Numerical range: (in) a matrix nutshell

Panayiotis J. Psarrakos* Michael J. Tsatsomeros†

August 12, 2002

*Department of Mathematics, National Technical University, Zografou Campus, Athens 15780, Greece.
†Mathematics Department, Washington State University, Pullman, Washington 99164-3113.

1 The definition and a name

The *numerical range*, *field of values*, *Wertovorrat*, *Hausdorff domain*, *range of values* have been, among others, competing names for the same notion. The first two have outlived the competition but the advantage goes decisively to the first: The *numerical range* of an operator T in a complex Hilbert space H. This is the term coined by Marshall Stone in 1932 [29] for the subset of the complex plane $\mathbb{C}$ given by

$$W(T) = \{\langle Tx, x\rangle \;:\; x \in H,\; \|x\| = 1\}.$$

In technical terms here, 'T is a bounded linear operator on a separable Hilbert space H with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|x\| = \sqrt{\langle x, x\rangle}$'. But to keep our lives simple[1], let's confine our discussion to finite dimensions and talk about the numerical range of an $n \times n$ complex matrix A (we write $A \in \mathcal{M}_n(\mathbb{C})$), where $\|x\| = \sqrt{x^* x}$ is the Euclidean length of $x \in \mathbb{C}^n$ and x^* its conjugate transpose; that is,

$$W(A) = \{x^* A x \;:\; x \in \mathbb{C}^n,\; x^* x = 1\}.$$

This is a remarkable set for many reasons. It succinctly captures information about A as a transformation, particularly about the eigenvalues and, perhaps more importantly, about the eigenspaces of A. The shape of $W(A)$ (more accurately its boundary) is also related to the combinatorial structure of A, by which we mean the pattern of signs of its entries or the pattern of its nonzero entries, sometimes represented by a graph. For example, as we shall see in Section 4, an 'irreducible' entrywise nonnegative matrix whose numerical range has 4 maximal elements as depicted on the left of Figure 1 must have 'cyclic index 4'. The numerical range on the right is a circle centered at the origin and thus has an infinite number of maximal elements; the corresponding matrix must also have a special combinatorial structure.

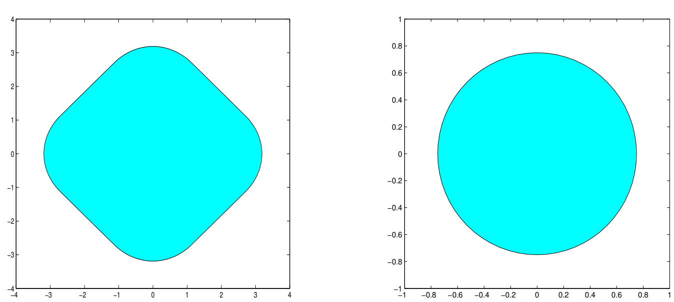

Figure 1: Numerical ranges with 4 and with infinite number of maximal elements.

Despite the conceptual simplicity of the definition of $W(A)$ (after all, it is the image of the Euclidean unit sphere under the continuous map $x \to x^* A x$), there are many interesting unanswered questions and fundamental issues surrounding it. Some are curiosity driven and some are driven by its ubiquitous nature in pure and applied mathematics. As a prelude, can you find necessary and sufficient conditions on the matrix A so that $0 \in W(A)$?

For the eager reader, more questions like the one above are contained in Section 8. For the enduring reader, we will first review the basic properties of the numerical range of a matrix and focus on results about its convexity and the geometry of its boundary (some proofs in Sections 3 and 4 are technical and can be skipped). In Section 6, we will see how can one 'draw' the boundary of the numerical range and compute the numerical radius, which is the distance from the origin to the point in the numerical range farthest away. We will then be in better position to discuss and appreciate the open problems. They are easy to state but, don't be fooled, they are quite challenging. The picture we paint of this mathematical area, our bibliography and collection of open problems are by no means comprehensive. They are only meant as a starting place and are certainly biased by our specific interests and point of view.

[1] Hilbert, David (1862-1943): *The art of doing mathematics consists in finding that special case which contains all the germs of generality.* In N. Rose Mathematical Maxims and Minims, Raleigh NC: Rome Press Inc., 1988.

2 Basic properties and some definitions

The numerical range of a matrix is known to be a **compact** and **convex** subset of $\mathbb{C}$. To be compact in this setting means two things: That the numerical range is **closed** (i.e., it contains its boundary points) and it is **bounded** (i.e., it can all fit in a disk of finite radius). To be convex means that if you draw a line segment between any two points of the numerical range, the whole line segment belongs to the numerical range.

The property of compactness follows from the fact that $W(A)$ is the image of a compact set (the unit sphere) under a continuous function, a well-known theorem in mathematical analysis. The convexity of $W(A)$ is the celebrated *Toeplitz-Hausdorff Theorem*. In May of 1918, Toeplitz [31] showed that the 'outer boundary of $W(A)$ is a convex curve', and in November of the same year, Hausdorff [15], using a different approach, proved that the numerical range is a convex set. Since then, many proofs of the Toeplitz-Hausdorff Theorem have surfaced. In Section 3, we will discuss a proof very close to the original proof of Hausdorff.

The **spectrum** of A is by definition the set $\sigma(A) = \{\lambda \in \mathbb{C} : \det(\lambda I_n - A) = 0\}$; that is, $\sigma(A)$ comprises what are known as the **eigenvalues** of A. The spectrum of A always lies in $W(A)$. This becomes clear by noticing that when x is a unit eigenvector corresponding to an eigenvalue λ (i.e., $Ax = \lambda x$ and $\|x\| = 1$), then $\lambda = x^* A x \in W(A)$.

