NORTH-HOLLAND

The Upper Numerical Range of a Quaternionic Matrix Is Not a Complex Numerical Range

Robert C. Thompson
Department of Mathematics
University of California
Santa Barbara, California 93106-3080

Submitted by Frank Hall

ABSTRACT

A recent result is that the quaternionic numerical range of a matrix with quaternion entries has a convex intersection with the upper half complex plane. This intersection is now shown to be generally not achievable as the upper half plane part of the complex numerical range of any complex matrix. A key step in the proof is that if a complex matrix has an elliptical arc as part of the boundary of its complex numerical range, then the full ellipse defined by the arc is also in its complex numerical range. © Elsevier Science Inc., 1997

1. INTRODUCTION

The well-studied numerical range $W(C)$ of a complex square matrix C is defined as the set of all complex numbers x^*Cx as x ranges over all complex unit column vectors: $x^*x = 1$. (As usual, x^* is the transpose of x with complex conjugation applied to each entry.) The famous Toeplitz-Hausdorff theorem asserts that the compact subset $W(C)$ of the complex plane is convex. See [9] for a comprehensive survey article.

Now let Q be a square matrix with quaternion entries. The quaternionic numerical range $W(Q)$ is the set of all quaternions of the form x^*Qx as x runs over all quaternion unit column vectors: $x^*x = 1$. (As usual, x^* is the transpose x with quaternion conjugation applied to each entry.) This set, which was first studied by R. Kippenhahn in an influential paper [7], lies in

655 Avenue of the Americas, New York, NY 10010

0024-3795/97/$17.00
PII S0024-3795(6)00041-9

real four dimensional space and is generally not convex. Kippenhahn introduced a set having an easier visual presentation, the *bild* of Q, denoted by $B(Q)$, and defined as the intersection of $W(Q)$ with the complex plane. This is a natural set to study, since, as Kippenhahn showed, a rotation group applied to $B(Q)$ fully specifies $W(Q)$. Unfortunately $B(Q)$ is less than fully satisfactory, since it is generally not convex.

A more desirable set to study in the quaternionic case is the intersection of $W(Q)$ with the upper complex plane (the half plane with nonnegative imaginary component). We call this set the *upper bild* of Q, or the *upper numerical range* of Q, and denote it by $B_+(Q)$. The same rotation group applied to $B_+(Q)$ fully specifies $W(Q)$, so it is an equally natural object to study, in fact more so, since it *is* convex. The convexity was proved by So, Thompson, and Zhang for normal quaternionic matrices [11] and by So and Thompson [12] for general quaternionic matrices. The proof in [11] was somewhat lengthy but was subsequently shortened in [2]. The proof in [12] is less lengthy but required rather delicate computations in a multivariable commutative polynomial ring modulo an ideal.

An attempt was made in [12] to reduce the convexity argument for $B_+(Q)$ to another proposition, namely that $B_+(Q)$ is realizable as the upper complex plane part of the complex numerical range of some complex matrix C. Example: In Figure 1 the oval region when truncated from below at the horizontal line is the upper bild of a certain quaternionic matrix Q, and the full region (convex hull of the oval and vertex u) supported by shaded territory is the complex numerical range of a certain complex matrix C. The

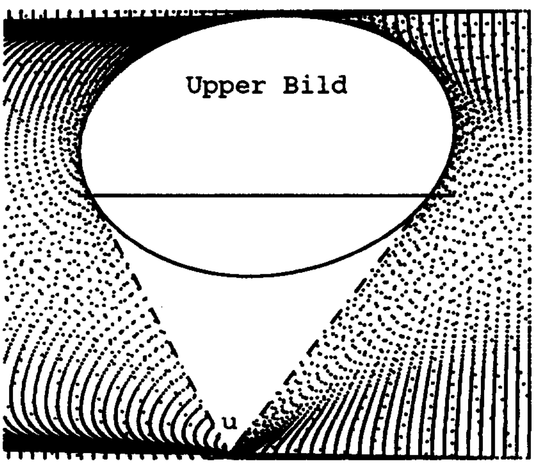

FIG. 1.

horizontal line is part of the real axis. The diagram appears to show that the upper bild of Q is the complex numerical range of C intersected with the upper complex plane, whence convex.

If this property were generally true, the Toeplitz-Hausdorff theorem applied to C would imply the convexity of $B_+(Q)$. However, no proof has been found to support this line of reasoning. Thus the convexity proof for the upper bild in [12] had to follow quite different lines.

