NOTES ON THE NUMERICAL RANGE

JOEL H. SHAPIRO

Abstract. This is an introduction to the notion of numerical range for bounded linear operators on Hilbert space. The main results are: determination of the numerical range for two by two matrices, the Toeplitz-Hausdorff Theorem establishing the convexity of the numerical range for any Hilbert space operator, and a detailed discussion of the relationship between the numerical range and the spectrum. The high points of this latter topic are: containment of the spectrum in the closure of the numerical range, and Hildebrandt's theorem which asserts that the intersection of the closures of the numerical ranges of all operators similar to a given one T is the precisely the convex hull of the spectrum of T.

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 $x \to \langle Tx, x\rangle$ associated with the operator. More precisely,

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

Thus the numerical range of an operator, like the spectrum, is a subset of the complex plane whose geometrical properties should say something about that operator. The goal of these notes is to give some idea of what this "something" might be.

A major theme will be to compare the properties and utility of the numerical range and the spectrum. In §2 we will 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 it is easy to see that only λ times the identity operator can have this set as its numerical range. 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 a standard result in Hilbert space theory asserts that the operator must be Hermitian! See, for example, [1, page 41, Theorem 2]

Date: September 19, 2003.

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) $W(T^*) = \{\overline{\lambda} : \lambda \in W(T)\}$,
(e) $W(I) = \{1\}$. *More generally, if α and β are complex numbers, and T a bounded operator on $\mathcal{H}$, then $W(\alpha T + \beta I) = \alpha W(T) + \beta$.*
(f) *If $\mathcal{H}$ is finite dimensional then $W(T)$ is compact.*

The last fact follows from the compactness of the unit sphere of $\mathcal{H}$ and the continuity of the quadratic form associated with T. In §2 we'll give examples which show that, by contrast, when $\mathcal{H}$ is infinite dimensional it supports bounded operators with non-closed numerical range.

It is part (c) that gets us started on the major theme of these notes. We will prove in §5 that the spectrum of an operator lies in the *closure* of its numerical range, and in §4 we'll prove the most famous result in the subject: the *Toeplitz-Hausdorff Theorem* [7],[2], which asserts that the numerical range is always *convex*. Thus the *convex hull* of the spectrum of an operator lies in the closure of the numerical range. By the similarity-invariance of the spectrum, its convex hull must, in fact, lie in the intersection of the closures of the numerical ranges of all the operators similar to T. This leads up to the beautiful theorem of Stephan Hildebrandt which asserts that this intersection is *precisely* the convex hull of the spectrum.

Along the way to proving the Toeplitz-Hausdorff Theorem we will obtain a complete description of the numerical ranges of two by two matrices; they are (possibly degenerate) elliptical discs with foci at the eigenvalues of the matrix.

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 $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$. 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:

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

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. $\square$

2.2. **Non-similarity invariance of the numerical range.** 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 is 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 called a "backward shift."

This two dimensional backward shift dramatically illustrates the non-similarity invariance of the numerical range. For this, let T_λ be the operator associated with the matrix $\begin{bmatrix} 0 & \lambda \\ 0 & 0 \end{bmatrix}$, so T_1 is the matrix of the "backward shift" we considered in §2.1, and $T_\lambda = \lambda T_1$. Thus $W(T_\lambda) = \lambda W(T_1)$, the closed disc of radius $|\lambda|$ centered at the origin, so all the different operators T_λ have different numerical ranges. But for $\lambda \neq 0$ *all these operators are similar*. Indeed, $S_\lambda := \begin{bmatrix} \lambda & 0 \\ 0 & 1 \end{bmatrix}$ is nonsingular and $S_\lambda T_1 S_\lambda^{-1} = T_\lambda$.

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:

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

(Note that B is the infinite dimensional version of the example of part (a)). 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

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

and therefore would have 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, with the scalar in question necessarily being λ.

Now an easy calculation shows that the only sequence x with $Bx = \lambda x$ is $x = x_\lambda$ which, because $|\lambda| = 1$ is not in ℓ^2. This contradiction shows that λ cannot belong to the numerical range of B. $\square$

Note that, relative to the standard unit-vector basis of ℓ^2, the backward shift B has matrix that is zero everywhere except on the first superdiagonal, where its entries are all 1.