The **numerical radius** of A is defined by

$$r(A) = \max\{|z| : z \in W(A)\},$$

and is always greater than or equal to the **spectral radius** of A,

$$\rho(A) = \max\{|z| : z \in \sigma(A)\}.$$

Also of interest, though less commonly considered, is the distance between the origin and the boundary $\partial W(A)$, called the **inner numerical radius**,

$$\hat{r}(A) = \min\{|z| : z \in \partial W(A)\}.$$

The concepts of numerical range and numerical radius have been studied extensively the last few decades. This is because they are very useful in studying and understanding the role of matrices and operators (see e.g., [3, 4, 14, 16]) in applications such as numerical analysis and differential equations (see e.g., [1, 5, 8, 10, 12, 20, 24, 27] and the discussion in Section 7).

Let's now recall a few more definitions and notation. By A^T and A^* we denote the transpose and conjugate transpose of A, respectively; also $\overline{\cdot}$ denotes complex conjugation. The **convex hull** of a set G is

$$\text{convex hull}\,\{G\} = \{tg_1 + (1-t)g_2 : t \in [0,1] \;\text{ and }\; g_1, g_2 \in G\};$$

that is, it consists of all the line segments between points in G.

Based on the definition of the numerical range, one can now fairly easily deduce the following basic properties; for details see primarily [16, Chapter 1] but also [13].

(P$_1$) For any $A \in \mathcal{M}_n(\mathbb{C})$ and for any $a, b \in \mathbb{C}$, $W(aA + bI_n) = aW(A) + b$.

(P$_2$) For any $A, B \in \mathcal{M}_n(\mathbb{C})$, $W(A + B) \subseteq W(A) + W(B)$.

(P$_3$) If $H(A) = (A + A^*)/2$ and $S(A) = (A - A^*)/2$ are the **Hermitian part** and the **skew-Hermitian part** of $A \in \mathcal{M}_n(\mathbb{C})$, respectively, then

$$\operatorname{Re} W(A) = W(H(A)) \quad \text{and} \quad \operatorname{Im} W(A) = W(S(A)).$$

(Re and Im are used to denote the real and imaginary parts of a set, respectively.)

(P$_4$) For any $A \in \mathcal{M}_n(\mathbb{C})$, $W(A^T) = W(A)$ and $W(A^*) = W(\overline{A}) = \overline{W(A)}$. As a consequence, if $A \in \mathcal{M}_n(\mathbb{R})$, then $W(A)$ is symmetric with respect to the real axis.

(P$_5$) For any $A \in \mathcal{M}_n(\mathbb{C})$ and any principal submatrix A' of A, $W(A') \subseteq W(A)$. (Recall that a principal submatrix is obtained by striking out some columns and the corresponding rows.)

(P$_6$) For any $A \in \mathcal{M}_n(\mathbb{C})$ and $B \in \mathcal{M}_m(\mathbb{C})$, convex hull $\{W(A) \cup W(B)\} = W(C)$, where

$$C = A \oplus B := \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \in \mathcal{M}_{2n}(\mathbb{C}) \ \text{ is the } \textbf{direct sum} \text{ of } A \text{ and } B.$$

(P$_7$) Let X be an $n \times m$ $(1 \leq m \leq n)$ complex matrix such that $X^*X = I_m$. Then for any $A \in \mathcal{M}_n(\mathbb{C})$, $W(X^*AX) \subseteq W(A)$. Equality holds if $m = n$; i.e., as we say, $W(A)$ is 'invariant under unitary similarities' (a square matrix X is called **unitary** if $X^* = X^{-1}$).

(P$_8$) For any $A \in \mathcal{M}_n(\mathbb{C})$, $W(A)$ contains the convex hull of the eigenvalues of A. If A is **normal**, i.e., $A^*A = AA^*$, then $W(A)$ equals the convex hull of $\sigma(A)$.

(P$_9$) For any $A \in \mathcal{M}_n(\mathbb{C})$, $W(A) \subset \mathbb{R}$ if and only if A is **Hermitian**, i.e., $A^* = A$; in this case, the endpoints of $W(A)$ coincide with the minimum and the maximum eigenvalues of A. Furthermore, $W(A)$ is a line segment in the complex plane if and only if the matrix A is normal and has collinear eigenvalues; or equivalently, if and only if $A = aH + bI_n$ for some $a, b \in \mathbb{C}$ and an Hermitian matrix H.

Example 1 In Figure 2, we have sketched first the numerical range of a normal matrix with eigenvalues -1, $1 \pm \mathrm{i}$, $2 \pm \mathrm{i}$. The numerical range to the right must be that of a non-real (complex) matrix because it is not symmetric with respect to the real axis.

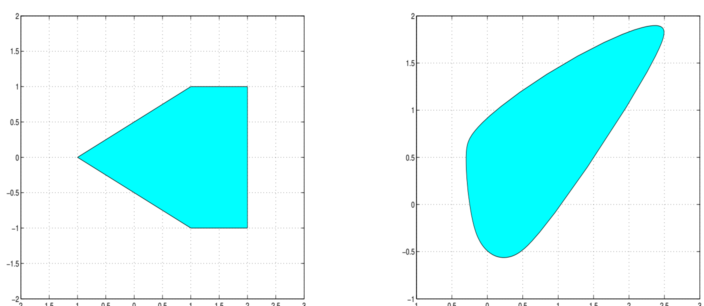

Figure 2: Numerical range of a normal and a non-real matrix.

3 Convexity

Most known proofs of the Toeplitz-Hausdorff Theorem (see [6, 7, 16, 22]) begin by observing that the problem reduces to its two-dimensional special case. The original proof of Hausdorff [15] does not use this reduction and is based on the following lemma.

Lemma 2 *Let $H \in \mathcal{M}_n(\mathbb{C})$ be an Hermitian matrix and let $\mu \in W(H)$. Then the set $\mathcal{L}_H(\mu) = \{x \in \mathbb{C}^n : x^*x = 1,\ x^*Hx = \mu\}$ is path connected* (i.e., any two points in this set can be connected by a continuous curve contained entirely in the set).

