NOTES ON THE NUMERICAL RANGE

JOEL H. SHAPIRO

Abstract. The inner product of a Hilbert space associates to each bounded linear operator thereon a continuous complex-valued quadratic form which—thanks to the polarization identity—uniquely determines the operator. The operator's *numerical range* is the image under this quadratic form of the surface of the Hilbert-space unit ball; it is a set of complex numbers that contains its operator's eigenvalues (if there are any), and whose closure contains the operator's spectrum. A century-old result of Toeplitz and Hausdorff shows that: *The numerical range is always convex*! From this remarkable result has grown an enormous body of work seeking to link the properties of matrices and operators with the geometry of their numerical ranges.

In these notes we'll establish the basic properties of the numerical range, prove the Toeplitz-Hausdorff convexity theorem, and explore further connections between the numerical range and the spectrum. We'll show how convexity leads to an effective way of computing the boundary of the numerical range, and will finish up with a (very brief) survey of a (very small) sample of recent literature.

Contents

Date: May 5, 2017.

An earlier draft of this work, based on lectures I gave while at Michigan State University, was partially supported by the National Science Foundation. I am indebted to Paul Bourdon for catching several errors in this earlier draft, and to Ryszard Szwarc for helpful communications regarding Hildebrandt's Intersection Theorem.

1. Introduction

For a bounded linear operator T on a Hilbert space $\mathcal{H}$, the numerical range $W(T)$ is the image of the unit sphere of $\mathcal{H}$ under the quadratic form $x \to \langle Tx, x\rangle$, i.e.,

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

Thus the numerical range of an operator, like the spectrum, is a subset of the complex plane; its properties should say something about that operator. In these notes we'll try to get some idea of what this "something" might be.

A major theme will involve connecting the numerical range with the spectrum, however in §2 we'll see that—unlike the spectrum—the numerical range is almost never invariant under similarity. This apparent disadvantage is really something good, since it gives the numerical range a chance to say something about individual operators, whereas the spectrum can only refer to whole similarity classes.

For example, many operators can have spectrum equal to the single point $\{\lambda\}$, but this singleton can be the numerical range of *only* λ times the identity operator (Corollary 3.2, page 9 below). A little less trivially: if the spectrum of an operator lies in the real line, we know little about the operator, but if its *numerical range* is real, then from basic Hilbert space theory we know that the operator must be self-adjoint (see, e.g., Corollary 4.1 below).

Very little about the numerical range is obvious; here is a more-or-less complete list of what is.

1.1. **Proposition.** *For an operator T on a Hilbert space $\mathcal{H}$:*

- (a) *$W(T)$ is invariant under unitary similarity.*
- (b) *$W(T)$ lies in the closed disc of radius $\|T\|$ centered at the origin.*
- (c) *$W(T)$ contains all the eigenvalues of T.*
- (d) *$W(T^*) = \{\overline{\lambda} : \lambda \in W(T)\}$.*[2]
- (e) *$W(I) = \{1\}$.*
- (f) *If α and β are complex numbers, and T a bounded operator on $\mathcal{H}$, then $W(\alpha T + \beta I) = \alpha W(T) + \beta$.*
- (g) *If $\mathcal{H}$ is finite dimensional then $W(T)$ is compact.*

Part (g) follows from the continuity of the quadratic form associated with T and the compactness of the unit sphere of (finite dimensional) $\mathcal{H}$. However if $\mathcal{H}$ is infinite dimensional then $W(T)$ need not be compact, even if T is a compact operator! (See Theorem 2.4 and Example 2.7 below.)

[1] The symbol ":=" means "is defined as".

[2] The overline means "complex conjugate".

1.2. **Outline of what follows.** In the next section we'll give some elementary examples which illustrate the convexity and similarity non-invariance of the numerical range, and show that when $\mathcal{H}$ is infinite dimensional there are bounded (even compact) operators with non-closed numerical range.

In Section 3 we'll study the properties of the quadratic form,

$$Q_T(x) := \langle Tx,\, x\rangle \qquad (x \in \mathbb{C}) \tag{1}$$

that creates the numerical range of the operator T by mapping the Hilbert-space unit sphere into the complex plane. We'll discuss the *polarization identity* that establishes the connection between Q_T and its bilinear cousin

$$(x, y) \to \langle Tx,\, y\rangle \qquad (x, y \in \mathbb{C}), \tag{2}$$

and will see how polarization relates the numerical range to the operator norm, especiallyx when the operator is self-adjoint.

In §5 we'll turn to a detailed analysis of the numerical range of a two-by-two matrix, establishing that it is always a possibly degenerate[3] elliptical "disc" with foci at the eigenvalues. Our proof will proceed by reducing the general case to the consideration of matrices with trace zero. The two-by-two result, along with a compression argument, will then make short work of the subject's most famous result, the *Toeplitz-Hausdorff Theorem* ([21, 1918], [8, 1919]) establishing the convexity of the numerical range. We'll prove this in §6, and observe that it leads to a higher dimensional generalization of the trace-zero result used to establish the two dimensional ellipse theorem.

The Toeplitz-Hausdorff Theorem and Proposition 1.1(c) above will get us started on the second major theme of these notes: the connection between the numerical range and the spectrum. In §7 we will observe that the spectrum of an operator lies in the *closure* of its numerical range, hence by the Toeplitz-Hausdorff theorem the same is true of the *convex hull* of the spectrum. By the similarity-invariance of the spectrum, its convex hull must therefore lie in the intersection of the closures of the numerical ranges of all the operators similar to T. This will lead to the Stephan Hildebrandt's beautiful theorem which asserts that the convex hull of the spectrum is *precisely* this intersection.

In §8 we discuss the connection between eigenvalues and points on the *boundary* of the numerical range.We;ll see that "corner points" of the numerical range must be eigenvalues, and thanks to another theorem of Hildebrandt we'll see that eigenvalues on the boundary of the numerical range behave just like eigenvalues of normal operators. Taken together, these two results assert that corner points of the numerical range are always "normal" eigenvalues.

In §9 we'll outline a method that exploits the Toeplitz-Hausdorff Theorem in the computation of the numerical range of a matrix, after which a final section will provide a few suggestions for further reading.

[3]Meaning: a line segment, or even a single point.

2. Elementary examples.

2.1. **A "finite" backward shift.** *Let T be the operator on $\mathbb{C}^2$ whose matrix with respect to the standard basis is $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$. Then $W(T)$ is the closed disc of radius 1/2, centered at the origin.*

Proof. We parameterize the unit (column) vectors of $\mathbb{C}^2$ as follows:

$$x = x(\theta, \varphi, t) = e^{i\varphi}[t, e^{i\theta}\sqrt{1-t^2}\,]', \tag{3}$$

where the prime symbol "$'$" denotes "transpose", θ and φ are real, and $0 \le t \le 1$. Now a little calculation shows that

$$\langle Tx,\, x\rangle = e^{i\theta} t\sqrt{1-t^2},$$

which, as θ traverses the real line, describes the circle of radius $t\sqrt{1-t^2}$, centered at the origin. Thus $W(T)$ is the union of all these circles as t runs over the closed unit interval, i.e. it is the disc of radius

$$\max_{0\le t\le 1} t(1-t^2)^{1/2} = 1/2,$$

centered at the origin. □

2.2. **Similarity "non-invariance".** From now on we will identify operators on finite dimensional Hilbert spaces with the matrices that represent them relative to convenient orthonormal bases for that space. For the previous example, the action of T relative to the standard unit-vector basis $\{e_1, e_2\}$ of $\mathbb{C}^2$ is to take e_1 to the zero-vector, and e_2 to e_1. This is why it's being called a "backward shift."

This two dimensional backward shift dramatically illustrates the non-similarity invariance of the numerical range. For each complex number λ, consider the operator λT, whose matrix is $\left[\begin{smallmatrix} 0 & \lambda \\ 0 & 0 \end{smallmatrix}\right]$. By Proposition 1.1(f) we know that $W(\lambda T) = \lambda W(T)$, the closed disc of radius $|\lambda|$ centered at the origin, so if $|\lambda| \ne 1$ then λT has numerical range different from that of T. However: for $\lambda \ne 0$ *all the operators λT are similar* since the matrix $S_\lambda := \left[\begin{smallmatrix} \lambda & 0 \\ 0 & 1 \end{smallmatrix}\right]$ is nonsingular and $S_\lambda T S_\lambda^{-1} = \lambda T$.

2.3. **The backward shift on ℓ^2.** The infinite dimensional version of the two dimensional backward shift discussed above is one of the most important examples in operator theory. This is the operator B defined on ℓ^2 by:

$$B(\xi_0, \xi_1, \xi_2 \ldots) = (\xi_1, \xi_2, \ldots) \qquad ((\xi_0, \xi_1, \xi_2 \ldots) \in \ell^2).$$

Clearly B has norm one, so by Proposition 1.1(b) its numerical range is contained in the closed unit disc $\overline{\mathbb{U}} := \{\lambda \in \mathbb{C} : |\lambda| \le 1\}$.

2.4. **Proposition.** $W(B) = \mathbb{U}$ (the open unit disc).

