The shell of a Hilbert-space operator

By CHANDLER DAVIS in Toronto (Canada)

For an arbitrary closed linear operator in Hilbert space, I will define a subset of real 3-space which summarizes much information about it: its point spectrum, its numerical range, many of its spectral sets, and more besides.

1. Notations and principal ideas

Let $\mathfrak{H}$ be a complex Hilbert space. For any $x\in\mathfrak{H}$, let the corresponding linear functional be x^*; thus y^*x is the inner product of x by y.

Let $\mathfrak{A}$ be a closed linear relation in $\mathfrak{H}$; that is, a closed linear subspace of $\mathfrak{H}_1\oplus\mathfrak{H}_2$, where each $\mathfrak{H}_i$ is a replica of $\mathfrak{H}$. (No distinction is to be made between $\mathfrak{H}_1$ and $\mathfrak{H}_1\oplus\{0\}$, or between $\mathfrak{H}_2$ and $\{0\}\oplus\mathfrak{H}_2$.) The most important case is that of an operator A, i.e. when $(y, x)\in\mathfrak{A}$ means that $y=Ax$. The 'domain' of $\mathfrak{A}$ is $\{x\colon (\exists y)(y, x)\in\mathfrak{A}\}$, its 'range' is $\{y\colon (\exists x)(y, x)\in\mathfrak{A}\}$. (This reversal of the customary order in the notation for relations will save me, in § 5, from having to reverse order in a more troublesome way.)

Before giving the novel ideas I must also fix the notations for stereographic projection. Let $\boldsymbol{C}$ denote the complex plane and $\overline{\boldsymbol{C}}=\boldsymbol{C}\cup\{\infty\}$. Define $\tau\colon \overline{\boldsymbol{C}}\to\boldsymbol{R}^3$ by

$$\tau(z)=\left(\frac{2z}{1+|z|^2}, \frac{-1+|z|^2}{1+|z|^2}\right) \quad (z\in\boldsymbol{C}), \quad \text{and} \quad \tau(\infty)=(0,1). \tag{1.1}$$

The first two co-ordinates in $\boldsymbol{R}^3$ are here collapsed into a single complex number; this will be done frequently throughout. In this notation, the Riemann sphere $\tau(\overline{\boldsymbol{C}})$ is the unit sphere $S=\{(\zeta, h)\in\boldsymbol{R}^3\colon |\zeta|^2+h^2=1\}$. Letting $B=\{(\zeta, h)\in\boldsymbol{R}^3\colon |\zeta|^2+h^2\leqq 1\}$, the unit ball, define $\pi\colon B\to\overline{\boldsymbol{C}}$ by taking $\pi(\zeta, h)$ to be that $z\in\boldsymbol{C}$ for which (ζ, h) is on the line joining $(z, 0)$ to $(0, 1)$; explicitly,

$$\pi(\zeta, h)=\frac{\zeta}{1-h} \quad (h\neq 1), \quad \text{and} \quad \pi(0,1)=\infty. \tag{1.2}$$

For any subset $E \subseteq B$, $\pi(E)$ will be called the 'shadow' of E. On S, π coincides with τ^{-1}.

Definition 1.1. The 'shell' of the relation $\mathfrak{A}$, denoted $s(\mathfrak{A})$, is defined as the set of points

$$\varphi(y, x) = \left(\frac{2x^* y}{\|x\|^2 + \|y\|^2}, \frac{-\|x\|^2 + \|y\|^2}{\|x\|^2 + \|y\|^2} \right) \tag{1.3}$$

in $\boldsymbol{R}^3$, where (y, x) runs over all non-zero elements of $\mathfrak{A}$.

In case of an operator, I will write alternatively $s(A)$. This is the case where $(0, 1)$ does not belong to the set. $s(\mathfrak{A})$ is void if and only if $\mathfrak{A} = \{(0, 0)\}$.

One finds by direct computation that $s(\mathfrak{A}) \subseteq B$. Further geometric properties of the set, in relation to spectral properties of $\mathfrak{A}$, will be developed below, especially in §§ 2—3. Examples are treated in § 4. In § 5, I will state the essential facts on the transformation of $s(\mathfrak{A})$ when $\mathfrak{A}$ is subjected to a Möbius transformation.

Next, consider spectral sets [16], [17, § 154].

Let us adopt the following terminology:

(i) a set of the form $\{z \in \boldsymbol{C}: |z - z_0| \leqq r\}$ $(z_0 \in \boldsymbol{C}, r > 0)$ will be called a 'finite disk';

(ii) a set of the form $\{z \in \boldsymbol{C}: |z - z_0| \geqq r\} \cup \{\infty\}$ $(z_0 \in \boldsymbol{C}, r > 0)$ will be called a 'complementary disk';

(iii) a set of the form $\{z \in \boldsymbol{C}: \operatorname{Re}(\bar{\zeta} z) \geqq a\} \cup \{\infty\}$ $(|\zeta| = 1, a \in \boldsymbol{R})$ will be called a 'half-plane';

(iv) a set of any of the types (i)—(iii) will be called a 'disk'. Thus the disks are exactly those subsets X of $\overline{\boldsymbol{C}}$ for which $\tau(X)$ is a proper spherical cap.