Proof By Property (P_1) and the second part of Property (P_7), without loss of generality, we can assume that $\mu = 0$ and H is a real diagonal matrix, namely, $H = \mathrm{diag}\{d_1, d_2, \ldots, d_n\}$. Then

$$W(H) = \left\{ \sum_{j=1}^{n} d_j |x_j|^2 : x_1, x_2, \ldots, x_n \in \mathbb{C},\ \sum_{j=1}^{n} |x_j|^2 = 1 \right\}.$$

Suppose now that $x = (x_j), y = (y_j) \in \mathcal{L}_H(0)$, i.e., x, y are unit vectors such that $\sum_{j=1}^{n} d_j |x_j|^2 = \sum_{j=1}^{n} d_j |y_j|^2 = 0$. We will show that there is a continuous path in $\mathcal{L}_H(0)$ connecting x and y. Since every vector $[\, r_1 e^{\mathrm{i}\theta_1}\ r_2 e^{\mathrm{i}\theta_2}\ \cdots\ r_n e^{\mathrm{i}\theta_n}\,]^T \in \mathcal{L}_H(0)$ $(r_j \geq 0,\ \theta_j \in [0, 2\pi)\ ;\ j = 1, 2, \ldots, n)$ is connected to the real vector $[\, r_1\ r_2\ \cdots\ r_n\,]^T \in \mathcal{L}_H(0)$ by the continuous curve

$$[\, r_1 e^{\mathrm{i}\theta_1(1-t)}\ r_2 e^{\mathrm{i}\theta_2(1-t)}\ \cdots\ r_n e^{\mathrm{i}\theta_n(1-t)}\,]^T\ ;\quad t \in [0,1]$$

in $\mathcal{L}_H(0)$, we can assume that x and y are real and entrywise nonnegative. Then the continuous curve

$$u(t) = (u_j(t)) = \left(\sqrt{(1-t)x_j^2 + t\, y_j^2} \right) \in \mathcal{L}_H(0) \cap \mathbb{R}^n\ ;\quad t \in [0,1]$$

satisfies $u(0) = x$ and $u(1) = y$, completing the proof. $\square$

Of course, the above lemma holds for every scalar multiple of an Hermitian matrix. Moreover, for a diagonal real matrix $H = \mathrm{diag}\{d_1, d_2, \ldots, d_n\}$ with $d_1 \geq d_2 \geq \cdots \geq d_n$, the quantity $\sum_{j=1}^{n} d_j |x_j|^2$ (where $x_1, x_2, \ldots, x_n \in \mathbb{C}$, $\sum_{j=1}^{n} |x_j|^2 = 1$) attains its maximum, d_1, when $|x_1| = 1$ and $x_2 = x_3 = \cdots = x_n = 0$; it attains its minimum, d_n, when $|x_n| = 1$ and $x_1 = x_2 = \cdots = x_{n-1} = 0$. As a consequence, the following holds.

Proposition 3 *Let $H \in \mathcal{M}_n(\mathbb{C})$ be an Hermitian matrix and let μ be either its minimum or its maximum eigenvalue. Then the set $\mathcal{L}_H(\mu) = \{x \in \mathbb{C}^n : x^*x = 1,\ x^*Hx = \mu\}$ comprises all unit eigenvectors of H corresponding to μ.*

Theorem 4 [Toeplitz-Hausdorff Theorem] *For any $A \in \mathcal{M}_n(\mathbb{C})$, $W(A)$ is convex.*

Proof We need to prove that for any distinct points $\mu, \nu \in W(A)$, the line segment $[\mu, \nu]$ lies in $W(A)$. By Property (P_1), without loss of generality, we may assume that $\mu = 0$ and $\nu = 1$. Let $x, y \in \mathbb{C}^n$ be two unit vectors such that $x^*Ax = 0$ and $y^*Ay = 1$, and consider the Hermitian part of A, $H(A)$, and the skew-Hermitian part of A, $S(A)$, as in Property (P_3). By Lemma 2, the set $\mathcal{L}_{S(A)}(0) = \{x \in \mathbb{C}^n : x^*x = 1,\ x^*S(A)x = 0\}$ is path connected. Since $x, y \in \mathcal{L}_{S(A)}(0)$, there is a continuous vector function $z(t) : [0,1] \to \mathcal{L}_{S(A)}(0)$ such that $z(0) = x$ and $z(1) = y$. Hence, the function

$$\begin{aligned} z(t)^*Az(t) &= z(t)^*H(A)z(t) + z(t)^*S(A)z(t) \\ &= z(t)^*H(A)z(t) \end{aligned}$$

is real and continuous with respect to the variable t, and satisfies $z(0)^*Az(0) = x^*Ax = 0$ and $z(1)^*Az(1) = y^*Ay = 1$. Thus, $[0,1] \subseteq W(A)$ and the proof is complete. $\square$

Remark As it has been proven recently by Lyubich and Markus [23], Lemma 2 is true for every $n \times n$ complex matrix $(n \geq 3)$.

It is worth recording in the next theorem that in the special case of 2×2 matrices, the numerical range is always an elliptical disk (with possibly degenerate interior). Recall that the **trace** of a matrix is simply the sum of its diagonal entries.

Theorem 5 [Elliptical Range Theorem] *If $A \in \mathcal{M}_2(\mathbb{C})$ with eigenvalues λ_1 and λ_2, then the numerical range $W(A)$ is an elliptical disk centered at $(1/2)\mathrm{trace}(A)$, has foci λ_1 and λ_2, and minor axis length equal to $\sqrt{\mathrm{trace}(A^*A) - |\lambda_1|^2 - |\lambda_2|^2}$.*

4 Boundary and geometry

Since the numerical range of every $A \in \mathcal{M}_n(\mathbb{C})$ is a compact subset of $\mathbb{C}$, the boundary $\partial W(A)$ and its properties are naturally of special interest. Let's first look at the case where $W(A)$ has no interior points.