Proof. For $\lambda \in \mathbb{U}$ the vector $x_\lambda := (1, \lambda, \lambda^2, \lambda^3, \ldots)$ belongs to ℓ^2, and $Bx_\lambda = \lambda x_\lambda$, i.e., each $\lambda \in \mathbb{U}$ is an eigenvalue of B with eigenvector x_λ. Thus $\mathbb{U} \subset W(B)$. We have already observed that $W(B) \subset \overline{\mathbb{U}}$, so it's enough to show that no point on the unit circle belongs to $W(B)$. Suppose, for the sake of contradiction, that some λ of modulus one belonged to the numerical range of B. Then there would be a unit vector x in $\mathcal{H}$ with $\lambda = \langle Bx, x\rangle$. Since $\|B\| = 1$ we would then have

$$1 = |\lambda| = |\langle Bx, x\rangle| \leq \|Bx\|\,\|x\| \leq \|x\|\,\|x\| = 1,$$

i.e., there would be equality in the Cauchy-Schwarz inequality (the inequality in the middle of the above display). It would follow that Bx is a scalar multiple of x, the scalar in question necessarily being λ. Now an easy calculation shows that since $Bx = \lambda x$ and $|\lambda| = 1$, the complex sequence x is not in ℓ^2, a contradiction that shows λ cannot belong to $W(B)$. □

Unitarily diagonalizable operators. Let us call a bounded operator T on a Hilbert space $\mathcal{H}$ *unitarily diagonalizable* if it has diagonal matrix relative to some orthonormal basis, i.e., if there exists an orthonormal basis $\{e_n\}_0^\infty$ for $\mathcal{H}$ consisting of eigenvectors of T. All normal operators on a finite dimensional Hilbert space, and more generally, all compact normal operators on a separable Hilbert space, are unitarily diagonalizable.[4]

2.5. **Theorem.** *The numerical range of a unitarily diagonalizable operator is the convex hull of its eigenvalues.*

Proof. Let's begin with the finite-dimensional case, where we can identify our operator with an $n \times n$ matrix A which, by the diagonalizability hypothesis, is unitarily equivalent to a diagonal matrix $D = \text{diag}\{\lambda_1, \lambda_2, \ldots \lambda_n\}$. Thus $W(A) = W(D)$, the set of complex numbers $\langle Dv, v\rangle$ as v runs through the unit vectors of $\mathbb{C}^n$. Each such v is a column vector with j-th coordinate ξ_j, where $\sum_j |\xi_j|^2 = 1$, and one calculates that $\langle Dv, v\rangle = \sum_j \lambda_j |\xi_j|^2$. Thus, as v runs through the unit vectors of $\mathbb{C}^n$, the points of the numerical range of A run through the convex hull of its eigenvalues.

More generally, let T be a diagonalizable operator on a Hilbert space $\mathcal{H}$. By hypothesis there is an orthonormal basis $\{e_n\}$ for $\mathcal{H}$ and a sequence $\{\lambda_n\}$ of complex numbers such that $Te_n = \lambda_n e_n$ for every non-negative integer n.

[4] As are compact normal operators on an arbitrary Hilbert space, under suitable interpretation of the concept of "orthonormal basis."

Thus

$$\begin{aligned} W(T) &:= \{\langle Tf, f\rangle : f \in \mathcal{H}, \|f\| = 1\} \\ &= \left\{\sum_n \lambda_n |\langle f, e_n\rangle|^2 : f \in \mathcal{H}, \|f\| = 1\right\} \\ &= \left\{\sum_n \lambda_n a_n : 0 \le a_n \le 1, \sum_n a_n = 1\right\} \\ &:= \operatorname{conv}_\infty(\Lambda), \end{aligned}$$

where $\Lambda = \{\lambda_n\}$ is the collection of eigenvalues of T. Thus we need only prove:

2.6. **Proposition.** *For any countable set* $\Lambda = \{\lambda_n\}$ *of complex numbers,* $\operatorname{conv}_\infty(\Lambda) = \operatorname{conv}(\Lambda)$*, the convex hull of* Λ*.*

Proof of Proposition. Clearly $\operatorname{conv}(\Lambda) \subset \operatorname{conv}_\infty(\Lambda)$, and $\operatorname{conv}_\infty(\Lambda)$ is convex. We want to show that if $p \in \operatorname{conv}_\infty(\Lambda)$ then p is an honest convex combination of points of Λ. Now

$$\operatorname{conv}(\alpha\Lambda + \beta) = \alpha \operatorname{conv}(\Lambda) + \beta \qquad \forall \alpha, \beta \in \mathbb{C}, \tag{4}$$

and the same is true of $\operatorname{conv}_\infty(\Lambda)$, hence we may, upon replacing Λ by $\Lambda - p$, assume that $p = 0$. Suppose $0 \notin \operatorname{conv}(\Lambda)$. Then there is a half-plane H that contains $\operatorname{conv}(\Lambda)$ and whose boundary contains 0. By rotating about the origin (again using (4)) we may assume that Λ, and hence both $\operatorname{conv}(\Lambda)$ and $\operatorname{conv}_\infty(\Lambda)$ lie in the closed upper half-plane.

We are assuming that $0 \in \operatorname{conv}_\infty(\Lambda)$, hence there exist numbers a_n between 0 and 1 such that $0 = \sum_n a_n \lambda_n$. Since $0 \notin \operatorname{conv}(\Lambda)$, infinitely many of the a_n are nonzero. Now $0 = \sum_n a_n \operatorname{Im}(\lambda_n)$, and since $\operatorname{Im}\lambda_n \ge 0$ for each n, we must have λ_n real for each non-zero a_n. Thus there must be some λ_n that is real and negative, and another λ_m that is real and positive. Then the origin lies on the line segment between these two numbers, and hence belongs to the convex hull of Λ. But we assumed $0 \notin \operatorname{conv}(\Lambda)$. This contradiction shows $0 \in \operatorname{conv}(\Lambda)$. This completes the proof of the Proposition, and therefore also the proof of the Theorem. □

2.7. **A compact operator with non-closed numerical range.** The backward shift on ℓ^2 showed us that the numerical range of an operator on infinite dimensional Hilbert space need not be closed. Theorem 2.5 lets us produce examples of *compact* operators with this property. Let $\{\lambda_n\}$ be positive real numbers with $\lambda_n \searrow 0$. Let T be the operator on ℓ^2 whose matrix with respect to the standard orthonormal basis is $\operatorname{diag}\{\lambda_n\}$. It's an elementary exercise to show that this operator is compact, and by Theorem 2.5 its numerical range of T is the half-open interval $(0, \lambda_0]$. For more on the question of numerical-range compactness, see [3, 1972].

Direct sums provide a convenient way to create further examples of numerical ranges. The underlying principal is:

2.8. **Proposition.** *Suppose $\mathcal{H}_1$ and $\mathcal{H}_2$ are Hilbert spaces on which live bounded operators T_1 and T_2 respectively. Then $W(T_1 \oplus T_2)$ is the convex hull of $W(T_1) \cup W(T_2)$.*

Proof. We'll think of $\mathcal{H}_1$ and $\mathcal{H}_2$ as orthogonal subspaces of a Hilbert space $\mathcal{H}$ that span $\mathcal{H}$, and $T := T_1 \oplus T_2$ to be a bounded operator on $\mathcal{H}$ leaving invariant $\mathcal{H}_1$ and $\mathcal{H}_2$, with T_1 and T_2 denoting the restriction of T to these respective subspaces.

Suppose w is a point of the convex hull of $W(T_1) \cup W(T_2)$. We wish to show that $w \in W(T)$. Since numerical ranges are convex, there exist $w_1 \in W(T_1)$, $w_2 \in W(T_2)$ and $s \in [0, 1]$ such that $w = sw_1 + (1-s)w_2$. For $j = 1, 2$, there exists a unit vector $v_j \in \mathcal{H}_j$ such that $w_j = \langle T_j v_j,\, v_j\rangle$. Let

$$v = \sqrt{s}\, v_1 + \sqrt{1-s}\, v_2 \ . \tag{5}$$

Since the summands on the right-hand side of (5) are orthogonal, v is a unit vector in $\mathcal{H}$: $\|v\|^2 = s\|v_1\|^2 + (1-s)\|v_2\|^2 = 1$. Moreover—again by orthogonality,

$$\begin{aligned} \langle Tv,\, v\rangle &= \langle T_1(\sqrt{s}\, v_1),\, \sqrt{s}\, v_1\rangle + \langle T_2(\sqrt{1-s}\, v_2),\, \sqrt{1-s}\, v_2\rangle \\ &= s\langle T_1 v_1,\, v_1\rangle + (1-s)\langle T_2 v_2,\, v_2\rangle \\ &= sw_1 + (1-s)w_2 \\ &= w \end{aligned}$$

Thus $w \in W(T)$. *Conclusion*: The convex hull of $W(T_1) \cup W(T_2)$ is a subset of $W(T_1) \oplus W(T_2)$.

For the opposite inclusion, suppose $w \in W(T)$. We wish to show that w is a convex combination of points in $W(T_1)$ and $W(T_2)$. We may assume that w does not belong to either of these numerical ranges; otherwise there is nothing to prove.