The objective of the present note may now be stated: to exhibit a quaternionic matrix Q for which $B_+(Q)$ can be proved *not* to be the upper complex plane part of the complex numerical range of *any* complex matrix C. Thus some of the sets obtained as upper bilds are new to numerical range studies. The proof that supports our example will rely on a standard theorem in analytic perturbation theory for Hermitian matrices.

2. A PRELIMINARY LEMMA

Henceforth $W(C)$ will denote the complex numerical range of a complex matrix C. The following lemma will be our key tool.

LEMMA. *Let C be a complex matrix for which the boundary of $W(C)$ contains an elliptical arc of positive length. Then the full boundary and interior of the ellipse defined by this arc lie in $W(C)$.*

Proof. The hypotheses are to imply that the elliptical arc is not a straight line segment. We review the algorithm implicitly stated in [7] (see also [5, 10]) for the computation of the boundary of the numerical range of a complex matrix C. Write C in terms of Hermitian components as $C = H + iK$ where H and K are Hermitian. Let $\lambda_{\max}$ be the maximal eigenvalue of H. Then the line passing through $\lambda_{\max}$ parallel to the imaginary axis is a line tangent to $W(C)$. Let a unit eigenvector of H belonging to $\lambda_{\max}$ be $e_{\max}$. Then the point at which this tangency occurs is $e^*_{\max}Ce_{\max} = \lambda_{\max} + ie^*_{\max}Ke_{\max}$. Now let ϕ be an angle, and rotate $W(C)$ through angle $-\phi$ by considering $e^{-i\phi}C = \tilde{H} + i\tilde{K}$, where $\tilde{H} = H\cos\phi + K\sin\phi$ and $\tilde{K}$ are Hermitian. Denote the maximal eigenvalue of $\tilde{H}$ by $\tilde{\lambda}_{\max}$, with unit eigenvector $\tilde{e}_{\max}$. Then a tangency point for $W(e^{-\phi}C)$ is $\tilde{e}^*_{\max}(e^{-i\phi}C)\tilde{e}_{\max}$. Rotating through angle ϕ, we obtain the point $\tilde{e}^*_{\max}C\tilde{e}_{\max}$ on the boundary of $W(C)$. A plot of these points as ϕ ranges over $0 \leqq \phi \leqq 2\pi$ then yields the boundary curve of $W(C)$. The tangency (support) lines are rotated through the same angles, so we obtain tangency lines, each comprising the points (x, y) for which

$$x \cos \phi + y \sin \phi - \tilde{\lambda}_{\max} = 0.$$

The Matlab program published in [9] exhibits $W(C)$ by plotting these support lines for a large selection of values of ϕ spread throughout $[0, 2\pi]$. An even simpler technique (requiring very little Matlab code) is to exhibit the points $\tilde{e}^*_{\max} C \tilde{e}_{\max}$ for a large selection of ϕ values spread throughout $[0, 2\pi]$, where $e_{\max}$ is a unit eigenvector of $\tilde{H}$ belonging to its maximal eigenvalue. In this manner a plot of the boundary of $W(C)$ is directly found.

Fix angle ϕ, and let a new coordinate frame be obtained by rotating the real and imaginary axes through ϕ (fixing the origin) to obtain new real and imaginary axes. The construction of the last paragraph may be described in this new coordinate frame and works in this way: on the new real axis plot the point P with coordinate $\tilde{\lambda}_{\max}$; then on the line perpendicular to the new real axis through P plot the point with perpendicular coordinate $\tilde{e}^*_{\max} \tilde{K} \tilde{e}_{\max}$, where $\tilde{\lambda}_{\max}$ is the maximum eigenvalue of $\tilde{H} = H \cos \phi + K \sin \phi$ with associated unit eigenvector $\tilde{e}_{\max}$, and $\tilde{K} = -H \sin \phi + K \cos \phi$. This perpendicular line through P supports $W(C)$, meeting $W(C)$ in a single point or an interval. For convenience we call this procedure the *tangency construction*.

Thus a point on the boundary of $W(C)$ is $\tilde{e}^*_{\max} C \tilde{e}_{\max}$, where $\tilde{e}_{\max}$ is a unit eigenvector of $\tilde{H}(\phi) = H \cos \phi + K \sin \phi$ for its maximal eigenvalue $\tilde{\lambda}_{\max}$. Here we have an eigenvalue and associated unit eigenvector of a Hermitian matrix $\tilde{H}$ depending analytically on a real parameter ϕ. We shall apply to $\tilde{\lambda}_{\max}$ and to $\tilde{e}_{\max}$ the basic fact from perturbation theory [4, 6] that a Hermitian matrix function analytic in a real parameter ϕ has eigenvalues and associated orthonormal eigenvectors depending analytically on ϕ. Thus $\tilde{e}_{\max}$ may be regarded as analytically dependent on ϕ for a suitably small but nonempty range of values of ϕ.