Proposition 6 *Let $A \in \mathcal{M}_n(\mathbb{C})$. Then the following hold:*

(i) $W(A) = \{\mu\}$ *if and only if* $A = \mu I_n$.

(ii) $W(A)$ *has empty interior (i.e.,* $W(A) = \partial W(A)$ *is a line segment) if and only if*

$$\frac{1}{n}\,\mathrm{trace}\,(A) \in \partial W(A).$$

A boundary point μ of a closed region $\Omega \subset \mathbb{C}$ is said to be a **corner** of Ω if there is $\epsilon > 0$ such that the intersection of Ω with the circular disk $S(\mu, \epsilon) = \{z \in \mathbb{C} : |z - \mu| \leq \epsilon\}$ is contained in a sector of $S(\mu, \epsilon)$ of angle less than π. See Figure 3 for an illustration of corners and of the following result.

Theorem 7 *For any $A \in \mathcal{M}_n(\mathbb{C})$, if μ is a corner of $W(A)$, then μ is an eigenvalue of A.*

Proof We use a method similar to the one presented in [21]. By Property (P_1), without loss of generality, we assume that $\mu = 0$. Let $x \in \mathbb{C}^n$ be a unit vector such that $x^*Ax = 0$, and let $y \in \mathbb{C}^n$ be an arbitrary unit vector. Consider the function

$$\begin{aligned} G_y(\lambda, z) &= (x + \overline{z}y)^*(\lambda I_n - A)(x + zy) \\ &= (z^2 + 2\,\mathrm{Re}(x^*y)z + 1)\lambda - ((y^*Ay)z^2 + 2\,\mathrm{Re}(y^*Ax)z) \\ &= (\lambda - \lambda_y(z))R_y(z)\ ; \qquad \lambda, z \in \mathbb{C}, \end{aligned}$$

where

$$\lambda_y(z) = \frac{(y^*Ay)z^2 + 2\,\mathrm{Re}(y^*Ax)z}{z^2 + 2\,\mathrm{Re}(x^*y)z + 1}$$

is analytic in a neighborhood of the origin and satisfies $\lambda_y(0) = 0$, and

$$R_y(z) = z^2 + 2\,\mathrm{Re}(x^*y)z + 1$$

is analytic and does not vanish in a neighborhood of the origin. Clearly, there exists $\varepsilon > 0$ such that for every $t \in (0, \varepsilon)$,

$$0 \;=\; G_y(\lambda_y(t), t) \;=\; (x + t\,y)^*(\lambda_y(t)I_n - A)(x + t\,y).$$

As a consequence, for every $t \in (0, \varepsilon)$,

$$\lambda_y(t) = \frac{[(y^*Ay)\,t + 2\,\mathrm{Re}(y^*Ax)]\,t}{t^2 + 2\,\mathrm{Re}(x^*y)\,t + 1} \;\in\; W(A), \tag{1}$$

and since the origin is a corner of $W(A)$, the curve $\lambda_y(t)$ $(t \in (0, \varepsilon))$ lies in a cone

$$\mathcal{K}_0 \;=\; \{\omega \in \mathbb{C} \;:\; \theta_1 \leq \mathrm{Arg}\,\omega \leq \theta_2,\;\; 0 < \theta_2 - \theta_1 \leq \psi_0 < \pi\}.$$

For sufficiently small ε, the ratio $t/(t^2 + 2\,\mathrm{Re}(x^*y)\,t + 1)$ is positive for every $t \in (0, \varepsilon)$. Then, by (1), the curve $\lambda_y(t)$ $(t \in (0, \varepsilon))$ lies in $\mathcal{K}_0$ for every unit $y \in \mathbb{C}^n$ (recall that y is arbitrary) if and only if $Ax = 0$. Thus, 0 is an eigenvalue of A and x is a corresponding eigenvector. $\square$

Not every eigenvalue of A on the boundary of $W(A)$ is necessarily a corner of $W(A)$. What we can say, however, is that every eigenvalue μ of A on $\partial W(A)$ is **semisimple** (i.e., the geometric multiplicity and the algebraic multiplicity of μ are equal), and that every eigenvector of A corresponding to μ is orthogonal to the eigenvectors of A corresponding to the rest of the eigenvalues of A.

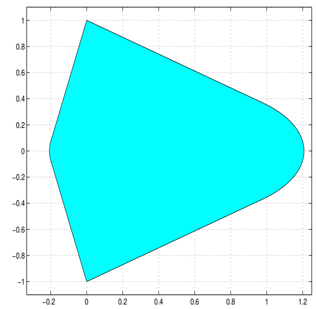

Figure 3: The imaginary numbers i and −i (corners) must be eigenvalues.

Theorem 8 *Let $A \in \mathcal{M}_n(\mathbb{C})$ and let μ be an eigenvalue of A on the boundary $\partial W(A)$. Then A is unitarily similar to a matrix of the form $\mu I_m \oplus B$ $(m \leq n)$ such that $\mu \notin \sigma(B)$.*

Proof As Schur's Theorem states, A is unitarily similar to an upper triangular matrix T whose k first diagonal elements equal to μ, where k is the algebraic multiplicity of the eigenvalue μ. If we assume that in the first k rows of T there is a nonzero entry ξ off the main diagonal, then T has a principal submatrix of the form $T_2 = \begin{bmatrix} \mu & \xi \\ 0 & \mu \end{bmatrix}$. By Theorem 5, $W(T_2)$ is the (nondegenerate) circular disk with center at μ and radius $(1/2)|\xi|$. By Properties (P_5) and (P_7), it follows that $\mu \in \mathrm{Int}W(T_2) \subset W(A)$ is an interior point of $W(A)$; a contradiction that completes the proof. □

Using the results presented so far, one can characterize those complex matrices whose numerical ranges are convex polygons. By the second part of Property (P_8), we know that the numerical range of a normal matrix is always a convex polygon. The converse is not always true. Instead, the following holds.