Thus we're given a unit vector $v \in \mathcal{H}$ such that $w = \langle Tv,\, v\rangle$, where v has unique representation $v = v_1 + v_2$ with $v_1 \in \mathcal{H}_1$ and $v_2 \in \mathcal{H}_2$, and neither of these component vectors is zero. By orthogonality: $\|v_1\|^2 + \|v_2\|^2 = 1$. Let $\|v_1\|^2 = s$, so $0 < s < 1$ and $\|v_2\|^2 = 1 - s$. Let $u_1 = v_1/\sqrt{s}$ and $u_2 = v_2/\sqrt{1-s}$. Then u_j is a unit vector in $\mathcal{H}_j$ $(j = 1, 2)$ and

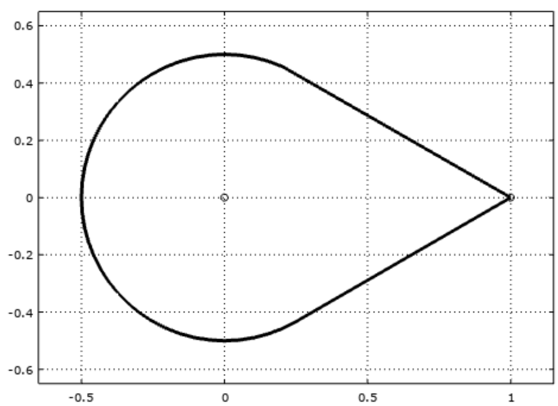
$$\begin{aligned} w = \langle Tv,\, v\rangle &= \langle T_1 v_1,\, v_1\rangle + \langle T_2 v_2,\, v_2\rangle \\ &= \langle T_1(\sqrt{s}\, u_1),\, \sqrt{s}\, u_1\rangle + \langle T_2(\sqrt{1-s}\, u_2),\, \sqrt{1-s}\, u_2\rangle \\ &= s\underbrace{\langle T_1 u_1,\, u_1\rangle}_{\in W(T_1)} + (1-s)\underbrace{\langle T_2 u_2,\, u_2\rangle}_{\in W(T_2)} . \end{aligned}$$

Thus w lies in the convex hull of $W(T_1) \cup W(T_2)$, as desired. □

2.9. **Example.** The matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

represents the operator $T_1 \oplus T_2$ on $\mathcal{H} = \mathbb{C}^3 = \mathbb{C} \oplus \mathbb{C}^2$, where T_1 is the identity operator on the one dimensional Hilbert space $\mathbb{C}$ and T_2 is the two dimensional "backward shift" of §2.1. Thus $W(A)$ is the "ice-cream cone" formed by the convex hull of $W(T_1) =$ the singleton $\{1\}$ and $W(T_2) =$ the closed disc of radius $1/2$, centered at the origin. The picture is shown below, with the eigenvalues of T (0 and 1) represented by small circles.



3. Quadratic considerations.

Suppose for the moment that $\mathcal{H}$ is just a complex inner-product space, and that T is a linear transformation on $\mathcal{H}$; in particular, $\|x\|^2 = Q_I(x)$ for each vector $x \in \mathcal{H}$. Thanks to the bilinearity of the inner product we see that for any vectors x and y in $\mathcal{H}$:

$$Q_T(x+y) = Q_T(x) + \langle Tx, y\rangle + \langle Ty, x\rangle + Q_T(y), \tag{6}$$

and substitution of $-y$ for y yields the companion identity

$$Q_T(x-y) = Q_T(x) - \langle Tx, y\rangle - \langle Ty, x\rangle + Q_T(y). \tag{7}$$

Subtraction of (7) from (6) yields for each pair $x, y \in \mathcal{H}$,

$$Q_T(x+y) - Q_T(x-y) = 2\langle Tx, y\rangle + 2\langle Ty, x\rangle,$$

and substitution of iy for y in this last identity provides, after a little manipulation,

$$iQ_T(x+iy) - iQ_T(x-iy) = 2\langle Tx, y\rangle - 2\langle Ty, x\rangle,$$

so addition of the last two identities yields the *Polarization Identity*:

$$Q_T(x+y) - Q_T(x-y) + iQ_T(x+iy) - iQ_T(x-iy) = 4\langle Tx, y\rangle.$$

Using the fact that

$$Q_T(ax) = |a|^2 Q_T(x) \tag{8}$$

for each pair x, y of vectors in $\mathcal{H}$, we can rewrite the Polarization Identity in a way that's perhaps easier to remember:

$$\langle Tx,\, y\rangle = Q_T\Big(\frac{x+y}{2}\Big) - Q_T\Big(\frac{x-y}{2}\Big) + iQ_T\Big(\frac{x+iy}{2}\Big) - iQ_T\Big(\frac{x-iy}{2}\Big), \tag{9}$$

valid for all $x, y \in \mathcal{H}$.

3.1. **Corollary.** *Suppose $\mathcal{H}$ is a complex inner-product space, with S and T linear transformations thereon. If $Q_T = Q_S$ on $\mathcal{H}$, then $S = T$.*

Proof. We're assuming that $Q_T(x) = Q_S(x)$ for each $x \in \mathcal{H}$. By the Polarization Identity, this means that $\langle Tx,\, y\rangle = \langle Sx,\, y\rangle$ i.e., that $\langle (T-S)x,\, y\rangle = 0$, for each pair of vectors $x, y \in \mathcal{H}$. This implies, upon setting $y = (T-S)x$, that $\|(T-S)x\|^2 = 0$ for each $x \in \mathcal{H}$, i.e., that $T - S = 0$. □

Remark. If, instead of subtracting (7) from (6), we *add* it, there results

$$Q_T(x+y) + Q_T(x-y) = 2Q_T(x) + 2Q_T(y).$$

For the special case $T =$ identity map on $\mathcal{H}$, this is the *Parallelogram Law*:

$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2 \qquad (x, y \in \mathcal{H}). \tag{10}$$

3.2. **Corollary.** *If T is a bounded operator on a Hilbert space $\mathcal{H}$ and $W(T)$ is the singleton $\{\lambda\}$, then $T = \lambda I$.*

Proof. We're assuming that for each unit vector $x \in \mathcal{H}$:

$$Q_T(x) = \lambda = Q_{\lambda I}(x)$$

i.e., that $Q_T = Q_{\lambda I}$. Thus $T = \lambda I$ by Corollary 3.1. □

3.3. **The Numerical Radius.** For $\mathcal{H}$ a Hilbert space and T a bounded operator on $\mathcal{H}$, define the *numerical radius*, $w(T)$, of T to be the supremum of the absolute values of points in the numerical range of T, i.e.,

$$w(T) := \sup\{|Q_T(x) \colon x \in \mathcal{H}, \|x\| = 1|\}.$$

Thanks to the fact that $Q_T(ax) = |a|^2 Q_T(x)$ for each $x \in \mathcal{H}$ and $a \in \mathbb{C}$, we have for each bounded operator T on $\mathcal{H}$:

$$|Q_T(x)| \le w(T)\|x\|^2 \qquad (x \in \mathcal{H}). \tag{11}$$

Proof. We need only consider $x \in \mathcal{H}\backslash\{0\}$, in which case we can form $v = x/\|x\|$, the unit vector in the direction of x.

$$|Q_T(x)| = |Q_T(\|x\|v)| = \|x\|^2 |Q_T(v)| \le \|x\|^2 w(T)\,,$$

as desired. □

3.4. **Proposition.** *For each bounded operator T on $\mathcal{H}$:*

$$w(T) \le \|T\| \le 2w(T).$$

Proof. The first inequality follows immediately from the Cauchy-Schwartz inequality and the fact that $\|T\|$ is the supremum of all numbers $|\langle Tx, y\rangle|$ where x and y range over all unit vectors in $\mathcal{H}$.

For the second inequality, fix $x, y \in \mathcal{H}$ and apply the triangle inequality to the Polarization Identity (9):

$$|\langle Tx, y\rangle| \le \left|Q_T\Big(\frac{x+y}{2}\Big)\right| + \left|Q_T\Big(\frac{x-y}{2}\Big)\right| + \left|Q_T\Big(\frac{x+iy}{2}\Big)\right| + \left|Q_T\Big(\frac{x-iy}{2}\Big)\right|$$

$$\le w(T)\left[\left\|\frac{x+y}{2}\right\|^2 + \left\|\frac{x-y}{2}\right\|^2\right] + w(T)\left[\left\|\frac{x+iy}{2}\right\|^2 + \left\|\frac{x-iy}{2}\right\|^2\right].$$

By the Parallelogram Law (10), each of the terms in square brackets is equal to $\|x\|^2/2 + \|y\|^2/2$, Thus

$$|\langle Tx, y\rangle| \le w(T)\big(\|x\|^2 + \|y\|^2\big),$$

from which the desired inequality follow upon taking the supremum of both sides of the one above over all unit vectors x and y in $\mathcal{H}$. □

The inequalities of Proposition 3.4 cannot be improved. The identity transformation (on any Hilbert space) establishes the best-possible nature of the constant "1" in the first inequality, We'll see in §2.1 that the matrix $\left[\begin{smallmatrix}0 & 1\\ 0 & 0\end{smallmatrix}\right]$, viewed as a linear transformation on $\mathbb{C}^2$, has numerical radius 1/2. Since its operator norm is 1 (easy exercise), this example shows that the constant "2" in the second inequality of Proposition 3.4 is also best-possible.

We'll see in the next section that for restricted classes of operators it may be possible to say more about the relationship between norm and numerical radius. In particular, for any self-adjoint operator these two quantities are *equal*.

4. Self-adjoint operators.

Recall that each bounded operator T on a Hilbert space $\mathcal{H}$ has a companion *adjoint* operator T^* defined on $\mathcal{H}$ by

$$\langle T^*x, y\rangle = \langle x, Ty\rangle \qquad (x, y \in \mathcal{H}).$$

T^* is a bounded operator on $\mathcal{H}$, with the same norm as T.

If $T = T^*$, then T is called (not surprisingly) *self-adjoint*, or equivalently, *hermitian*. In this section we'll see that the numerical range has special significance for self-adjoint operators. Throughout, T will denote a bounded linear operator on a Hilbert space $\mathcal{H}$.

4.1. **Theorem.** *T is self-adjoint if and only if $W(T) \subset \mathbb{R}$.*

Proof. If $T = T^*$ then for each $x \in \mathcal{H}$:

$$\langle Tx, x\rangle = \langle x, T^*x\rangle = \langle x, Tx\rangle = \overline{\langle Tx, x\rangle},$$

where the line over the last term means "complex conjugate." Thus for each $x \in \mathcal{H}$ the complex number $Q_T(x) := \langle Tx, x\rangle$ coincides with its complex conjugate, so is real. Thus $W(T) \subset \mathbb{R}$.