The main result of the paper, Theorem 7.2, may be stated roughly as follows: A disk X is a spectral set for A if and only if $s(A)$ is contained in the convex hull of $\tau(X)$. That is, the support planes of (the convex closure of) $s(A)$ correspond naturally one-one to the minimal disk spectral sets of A. This is exploited in § 8 to give a description in terms of the shell of the operator classes occurring in [22].

In order to formulate my results for arbitrary closed linear relations, I had to supplement the basic results of the paper [1] [1]) of ARENS with a study of the spectrum (§ 2, below) and of the rational functional calculus (§ 6).

My grateful acknowledgement is due to H. S. M. COXETER, C. FOIAŞ, G. KALISCH, and C. A. MCCARTHY, for stimulating conversations.

[1]) The reader is warned of the uncommonly pesky misprints in the article [1]. Professor ARENS has pointed out also that Theorem 3.7 of [1] is not true in quite the generality claimed.

2. The shell and the spectrum

It has already been remarked that $s(\mathfrak{A}) \subseteq B$. The shell may have points on the boundary, the sphere S; this depends on spectral properties of $\mathfrak{A}$, see Theorem 2. 2 below. I define the spectrum in a form convenient for the purpose; cf. [13, § 2. 16], [1, § 2]. (In the following definitions, recall that it is $\mathfrak{H}_2$ which contains the domain.)

Definition 2. 1. $0 \in \sigma_p(\mathfrak{A})$, the 'point spectrum' of $\mathfrak{A}$, in case $\mathfrak{A} \cap \mathfrak{H}_2 \neq \{(0, 0)\}$. The 'null-space' $\mathfrak{N}(\mathfrak{A})$ is the set of $x \in \mathfrak{H}$ such that $(0, x) \in \mathfrak{A}$, thus $0 \in \sigma_p(\mathfrak{A})$ if and only if $\mathfrak{N}(\mathfrak{A}) \neq \{0\}$.

Definition 2. 2. $0 \in \sigma_c(\mathfrak{A})$, the 'continuous spectrum' of $\mathfrak{A}$, in case there exist $x_n \in \mathfrak{H} \ominus \mathfrak{N}(\mathfrak{A})$, $y_n \in \mathfrak{H}$, such that $(y_n, x_n) \in \mathfrak{A}$, $\|x_n\| = 1$, and $y_n \to 0$.

Definition 2. 3. The 'range' $\mathfrak{R}(\mathfrak{A})$ is the set of $y \in \mathfrak{H}$ such that for some $x \in \mathfrak{H}$, $(y, x) \in \mathfrak{A}$. $0 \in \sigma_r(\mathfrak{A})$, the 'residual spectrum' of $\mathfrak{A}$, in case $\overline{\mathfrak{R}(\mathfrak{A})} \neq \mathfrak{H}$.

According to these definitions, 0 may belong to any one of σ_p. σ_c, σ_r, independently of whether it belongs to the others.

Let $\mathfrak{I}$ denote the identity relation: $(y, x) \in \mathfrak{I}$ if and only if $x = y$.

Definition 2. 4. For $z \in \boldsymbol{C}$, $z \in \sigma_p(\mathfrak{A})$ if and only if $0 \in \sigma_p(\mathfrak{A} - z\mathfrak{I})$, and similarly for σ_c, σ_p. $\infty \in \sigma_p(\mathfrak{A})$ if and only if $0 \in \sigma_p(\mathfrak{A}^{-1})$, and similarly for σ_c, σ_r.

Thus $\infty \in \sigma_r(\mathfrak{A})$ if and only if the domain $\mathfrak{D}(\mathfrak{A})$ is not dense.

Definition 2. 5. The 'approximate point spectrum' $\sigma_\pi(\mathfrak{A})$ is $\sigma_p(\mathfrak{A}) \cup \sigma_c(\mathfrak{A})$. The 'approximate residual spectrum' $\sigma_\varrho(\mathfrak{A})$ is $\sigma_c(\mathfrak{A}) \cup \sigma_r(\mathfrak{A})$. The 'spectrum' $\sigma(\mathfrak{A})$ is $\sigma_p(\mathfrak{A}) \cup \sigma_c(\mathfrak{A}) \cup \sigma_r(\mathfrak{A})$.