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 $\{e_n\}$ for $\mathcal{H}$ consisting of eigenvectors of T. All normal operators on a finite dimensional space, and more generally, all compact normal operators, are unitarily diagonalizable.

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

Proof. Let T be the operator and $\mathcal{H}$ the Hilbert space on which it acts. 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. Thus

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

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,* $\operatorname{conv}_\infty(\Lambda) = \operatorname{conv}(\Lambda)$.

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

$$\operatorname{conv}(\alpha\Lambda + \beta) = \alpha \operatorname{conv}(\Lambda) + \beta \qquad \forall \alpha, \beta \in \mathbb{C}, \tag{2}$$

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 line that separates 0 from $\operatorname{conv}(\Lambda)$, so by a rotation about the origin (again using (2)) 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 $0 \in \operatorname{conv}_\infty(\Lambda)$, i.e. that there exist numbers a_n between 0 and 1 such that $0 = \sum_{n=0}^{\infty} a_n \lambda_n$. We may as well assume infinitely many of the a_n are nonzero, else trivially $0 \in \operatorname{conv}(\Lambda)$. Now $0 = \sum_{n=0}^{\infty} a_n \operatorname{Im}(\lambda_n)$, 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. □

This gives us another way to produce examples of non-closed numerical ranges: 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 $\operatorname{diag}\{\lambda_n\}$. Then the numerical range of T is the half-open interval $(0, \lambda_0]$. (Admittedly, we don't need the Theorem to prove this simple fact).

3. The numerical range of a two by two matrix

In this section we prove that the numerical range of a two by two matrix (i.e. an operator on a two dimensional Hilbert space) assumes one of the following three forms:

(a) A single point, if the operator is a scalar multiple of the identity,

(b) a line segment joining the eigenvalues, if the operator is normal with two distinct eigenvalues, or

(c) an elliptical disc with foci at the eigenvalues, if the operator has distinct eigenvalues, but is not normal.

In other words:

> *The numerical range of an operator on a two dimensional Hilbert space is a (possibly degenerate) elliptical disc with foci at the eigenvalues.*

Part (a) is trivial. For (b) note that for finite square matrices "normal" means "unitarily diagonalizable" so the result follows from the work of §2.4; $W(T)$ is the line segment joining the eigenvalues. For part (c) we just worked out a special case, namely:

$$W\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = \tfrac{1}{2}\overline{\mathbb{U}}.$$

More generally, Schur's theorem asserts that any square matrix is unitarily equivalent to an upper triangular one, so if T is a two by two matrix with just one eigenvalue λ, then it's unitarily equivalent to a matrix of the form $\lambda I + \mu N$, with $N = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, the matrix just discussed above. Thus by Proposition 1.1(e), $W(T) = \lambda + \mu\, W(N)$, the closed disc of radius $|\mu|/2$, centered at λ.

In order to get serious about part (c) we need to work out a more general class of examples. Let us call a complex matrix with entries zero everywhere except possibly except on the major cross-diagonal a "cross-diagonal" matrix.

3.1. **Proposition.** *The numerical range of a two by two cross-diagonal matrix is either an elliptical disc with foci at the eigenvalues, or a line segment joining the eigenvalues.*

Proof. We have $T = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}$ where a and b are complex numbers. First suppose a and b are positive. We may assume $0 < b \leq a$, else take adjoints and use Proposition 1.1(d). Then employing the parameterization (1) and doing some computation:

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

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

- The line segment $[-a, a]$ if $a = b$ (in which case $\pm a$ are the eigenvalues of T),or
- The ellipse with center at the origin, horizontal major axis of length $a+b$ and vertical minor axis of length $a - b$ if $a \neq b$.

Now from analytic geometry we know that for an ellipse with foci $\pm F$ on the real axis, major semi-axis of length M and minor semi-axis of length m, we have $F^2 + m^2 = M^2$. 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 $W(T)$ are the eigenvalues of T.