Conversely, suppose $W(T) \subset \mathbb{R}$, so by Eqn.(8), $Q_T(x)$ is real for every vector in $x \in \mathcal{H}$. Fix vectors x and y in $\mathcal{H}$, and rewrite the Polarization Identity (9) as

$$4\langle Tx, y\rangle = [Q_T(x+y) - Q_T(x-y))] + i[Q_T(x+iy) - Q_T(x-iy)].$$

Since we're assuming that Q_T takes only real values, the right-hand side of this equation is the decomposition of the left-hand side into real and imaginary parts, i.e.,

$$4\,\mathrm{Re}\,\langle Tx, y\rangle = Q_T(x+y) - Q_T(x-y) \tag{12}$$

and

$$4\,\mathrm{Im}\,\langle Tx, y\rangle = Q_T(x+iy) - Q_T(x-iy). \tag{13}$$

We wish to show that $\langle Tx, y\rangle = \langle x, Ty\rangle$, i.e., that

$$\mathrm{Re}\,\langle Tx, y\rangle = \mathrm{Re}\,\langle Ty, x\rangle \tag{14}$$

and

$$\mathrm{Im}\,\langle Tx, y\rangle = \mathrm{Im}\,\langle x, ty\rangle\,. \tag{15}$$

Upon interchanging x and y in Eqn.(12) we see obtain

$$4\mathrm{Re}\,\langle Ty, x\rangle = Q_T(y+x) - Q_T(y-x).$$

Now thanks to Eqn.(8),

$$Q_T(y-x) = Q_T((-1)(x-y))) = Q_T(x-y),$$

which establishes Eqn.(14).

For the second of these equations we use Eqn.(8) in the form $Q_T(\pm iv) = Q_T(v)$ to obtain:

$$\begin{aligned} 4\,\mathrm{Im}\,\langle Tx, y\rangle &= Q_T(x+iy) - Q_T(x-iy) \\ &= Q_T(i(y-ix)) - Q_T((-i)(y+ix)) \\ &= Q_T(y-ix) - Q_T(y+ix) \\ &= -4\,\mathrm{Im}\,\langle Ty, x\rangle, \end{aligned}$$

with the last equality following from Eqn.(13). Thus

$$\mathrm{Im}\,\langle Tx, y\rangle = -\mathrm{Im}\,\overline{\langle x, Ty\rangle} = \mathrm{Im}\,\langle Tx, y\rangle,$$

so we've established Eqn.(15), and finished the proof. □

Our identification (12) of the real part of Q_T when T is self-adjoint leads to the following useful "hermitian" improvement of Proposition 3.4.

4.2. **Theorem.** *If T is self-adjoint then $\|T\| = w(T)$.*

Proof. We already know that $w(T) \leq \|T\|$, so we need only prove the opposite inequality. Fix unit vectors x and y in $\mathcal{H}$, and use Eqn.(8) to rewrite Eqn.(12) as:

$$\operatorname{Re}\langle Tx,\, y\rangle = Q_T\Big(\frac{x+y}{2}\Big) - Q_T\Big(\frac{x-y}{2}\Big).$$

Thus, the triangle inequality:

$$\begin{aligned}|\operatorname{Re}\langle Tx,\, y\rangle| &\leq \Big|Q_T\Big(\frac{x+y}{2}\Big)\Big| + \Big|Q_T\Big(\frac{x-y}{2}\Big)\Big| \\ &\leq \tfrac{w(T)}{4}\big[\|x+y\|^2 + \|x-y\|^2\big].\end{aligned}$$

By the Parallelogram Law (10) and our assumption that x and y are unit vextors, the term in square brackets on the right is just $2\|x\|^2 + 2\|y\|^2$, which is ≤ 4. Thus $|\operatorname{Re}\langle Tx,\, y\rangle| \leq w(T)$.

If we can prove that this last inequality continues to hold when we erase the symbol "Re" on its left-hand side, then upon sup-ing the left-hand side of the result over all unit vectors x and y we'll obtain the desired estimate: $\|T\| \leq w(T)$.

To this end, fix—once again—a pair x, y of unit vectors, and this time assume (without loss of generality) that $\langle Tx,\, y\rangle \neq 0$. Then we can form $\omega := \langle Tx,\, y\rangle / |\langle Tx,\, y\rangle|$, a complex number of unit modulus. Since the inner product is conjugate-homogeneous in the second variable:

$$|\langle Tx,\, y\rangle| = \langle Tx,\, \alpha y\rangle = \operatorname{Re}\langle Tx,\, \alpha y\rangle \leq w(T)\,,$$

as desired. □

5. The numerical range of a two-by-two matrix

In this section[5] we'll prove that the numerical range of a two-by-two matrix (i.e., an operator on a two dimensional complex Hilbert space) is either a point, a line, or an elliptical (perhaps cirular) disc.

Suppose first that our matrix A is *normal*, i.e., that it commutes with its adjoint (the complex conjugate of its transpose). Then A is unitarily equivalent to a diagonal matrix, so if it has just a single eigenvalue λ, it's clearly equivalent to, and therefore equal to, λ times the identity matrix. Thus by Corollary 3.2, $W(A) = \{\lambda\}$. If, on the other hand, our normal matrix A has two distinct eigenvalues, then it's unitarily equivalent to the diagonal matrix having these eigenvalues as its main diagonal, so by Theorem 2.5, $W(A)$ is the convex hull of these eigenvalues, i.e., the line segment joining them. *Summary*: If A is a two-by-two normal matrix then its numerical range is either a point or a line segment.

[5]The material of this section follows closely that of [10], Sec. 1.3, pp. 17–28

Suppose now that A is not normal. In §2.1 we worked out a special case $N = \left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$, for which $W(N) = \frac{1}{2}\overline{\mathbb{U}}$, the closed disc of radius $\frac{1}{2}$ centered at the origin. More generally, suppose our two-by-two matrix A has just just one eigenvalue λ. Let $v \in \mathbb{C}^2$ be a unit eigenvector for this eigenvalue, and $w \in \mathbb{C}^2$ a unit vector orthogonal to v. Then relative to the basis (v, w) for $\mathbb{C}^2$, the linear transformation represented by A has an upper triangular matrix B; to which A is unitarily equivalent.[6] In particular, B inherits "single eigenvalue" property of A, and therefore has λ as both entries on its main diagonal, so $B = \left[\begin{smallmatrix} \lambda & \mu \\ 0 & \lambda \end{smallmatrix}\right] = \lambda I + \mu N$. Thus by Proposition 1.1(e), $W(A) = \lambda + \mu W(N)$, i.e., the closed disc of radius $|\mu|/2$, centered at the eigenvalue λ.

It remains to understand the numerical ranges of non-normal two-by-two matrices that have two distinct eigenvalues. For this we'll need to work out a more general class of examples. Let's call a complex matrix with entries zero everywhere except possibly on the major cross-diagonal a "cross-diagonal" matrix.

5.1. **Proposition.** *The numerical range of a non-normal, two-by-two cross-diagonal matrix is an elliptical disc with foci at the eigenvalues.*

Proof. We have $A = \left[\begin{smallmatrix} 0 & a \\ b & 0 \end{smallmatrix}\right]$, where a and b are complex numbers. First suppose a and b are positive. We may assume $0 < b \le a$, else take adjoints and use Proposition 1.1(d). Then employing the parameterization (3) and doing some computation:

$$\begin{aligned} W(A) &= \left\{ t\sqrt{1-t^2}\,[(ae^{i\theta} + be^{-i\theta}] : \theta \in \mathbb{R}, 0 \le t \le 1 \right\} \\ &= \left\{ t\sqrt{1-t^2}\,[(a+b)\cos\theta + i(a-b)\sin\theta] : \theta \in \mathbb{R}, 0 \le t \le 1 \right\}, \end{aligned}$$

which (because $\max_{0 \le t \le 1} t\sqrt{1-t^2} = 1/2$) describes either:

(a) The line segment $[-a, a]$ if $a = b$ (in which case A is, by Theorem 4.1 self-adjoint, and therefore normal, with eigenvalues $\pm a$, or

(b) The elliptical disc with center at the origin, whose boundaryn has horizontal major axis of length $a+b$ and vertical minor axis of length $a - b$ if $a \ne b$.

Since we're assuming A is not normal, its numerical range is, indeed, an elliptical disc. Let's recall from high-school geometry that if E is an ellipse in $\mathbb{R}^2$ with foci F_1 and F_2 (not necessarily distinct), then there a positive number k such that E is the set of points $P \in \mathbb{R}^2$ the sum of whose distances from the foci equals k. From this it's easy to check that if M is the length of the major semi-axis of E and m the length of the minor semi-axis, and d the distance between the center of E and either of its foci, then $M^2 = d^2 + m^2$.

[6] This is a special case of a famous result if Issai Schur, which asserts that: *Every square complex matrix is unitarily equivalent to an upper triangular matrix.*

In our case $M = (a+b)/2$ and $m = (a-b)/2$, so

$$F = \pm\sqrt{M^2 - m^2} = \pm\sqrt{ab},$$

i.e. *the foci of $\partial W(T)$ are the eigenvalues of A.* Thus:

> *For any non-negative numbers a and b, the numerical range of the matrix $\left[\begin{smallmatrix} 0 & a \\ b & 0 \end{smallmatrix}\right]$ is a (possibly degenerate) elliptical disc with foci at the eigenvalues.*

Now suppose a and b are arbitrary complex numbers. Write both in polar form: $a = |a|e^{i\alpha}$ and $b = |b|e^{i\beta}$, and observe that

$$\textit{if} \quad S = \begin{bmatrix} 1 & 0 \\ 0 & e^{i\frac{\alpha-\beta}{2}} \end{bmatrix} \quad \textit{then} \quad SAS^{-1} = e^{i\frac{\alpha+\beta}{2}} \begin{bmatrix} 0 & |a| \\ |b| & 0 \end{bmatrix}.$$

Thus A is unitarily equivalent to a cross-diagonal matrix whose non-zero entries are positive. The Proposition now follows from the work just done on this special case. □

The next result shows that Proposition 5.1 is not quite as special as it seems.