Proposition 2. 1. *$0 \in \sigma_\pi(\mathfrak{A})$ if and only if there exist $x_n, y_n \in \mathfrak{H}$ such that $(y_n, x_n) \in \mathfrak{A}$, $\|x_n\| = 1$, and $y_n \to 0$. $0 \in \sigma_\varrho(\mathfrak{A})$ if and only if $\mathfrak{R}(\mathfrak{A}) \neq \mathfrak{H}$.*

The first property is the familiar justification for the term 'approximate point spectrum', cf. [10]. The second has no counterpart under the usual definitions.

Proposition 2. 2. *$\sigma_\pi(\mathfrak{A})$ is closed.*

I give only the key step in the (familiar) proof: suppose we have chosen, as we may if $0 \in \overline{\sigma_\pi(\mathfrak{A})}$, numbers $z_n \in \sigma_\pi(\mathfrak{A})$ with $z_n \to 0$, and elements $(y_n, x_n) \in \mathfrak{A} - z_n\mathfrak{I}$ with $\|x_n\| = 1$ and $\|y_n\| \to 0$. Then $0 \in \sigma_\pi(\mathfrak{A})$ is established by using, as the sequence in Prop. 2. 1, $(y_n + z_n x_n, x_n) \in \mathfrak{A}$.

Definition 2. 6. The 'adjoint' $\mathfrak{A}^*$ of a relation $\mathfrak{A}$ is $(-\mathfrak{U}^{-1})^\perp$; that is, $(w, z) \in \mathfrak{A}^*$ if and only if, for all $(y, x) \in \mathfrak{A}$, $w^* x = z^* y$.

It follows easily that $(a\mathfrak{A} + b\mathfrak{I})^* = \bar{a}\mathfrak{A}^* + \bar{b}\mathfrak{I}$.

The following property of σ_p, σ_c, σ_r may justify the peculiar way I have defined the types of spectral point.

Theorem 2. 1. *$z\in\sigma_p(\mathfrak{A})$ if and only if $\bar{z}\in\sigma_r(\mathfrak{A}^*)$. $z\in\sigma_c(\mathfrak{A})$ if and only if $\bar{z}\in\sigma_c(\mathfrak{A}^*)$. Similar assertions hold for ∞ in place of z.*

Proof. The first statement follows at once from the definitions. As for ∞, it presents no special problem: merely consider $\mathfrak{A}^{-1}$ in place of $\mathfrak{A}$. It remains to prove the assertion concerning $z\in\sigma_c$, and we may evidently take $z=0$. The proof is essentially familiar.

Let P denote the projector of $\mathfrak{H}_1\oplus\mathfrak{H}_2$ onto $\mathfrak{H}_2$, and let Q denote the projector of $\mathfrak{H}_1\oplus\mathfrak{H}_2$ onto $\mathfrak{A}$. The pairs (y_n, x_n) occurring in Definition 2. 2 will be sums of $(y_n, 0)\in\mathfrak{H}_1\perp\mathfrak{H}_2$ and $(0, x_n)\in\mathfrak{H}_2$; because $x_n\perp\mathfrak{N}(\mathfrak{A})$, we have also $(0, x_n)\perp\mathfrak{H}_2\cap\mathfrak{A}$; hence $(y_n, x_n)\perp\mathfrak{H}_2\cap\mathfrak{A}$. Also it is clear that both (y_n, x_n) and $(0, x_n)$ are orthogonal to $\mathfrak{H}_1\cap\mathfrak{A}^\perp$. We now confine attention for the moment to a certain subspace $\mathfrak{K}$ of $\mathfrak{H}_1\oplus\mathfrak{H}_2$, namely,

$$\mathfrak{K}=(\mathfrak{H}_2\cap\mathfrak{A}\oplus\mathfrak{H}_1\cap\mathfrak{A}^\perp)^\perp=\mathfrak{N}(P-Q)^\perp. \tag{2. 1}$$

It reduces both P and Q.

We have seen that $(y_n, x_n)\in Q\mathfrak{K}$ and $(0, x_n)\in P\mathfrak{K}$, though both of norm $\geqq 1$, satisfy $\|(y_n, x_n)-(0, x_n)\|=\|y_n\|\to 0$. It is required to prove the existence of w_n, z_n such that $(z_n, w_n)\in(1-Q)\mathfrak{K}$ and $(z_n, 0)\in(1-P)\mathfrak{K}$, both (z_n, w_n) and $(z_n, 0)$ have norm $\geqq 1$, and yet they satisfy $\|(z_n, w_n)-(z_n, 0)\|=\|w_n\|\to 0$; for then the sequence of pairs $(-w_n, z_n)\in\mathfrak{A}^*$ will show $0\in\sigma_c(\mathfrak{A}^*)$ from Def. 2. 2.