Our hypothesis is that part of the boundary of $W(C)$ is a segment of a nondegenerate ellipse. Let $\mathscr{E}$ denote the ellipse containing this boundary segment. Placing the coordinate origin at the center of $\mathscr{E}$ by translation, and by rotation aligning the x axis with one of the semiaxes of $\mathscr{E}$, we may assume that $\mathscr{E}$ has equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0.$$

(This choice of coordinate frame means that C is replaced with $e^{i\gamma}(zI + C)$ for a suitable complex z and real γ.) Thus there is an open nonempty interval $I \subset [0, 2\pi]$ of values of ϕ such that the coordinates of the analytic point $\tilde{e}^*_{\max} C \tilde{e}_{\max}$ satisfy the equation of $\mathscr{E}$. By perturbation theory, for values of ϕ inside or outside I, there are an analytic eigenvalue $\tilde{\lambda}$ of $\tilde{H}$ and a corresponding analytic unit eigenvector of $\tilde{e}$ of $\tilde{H}$ which for ϕ in I become $\tilde{\lambda}_{\max}$ and $\tilde{e}_{\max}$. Substituting the coordinates of $\tilde{e}^* C \tilde{e}$ into the left side of the formula for the ellipse, this left side becomes an analytic function of ϕ which vanishes when $\phi \in I$. By analyticity it must therefore vanish for all ϕ. Thus the analytic continuation of $\tilde{e}^*_{\max} C \tilde{e}_{\max}$ necessarily yields points on $\mathscr{E}$.

By assumption, for $\phi \in I$, the line equation

$$x \cos \phi + y \sin \phi - \tilde{\lambda}_{\max} = 0$$

is tangent to the ellipse $\mathscr{E}$, the tangency point being $\tilde{e}^*_{\max} C \tilde{e}_{\max}$. Analytically continuing this point to become $\tilde{e}^* C \tilde{e}$, and analytically moving the tangency line so that its equation is $x \cos \phi + y \sin \phi - \tilde{\lambda} = 0$, we wish to show that the analytically moved line is still tangent to the analytically prolonged elliptical arc, the tangency point being $\tilde{e}^* C \tilde{e}$. The idea of this argument is that if two lines in the plane pass through a common point (x_0, y_0), and if they have nonzero linearly dependent normal vectors, then the lines are the same.

The line with equation $x \cos \phi + y \sin \phi - \tilde{\lambda} = 0$ contains the point $x_0 + iy_0 = \tilde{e}^* C \tilde{e}$. This is because $x \cos \phi + y \sin \phi - \tilde{\lambda}$ vanishes, by hypothesis, when $x = x_0$, $y = y_0$, $\phi \in I$, with $\tilde{\lambda}_{\max}$ in place of $\tilde{\lambda}$ and $\tilde{e}_{\max}$ in place of $\tilde{e}$. By analyticity it therefore continues to vanish when ϕ is outside I, with $\tilde{e}^*_{\max} C \tilde{e}_{\max}$ becoming $\tilde{e}^* C \tilde{e}$.

The tangent line to the ellipse passing through a point (x_0, y_0) on the ellipse has equation

$$\frac{xx_0}{a^2} + \frac{yy_0}{b^2} - 1 = 0.$$

The point $x_0 + y_0 i = \tilde{e}^* C \tilde{e}$ is on the ellipse, as already observed. Therefore (x_0, y_0) satisfies the equation of the ellipse tangent line passing through itself.

A normal vector to the line with equation $x \cos \phi + y \sin \phi - \tilde{\lambda} = 0$ is $[\cos \phi, \sin \phi]$. A normal vector to the ellipse tangent line is $[x_0/a^2, y_0/b^2]$. By hypothesis two lines are the same when $\phi \in I$, and therefore for $\phi \in I$

$$\det \begin{bmatrix} \cos \phi & \sin \phi \\ x_0/a^2 & y_0/b^2 \end{bmatrix}$$