5.2. **Lemma.** *If A is a two-by-two complex matrix with trace zero then A is unitarily equivalent to a cross-diagonal matrix.*[7]

Proof. Suppose A has just one eigenvalue. We've already noted that any two-by-two matrix is unitarily equivalent to an upper triangular one B which, since A has trace zero, must only zeros on its diagonal. Thus $B = \left[\begin{smallmatrix} 0 & a \\ 0 & 0 \end{smallmatrix}\right]$.

Next, suppose A has two distinct eigenvalues, necessarily $\pm\lambda \neq 0$. Let u be a unit eigenvector for λ, and v a unit eigenvector for $-\lambda$. Then the pair $\{u, v\}$ is linearly independent, so for each real θ the vector

$$w_\theta := u + e^{i\theta}v$$

is non-zero, as is its image

$$Aw_\theta = Au + e^{i\theta}Av = \lambda(u - e^{i\theta}v).$$

A little calculation shows that

$$\langle Aw_\theta, w_\theta\rangle = 2i\lambda\,\mathrm{Im}\,\{e^{-i\theta}\langle u, v\rangle\}, \tag{16}$$

so upon choosing θ so that $e^{i\theta} = \frac{\langle u, v\rangle}{|\langle u, v\rangle|}$ (where the denominator is not zero because u and v are not orthogonal), we see that $e^{-i\theta}\langle u, v\rangle = |\langle u, v\rangle|$, hence the right-hand side of Eqn.(16) is zero. Thus Aw_θ and w_θ are (non-zero) orthogonal vectors. Let z and w be the unit vectors in the direction of w_θ and Aw_θ, respectively, so that the pair (z, w) is an orthonormal basis for $\mathbb{C}^2$. With respect to this new basis the linear transformation represented

[7] More generally, every square trace-zero matrix is unitarily equivalent to a zero-diagonal matrix. In Theorem 6.2 below, we'll give a proof based on the Toeplitz-Hausdorff Theorem!

by A has matrix B that is unitarily equivalent to A, and has $(1,1)$-entry $\langle Az, z\rangle$, which is a scalar multiple of $\langle Aw_\theta, w_\theta\rangle = 0$. Thus the $(1,1)$-entry of B is zero. Being unitarily equivalent to A, the matrix B has the same trace, namely zero, hence its (2,2) entry must also be zero.

Conclusion: A is unitarily equivalent to a matrix with zero diagonal. □

The last result reduces the determination of $W(A)$ for general two-by-two complex matrices to the trace-zero case, which is handled by Proposition **??**. Thus we have proved:

5.3. **The Elliptical Range Theorem.** *If A is a two-by-two complex matrix, then $W(A)$ is a (possibly degenerate[8]) elliptical disc with foci at the eigenvalues of A and center at* $(\text{trace } A)/2$.

5.4. **More on the dimensions of the ellipse.** Our work on cross-diagonal matrices shows that for *any* two-by-two matrix, the foci (suitably interpreted in degenerate cases) of the numerical range are the eigenvalues. Having noted this, it makes sense to seek an intrinsic expression for the lengths of the major and minor axes.

Let's restrict attention to matrices whose numerical ranges are actual (non-degenerate) elliptical discs. As before, it is enough to work with trace-zero matrices, and by unitary equivalence, with cross-diagonal matrices, and finally, with the special case $A = \left[\begin{smallmatrix} 0 & a \\ b & 0 \end{smallmatrix}\right]$ with (since we do not want degeneracies) $0 < b \le a$. Here the eigenvalues are $\lambda_1 = \sqrt{ab}$ and $\lambda_2 = -\sqrt{ab}$, and unit eigenvectors for λ_1 and λ_2 respectively are:

$$f_1 = \frac{1}{\sqrt{a+b}}\begin{bmatrix} \sqrt{a} \\ \sqrt{b} \end{bmatrix} \quad \text{and} \quad f_2 = \frac{1}{\sqrt{a+b}}\begin{bmatrix} \sqrt{a} \\ -\sqrt{b} \end{bmatrix} .$$

Thus

$$\gamma := \langle f_1, f_2\rangle = \frac{a-b}{a+b} = \frac{\text{length of minor axis}}{\text{length of major axis}} ,$$

and

$$\sqrt{1-\gamma^2} = \frac{2\sqrt{ab}}{a+b} = \frac{\lambda_1 - \lambda_2}{\text{length of major axis}} = \frac{\text{distance between foci}}{\text{length of major axis}} ,$$

hence:

$$\text{length of major axis} = \frac{\text{distance between foci}}{\sqrt{1-\gamma^2}} ,$$

and

$$\text{length of minor axis} = \gamma \times \text{length of major axis}.$$

Thus we have shown that every two-by-two cross-diagonal matrix has as its numerical range a (possibly degenerate) elliptical disc with foci at the eigenvalues, and—in the event the entries are non-negative—have calculated the lengths of the major and minor axes in terms of quantities involving only the matrice's eigenvectors and eigenvalues. We have also shown that

[8]Meaning: a line segment.

every cross-diagonal two-by-two matrix is unitarily similar to a unimodular multiple of one with non-negative entries. This leads to the following result for general two-by-two matrices, where absolute values take care of the fact that the order in which eigenvalues occur is arbitrary:

5.5. **Theorem** (Elliptical range theorem; the "full story")**.** *Suppose A is a two-by-two matrix with distinct eigenvalues λ_1 and λ_2, to which correspond unit eigenvectors f_1 and f_2. Let $\gamma = |\langle f_1, f_2\rangle|$. Then:*

(a) *$W(A)$ is a (possibly degenerate) elliptical disc whose boundary has its foci at λ_1 and λ_2.*
(b) *The major axis of $\partial W(A)$ has length $\frac{|\lambda_1-\lambda_2|}{\sqrt{1-\gamma^2}}$.*
(c) *The minor axis of $\partial W(A)$ has length γ times that of the major axis.*

The limiting cases are:

- *The eigenvalues are distinct and the eigenvectors orthogonal*, in which case the Theorem predicts that $W(A)$ is an "elliptical disc" with major axis of has length equal to the distance between the eigenvalues and the minor axis has length zero, i.e., $W(A)$ is the line segment joining the eigenvalues. In this case our matrix A is unitarily equivalent to the 2×2 diagonal matrix with diagonal (λ_1, λ_2), so the result of the Theorem is in accordance with Theorem 2.5.
- *The eigenvalues are not distinct*, in which case the eigenvectors are parallel, hence $\gamma = 1$, so part (b) of the Theorem makes no sense, but part (c) predicts that the major and minor axes of $\partial W(A)$ have the same length, i.e., that $W(A)$ is a "circular" disc. This is what we found, e.g., in §2.1.

6. The Toeplitz-Hausdorff Theorem

This is the "Fundamental Theorem of the Numerical Range:"

6.1. **Theorem** (O. Toeplitz [21, 1918] and F. Hausdorff [8, 1919])**.** *For each bounded linear operator on a Hilbert space, the numerical range is convex.*[9]

Proof. The idea is to "compress" the problem to two dimensions. More precisely, suppose $\mathcal{H}$ is a Hilbert space, $\mathcal{M}$ a (closed linear) subspace of $\mathcal{H}$, and $P_{\mathcal{M}}$ the orthogonal projection of $\mathcal{H}$ onto $\mathcal{M}$. For a bounded linear operator T on $\mathcal{H}$ the *compression $T_{\mathcal{M}}$ of T to $\mathcal{M}$* is the restriction to $\mathcal{M}$ of the operator $P_{\mathcal{M}}T$.

Now suppose $x \in \mathcal{M}$. Then

$$\langle T_{\mathcal{M}}x, x\rangle = \langle P_{\mathcal{M}}Tx, x\rangle = \langle Tx, P_{\mathcal{M}}^*x\rangle = \langle Tx, P_{\mathcal{M}}x\rangle = \langle Tx, x\rangle$$

where in the third equality we use the fact that the projection $P_{\mathcal{M}}$ is self-adjoint (and in the fourth one the fact that $x \in \mathcal{M}$). In particular,

[9]Toeplitz proved the "outer boundary" of the numerical range to be a convex curve; Hausdorff subsequently proved that there's no "inner boundary."

> *The numerical range of a bounded linear operator on a Hilbert space contains the numerical ranges each of its compressions.*

Given our work on two dimensional operators, the Toeplitz-Hausdorff Theorem follows easily from this. Suppose T is a bounded linear operator on the Hilbert space $\mathcal{H}$. Suppose λ and μ are two distinct points of $W(T)$. We desire to show that the line segment $[\lambda, \mu]$ lies entirely in $W(T)$. We have unit vectors f and g with $\lambda = \langle Tf, f\rangle$ and $\mu = \langle Tg, g\rangle$. These vectors are linearly independent (else $\lambda = \mu$), hence they span a two dimensional (closed) subspace $\mathcal{M}$ of $\mathcal{H}$. By our results for dimension two, the compression of T to $\mathcal{M}$ has convex numerical range (either an elliptical disc, or a line segment), which contains both λ and μ, and therefore the segment $[\lambda, \mu]$. As observed in the paragraph above, this segment also lies in the numerical range of T. □

The key step in our proof was the observation that every two-by-two complex matrix with trace zero is unitarily equivalent to one with main diagonal identically zero. Fittingly, the Toeplitz-Hausdorff Theorem provides an n-dimensional generalization.