To prove this, define, as in [7] and [14, I. 4. 6 and I. 6. 8],

$$C=(P+Q-1)^2,\quad S=(P-Q)^2$$

(commuting operators $\geqq 0$). We are already restricted (see (2. 1)) to $\mathfrak{K}=\mathfrak{N}(S)^\perp$, and it is easy to see that there is no loss in restricting further to $\mathfrak{L}=\mathfrak{K}\ominus\mathfrak{N}(C)$. $(\mathfrak{N}(C)=\mathfrak{H}_1\cap\mathfrak{A}\oplus\mathfrak{H}_2\cap\mathfrak{A}^\perp.)$ Then

$$(CS)^{-\frac{1}{4}}(QP-PQ) \tag{2. 2}$$

turns out to be a unitary operator $\mathfrak{L}\to\mathfrak{L}$ taking $P\mathfrak{L}$ to $(1-P)\mathfrak{L}$ and QL to $(1-Q)\mathfrak{L}$. Therefore we are able to specify (z_n, w_n) and $(z_n, 0)$ having the properties desired: the images under (2. 2) of (y_n, x_n) and $(0, x_n)$, respectively.

The proof is complete.

Proposition 2. 3. *$\sigma(\mathfrak{A})$ is closed.*

This is immediate from Prop. 2. 2 and Thm. 2. 1.

The main relevance of mentioning the duality between point and residual spectra in the present connection is that, as will now be explained, only σ_π is involved n matters concerning the shell.

Theorem 2. 2. *The set $S\cap s(\mathfrak{A})$ consists exactly of the image of $\sigma_p(\mathfrak{A})$ under the stereographic projection τ.*

Indeed, for C this follows easily from (1. 1) and (1. 3). As for ∞, $\infty \in \sigma_p(\mathfrak{A})$ is equivalent to $(y, 0) \in \mathfrak{A}$ for some non-zero y, which has already been remarked to be equivalent to $(0, 1) \in s(\mathfrak{A})$.

Theorem 2. 3. *The set* $S \cap \overline{s(\mathfrak{A})}$ *consists exactly of the image of* $\sigma_\pi(\mathfrak{A})$ *under the stereographic projection* τ.

Proof. Part I. Let $z \in C \cap \sigma_c(\mathfrak{A})$; it is to be proved that $\tau(z) \in \overline{s(\mathfrak{A})}$. We can choose $(y_n, x_n) \in \mathfrak{A}$ such that $\|x_n\| = 1$ and $y_n - z x_n \to 0$. For such a sequence, it is easy to compute that $\varphi(y_n, x_n) \to \tau(z)$.

Let $\infty \in \sigma_c(\mathfrak{A})$; it is to be proved that $\tau(\infty) \in \overline{s(\mathfrak{A})}$. The above paragraph (for $z = 0$) gives a proof by exchanging x with y in (1. 3).

Part II. Assume $z \in C$, $\tau(z) \in \overline{s(\mathfrak{A})}$. This means that there exist $(y_n, x_n) \in \mathfrak{A}$ such that $\varphi(y_n, x_n) \to \tau(z)$.

If $\|x_n\| \neq o(\|y_n\|)$, there is no loss of generality in assuming that no x_n is 0; replacing (y_n, x_n) by $(y_n/\|x_n\|, x_n/\|x_n\|)$ gives still a point of $\mathfrak{A}$, having the same image under φ, so in this case there is no loss of generality in assuming that $\|x_n\| = 1$. Now

$$\frac{-1 + \|y_n\|^2}{1 + \|y_n\|^2} \to \frac{-1 + |z|^2}{1 + |z|^2}$$

implies $\|y_n\| \to |z|$; then also $x_n^* y_n \to z$. This implies that $\|y_n - z x_n\|^2 \to 0$. Hence $z \in \sigma_c(\mathfrak{A})$ (provided, to be sure, that $x_n \perp \mathfrak{N}(\mathfrak{U} - z\mathfrak{I})$; but if $\mathfrak{N}(\mathfrak{A} - z\mathfrak{I})$ is non-trivial then $z \in \sigma_p(\mathfrak{A})$, which is also all right).

If, on the other hand, $\|x_n\| = o(\|y_n\|)$, then

$$\frac{-\|x_n\|^2 + \|y_n\|^2}{\|x_n\|^2 + \|y_n\|^2} \to 1,$$

$\varphi(y_n, x_n) \to (0, 1)$, and we must be in the remaining case, $\tau(\infty) \in \overline{s(\mathfrak{A})}$. This is easily disposed of, and the proof is complete.

3. The shell and the numerical range

First, the following key observation, which follows immediately from the definitions.