To summarize our work to this point:

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

Precisely the same result holds if a and b are arbitrary complex numbers. Indeed, write both in polar form: $a = |a|e^{i\alpha}$ and $b = |b|e^{i\beta}$, and observe that

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

From Proposition 1.1(e) and the result just proved for non-negative a and b we see that $W(T)$ is an ellipse with foci at

$$\pm\sqrt{|a||b|}\, e^{i\frac{\alpha+\beta}{2}} = \pm\sqrt{ab} = \text{ the eigenvalues of } T.$$

Taking into account the lengths of the axes of our ellipse, we can summarize the work to this point as follows:

3.2. **Proposition.** *If* $T = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}$ *where* $a, b \in \mathbb{C}$*, then* $W(T)$ *is the (possibly degenerate) ellipse with center at the origin, whose foci are the eigenvalues* $\pm\sqrt{ab}$ *of* T*, whose major axis has length* $|a| + |b|$ *and whose minor axis has length* $||a| - |b||$*.*

We will return to the discussion of the dimensions of $W(T)$ in terms of quantities intrinsic to the operator T before long. However right now let's complete the characterization of numerical ranges of operators on two dimensional Hilbert space.

3.3. **The Ellipse Theorem.** *If* T *is a linear transformation on* $\mathbb{C}^2$*, then* $W(T)$ *is a (possibly degenerate) elliptical disc.*

Proof. It is enough to consider T with trace zero (else replace T by $T - (\operatorname{trace} T/2)I$, and use the transformation law (c) of Proposition 1.1). In view of what we've done with cross diagonal matrices, the Ellipse Theorem will follow immediately from the following result, which is interesting in its own right:

> *If* T *is a two by two complex matrix with trace zero then* T *is unitarily equivalent to a matrix with zero-diagonal.*

For the proof, suppose first that T has only one eigenvalue λ. As in the discussion at the beginning of §3, we may (thanks to Schur's Theorem), assume that T is upper triangular with both diagonal elements equal to λ. But since T has trace zero, we must have $\lambda = 0$, hence T is unitarily equivalent to an upper triangular matrix with zero-diagonal, i.e. to a matrix of the type considered in §2.2: $\begin{bmatrix} 0 & a \\ 0 & 0 \end{bmatrix}$. Thus if T has only one eigenvalue its numerical range is a closed disc (degenerate in case T is a scalar multiple of the identity) centered at that eigenvalue.

Next, suppose T has two distinct eigenvalues, necessarily $\pm\lambda \neq 0$. Let v be a unit eigenvector for λ_1, and w a unit eigenvector for $-\lambda_1$. If $v \perp w$ then T is diagonalizable hence by Theorem 2.5 $W(T)$ is the line segment joining the eigenvalues.

Suppose v is not orthogonal to w. Then for each $\theta \in \mathbb{R}$ set $z_\theta = v + e^{i\theta}w$, so that $\langle Tz_\theta, z_\theta\rangle = 2i\lambda \operatorname{Im}\{e^{-i\theta}\langle z, w\rangle\}$. Upon setting θ equal to any value of the argument of the nonzero complex number $\langle z, w\rangle$, we see that $\langle Tz_\theta, z_\theta\rangle = 0$. Now clearly $z_\theta \neq 0$ (by the linear independence of v and w); set $w = z_\theta/\|z_\theta\|$, so w is a unit vector with $\langle Tw, w\rangle = 0$. Choose u any unit vector in $\mathbb{C}^2$ that is orthogonal to w. Then $\{v, u\}$ is an orthonormal basis for $\mathbb{C}^2$, relative to which the matrix of T has $(1,1)$ entry zero. But since this matrix must have trace zero, the $(2,2)$ entry is also zero, hence the diagonal is identically zero, as desired. This completes the proof of the Ellipse Theorem. $\square$

3.4. **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, one wonders if there is some kind of intrinsic expression for the lengths of the major and minor axes. This is indeed the case.