6.2. **Theorem** (W. V. Parker [15, 1948]). *Every square, trace-zero, complex matrix is unitarily equivalent to a matrix whose main diagonal is identically zero.*

Proof. Let A be an $n \times n$ complex matrix. By Schur's Theorem[10] the trace of A is the sum of its eigenvalues, each eigenvalue occurring in the sum as many times as its algebraic multiplicity[11]. Thus $0 = \operatorname{trace} A/n$ is a convex combination of the eigenvalues of A, each of which lies in $W(A)$. By the Toeplitz-Hausdorff Theorem, $0 \in W(A)$, so there is a unit vector u_1 with $\langle Au_1, u_1\rangle = 0$. Let M be the one dimensional subspace spanned by u_1, so in the decomposition of $\mathbb{C}^n$ into the orthogonal direct sum of M and the $n-1$ dimensional subspace $M^\perp$ we may view A (or rather the linear transformation of $\mathbb{C}^n$ represented by A) as an operator matrix of the form $\left[\begin{smallmatrix} 0 & * \\ * & A_M \end{smallmatrix}\right]$, where A_M is the compression of A to M, "0" is a one-by-one matrix, and the asterisks denote matrices of dimensions $(n-1) \times 1$ and $1 \times (n-1)$ respectively. The argument can now be repeated on A_1, producing a vector u_2 orthogonal to u_1 with $0 = \langle A_1 u_2, u_2\rangle = \langle Au_2, u_2\rangle$. Then the dimension reduction argument can be repeated, with M now the span of u_1 and u_2, and the Toeplitz-Hausdorff Theorem used to produce a unit vector $u_3 \in M^\perp$ for which $\langle Tu_3, u_3\rangle = 0$. The process ends after n repetitions, producing an orthonormal basis for $\mathbb{C}^n$ relative to which the matrix of the operator represented by A has zero diagonal. □

[10] *Every square complex matrix is unitarily similar to an upper-triangular matrix.*

[11] Its multiplicity as a root of the characteristic polynomial

7. The numerical range and the spectrum

We have observed that the numerical range of an operator contains all its eigenvalues. What connection exists between the numerical range and the spectrum? Since the spectrum is closed and the numerical range need not be (if $\mathcal{H}$ is infinite dimensional), we can't expect the spectrum to lie in the numerical range. This is dramatically illustrated by the elementary example $T = \operatorname{diag}\{1/n\}_1^\infty$ on ℓ^2, for which the numerical range is the half-open interval $(0,1]$, while the spectrum is the countable set $\{\frac{1}{n}\}_1^\infty \cup \{0\}$ which is almost, but not quite, not contained in $(0,1]$. This example, nevertheless, points to a general truth:

7.1. **Theorem.** *If T is a bounded linear operator on a Hilbert space $\mathcal{H}$, then the spectrum of T is contained in the* closure $\overline{W}(T)$ *of $W(T)$.*

Proof. Because both the spectrum and the numerical range transform properly under affine mappings of operators, it is enough to prove that if $0 \in \sigma(T)$ then $0 \in \overline{W}(T)$. So suppose $0 \in \sigma(T)$, i.e., that T is not invertible. There are two possibilities: T is not bounded below, or T is bounded below (i.e., has closed range) but is not onto.

(a) If T is not bounded below then there exist unit vectors $f_n \in \mathcal{H}$ such that $\langle Tf_n, f_n\rangle \to 0$, thus exhibiting 0 as a limit of points in $W(T)$, and so placing 0 in $\overline{W}(T)$.
(b) It's even better if T is bounded below but not onto. For in this case, $\{0\} \neq (\operatorname{ran} T)^\perp = \ker T^*$, hence $0 \in W(T^*)$, the complex conjugate of $W(T)$. Thus 0 actually lies in the numerical range of T (not just its closure). □

7.2. **Remark.** Thanks to the Toeplitz-Hausdorff Theorem and the similarity-invariance of the spectrum, we see that

$$\operatorname{conv}\{\sigma(T)\} \subset \bigcap\{\overline{W}(VTV^{-1})\colon V \text{ invertible on } \mathcal{H}\}.^{12}$$

Our next major result, due to Stephan Hildebrandt [9, 1996] asserts that this set containment is actually *equality*. We'll present a strikingly short proof of Hildebrandt's theorem due to James Williams [22, 1969]. This proof will require a result of Gian-Carlo Rota [18, 1960] that is interesting in its own right. Rota's theorem asserts that every strict contraction on a Hilbert space is similar to "part" of a backward shift. More precisely, given a Hilbert space $\mathcal{H}$ let $\ell^2(\mathcal{H})$ denote the space of sequences $x = (\xi_n)_0^\infty$ with each $\xi_n \in \mathcal{H}$ and

$$\|x\|^2 = \sum_{n=0}^\infty \|x_n\|_{\mathcal{H}}^2.$$

[12] Here "conv" denotes "convex hull."

The functional $\|\cdot\|$ turns $\ell^2(\mathcal{H})$ into a Hilbert space, on which acts the *backward shift* B, defined, by

$$B(x_0, x_1, x_2, \ldots) = (x_1, x_2, x_3, \ldots);$$

it is a bounded linear operator on $\ell^2(\mathcal{H})$ of norm one.

7.3. **Rota's Theorem.** *Suppose T is a bounded linear operator on a Hilbert space $\mathcal{H}$, and that the spectrum of T lies in the open unit disc. Then there is a B-invariant subspace $\mathcal{M}$ of $\ell^2(\mathcal{H})$ and an isomorphism W of $\mathcal{H}$ onto $\mathcal{M}$ such that $T = W^{-1}BW$, i.e., T is similar to the restriction of B to an invariant subspace.*

Proof. Our assumption on T is that its spectral radius $r(T)$ is less than 1, so by the spectral radius formula, $\lim_n \|T^n\|^{1/n} = r(T) < 1$. Thus for (say) $\rho := (1 + r(T))/2 < 1$ we have $\|T^n\| < \rho^n$ for all sufficiently large n, and so $\sum \|T^n\|^2 := M < \infty$. Define the map W from $\mathcal{H}$ into $\mathcal{H} \times \mathcal{H} \times \ldots$ by:

$$W(x) = (x, Tx, T^2x, \ldots) \qquad (x \in \mathcal{H}),$$

so that for each $x \in \mathcal{H}$,

$$\|Wx\|^2 \leq \|x\|^2 \sum_{n=0}^{\infty} \|T^n\|^2 = M\|x\|^2,$$

hence W is a bounded linear operator from $\mathcal{H}$ into $\ell^2(\mathcal{H})$. Clearly $BWx = WTx$ for each $x \in \mathcal{H}$, so $BW = WT$ on $\mathcal{H}$. It's also clear that $\|Wx\| \geq \|x\|$ for every $x \in \mathcal{H}$, so $\mathcal{M} := W(\mathcal{H})$ is a closed subspace of $\ell^2(\mathcal{H})$ and W is an isomorphism of $\mathcal{H}$ onto $\mathcal{M}$. The intertwining relationship $BW = WT$ guarantees that $\mathcal{M}$ is B-invariant, and this completes the proof. □

Before turning to the proof of Hildebrandt's theorem, let's show how Rota's Theorem gives a quick proof of an interesting result that appears to be weaker than, but actually implies, Hildebrandt's Theorem.

7.4. **Corollary.** *If T is a bounded linear operator on a Hilbert space $\mathcal{H}$, then the spectral radius of T is* $\inf\{\|VTV^{-1}\| \colon V \text{ invertible on } \mathcal{H}\}$.

Proof. Let $s(T)$ denote the infimum in question. We all know that $r(T)$ is $\leq \|T\|$ for any operator T on a Hilbert space, and that spectral radius is a similarity invariant. Thus $r(T) \leq s(T)$.

We'll use Rota's theorem to get the opposite inequality. Since this requires an operator of spectral radius < 1, for each $\varepsilon > 0$ we'll set

$$T_\varepsilon = [r(T) + \varepsilon]^{-1}\, T\,.$$

Then $r(T_\varepsilon) < 1$, so Rota's theorem applies and produces a B-invariant subspace $\mathcal{M}$ of $\ell^2(\mathcal{H})$ and an isomorphism W of $\mathcal{H}$ onto $\mathcal{M}$ for which the restriction of B to $\mathcal{M}$ is $WT_\varepsilon W^{-1}$. Thus

$$[r(T) + \varepsilon]^{-1} \|WTW^{-1}\| = \|WT_\varepsilon W^{-1}\| = \|B|_{\mathcal{M}}\| \leq 1,$$

i.e. $\|WTW^{-1}\| \leq [r(T) + \varepsilon]^{-1}$. Now the dimension of $\mathcal{M}$ is the same as that of $\mathcal{H}$, so we may regard W as an isomorphism of $\mathcal{H}$. Thus the last estimate shows that for every $\varepsilon > 0$ the infimum in the statement of the result we are trying to prove is $\leq r(T) + \varepsilon$, and therefore it is $\leq r(T)$. □

Finally, we use Corollary 7.4 and the Toeplitz-Hausdorff Theorem to get a quick proof of the main result of this section.