vanishes. By analytic continuation it will still vanish for ϕ outside I. The normal vector $[\cos\phi, \sin\phi]$ for the first line is clearly not 0. Neither is the normal vector $[x_0/a^2, y_0/b^2]$ for the second line, since (x_0, y_0) lies on $\mathscr{E}$, and $\mathscr{E}$ does not pass through the origin. Therefore the line with equation $x\cos\phi + y\sin\phi - \tilde{\lambda} = 0$ will, for each ϕ, be tangent to the ellipse $\mathscr{E}$ at the point $x_0 + iy_0 = \tilde{e}^* C\tilde{e}$. The upshot of this reasoning is that the tangency construction [which for $\phi \in I$ yields points on the part of the boundary of $W(C)$ belonging to $\mathscr{E}$] prolongs in an analytic way to arbitrary $\phi \in [0, 2\pi]$ to give the point $\tilde{e}^* C\tilde{e}$ on $\mathscr{E}$ incident with the tangent line to $\mathscr{E}$ having equation $x\cos\phi + y\sin\phi - \tilde{\lambda} = 0$. This line is orthogonal to the axis with inclination angle ϕ relative to the original real axis.

A simple geometric diagram shows that the tangency point created by the tangency construction must move smoothly with monotone rotation as ϕ smoothly and monotonically covers $[0, 2\pi]$. Thus the tangency point must rotate fully once around $\mathscr{E}$ as ϕ monotonically covers $[0, 2\pi]$. Applying this to the analytic point $\tilde{e}^* C\tilde{e}$, we now know that it must rotate fully around $\mathscr{E}$ as ϕ covers $[0, 2\pi]$. In particular, every point of $\mathscr{E}$ is obtained by analytic continuation (using analytic eigenvalues and eigenvectors) of the portion of $\mathscr{E}$ forming part of the boundary of $W(C)$.

Choosing a new coordinate frame by rotating the given coordinate frame does not affect these conclusions: ϕ is replaced by $\phi + \theta$ for a constant angle θ.

Choosing a new coordinate frame by translating the origin also does not affect these conclusions. Let the new origin be at $z = a + bi$, a and b real. We replace C with $C - zI = (H - aI) + i(K - bI)$, so that the $\tilde{H}$ and $\tilde{K}$ in the old and new coordinate frames are $\tilde{H}_{\text{new}} = \tilde{H}_{\text{old}} - (a\cos\phi + b\sin\phi)I$, $\tilde{K}_{\text{new}} = \tilde{K}_{\text{old}} - (-a\sin\phi + b\cos\phi)I$. Thus $\tilde{\lambda}_{\max,\text{new}} = \tilde{\lambda}_{\max,\text{old}} - (a\cos\phi + b\sin\phi)$. So $\tilde{\lambda}_{\max}$ translates (by an amount dependent on ϕ), $\tilde{e}_{\max}$ is unchanged, and $\tilde{e}^*_{\max} C\tilde{e}_{\max}$ translates. Geometrically, the tangency point on the ellipse belonging to angle ϕ is unchanged, with the tangency construction reaching it by following a line from the new origin parallel to the line followed from the old origin, until a tangency line perpendicular to both lines is reached, then moving along the tangency line to the same tangency point.

The above descriptions of a coordinate frame change are unusually detailed because a coordinate frame change is used in the next paragraph. The point is that the tangency construction works in any coordinate frame, selecting the maximal eigenvalue $\tilde{\lambda}_{\max}$ and corresponding eigenvector $\tilde{e}_{\max}$ of $\tilde{H}$ to obtain boundary points of $W(C)$ and some eigenvalue $\tilde{\lambda}$ with corresponding eigenvector $\tilde{e}$ to obtain points on $\mathscr{E}$.