Theorem 3. 1. *The shadow of the shell is the numerical range.*

The definitions to be consulted are those in § 1, and the following

Definition 3. 1. The 'numerical range' of $\mathfrak{A}$, denoted $w(\mathfrak{A})$, is

$$\{x^* y : \|x\| = 1, (y, x) \in \mathfrak{A}\},$$

with ∞ adjoined in case $\infty \in \sigma_p(\mathfrak{A})$.

In light of the last theorem, the following may be regarded as a sharpening of the Hausdorff—Toeplitz theorem, which asserts that the numerical range of an operator is convex:

Theorem 3. 2. *For every pair of points of $s(\mathfrak{A})$, there is an ellipsoid (perhaps degenerate) containing them and lying in $s(\mathfrak{A})$.*

Proof. Let us normalize every vector (y, x) entering in (1. 3), $\|(y, x)\|^2 = \|x\|^2 + \|y\|^2 = 1$, so that (1. 3) reads simply $\varphi(y, x) = (2x^*y, -\|x\|^2 + \|y\|^2)$.

Now argue as in the proof of the Hausdorff—Toeplitz theorem [9]. The two given points of $s(\mathfrak{A})$ come from two unit vectors (y_1, x_1), (y_2, x_2), spanning a space $\mathfrak{S} \subseteqq \mathfrak{A}$. The set of points

$$(3.1) \qquad \left(\mathrm{Re}(2x^*y),\quad \mathrm{Im}(2x^*y),\quad -x^*x + y^*y\right)$$

in $\boldsymbol{R}^3$, as (y, x) ranges over all unit vectors in the 2-dimensional space $\mathfrak{S}$, will be shown to be an ellipsoid, and it is $\subseteqq s(\mathfrak{A})$ by definition.

Each component in (3. 1) is a hermitian form in the variable (y, x). By suitably choosing co-ordinates in $\boldsymbol{R}^3$ we may assume all have trace zero; and then by suitable choice of co-ordinates in $\mathfrak{S}$ we may write (3. 1) as

$$(3.2) \qquad \left(a_1(|\xi|^2 - |\eta|^2) + 2\,\mathrm{Re}\,(b_1\bar{\xi}\eta),\quad a_2(|\xi|^2 - |\eta|^2) + \right.$$

$$\left. + 2\,\mathrm{Re}\,(b_2\bar{\xi}\eta),\quad a_3(|\xi|^2 - |\eta|^2)\right) \quad (|\xi|^2 + |\eta|^2 = 1),$$

with a_1, a_2, a_3 real. It is elementary to express the set of points (3. 2) explicitly as an ellipsoid, and the proof is complete.

Thus the shell is in general not convex, as the examples immediately following will illustrate, but it becomes convex if its "holes are filled up". That is, the unbounded component of the complement of the shell, is the complement of a convex set.

Proposition 3. 1. *Assume $\mathfrak{A}$ is not densely defined. It may be extended to $\mathfrak{B} = \mathfrak{A} \oplus (\mathfrak{H}_2 \ominus \mathfrak{D}(\mathfrak{A}))$ (that is, $\mathfrak{B}$ is the zero operator on $\mathfrak{D}(\mathfrak{A})^\perp$). Then $s(\mathfrak{B})$ is a union of (perhaps degenerate) ellipsoids joining points of $s(\mathfrak{A})$ with $(0, -1)$.*

Proof. Of course $s(\mathfrak{B}) \supseteqq s(\mathfrak{A})$ just because $\mathfrak{B} \supseteqq \mathfrak{A}$. The general point of $\mathfrak{B}$ is $(y, x + x')$, where $(y, x) \in \mathfrak{A}$ and $x' \in \mathfrak{D}(\mathfrak{A})^\perp$. Now use (y, x) and $(0, x')$ as the vectors spanning $\mathfrak{S}$, in the construction of Thm. 3. 2.

As a corollary of Thm. 3. 2, we have

Theorem 3. 3. *Unless $\sigma_\pi(\mathfrak{A})$ is void, every point of $\overline{s(\mathfrak{A})}$ lies on a (perhaps degenerate) ellipsoid lying in $\overline{s(\mathfrak{A})}$ and intersecting S.*

Proof. Imbed $\mathfrak{H}$ in a space $\mathfrak{H}^0$ of approximate proper vectors of $\mathfrak{A}$. This is done in the same manner as given by BERBERIAN [3] for the case of an operator. Let $\mathfrak{A}^0$ denote the corresponding relation in $\mathfrak{H}^0 \oplus \mathfrak{H}^0$. It is easy to see that $s(\mathfrak{A}^0) = \overline{s(\mathfrak{A})}$. Now $s(\mathfrak{A}^0) \cap S$ is just $\tau(\sigma_p(\mathfrak{A}^0)) = \tau(\sigma_\pi(\mathfrak{A}))$, which is not empty. Given any point of $s(\mathfrak{A}^0)$, choose an arbitrary point of $s(\mathfrak{A}^0) \cap S$, and invoke Thm. 3. 2 to show there is an ellipsoid in $s(\mathfrak{A}^0)$ joining them, as desired.