7.5. **Hildebrandt's Intersection Theorem.** *If T is a bounded linear operator on a Hilbert space, then the convex hull of its spectrum is equal to*

$$\bigcap\{\overline{W}(VTV^{-1}) : V \text{ invertible on } \mathcal{H}\}.$$

Proof. We have already observed the containment"$\subset$", so it remains only to go the other way. For this, suppose λ is not in the convex hull of the spectrum of T. We wish to show that λ is not in the intersection of numerical-range closures of operators in the similarity orbit of T, i.e., that there is an invertible operator V on $\mathcal{H}$ such that $\lambda \notin \overline{W}(V^{-1}TV)$. Because $\operatorname{conv}\{\sigma(T)\}$ is compact, there is an open disc Δ that contains it, but whose closure does not contain λ. Because both the numerical range and spectrum behave properly relative to affine mappings of operators, we may assume without loss of generality that Δ is the open unit disc, so in particular $r(T) < 1$. Corollary 7.4 thus provides an invertible operator V on $\mathcal{H}$ such that $\|V^{-1}TV\| \leq (1 + r(T))/2 < 1$, hence $\overline{W}(V^{-1}TV) \subset \Delta$ and therefore $\lambda \notin \overline{W}(V^{-1}TV)$. □

8. Eigenvalues in the boundary of the numerical range

8.1. **Definition.** A *corner point* of a convex set C is a point on the boundary of C which lies at the vertex of a sector that contains C and has angular opening less that π radians.

8.2. **The Corner-Point Theorem.** *If T is a bounded linear operator on $\mathcal{H}$, and $\lambda \in W(T)$ is a corner point of $W(T)$, then λ is an eigenvalue of T.*

Proof. For the corner point λ we have

$$\lambda = \langle Tf, f\rangle \tag{17}$$

for some unit vector f. We'll show that $Tf = \lambda f$, which will prove the theorem.

Suppose, for the sake of contradiction, this is not the case. Then $g = Tf$ is not a scalar multiple of f (by Eqn.(17), if it were a scalar multiple, that scalar would have to be λ), so the linear subspace $\mathcal{M}$ of $\mathcal{H}$ spanned by f and g is two dimensional. Thus the numerical range of the compression $T_{\mathcal{M}}$ of T to $\mathcal{M}$ is, by Theorem 5.5 a possibly degenerate ellipse containing λ and contained in $W(T)$. Since it is a boundary point of $W(T)$, λ must lie on the boundary of this ellipse. But λ is a *corner point* of $W(T)$, so $W(T)$ cannot contain a non-degenerate ellipse with λ on its boundary, hence $W(T_{\mathcal{M}})$ must be either a line segment with λ as an endpoint, or just the singleton $\{\lambda\}$. In the latter case, $T_{\mathcal{M}}$ is λ times the identity map on $\mathcal{M}$, in which case it's clear that $Tf = \lambda f$, contradicting our assumption that this was not the case. In the former case ($W(T_{\mathcal{M}})$ a line segment with λ as an endpoint) there is another endpoint μ, and by Theorem 5.5 both λ and μ, being foci of the degenerate ellipse (i.e., "endpoints of the line segment $[\mu, \lambda]$"), are eigenvalues of $T_{\mathcal{M}}$, and their corresponding eigenvectors are, by Theorem 5.5, *orthogonal.* It follows that the matrix of $T_{\mathcal{M}}$ with respect to these eigenvectors (now normalized to have unit length) is diagonal, from which it's easy to prove that (17) implies $Tf = \lambda f$. □

8.3. **A word of warning.** It's crucial in the above theorem that the corner point λ already be in the numerical range of T. If λ is merely on the boundary of $W(T)$ then it may well not be an eigenvalue, as the example $T = \operatorname{diag}\{1/n : n = 1, 2, \ldots\}$, acting on ℓ^2, shows. Here $W(T) = (0, 1]$, so 0, which is clearly *not* an eigenvalue, is nevertheless a corner point on the boundary of $W(T)$. The point is, of course, that $0 \notin W(T)$.

The story on corner points does not end here. It turns out that if $\lambda \in W(T)$ is a corner point of $W(T)$, then each T-eigenvector for λ is, in addition, a T^*-eigenvector for $\overline{\lambda}$. This follows from a more general result, also due to Hildebrandt, for which we introduce the following terminology.

8.4. **Definition.** An eigenvalue λ for a linear operator T is said to be *normal* if $\ker\{T - \lambda I\} = \ker\{T^* - \overline{\lambda} I\}$.

The terminology reflects the fact that eigenvalues of normal operators are normal in the sense defined above. The importance of normal eigenvalues derives from the fact that their eigenspaces *reduce* the operator (i.e., both the eigenspace and its orthogonal complement are invariant for the operator). Another important—and easily proved—property of normal operators is that the eigenvectors corresponding to different eigenvalues must be orthogonal. Indeed, if T is a normal operator with eigenvalues $\lambda \neq \mu$ and corresponding eigenvectors v and w, then

$$\lambda\langle v, w\rangle = \langle Tv, w\rangle = \langle v, T^*w\rangle = \langle v, \overline{\mu}w\rangle = \mu\langle v, w\rangle\,.$$

Since $\lambda \neq \mu$ we conclude that $\langle v, w\rangle = 0$, as promised. □

Note that this argument did not require the full strength of normality for T, just that λ be a normal eigenvalue. Thus:

8.5. **Proposition.** *Suppose T is a bounded operator on the Hilbert space $\mathcal{H}$, that λ is a normal eigenvalue of T, and that μ is any eigenvalue of T not equal to λ. Then every T-eigenvector for λ is orthogonal to every T-eigenvector for μ.*

Here is Hildebrandt's theorem about boundary eigenvalues.

8.6. **Theorem (Hildebrandt** [9, 1966]**).** *Every eigenvalue in the boundary of the numerical range is a* normal *eigenvalue.*

Proof. The proof comes from [4, 2002], and is somewhat different from Hildebrandt's. We organize it into two stages, the first of which is an interesting application of the Elliptical Range Theorem.

LEMMA. *If $\|f\| = 1$ and $\langle Tf,\, f\rangle \in \partial W(T)$, then T^*f is a linear combination of f and Tf.*

Proof of Lemma. Suppose $T^*f \notin \operatorname{span}\{f, Tf\}$. We'll prove that λ cannot be in the boundary of $W(T)$. We begin by choosing a unit vector $g \in \mathcal{H}$ orthogonal to f and Tf, but not orthogonal to T^*f. Let $\mathcal{M}$ be the linear span of f and g. Then relative to the orthonormal basis $\{f, g\}$ for $\mathcal{M}$ the matrix of the compression $T_{\mathcal{M}}$ of T to $\mathcal{M}$ is:

$$A := \begin{bmatrix} \langle Tf,\, f\rangle & \langle Tg,\, f\rangle \\ \langle Tf,\, g\rangle & \langle Tg,\, g\rangle \end{bmatrix} = \begin{bmatrix} \lambda & b \\ 0 & d \end{bmatrix},$$

where $b = \langle Tg,\, f\rangle = \langle g,\, T^*f\rangle \neq 0$. By a straightforward calculation, $AA^* \neq A^*A$, i.e., A is not a normal matrix. Consequently A does not possess an orthogonal pair of eigenvectors, so Theorem 5.5 guarantees that $W(A)$ must be a nondegenerate ellipse having λ as a focus. Since $W(A) \subset W(T)$, this implies $\lambda \notin \partial W(T)$, thus completing the proof of the Lemma. □

Proof of Theorem. Suppose λ is an eigenvalue of T that lies on the boundary of $W(T)$. Fix a unit vector $f \in \mathcal{H}$ with $\langle Tf,\, f\rangle = \lambda$. By the Lemma there exist scalars a and b such that

$$T^*f = af + bTf = \underbrace{(a + b\lambda)}_{:=\mu} f = \mu f,$$

so f is an eigenvector of T^*, and moreover

$$\lambda = \langle Tf,\, f\rangle = \langle f,\, T^*f\rangle = \langle f,\, \mu f\rangle = \overline{\mu},$$

hence $T^*f = \overline{\lambda} f$. This shows that $\ker\{T - \lambda I\} \subset \ker\{T - \overline{\lambda} I\}$; the reverse containment follows upon substituting T^* for T. □

Hildebrandt's theorem, along with Theorem 8.2 on corner points, yields:

8.7. **Corollary.** *If $\lambda \in W(T)$ is a corner point of $W(T)$ then λ is a normal eigenvalue of T.*

A small modification of Example 2.9 shows that normal eigenvectors need not always be corner points. Consider the matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix}$$

which represents the operator $T = I \oplus T_2$ on $\mathbb{C}^3 = \mathbb{C} \oplus \mathbb{C}^2$, where T_2 is twice the "finite backward shift" of the first paragraph §2.2, and I is the identity map on $\mathbb{C}$. In S2.2 we saw that this backward shift had numerical range equal to the closed disc of radius 1/2, centered at the origin. Thus $W(T_2)$ is the closed unit disc. By Proposition 2.8 the numerical range of T is the convex hull of that of I (the singleton $\{1\}$) and that of T_2 (the closed unit disc $\overline{U}$). Since 1 lies in the closed unit disc, $W(T) = \overline{U}$. It's easy to check that $1 \in \partial W(T)$ is a normal eigenvalue of T, and it's certainly not a corner point of $W(T)$. □

9. Computing the numerical range

The Toeplitz-Hausdorff Theorem provides an elegant method for "computing" the numerical range of a square complex matrix A.[13] Observe first that the matrix

$$\operatorname{Re} A := \tfrac{1}{2}(A + A^*)$$

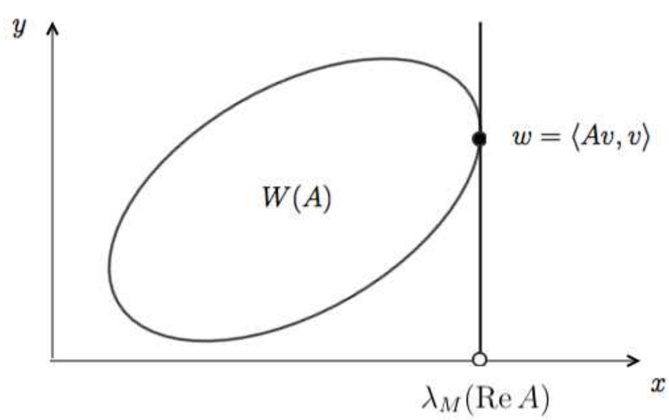
is self-adjoint, so its numerical range:

(a) lies on the real line (the easy part of Theorem 4.1), and
(b) is the convex hull of its eigenvalues (Proposition 2.5)—in this case, the closed interval between the maximum and minimum eigenvalues.