Theorem 9 *Suppose that the numerical range of $A \in \mathcal{M}_n(\mathbb{C})$ is a convex polygon. Then the following hold:*

(i) *If $n \leq 4$, then A is normal.*

(ii) *If $n > 4$, then either A is normal, or A is unitarily similar to a matrix of the form $B \oplus C$, where B is normal and $W(C) \subset W(B)$.*

(iii) *If $W(A)$ has n or $n-1$ vertices, then A is normal.*

Next, we mention some results that capture the essence of what is known about numerical ranges with rotational symmetries. Some more terminology is needed first.

The matrix A is called **irreducible** if there does not exist a permutation matrix P such that

$$PAP^T = \begin{bmatrix} A_{1,1} & A_{1,2} \\ 0 & A_{2,2} \end{bmatrix},$$

where $A_{1,1}$ and $A_{2,2}$ are square, non-vacuous matrices. The matrix A is called **k-cyclic** if for some permutation matrix P (think of P as the identity matrix with its rows scrambled),

$$PAP^T = \begin{bmatrix} 0 & A_{1,2} & 0 & \dots & \dots & 0 \\ 0 & 0 & A_{2,3} & 0 & \dots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & 0 & \dots & 0 & A_{k-1,k} \\ A_{k,1} & 0 & 0 & \dots & 0 & 0 \end{bmatrix}, \tag{2}$$

where the zero blocks along the diagonal are square. It is not hard to see that a k-cyclic matrix is also m-cyclic for any divisor m of k. The largest positive integer k for which A is k-cyclic is referred to as the *cyclic index* of A. When $A_{k,1} = 0$ and $k \geq 2$ in (2), A is said to be permutationally similar to a **block-shift matrix**.

The following result, found in [24], is an extension of a result that was originally observed in the unpublished doctoral thesis of J. N. Issos [17][2]. It refers to entrywise nonnegative matrices and parallels an essential part of what is known as the Perron-Frobenius Theorem on the eigenvalues of such matrices.

Theorem 10 *Let A be a square entrywise nonnegative matrix and assume that $H(A)$ (the Hermitian part of A) is irreducible. Assume also that $W(A)$ is not a circular disk centered at the origin. Then there exists a positive integer $k \leq n$ such that the maximal elements of $W(A)$ are exactly of the form $r(A)\, e^{i\frac{2t\pi}{k}}$ $(t = 0, 1, \dots, k-1)$. In fact, k coincides with the cyclic index of A.*

This result is illustrated in Figure 1 (left) in the introduction: a matrix satisfying the conditions of Theorem 10 is '4-cyclic' if and only if its numerical range has exactly 4 maximal elements.

A folk result attributed to J. Anderson states that for any $A \in \mathcal{M}_n(\mathbb{C})$ whose numerical range is included in a circular disk $\mathcal{D}$, if $W(A)$ meets the boundary of $\mathcal{D}$ at $n+1$ or more points, then $W(A)$ coincides with $\mathcal{D}$. Based on this result, it is shown in [30] that A is permutationally similar to a block-shift matrix if and only if the numerical range of any matrix with same zero pattern as A is a circular disk centered at the origin. Characterizing matrices whose numerical ranges are circular disks (not necessarily centered at the origin) remains an open problem.

5 Similarity transformations

Let $A, B \in \mathcal{M}_n(\mathbb{C})$ be two similar matrices, i.e., $A = S^{-1}BS$ for some invertible $S \in \mathcal{M}_n(\mathbb{C})$. If S is unitary (i.e., $S^{-1} = S^*$), then as we have seen $W(A) = W(B)$. In general, however, there is no tractable relation between $W(A)$ and $W(B)$, except of course that they both contain the convex hull of the common spectrum of A and B (see Property (P_8)). On the contrary, it is typical that the shape of the numerical range of a matrix changes dramatically under similarity (see Figure 4). In particular, for every $\mu \in W(A) \setminus \text{convex hull}\{\sigma(A)\}$, there exists an invertible matrix S_μ such that $\mu \notin W(S_\mu^{-1} A S_\mu)$. This is evident from the following theorem of Givens [11].

Theorem 11 *For any $A \in \mathcal{M}_n(\mathbb{C})$,* $\text{convex hull}\{\sigma(A)\} \;=\; \bigcap_{\det S \neq 0} W(S^{-1}AS)$.

Remarks

(i) When a matrix $A \in \mathcal{M}_n(\mathbb{C})$ is diagonalizable, ie., $S^{-1}AS$ is a diagonal matrix for some invertible matrix $S \in \mathcal{M}_n(\mathbb{C})$, then $W(S^{-1}AS) = \text{convex hull}\{\sigma(A)\}$.

(ii) The numerical range of a $k \times k$ 'Jordan block'

$$J_k(\mu) = \begin{bmatrix} \mu & 1 & \cdots & 0 \\ 0 & \mu & \ddots & \vdots \\ \vdots & \vdots & \ddots & 1 \\ 0 & 0 & \cdots & \mu \end{bmatrix}$$

is a circular disk centered at μ with radius $r(J_k(0)) = \cos(\pi/(k+1))$ (see e.g., [32]). Hence, by Property (P_6), the numerical range of a matrix J in Jordan Canonical Form coincides with the convex hull of the circular disks that represent the numerical ranges of its Jordan blocks (see Figure 4).

[2] James Nestor Issos was a student of Emily Haynsworth at Auburn University in the early to mid 60's.