Now we show that the full ellipse $\mathscr{E}$ must be within or on the boundary of $W(C)$. Suppose that $\mathscr{E}$ extends outside $W(C)$. For a point p on $\mathscr{E}$ but outside $W(C)$ define the distance $d(p)$ to $W(C)$ in this way. Draw a circle centered at p of large radius r, then continuously shrink r. There will be a minimum value of r at which the circle boundary meets $W(C)$. Let $d(p)$ denote this minimum r, and let w be a point on the boundary of $W(C)$ at which this minimal circle meets $W(C)$.The line segment from w to p then must be perpendicular to a tangent line to $W(C)$ at w. And p lies on one side of this tangent line, $W(C)$ on the other, since $W(C)$ is convex. Since the distances $d(p)$ of p from $W(C)$ are bounded above as p moves, there will be a point p on $\mathscr{E}$ outside $W(C)$ for which $d(p)$ has a largest possible value. Let this point be p_0, and the corresponding point on the boundary of $W(C)$ be w_0. As already observed, the line segment $w_0 p_0$ is perpendicular to a support line to $W(C)$ at w_0. However, the tangent line to $\mathscr{E}$ at p_0 must also be perpendicular to segment $w_0 p_0$. The reason for this is that otherwise a point on $\mathscr{E}$ near p_0 (move from p_0 on $\mathscr{E}$ in one of the two possible directions) would have larger distance from the support line through w_0 than p_0 has, and therefore larger distance from $W(C)$ than p_0 has. Thus the line segment $w_0 p_0$ is perpendicular to a support line to $W(C)$ and also perpendicular to a support line to $\mathscr{E}$, these support lines passing though w_0 and p_0 respectively. Now choose our coordinate system so that the origin is placed at w_0 and the positive x axis lies along the ray $w_0 p_0$. The tangency construction using a maximal eigenvalue and corresponding eigenvector of the $\tilde{H}$ relative to the present coordinate frame produces points on the boundary of $W(C)$, and we have proved that the tangency construction using an analytic eigenvalue and corresponding analytic eigenvector of $\tilde{H}$ produces any point on $\mathscr{E}$. Because the line perpendicular to $w_0 p_0$ through w_0 is tangent to the boundary of $W(C)$, $w_0 = 0$ is the maximal eigenvalue of the Hermitian matrix H, that is, of $\tilde{H}$ with $\phi = 0$. However, because the line through p_0 perpendicular to $w_0 p_0$ is tangent to $\mathscr{E}$ at p_0, p_0 is an eigenvalue of $\tilde{H}$. Thus p_0 is a positive eigenvalue of H with H having 0 as its maximal eigenvalue. This is a contradiction, so $\mathscr{E}$ cannot extend outside of $W(C)$. ∎

3. THE MAIN RESULT

We let

$$G = \begin{bmatrix} k_1 i & \gamma j \\ \gamma j & 1 + k_2 i \end{bmatrix},$$

where k_1, k_2, γ are positive real numbers, and i, j, k are the usual quaternionic units ($i^2 = -1$, $j^2 = -1$, $k^2 = -1$, $ij = -ji = k$, $ik = -ki = j$,

$jk = -kj = i$). The upper bild of this 2×2 quaternionic matrix was calculated by W. So, and appears in [12]. See Figure 2. In this figure $B_+(G)$ is outlined in bold, and is the union of four regions labeled p, r, u, v.

We claim that $B_+(G)$, the bold outlined region in Figure 2, cannot be realized as the upper plane part of $W(C)$ for any complex matrix C.

THEOREM. *There is no complex matrix C for which the upper complex plane part of its complex numerical range $W(C)$ is the upper bild $B_+(G)$ of G.*

Proof. It was proved in [12] that the oval curve in Figure 2 with endpoints ik_1 and $1 - ik_2$ is an ellipse. So also is the oval curve with endpoints $-ik_1$ and $1 + ik_2$. Suppose there is a complex matrix C for which the upper complex plane part of $W(C)$ has as boundary the bold curve shown in Figure 2. Then part of the boundary of $W(C)$ is a portion of the ellipse with endpoints ik_1 and $1 - ik_2$. Therefore the full ellipse must be in $W(C)$, by the lemma proved in the last section. Consequently $1 - ik_2$ is in $W(C)$. Because $W(C)$ is convex and $1 + ik_2$ is by hypothesis on its boundary, the line segment joining $1 \pm ik_2$ must be in $W(C)$. However, the point T is also part of the $W(C)$ boundary, by hypothesis, and the convex hull of these three points is the triangle with vertices $T, 1 \pm ik_2$. But, by assumption, the elliptical arc joining $1 + ik_2$ and T is part of the boundary of $W(C)$, contradicting the fact that it is now known to be interior to $W(C)$. ∎

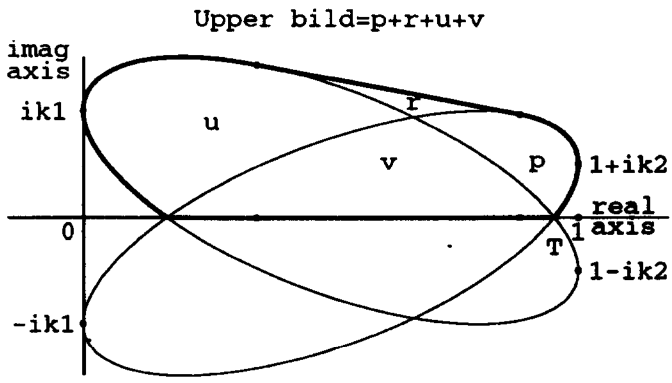

FIG. 2.