The proof could have been accomplished by approximation without mentioning $\mathfrak{H}^0$, as was done in Thm. 2. 3.

Many theorems are known drawing conclusions on the structure of an operator from failure of its numerical range to have a smooth boundary [9], [18], [12]. The key seems to be the degeneracy of the ellipsoid in Thm. 3. 2, and these theorems should have sharper forms in which the shell would figure in place of the numerical range. There should also be theorems relating the shell with dilations. A small beginning is made in § 8 of this paper.

4. Examples

In this section various results are listed, illustrating the sort of shells which occur. The justifications of the results are mostly easy, and are not given.

Example 1. Let A be a normal matrix with eigenvalues $\lambda_1, \cdots, \lambda_n$. Then $s(A)$ is the convex hull of the points $\tau(\lambda_j) \in S$.

Example 2. Let $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Then $s(A)$ is the set of $(\zeta, h) \in \boldsymbol{R}^3$ satisfying $\frac{1}{2}|\zeta|^2 + (h + \frac{1}{2})^2 = \frac{1}{4}$. This is a non-degenerate ellipsoid tangent to the unit sphere at the point corresponding to the eigenvalue of A, viz., $(0, -1) = \tau(0)$.

Example 3. Let $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$. Then $s(A)$ is the set of $(\xi, \eta, h) \in \boldsymbol{R}^3$ satisfying $\xi^2 + \eta^2 - \xi h + h^2 - \xi + h = 0$. This is a non-degenerate ellipsoid tangent to the unit spere at the two points corresponding to the eigenvalues of A, viz., $(1,0) = \tau(1)$ and $(0, -1) = \tau(0)$.

From these results, $s(A)$ can be found for all 2×2 matrices by using the results of § 5. In particular, the shell of a 2×2 matrix is a degenerate ellipsoid if and only if the matrix is normal; otherwise, a non-degenerate ellipsoid.

Example 4. Let A be a normal operator. Then $s(A)$ is the convex hull of $\tau(\sigma(A))$, except that the points $\tau(\sigma(A) \setminus \sigma_p(A))$ are not in $s(A)$. In particular, the shell of the bilateral shift is $\{(\zeta, 0): |\zeta| < 1\}$.

Example 5. Let V denote the unilateral shift. Then $s(V) = \{(\zeta, 0): |\zeta| < 1\}$, but $s(V^*) = \{(\zeta, h) \in B: |\zeta| < 1, h \leqq 0\}$. (To prove all the points with $h < 0$ are in

$s(V^*)$, it suffices to consider sequences of the form $(\delta, \zeta, \zeta^2, \zeta^3, \cdots).$] Note that V^{-1} is a not-everywhere-defined, bounded operator. It happens also that V and V^{-1} have the same shell. Nonetheless, the two operators are quite different: $\sigma_r(V) = \{z: |z|<1\}$, while $\sigma_r(V^{-1}) = \{z: |z|>1\} \cap \{\infty\}$. This is a pleasingly simple instance of Thm. 5. 1 and Prop. 6. 4 below. Note also that V^* happens to be the same as the extension of V^{-1} obtained as in Prop. 3. 1; the relationship between their shells is as described there.

The preceding examples show some empty σ_p, hence some shells disjoint from S. It is easy to see from familiar results that if $\mathfrak{D}(\mathfrak{A})$ is dense then $\sigma_\pi(\mathfrak{A})$ is not empty, hence $\overline{s(\mathfrak{A})}$ is not disjoint from S. However, we have

Example 6. In 2-space, let $Ax_2 = x_1 \perp x_2$, and let A be otherwise undefined. Then $s(A) = \{(0, 0)\}$, $\sigma_\pi(A)$ is void, $\sigma(A) = \sigma_r(A) = \bar{C}$.

5. Transformation properties of the shell

The facts to be presented will emerge as immediate consequences of the definitions, once these have been expressed in the notations appropriate to the purpose.