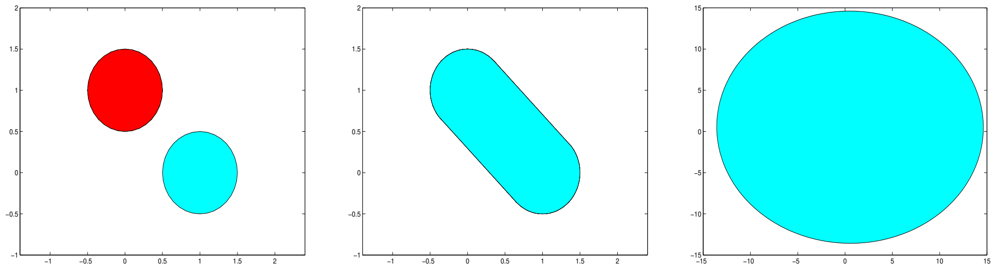

Figure 4: Numerical ranges of two Jordan blocks, their direct sum, and a similarity.

6 Numerical approximation and sketching

It is now time to explain how to sketch the numerical range of a matrix as in our figures. The 'brute force' approach would be to plot x^*Ax for lots and randomly chosen unit vectors x. But that would be too costly, and it would probably not accurately depict the boundary, especially if it is not smooth. Instead, since we know that the numerical range is compact (i.e., closed and bounded), we can try to sketch its boundary (and shade the interior). For that purpose, consider $A \in \mathcal{M}_n(\mathbb{C})$ and observe that since $W(A)$ is convex, the boundary of $W(A)$ can be approximated with convex polygons determined by supporting lines of $W(A)$. This approach has indeed a long history [18, 19, 25]) and is based on the following trick: As we have already seen, the numerical range of the Hermitian part of A, $H(A) = (A + A^*)/2$, is the closed real interval

$$W(H(A)) = [\lambda_{\min}(H(A)), \lambda_{\max}(H(A))],$$

where $\lambda_{\min}(H(A))$ and $\lambda_{\max}(H(A))$ are the minimum and maximum eigenvalues of $H(A)$, respectively. Moreover, by Property (P_3), $\operatorname{Re} W(A) = W(H(A))$. Consequently, for any unit $x \in \mathbb{C}^n$, the point $\mu = x^*Ax$ satisfies $\operatorname{Re}\mu = \max\{\operatorname{Re}\lambda : \lambda \in W(A)\}$ if and only if

$$\operatorname{Re}\mu = \operatorname{Re}(x^*Ax) = x^*H(A)x = \lambda_{\max}(H(A)).$$

Note also that by Proposition 3, $x^*H(A)x = \lambda_{\max}(H(A))$ if and only if x is a unit eigenvector of $H(A)$ corresponding to $\lambda_{\max}(H(A))$. As a consequence, if x is a unit eigenvector of $H(A)$ corresponding to $\lambda_{\max}(H(A))$, then $\mu = x^*Ax$ is a point farthest to the right of $W(A)$ and so the line $\{u + \mathrm{i}\, v : u, v \in \mathbb{R},\ u = \lambda_{\max}(H(A))\}$ is a supporting line of $W(A)$ at μ.

Now, for any $\theta \in [0, 2\pi)$, consider the matrices $e^{\mathrm{i}\theta}A = (\cos\theta + \mathrm{i}\sin\theta)A$, $H(e^{\mathrm{i}\theta}A) = \cos\theta\, H(A) + \sin\theta\,(\mathrm{i}S(A))$, the eigenvalue $\lambda_\theta = \lambda_{\max}(H(e^{\mathrm{i}\theta}A))$, and an associated unit eigenvector x_θ; i.e.,

$$H(e^{\mathrm{i}\theta}A)x_\theta = \lambda_\theta x_\theta \quad \text{and} \quad x_\theta^* x_\theta = 1.$$

Since the line $\{u + \mathrm{i}\, v : u, v \in \mathbb{R},\ u = \lambda_\theta\}$ is a supporting line of $W(e^{\mathrm{i}\theta}A)$ at the point $x_\theta^*(e^{\mathrm{i}\theta}A)x_\theta$, it follows that the line

$$e^{-\mathrm{i}\theta}\{u + \mathrm{i}\, v : u, v \in \mathbb{R},\ u = \lambda_\theta\} = \{u + \mathrm{i}\, v : u, v \in \mathbb{R},\ u\cos\theta - v\sin\theta = \lambda_\theta\}$$

is a supporting line of $W(A)$ at $x_\theta^* A x_\theta$. Therefore, choosing angles $0 = \theta_1 < \theta_2 < \cdots < \theta_\rho < \theta_{\rho+1} = 2\pi$, we can approximate $W(A)$ (see [16, 18]) by an inscribed convex polygon $\mathcal{Q}(\theta_1, \theta_2, \ldots, \theta_\rho)$ with vertices

$$q_j = x_{\theta_j}^* A x_{\theta_j}\ ; \quad j = 1, 2, \ldots, \rho$$

and by a circumscribed convex polygon $\mathcal{P}(\theta_1, \theta_2, \ldots, \theta_\rho)$ with vertices

$$p_j = e^{-\mathrm{i}\theta_j}\left[\lambda_{\theta_j} + \frac{\lambda_{\theta_j}\cos(\theta_{j+1} - \theta_j) - \lambda_{\theta_{j+1}}}{\sin(\theta_{j+1} - \theta_j)}\right]\ ; \quad j = 1, 2, \ldots, \rho.$$

Notice that the vertices of $\mathcal{P}(\theta_1, \theta_2, \ldots, \theta_\rho)$ are readily computable since they are the intersection points of the supporting lines of $W(A)$. We can refine the partition of $[0, 2\pi]$ by increasing ρ so that the area difference

$$\begin{aligned} &\text{area}\,(\mathcal{P}(\theta_1, \theta_2, \ldots, \theta_\rho)) - \text{area}\,(\mathcal{Q}(\theta_1, \theta_2, \ldots, \theta_\rho)) \\ &= \frac{1}{2}\,\mathrm{Im}\left[\sum_{j=1}^{\rho} \left(\overline{p}_j p_{j+1} - \overline{q}_j q_{j+1}\right)\right] \end{aligned} \tag{3}$$

(where it is assumed that $p_{\rho+1} = p_1$ and $q_{\rho+1} = q_1$) is satisfactorily small. This methodology is illustrated in the following algorithm and procedure.

