condition number for A. Let $y \neq 0$ be a given vector. For each integer ν define

$$\mu_\nu = y^*A^\nu y. \tag{3.13}$$

Then

$$1 \leqslant \frac{\mu_{\nu+1}\mu_{\nu-1}}{\mu_\nu^2} \leqslant \frac{(\chi + \chi^{-1})^2}{4}. \tag{3.14}$$

In this section we obtain a generalization of this result to normal matrices.

Let A be a normal matrix with the property that the smallest disc containing the eigenvalues of A does not contain the origin 0. For a given nonzero vector y, and each integer ν, define

$$\mu_{2\nu} = y^*(A^*)^\nu A^\nu y, \quad \mu_{2\nu+1} = y^*(A^*)^\nu AA^\nu y, \quad \mu_{2\nu+2} = y^*(A^*)^\nu A^*AA^\nu y.$$

THEOREM 3.2. *Let A be a normal matrix, and let Γ denote the smallest circular disc containing the eigenvalues of A. Let $c = (c_1, c_2)$ and r denote the center and radius of Γ, respectively. If $0 \notin \Gamma$ then*

$$1 \leqslant \frac{\mu_{2\nu+2}\mu_{2\nu}}{|\mu_{2\nu+1}|^2} \leqslant \frac{|c|^2}{|c|^2 - r^2}. \tag{3.15}$$

Proof. Let

$$x = \frac{A^{\nu} y}{\|A^{\nu} y\|}, \qquad \mu = x^* A x = \frac{\mu_{2\nu+1}}{\mu_{2\nu}}, \quad \text{and} \quad \sigma = \|Ax - \mu x\|.$$

By theorem 2.4 the point (μ, σ) lies in the region $\mathscr{R}$ of Figure 5. Let $\mu = \alpha + i\beta$. Then

$$\sigma^2 = x^* A^* A x - |\mu|^2 = \frac{\mu_{2\nu+2}}{\mu_{2\nu}} - (\alpha^2 + \beta^2).$$

This gives $\mu_{2\nu+2} = (\alpha^2 + \beta^2 + \sigma^2)\mu_{2\nu}$ and it follows that

$$\begin{aligned} \frac{\mu_{2\nu+2}\mu_{2\nu}}{|\mu_{2\nu+1}|^2} &= \frac{(\alpha^2 + \beta^2 + \sigma^2)\mu_{2\nu}^2}{|\mu_{2\nu+1}|^2} \\ &= \frac{\alpha^2 + \beta^2 + \sigma^2}{|\mu|^2} = \frac{\alpha^2 + \beta^2 + \sigma^2}{\alpha^2 + \beta^2} \geqslant 1. \end{aligned}$$

This proves the first inequality in (3.15).

To prove the second inequality we will show that

$$\frac{|c|^2}{|c|^2 - r^2}$$

is an upper bound on the value of the expression

$$\frac{\alpha^2 + \beta^2 + \sigma^2}{\alpha^2 + \beta^2} \tag{3.16}$$

for $(\alpha, \beta, \sigma) \in \mathscr{R}$. Consider maximizing (3.16) over the hemisphere

$$(\alpha - c_1)^2 + (\beta - c_2)^2 + \sigma^2 \leqslant r^2, \qquad \sigma \geqslant 0. \tag{3.17}$$

Note that $\mathscr{R}$ is contained in this hemisphere. Also (3.16) is bounded in this region, since $0 \notin \Gamma$.

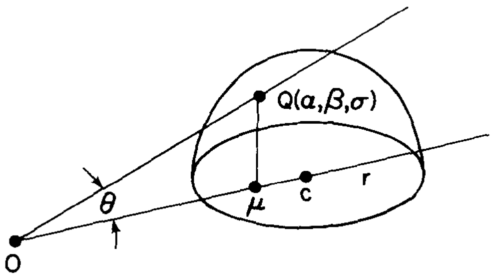

FIG. 5.

It is easy to see that the maximum of (3.16) over the hemisphere (3.17) occurs at a point $Q=(\alpha,\beta,\sigma)$ on the great circle whose diameter in the plane passes through c and $\mu=(\alpha,\beta)$. In fact, (3.16) is just $(\cos\theta)^{-2}$, where θ is the angle the line through 0 and $Q=(\alpha,\beta,\sigma)$ makes with the complex plane. Therefore (3.16) is maximized when θ is made as large as possible. It follows from Theorem 1.1, with $P=0$, that

$$|Q|^2=(|c|-r)(|c|+r)=|c|^2-r^2,$$

or

$$\alpha^2+\beta^2+\sigma^2=|c|^2-r^2. \tag{3.18}$$

Also, by taking $P=\mu$ in Theorem 1.1 we obtain

$$\begin{aligned}\sigma^2&=(r-|\mu-c|)(r+|\mu-c|)=r^2-|\mu-c|^2\\&=r^2-|c|^2+2\mu\cdot c-|\mu|^2.\end{aligned}$$

This, together with (3.18), implies that

$$\mu\cdot c=\alpha^2+\beta^2+\sigma^2=|c|^2-r^2.$$

But c is a positive multiple of μ, so $\mu\cdot c=|\mu||c|$ and therefore

$$|\mu|^2=\left(\frac{|c|^2-r^2}{|c|}\right)^2=\alpha^2+\beta^2.$$

This combines with (3.18) to give

$$\frac{\alpha^2+\beta^2+\sigma^2}{\alpha^2+\beta^2}=\frac{|c|^2}{|c|^2-r^2},$$

and this maximizes this expression subject to (3.17). This completes the proof of (3.15). ∎

If A is a positive definite Hermitian matrix with eigenvalues $0<\lambda_1\leqslant\cdots\leqslant\lambda_n$, we have

$$c=\left(\tfrac{1}{2}(\lambda_1+\lambda_n),0\right)\quad\text{and}\quad r=\tfrac{1}{2}(\lambda_n-\lambda_1).$$

We leave it to the reader to verify that in this case (3.15) reduces to (3.14).

REFERENCES

1 A. J. Hoffman and H. W. Wielandt, The variation of the spectrum of a normal matrix, *Duke Math. J.* 20:37–39 (1953).
2 E. R. Barnes and A. J. Hoffman, On bounds for eigenvalues of real symmetric matrices, *Linear Algebra Appl.* 40:217–223 (1981).
3 N. G. de Bruijn, Remarks on Hermitian matrices, *Linear Algebra Appl.* 32:201–208 (1980).
4 F. P. Preparata and M. I. Shamos, *Computational Geometry, an Introduction*, Springer-Verlag, New York, 1985.
5 Olga Taussky, Research Problem 2, *Bull. Amer. Math. Soc.* 66:275 (1960).
6 W. G. Strang, Eigenvalues of Jordan products, *Amer. Math. Monthly* 69:37–40 (1962).
7 R. Loewy, The minimal eigenvalue of the Lyapunov transform, *Linear and Multilinear Algebra* 1:213–229 (1973).
8 A. H. Schopf, On the Kantorovich inequality, *Numer. Math.* 2:344–346 (1960).
9 D. K. Faddeev and V. I. Faddeeva, *Computational Methods of Linear Algebra* (transl. by R. C. Williams), Freeman, San Francisco, 1963.

Received 18 February 1988; final manuscript accepted 1 September 1988