From Proposition 1.1 we see that

$$W(\operatorname{Re} A) = \operatorname{Re} W(A). \tag{18}$$

Let $\lambda_M(\operatorname{Re} A)$ denote largest eigenvalue of $\operatorname{Re} A$; it's the right-most point of the numerical range of $\operatorname{Re} A$, so by (18) above, the vertical line $x = \lambda_M(\operatorname{Re} A)$ intersects the boundary of $W(A)$, and $W(A)$ lies in the closed left half-plane bounded by this line. Moreover, if v is a unit eigenvector of $\operatorname{Re} A$ for the eigenvalue $\lambda_M(\operatorname{Re} A)$, then $w = \langle Av, v\rangle$ is a point of $W(A)$, and it's easy to check that, thanks to the linearity of the inner product in its first variable, $\operatorname{Re} w = \lambda_M(\operatorname{Re} A)$.



[13] The material of this section comes from [11, 1978], see also [10, pp. 33–39].

Now for simplicity, suppose that $W(A)$ is *strictly* convex, meaning that its boundary contains no line segments. Fix a point p in $\partial W(A)$. Since $W(A)$ is compact and convex, there is a line L_p that contains p and for which $W(A)$ lies entirely in one of the half-planes H_p determined by L_p. We call L_p a support line of $W(A)$; the last sentence asserts that $W(A)$ is the intersection of all the half-planes H_p as p runs through $\partial W(A)$. By our strict-convexity assumption, p is the *only* point at which L_p intersects $\partial W(A)$.

Let $e^{i\theta}$ be unit vector normal to L_p and pointing away from $W(A)$. Let $A_\theta = e^{-i\theta}A$; then $W(A_\theta)$ has support line $\operatorname{Re} z = \lambda_M(\operatorname{Re} A_\theta)$, so by the work of the last paragraph, if v_θ is a unit eigenvector for $\operatorname{Re} A_\theta$ corresponding to the eigenvalue $\lambda_M(\operatorname{Re} A_\theta)$, then $w_\theta = \langle A_\theta v,\, v\rangle$ is a point on the boundary of the numerical range of A_θ, and therefore $e^{i\theta}w_\theta$ is a point on L_p that intersects the boundary of $W(A)$. By strict convexity, this point is unique, i.e., must be p. Thus

$$p = e^{i\theta}\langle A_\theta v_\theta,\, v_\theta\rangle = \langle e^{i\theta}e^{-i\theta}Av_\theta,\, v_\theta\rangle = \langle Av_\theta,\, v_\theta\rangle.$$

Conversely, given $\theta \in [0, 2\pi)$ there is a support line for $W(A)$ intersecting its boundary, in exactly one point p_θ, thanks to strict convexity. According to what's just been done, this point is

$$p_\theta = \langle Av_\theta,\, v_\theta\rangle$$

where v_θ is a unit eigenvector (by strict convexity: unique up to $\pm$) for $\operatorname{Re}(e^{-i\theta}A)$. The entire boundary of $W(A)$ is therefore traced out by the points $\langle Av_\theta,\, v_\theta\rangle$ as θ runs through the interval $[0, 2\pi)$.

If $W(A)$ is *not* strictly convex, it will have supporting lines L_p that intersect the boundary in a nontrivial line segment; these will correspond to eigenvalues $\lambda_M\big(\operatorname{Re}(e^{-i\theta}A)\big)$ having eigenspace of dimension > 1. The corresponding line segment in $\partial W(A)$ will then be the convex hull of points $\langle Av,\, v\rangle$ where v runs through a unit-vector basis for the eigenspace of $\lambda_M\big(\operatorname{Re}(e^{-i\theta}A)\big)$.

The representation of the numerical range discussed here can be discretized into an easily implemented algorithm for approximating the numerical range of a matrix. For a discussion of the resulting algorithm and its accuracy, see [11, 1978]. For Matlab implementations see, e.g., [7] or [2], the latter of which also contains a discussion of the method.

10. Some further reading

The reader who finds this material interesting might also wish to consult [16, 2002], a set of introductory notes on the numerical range that places more emphasis on finite dimensional and computational aspects of the subject, and provides a proof of the Toeplitz-Hausdorff theorem closer in spirit to the original one of Hausdorff. Much of the material presented here on matrices comes from Chapter 1 of the book [10, 1991] by Horn and Johnson, which is an excellent source for further results and historical references.

The research literature on numerical ranges and numerical radii is enormous, and ongoing. Here's a micro-survey of just a tiny portion of what's out there.

- In [12, 1951] Rudolph Kippenhahn showed that the numerical range of a matrix is the convex hull of a certain algebraic curve associated with that matrix. Kippenhahn's "boundary generating curve" is still an active topic of research, and his paper is important enough that Zachlin and Hochstenback have provided an English translation [23, 2008].
- There are papers identifying classes of matrices that have specified geometric properties, e.g: circularity, elliptical-ness, boundaries contain flat portions. See, e.g., [17, 2013], [1, 2004], and the references therein.
- Corner-Point Theorem 8.2 is due to Kippenhahn [12] (see also [10, Thm. 1.6.3, pp. 50–51]); it has been generalized, with the notion of "corner point" replaced by "point of non-differentiability" (see [5, 1957]), and even point of "infinite curvature"; see, e.g.: [19, 2000], [20, 2000], and more recently, [6, 2015].
- Another fascinating subject is the question of *mapping theorems* for the numerical range. By the spectral mapping theorem we know that for each Hilbert-space operator T and holomorphic polynomial p (in one variable), $\sigma(p(T)) = p(\sigma(T))$. Is there a similar mapping theorem for the numerical range? The answer is NO, even for 2×2 matrices, but the search for a substitute has inspired much research. Two beautiful recent papers, [13, 2016] and [14, 2017], provide an excellent survey of, and strikingly simpler proofs for, much of what is known about this problem.

References

[1] E. S. Brown and I. M. Spitkovsky, *On matrices with elliptical numerical ranges*, Linear and Multilinear Algebra 52 (2004), 177–193.

[2] C. C. Cowen and E. Harel, *An effective algorithm for computing the numerical range,* Preprint and Matlab program downloadable from: `https://www.math.iupui.edu/~ccowen/Downloads/33NumRange.pdf`

[3] G. De Barra, J. R. Giles, and B. Sims, *On the numerical range of compact operators on Hilbert spaces,* J. London Math. Soc. (2), 5 (1972), 704–706.

[4] P. S. Bourdon and J. H. Shapiro, *When is zero in the numerical range of a composition operator?* Integral Equations and Operator Theory 44 (2002), 410–441.

[5] W. F. Donoghue, *On the numerical range of a bounded operator,* Michigan Mathematical Journal 4 (1957), 261–263.

[6] M. Hansmann, *An observation concerning boundary points of the numerical range*, Operators and Matrices, 9 (2015), 545–548.

[7] N. J. Higham, *The Matrix Computation Toolbox*, downloadable from `http://www.ma.man.ac.uk/~higham/mctoolbox`. The numerical range program is `fv.m`.

[8] F. Hausdorff, *Der Wertvorrat einer Bilinearform*, Math. Z. 3 (1919), 314–316.

[9] S. Hildebrandt, *Über den Numerischen Wertebereich eines Operators*, Math. Ann. 163 (1966), 230–247.

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

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

[12] R. Kippenhahn, *Über den Wertevorrat einer Matrix*, Math. Nachrichten 6 (1951) 193–228.

[13] H. Klaja, J. Mashreghi, and T. Ransford, *On mapping theorems for the numerical range*, Proc. Amer. Math. Soc. 144 (2016), 3009–3016.

[14] H. Klaja, J. Mashreghi, and T. Ransford, *Spectral sets for the numerical range*, preprint, arXiv.org, January 2017.

[15] W. V. Parker, *Sets of complex numbers associated with a matrix,* Duke Math J. 15 (1948), 711–715.

[16] Panayiotis, P. J. and Tsatsomeros, M. J. *Numerical range: (in) a nutshell,* Math. Notes from Washington State University, Vol 45, No. 2, 2002, Vol 46, No. 1, 2003. Available electronically at `www.math.wsu.edu/math/faculty/tsat/index.html`.

[17] L. Rodman and I. M. Spitkovsky, *On numerical ranges of rank-two operators*, Integr. Equ. Oper. Theory 77 (2013), 441–448.

[18] G-C. Rota, *On models for linear operators,* Comm. Pure Appl. Math. 13 (1960), 469–472.

[19] N. Salinas and M. V. Velasco, *Normal essential eigenvalues in the boundary of the numerical range*, Proc. Amer. Math. Soc. 129 (2000), 505–513.

[20] I. M. Spitkovsky, *On the non-round points of the boundary of the numerical range*, Linear and Multilinear Algebra, 47 (2000), 29–33.

[21] O. Toeplitz, *Das algebraische Analogon zu einem Satz von Fejér*, Math. Z. 2 (1918), 187–197.

[22] J. P. Williams, *Similarity and the numerical range*, J. Math. Anal. App. 26 (1969), 307–314.

[23] P. F. Zachlin and M. E. Hochstenbach, *On the numerical range of a matrix*, Linear and Multilinear Algebra 56 (2008), 185–225.

Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, Portland OR 97207

E-mail address: `joels314@gmail.com`