Let's restrict attention to matrices whose numerical ranges are 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 $T = \begin{bmatrix} 0 & a \\ b & 0 \end{bmatrix}$ 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 one calculates easily that unit eigenvectors for λ_1 and λ_2 respectively are:

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

It follows that

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

is the reciprocal of the eccentricity of $W(T)$, and

$$\sqrt{1-\gamma^2} = \frac{2\sqrt{ab}}{a+b} = \frac{2(\lambda_1 - \lambda_2)}{\text{length of major axis}}.$$

Thus:

$$\begin{aligned} \text{length of major axis} &= \frac{2(\lambda_1-\lambda_2)}{\sqrt{1-\gamma^2}}, \quad \text{and} \\ \text{length of minor axis} &= \gamma \times \text{length of major axis}. \end{aligned}$$

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 the eigenvectors and eigenvalues. We have also shown that *every* cross-diagonal two by two matrix is unitarily similar to a unimodular multiple 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:

3.5. **Numerical range of a two by two matrix: the full story.** *Suppose T 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) *$W(T)$ is a (possibly degenerate) elliptical disc with foci at λ_1 and λ_2.*
(b) *The eccentricity of $W(T)$ is $\frac{1}{\gamma}$.*
(c) *The major axis of $W(T)$ has length $\frac{2|\lambda_1-\lambda_2|}{\sqrt{1-\gamma^2}}$.*

Thus, given the eigenvalues, one increases the size of the numerical range by making similarity transformations that decrease the angle between the eigenvectors.

4. The Toeplitz-Hausdorff Theorem [7],[2]

4.1. **Theorem.** *The numerical range of every bounded linear operator T on a Hilbert space is convex.*

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 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,

> *The numerical range of a bounded linear operator on a Hilbert space contains the numerical ranges of all 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 two dimensional results, 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, as desired. $\square$

The key step in our proof of the Toeplitz-Hausdorff Theorem was the observation that every two by two complex matrix with trace zero is unitarily equivalent to one with diagonal identically zero. We close this section by showing how the Toeplitz-Hausdorff Theorem provides the n-dimensional generalization due to W. V. Parker [5].

4.2. **Theorem.** *Suppose A is an n by n matrix with complex entries whose trace is zero. Then A is unitarily equivalent to a matrix with diagonal identically zero.*

Proof. The trace of A is the sum of the eigenvalues of A, each eigenvalue being repeated in the sum as many times as its algebraic multiplicity—its multiplicity as a root of the characteristic polynomial (this statement is obvious for upper triangular matrices, and via Schur's theorem, any square matrix is unitarily equivalent to one that is upper triangular). Thus $0 = \text{trace}\, A/n$ 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 $\begin{bmatrix} 0 & * \\ * & A_1 \end{bmatrix}$, where $A_1 = A_M$ is the compression of A to M, 0 is a one by one matrix, and the two matrices "*" have dimensions $n-1$ by one and one by $n-1$. The argument can now be repeated on A_1, producing a vector u_2 orthogonal to u_1 with $0 = \langle A_1u_2, u_2\rangle = \langle Au_2, u_2\rangle$. 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. $\square$

Eigenvalues and "bad" boundary points. For a bounded operator T on a Hilbert space, let's say the boundary of $W(T)$ has "infinite curvature" at one of its points λ if there's no closed disc lying in $W(T)$ that contains λ. For example any "corner point" of the boundary is a point of infinite curvature.

4.3. **Folk Theorem.** *If $\lambda \in W(T)$ is a boundary point at which $\partial W(T)$ has infinite curvature, then λ is an eigenvalue of T.*

Proof. We have $\lambda = \langle Tf, f\rangle$ for some unit vector f. Let g be any unit vector orthogonal to f and consider the subspace $\mathcal{M}$ spanned by f and g. As in our proof of the Toeplitz-Hausdorff Theorem, the numerical range of the compression $T_{\mathcal{M}}$ of T to $\mathcal{M}$ is, by the work of §3.5, a possibly degenerate ellipse with λ on the boundary (because $\lambda \in \partial W(T)$). But $W(T_{\mathcal{M}}) \subset W(T)$, and since $W(T)$ contains no closed disc containing λ, it contains no closed (nondegenerate) elliptical disc containing λ. Thus $W(T_{\mathcal{M}})$ has to be a line segment with λ as an endpoint, from which it follows via the results of §3.5 that λ is an eigenvalue of $T_{\mathcal{M}}$, hence of T. □

5. The numerical range and the spectrum

We have observed that the numerical range of an operator contains all its eigenvalues. What about the spectrum? Since the spectrum is closed and the numerical range need not (if $\mathcal{H}$ is infinite dimensional) be closed, we can't expect the spectrum to lie in the numerical range, as illustrated by the 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 its closure, $[0, 1]$. Nevertheless, the next-best thing happens:

5.1. **Theorem.** *If T is a bounded linear operator on a Hilbert space $\mathcal{H}$, then the spectrum of T is contained in the* closure *of the numerical range of 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) If T is bounded below but not onto, then $\{0\} \neq (\operatorname{ran} T)^\perp = \ker T^*$, hence $0 \in W(T^*)$, and therefore $0 \in W(T)$. □

5.2. **Remark.** This proof shows that $\lambda \in \sigma(T)$ does *not* belong to $W(T)$ if and only if $T - \lambda I$ is one-to-one with range that is dense, but not closed.