4. ANOTHER APPROACH TO THE LEMMA

The well-known elliptical range theorem asserts that the numerical range of a 2×2 matrix is the boundary and interior of an ellipse (possibly degenerate). Thus a natural proof of our lemma would require showing that, if an arc from an ellipse $\mathscr{E}$ is part of the boundary of $W(C)$ (where C is a complex matrix), then there is a unitary conjugate U^*CU of C such that for the leading 2×2 block in U^*CU the elliptical range theorem produces $\mathscr{E}$ and its interior as numerical range. Our arguments yield no information on this point.

5. NOTES BY THE REFEREE

(1) The proof of the lemma in Section 2 may be simplified as follows: Suppose $C = H + iK$ with $H = (C + C^*)/2$. Then $\det(uH + vK + wI) = 0$ is the line equation of an algebraic curve $P(C)$ such that $W(C)$ is the convex hull of $P(C)$ (e.g., see [7], and also [13, 1.4]). Suppose the boundary of $W(C)$ contains an arc of the ellipse $\mathscr{E}$ that has positive length. Then there is a quadratic polynomial $f(u, v, w)$ such that the corresponding line equation of $\mathscr{E}$ equals $f(u, v, w) = 0$. Furthermore, $\det(uH + vK + wI)$ is divisible by $f(u, v, w)$, and the convex hull of $\mathscr{E}$ is a subset of the convex hull of $P(C)$, which is $W(C)$. The conclusion of the lemma follows.

(2) The alternative approach suggested in Section 4 may not work. For example, if

$$C = \sqrt{2}\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

then $W(C)$ is the unit disk centered at the origin. However, the matrix C cannot be unitarily similar to a matrix whose leading 2×2 block $\tilde{C}$ satisfies $W(\tilde{C}) = W(C)$. Otherwise, $\tilde{C}$ will be unitarily similar to

$$\begin{pmatrix} 0 & 2 \\ 0 & 0 \end{pmatrix},$$

and we have

$$\|C\| = \sqrt{2} < 2 = \|\tilde{C}\| \leqslant \|C\|,$$

which is a contradiction. (Here $\|X\|$ denotes the spectral norm of X.)

The author's work leading to this paper (and to [12]) was supported in part by the National Science Foundation.

REFERENCES

1 Y. H. Au Yeung, On the convexity of the numerical range in quaternionic Hilbert space, *Linear and Multilinear Algebra* 16:93–100 (1984).

2 Y. H. Au Yeung, A short proof of a theorem on the numerical range of a normal quaternionic matrix, 39:279–284 (1995).

3 Y. H. Au Yeung, On the eigenvalues and numerical range of a quaternionic matrix, pp. 19–30 in Five Decades as a Mathematician and Educator: On the 80th birthday of Professor Yung-Chow Wong, (editors: K.-Y. Chan and M.-C. Liu) World Scientific, New Jersey, 1996.

4 H. Baumgärtel, *Analytic Perturbation Theory for Matrices and Operators*, Birkhäuser, Boston, 1985.

5 C. R. Johnson, Numerical determination of the field of values of a general complex matrix, *SIAM J. Numer. Anal.* 15:595–602 (1978).

6 T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, New York, 1966.

7 R. Kippenhahn, Über die Wertvorrat einer Matrix, *Math. Nachr.* 6:193–228 (1951).

8 C. K. Li, C. H. Sung, and K. Tsing, c-Convex matrices: characterizations, inclusion relations, and normality, *Linear and Multilinear Algebra* 25:275–287 (1989).

9 C. K. Li, C-numerical ranges and C-numerical radii, *Linear and Multilinear Algebra* 37:51–82 (1994).

10 M. Marcus and I. Filippenko, Normality and the higher numerical range, *Linear and Multilinear Algebra* 30:419–430 (1978).

11 Wasin So, Robert C. Thompson, and Fuzhen Zhang, The numerical range of normal matrices with quaternion entries, *Linear and Multilinear Algebra* 37:175–195 (1994).

12 Wasin So and Robert C. Thompson, Convexity of the upper complex plane part of the numerical range of a quaternionic matrix, *Linear and Multilinear Algebra*, to appear.

13 M. Fiedler, Geometry of the numerical range of matrices, *Linear Algebra Appl.* 37:81–96 (1981).

Received 13 September 1995; final manuscript accepted 8 January 1996