Algorithm 1

Step 1 Choose $0 = \theta_1 < \theta_2 < \cdots < \theta_\rho < \theta_{\rho+1} = 2\pi$.

Step 2 For $j = 1, 2, \ldots, \rho$, repeat

- **(i)** Compute the eigenvalue $\lambda_j = \lambda_{\max}(H(e^{\mathrm{i}\theta_j} A))$.
- **(ii)** Compute a unit eigenvector $x_j = x_{\theta_j}$ corresponding to λ_j.
- **(iii)** Compute the j-th vertex of $\mathcal{Q}(\theta_1, \theta_2, \ldots, \theta_\rho)$, $q_j = x_j^* A x_j$.
- **(iv)** Compute the j-th vertex of $\mathcal{P}(\theta_1, \theta_2, \ldots, \theta_\rho)$,

$$p_j = e^{-\mathrm{i}\theta_j}\left[\lambda_{\theta_j} \frac{\lambda_{\theta_j}\cos(\theta_{j+1} - \theta_j) - \lambda_{\theta_{j+1}}}{\sin(\theta_{j+1} - \theta_j)}\right].$$

Step 3 Set $p_{\rho+1} = p_1$ and $q_{\rho+1} = q_1$, and compute the area difference in (3),

$$E = \text{area}\,(\mathcal{P}(\theta_1, \theta_2, \ldots, \theta_\rho)) - \text{area}\,(\mathcal{Q}(\theta_1, \theta_2, \ldots, \theta_\rho)).$$

Step 4 If E is 'small enough', then plot $\mathcal{P}(\theta_1, \theta_2, \ldots, \theta_\rho)$. Otherwise, go to Step 1, and choose a finer partition.

An implementation of the above algorithm in MATLAB is available at

`www.math.wsu.edu/math/faculty/tsat/files/matlab/nr.m`

The above methodology for sketching the boundary can also be formulated in terms of the minimum eigenvalue $\lambda_{\min}(H(e^{\mathrm{i}\theta} A))$ instead of $\lambda_{\max}(H(e^{\mathrm{i}\theta} A))$, without any essential changes. Moreover, the above discussion yields the following interesting result; see [9, 19, 25, 28] for relevant discussion.

Theorem 12 *For any $A \in \mathcal{M}_n(\mathbb{C})$, $\partial W(A)$ is the set of real points of the algebraic curve whose equation in line coordinates is* $\det\left[\, uI_n - vH(A) - w(-\mathrm{i}S(A)) \,\right] = 0$.

7 Numerical radius

Consider a matrix $A \in \mathcal{M}_n(\mathbb{C})$. The numerical radius $r(A) = \max\{|z| : z \in W(A)\}$ and the inner numerical radius $\hat{r}(A) = \min\{|z| : z \in \partial W(A)\}$ are of particular interest in many applications. For example, the numerical radius is frequently employed as a more reliable indicator of the rate of convergence of iterative methods than the spectral radius [1, 8]. The reason for this is revealed in the following theorem and corollary (for proofs and discussion see [2, 12, 14, 16, 26]).

Theorem 13 *For any $A \in \mathcal{M}_n(\mathbb{C})$, the following hold:*

(i) $(1/2)\|A\|_2 \leq r(A) \leq \|A\|_2$, *where $\|\cdot\|_2$ denotes the spectral (operator) norm induced on $\mathcal{M}_n(\mathbb{C})$.*

(ii) *For every positive integer* m*,* $r(A^m) \leq r(A)^m$ *.*

Corollary 14 *For any* $A \in \mathcal{M}_n(\mathbb{C})$ *and for any positive integer* m*,*

$$r(A^m)^{1/m} \;\leq\; \|A^m\|_2^{1/m} \;\leq\; 2^{1/m} r(A^m)^{1/m} \;\leq\; 2^{1/m} r(A).$$

The numerical radius also plays an important role in the stability analysis of finite difference approximations of solutions to hyperbolic initial value problems [12]. Furthermore, the inner numerical radius has recently been associated with stability issues of Hermitian generalized eigenproblems [5] and of higher order dynamical systems [27].

Using the notation of Section 6, the convexity of $W(A)$ yields directly two basic formulas for the numerical radius and the inner numerical radius of the matrix A.

Theorem 15 *For any* $A \in \mathcal{M}_n(\mathbb{C})$*,*

$$r(A) = \max_{\theta\in[0,2\pi]} \lambda_{\max}(H(e^{\mathrm{i}\theta}A)) \quad \textit{and} \quad \hat{r}(A) = \left| \min_{\theta\in[0,2\pi]} \lambda_{\max}(H(e^{\mathrm{i}\theta}A)) \right|.$$

It is now clear that we can compute $r(A)$ and $\hat{r}(A)$ using Step 1 and Step 2 of Algorithm 1 as follows.

Algorithm 2

Step 1 Choose $0 = \theta_1 < \theta_2 < \cdots < \theta_\rho < 2\pi$.

Step 2 For $j = 1, 2, \ldots, \rho,$ compute the eigenvalue $\lambda_j = \lambda_{\max}(H(e^{\mathrm{i}\theta_j}A))$.

Step 3 The numerical radius of A is $r(A) = \max\{\lambda_1, \lambda_2, \ldots, \lambda_\rho\}$, and the inner numerical radius of A is $\hat{r}(A) = |\min\{\lambda_1, \lambda_2, \ldots, \lambda_\rho\}|$.