5.3. **Remark.** Because of the Toeplitz-Hausdorff Theorem and the similarity-invariance of the spectrum, we see that

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

In the next section we will see that the set containment is actually *equality.* Thus the numerical range can serve as a device for locating the spectrum.

6. Hildebrandt's Theorem

In this section we prove the beautiful result of Stephan Hildebrandt (1985) asserting that for a bounded linear operator T on a Hilbert space, you get the convex hull of $\sigma(T)$ by intersecting the closures of the numerical ranges of all the operators similar to T [3].

We present a strikingly short proof to James Williams [8], which requires a result due to Gian-Carlo Rota [6] asserting 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 with entries in $\mathcal{H}$ that have square-summable norms. The backward shift B is defined on $\ell^2(\mathcal{H})$ in the usual way:

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

it is a bounded linear operator on $\ell^2(\mathcal{H})$ of norm one. The theorem of Rota can be stated precisely as follows:

6.1. **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 $x \in \mathcal{H}$, $\|Wx\|^2 \leq \|x\|^2 \sum \|T^n\|^2 = M\|x\|^2$, 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 we give an application of Rota's theorem that makes more precise the relationship between the norm and the spectral radius. We all know that $r(T) \leq \|T\|$ for any operator T on a Hilbert space, and that the spectral radius is a similarity invariant. Thus $r(T) \leq \inf\{\|VTV^{-1}\| : V \text{ invertible on } \mathcal{H}\}$, i.e., Thus the spectral radius behaves relative to the norm in a way that suggests the connection between the spectrum and the numerical range. The next result asserts that there is a Hildebrandt-type result for this situation.

6.2. **Corollary.** *If T is a bounded linear operator on a Hilbert space $\mathcal{H}$, then $r(T) = \inf\{\|VTV^{-1}\| : V \text{ invertible on } \mathcal{H}\}$.*

Proof. From the previous discussion we need only prove the inequality "$\geq$". In order to use Rota's theorem we need an operator of spectral radius < 1, so for each $\varepsilon > 0$ 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 $WT_\varepsilon T^{-1}$. 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 $\leq r(T)$, which is the desired result. □

Finally we can prove the main result of this section .

6.3. **Hildebrandt's Theorem.** *For every bounded linear operator on a Hilbert space, the convex hull of the 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 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, i.e., that there is an invertible operator V on $\mathcal{H}$ such that $\lambda \notin \overline{W}(V^{-1}TV)$. Because 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 6.2 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 $\lambda \notin \overline{W}(V^{-1}TV)$. $\square$

References

[1] P. R. Halmos, *Introduction to Hilbert Space,* Chelsea 1951.
[2] F. Hausdorff, *Der Wertvorrat einer Bilinearform*, Math. Z. 3 (1919), 314–316.
[3] S. Hildebrandt, *Über den Numerischen Wertebereich eines Operators*, Math. Ann. 163 (1966), 230–247.
[4] R. A. Horn and C. R. Johnson,*Topics in Matrix Analysis*, Cambridge Univ. Press 1991.
[5] W. V. Parker, *Sets of complex numbers associated with a matrix,* Duke Math J. 15 (1948), 711–715.
[6] G-C. Rota, *On models for linear operators,* Comm. Pure Appl. Math. 13 (1960), 469–472.
[7] O. Toeplitz, *Das algebraische Analogon zu einem Satz von Fejér*, Math. Z. 2 (1918), 187–197.
[8] J. P. Williams, *Similarity and the numerical range*, J. Math. Anal. App. 26 (1969), 307–314.

Michigan State University, East Lansing, MI 48824-1027, USA
E-mail address: shapiro@math.msu.edu