Möbius transformations $\bar{C} \to \bar{C}$ are most simply expressed in terms of complex-homogeneous co-ordinates. Parametrize $\bar{C}$ in terms of pairs (z_1, z_2), with z represented by $(z, 1)$ and ∞ by $(1, 0)$. Then the linear transformation

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \qquad (ad - bc \neq 0) \tag{5. 1}$$

of the space of pairs represents the Möbius transformation usually written

$$z \to \mu(z) = \frac{az + b}{cz + d}.$$

The Riemann sphere $S = \tau(\bar{C})$ may also be written in a way better suited to present purposes:

$$\tau(z_1, z_2) = (|z_1|^2, z_1\bar{z}_2, z_2\bar{z}_1, |z_1|^2). \tag{5. 2}$$

(To give these second and third components, not necessarily real but conjugate to each other, is more convenient than the equivalent procedure of giving the corresponding real and imaginary parts. Indeed, it makes possible the especially simple form of (5. 4) below.) These are real-homogeneous co-ordinates $(\delta_1, \delta_2, \delta_3, \delta_4)$ for the points of S, which will be the locus of non-zero quadruples with $\delta_2 = \delta_3$, $\delta_2\delta_3 = \delta_1\delta_4$. In terms of the previous co-ordinates (ζ, h), S was the locus of $|\zeta|^2 + h^2 = 1$. Composing (1. 2) with (5. 2) it turns out that

$$(\zeta, h) \quad \text{may be written} \quad (1 + h, \zeta, \bar{\zeta}, 1 - h). \tag{5. 3}$$

The representation (5. 3) can be applied to all of B, that is, to all (ζ, h) with $|\zeta|^2+h^2\leqq 1$. It gives all non-zero $(\delta_1, \delta_2, \delta_3, \delta_4)$ such that $\bar{\delta}_2=\delta_3$, $\delta_2\delta_3\leqq\delta_1\delta_4$. Planes, in the δ-co-ordinates, have homogeneous equations $\sum\bar{\Delta}_j\delta_j=0$, subject to Δ_1, Δ_4 real, and $\bar{\Delta}_2=\Delta_3$. Halfspaces have representations $\sum\bar{\Delta}_j\delta_j\geqq 0$, subject to the same conditions, with the additional one: $\delta_1+\delta_4$ non-negative. To handle linear inequalities we need positively homogeneous co-ordinates; Δ and $-\Delta$ will not give the same half-space.

Now the Möbius transformation μ above gives a transformation $S\to S$ taking $\tau(z_1, z_2)$ to $\tau(az_1+bz_2, cz_1+dz_2)$. Clearly the δ-co-ordinates are transformed by the matrix

$$\begin{pmatrix} a\bar{a} & a\bar{b} & b\bar{a} & b\bar{b} \\ a\bar{c} & a\bar{d} & b\bar{c} & b\bar{d} \\ c\bar{a} & c\bar{b} & d\bar{a} & d\bar{b} \\ c\bar{c} & c\bar{d} & d\bar{c} & d\bar{d} \end{pmatrix}. \tag{5.4}$$

The correspondence between (5. 1) and (5. 4) is a representation of $GL(2, \boldsymbol{C})$, viz., the tensor product of the natural representation and its complex conjugate. Specializing (5. 1) to have determinant 1 (as for present purposes we may), we give (5. 4) determinant 1 also. These matrices comprise the proper Lorentz group in its natural representation by linear transformations of $\boldsymbol{R}^4$ (though not in the customary co-ordinate system). Indeed this group is well known to be isomorphic to the group of Möbius transformations (e.g., [5, § 17]).

We are interpreting the δ_j as homogeneous co-ordinates, so for us the matrices (5. 4) give a representation ϱ of the group by non-linear transformations of $\boldsymbol{R}^3$. The invariant cone (or half-cone!) under the Lorentz group is for us replaced by S in $\boldsymbol{R}^3$, which was known to be invariant from the start. Now note that (corresponding to the fact that the "future" is invariant under Lorentz transformations) all the $\varrho(\mu)$ also take the ball B onto itself.

On $B\setminus S$ this gives a group of plane-preserving transformations — the group of congruences of the Beltrami model [4, § 16. 2] of hyperbolic 3-space. Cf. [19, § 15, ex. 5], [2, Abschnitt IV]. The rigid rotations of S are given in the particular case $d=\bar{a}$, $c=-\bar{b}$. Since all the $\varrho(\mu)$ are plane-preserving, those which are rigid rotations of S will also be ordinary rotations of all of B.

The one ingredient still lacking is the definition of a Möbius transformation of a relation. For μ as above, define

$$\mu(\mathfrak{A})=\{(ay+bx, cy+dx):(y, x)\in\mathfrak{A}\}. \tag{5.5}$$

This agrees with the usual definition $\mu(A)=(aA+b)(cA+d)^{-1}$ for the most important case, that in which A is an everywhere defined operator and $-d/c$ is not in its spectrum. On the other hand, it agrees for all $\mathfrak{A}$ with the usual definition of $\mathfrak{A}^{-1}=\{(x, y):(y, x)\in\mathfrak{A}\}$.

Theorem 5. 1. *The shell of* $\mu(\mathfrak{A})$ *is obtained from the shell of* $\mathfrak{A}$ *by the transformation* $\varrho(\mu)$.