8 Open problems

Problem 1 Let $A \in \mathcal{M}_n(\mathbb{C})$, and let $s_{\max}(A)$ and $s_{\min}(A)$ denote its maximum and minimum singular values (the square roots of the eigenvalues of A^*A), respectively. By the first part of Theorem 13, we know that $s_{\max}(A) = \|A\|_2 \geq r(A)$. Also, if A is Hermitian, then $s_{\min}(A) \geq \hat{r}(A) = 0$. Moreover, if A is normal, then the closest point of $\partial W(A)$ to the origin either coincides with a vertex of $W(A)$ or lies between two vertices of $W(A)$. Thus, it follows readily that $s_{\min}(A) \geq \hat{r}(A)$. So, the following question arises naturally:

Is it true that for every $A \in \mathcal{M}_n(\mathbb{C})$*,* $s_{\min}(A) \geq \hat{r}(A)$ *?*

Problem 2 One of the most important open problems on numerical ranges is the discovery of *necessary and/or sufficient conditions for the origin to be a point of* $W(A)$. More specifically, it would be interesting to discover conditions for the origin to belong to the boundary or to the topological interior of $W(A)$.

Problem 3 *What are tractable equivalent conditions so that the numerical range of a matrix is a circular disk?* This problem has been discussed at the end of Section 4. It is formally stated for nonnegative matrices (with connected undirected graphs) in [30], where it is solved only for disks centered at the origin.

References

[1] O. Axelsson, H. Lu, and B. Polman, On the numerical radius of matrices and its application to iterative solution methods, *Linear and Multilinear Algebra*, **37** (1994), pp. 225-238.

[2] C.A. Berger, On the numerical range of powers of an operator, *Notices of the American Mathematical Society*, **12** (1965), p. 590.

[3] F.F. Bonsall and Duncan, *Numerical Ranges I*, Cambridge University Press, Cambridge, 1971.

[4] F.F. Bonsall and Duncan, *Numerical Ranges II*, Cambridge University Press, Cambridge, 1973.

[5] S.-H. Cheng and N.J. Higham, The nearest definite pair for the Hermitian generalized eigenvalue problem, *Linear Algebra and Its Applications*, **302-303** (1999), pp. 63-76.

[6] C. Davis, The Toeplitz-Hausdorff Theorem explained, *Canadian Mathematical Bulletin*, **14** (1971), pp. 245-246.

[7] W. Donoghue, On the numerical range of a bounded operator, *Michigan Math. J.*, **4** (1957), pp. 261-263.

[8] M. Eiermann, Field of values and iterative methods, *Linear Algebra and Its Applications*, **180** (1993), pp. 167-197.

[9] M. Fiedler, Geometry of the numerical range of matrices, *Linear Algebra and Its Applications*, **37** (1981), pp. 81-96.

[10] M. Fiedler, Numerical range of matrices and Levinger's theorem, *Linear Algebra and Its Applications*, **220** (1995), pp. 171-180.

[11] W. Givens, Field of values of a matrix, *Proceedings of American Mathematical Society*, **3** (1952), pp. 206-209.

[12] M. Goldberg and E. Tadmor, On the numerical radius and its applications, *Linear Algebra and Its Applications*, **42** (1982), pp. 263-284.

[13] K.E. Gustafson and D.K.M. Rao, *Numerical Range*, Springer, New York, 1997.

[14] P.R. Halmos, *A Hilbert Space Problem Book*, Von Nostrand, New York, 1967.

[15] F. Hausdorff, Der Wertvorrat einer Bilinearform, *Math. Z.*, **3** (1919), pp. 314-316.

[16] R.A. Horn and C.R. Johnson, *Topics in Matrix Analysis*, Cambridge University Press, Cambridge, 1991.

[17] J. N. Issos. *The field of values of non-negative irreducible matrices*. Ph.D. Thesis, Auburn University, 1966.

[18] C.R. Johnson, Numerical determination of the field of values of a complex matrix, *SIAM J. Num. Analysis*, **15** (1978), pp. 595-602.

[19] R. Kippenhahn, Über den wertevorrat einer matrix, *Math. Nachr.*, **6** (1951), pp. 193-228.

[20] S. Kirkland, P. Psarrakos, and M. Tsatsomeros, On the location of the spectrum of hypertournament matrices, *Linear Algebra and Its Applications*, **323** (2001), pp. 37-49.

[21] H. Langer, A. Markus, and C. Tretter, Corners of numerical ranges, *Operator Theory: Advances and Applications*, **124** (2001), pp. 383-400.

[22] C.-K. Li, C-numerical ranges and C-numerical radii, *Linear and Multilinear Algebra*, **37** (1994), pp. 51-82.

[23] Y. Lyubich and A. Markus, Connectivity of level sets of quadratic forms and Hausdorff-Toeplitz type theorems, *Positivity*, **1** (1997), pp. 239-254.

[24] J. Maroulas, P. Psarrakos, and M. Tsatsomeros, Perron-Frobenius type results on the numerical range, *Linear Algebra and Its Applications*, **348** (2002), pp. 49-62.

[25] F.D. Murnaghan, On the field of values of a complex matrix, *Proc. Nat. Acad. Sci. USA*, **18** (1932), pp. 246-248.

[26] C. Pearcy, An elementary proof of the power inequality for the numerical radius, *Michigan Math. J.*, **13** (1960), pp. 289-291.

[27] P. Psarrakos and M. Tsatsomeros, On the stability radius of matrix polynomials, *Linear and Multilinear Algebra*, **50** (2002), pp. 151-165.

[28] H. Shapiro, A conjecture of Kippenhahn about the characteristic polynomial of a pencil generated by two Hermitian matrices II, *Linear Algebra and Its Applications*, **45** (1982), pp. 97-108.

[29] M. Stone, *Linear Transformations in Hilbert Space*, American Mathematical Society, Providence, 1932.

[30] B.-S. Tam and S. Yang, On matrices whose numerical ranges have circular or weak circular symmetry, *Linear Algebra and Its Applications*, **302/303** (1999), pp. 193-221.

[31] O. Toeplitz, Das algebraische Analogon zu einern Satze von Fejér, *Math. Z.*, **2** (1918), pp. 187-197.

[32] P.-Y. Wu, A numerical range characterization of Jordan blocks, *Linear and Multilinear Algebra*, **43** (1998), pp. 351-361.