Proof. Write everything in the δ-co-ordinates and this falls out. The shell of $\mathfrak{A}$ is

$$\{(\|y\|^2, x^*y, y^*x, \|x\|^2):(y, x)\in A\}.$$

Similarly, $s(\mu(\mathfrak{A}))$ is

$$\{(\|ay+bx\|^2, (\bar{c}y^*+\bar{d}x^*)(ay+bx), (\bar{a}y^*+\bar{b}x^*)(cy+dx), \|cy+dx\|^2):(y, x)\in\mathfrak{A}\}.$$

Clearly the quadruples in $s(\mu(\mathfrak{A}))$ are got from those in $s(\mathfrak{A})$ by applying the matrix (5. 4).

This proves the theorem as stated, but leaves open the question whether, for two Möbius transformations μ_1, μ_2, we need have $\mu_1(\mu_2(\mathfrak{A}))=\mu_1\circ\mu_2(\mathfrak{A})$. This is one of a whole class of questions which will be treated in the next section.

6. Rational functional calculus of relations

The main result of the paper was stated in the introduction for operators only. Before stating it for general closed linear relations $\mathfrak{A}$, spectral sets must be defined for them, and this means that we need discussion of rational functions of relations.

The definition of powers is standard:

$$\mathfrak{A}^2=\mathfrak{A}\circ\mathfrak{A}=\{(y, x):(\exists w)(y, w)\in\mathfrak{A}\ \&\ (w, x)\in\mathfrak{A}\},$$

etc. Similarly for linear combinations:

$$a_1\mathfrak{A}_1+a_2\mathfrak{A}_2=\{(a_1y_1+a_2y_2, x):(y_1, x)\in\mathfrak{A}_1\ \&\ (y_2, x)\in\mathfrak{A}_2\}.$$

This defines polynomials. ARENS [1] explains their properties. In particular, it is important to use only polynomials with leading coefficient non-zero, because $0\mathfrak{A}$ (say) may not be the zero relation: it may be properly contained in it. With this understanding, ARENS proves the following:

If p is a polynomial

$$p(z)=a_nz^n+\cdots+a_1z+a_0 \qquad (a_n\neq 0),$$

then $(y, x)\in p(\mathfrak{A})$ if and only if there exist $w_0=x, w_1, \cdots, w_n$ with $(w_j, w_{j-1})\in\mathfrak{A}$ and $y=\sum_{j=0}^{n}a_jw_j$. If p and q are polynomials then $p(\mathfrak{A})q(\mathfrak{A})=(pq)(\mathfrak{A})$ — this despite the fact that the distributive law fails in general. If p and q are polynomials then $p(q(\mathfrak{A}))=p\circ q(\mathfrak{A})$. If p and q are polynomials such that, in forming $p+q$, the leading terms do not cancel, then $(p+q)(\mathfrak{A})=p(\mathfrak{A})+q(\mathfrak{A})$.

We want to extend the ideas to rational functions. Again, care is required because (for instance) $\mathfrak{A}\mathfrak{A}^{-1}$ need not be $\mathfrak{I}$, but may be a proper subset. There are therefore two inequivalent possible definitions, of which I choose this one (already used in (5. 5) in the case of a Möbius transformation):

Definition 6. 1. Let $f=p/r$ be a rational function, where p and r are polynomials without common factor: $p(z)=a_nz^n+\cdots+a_1z+a_0$ $(a_n\neq 0)$, $r(z)=b_mz^m+\cdots+b_1z+b_0$ $(b_m\neq 0)$. Then $(y, x)\in f(\mathfrak{A})$ if and only if there exist w_k $(k=0, 1, \cdots, \max\{n, m\})$ with $(w_k, w_{k-1})\in\mathfrak{A}$, such that $y=\sum_{j=0}^{n} a_jw_j$ and $x=\sum_{j=0}^{m} b_jw_j$.

By ARENS's result just quoted, this gives the usual result in case f is a polynomial. $(r=1)$. For general polynomial r, the $f(\mathfrak{A})$ determined by Def. 6. 1 is $\subseteq p(\mathfrak{A})(r(\mathfrak{A}))^{-1}$ and the inclusion may be proper, though equality holds for $\mathfrak{A}$ an operator. We do have some of the expected relations.

Proposition 6. 1. *Let* $p=q_1r+r_1$, *with* $m_1=\text{degree}\,(r_1)<n=\text{degree}\,(p)$. *Then* $(p/r)(\mathfrak{A})=q_1(\mathfrak{A})+(r_1/r)(\mathfrak{A})$.

(The hypothesis implies that $m=\text{degree}\,(r)\leqq n$, but not that $m_1<m$.)

Proof. Let $r_1(z)=\sum_{j=0}^{m_1}c_jz^j$, so that $(q_1r)(z)=\sum_{j=0}^{n}(a_j-c_j)w_j$. The general pair in $q_1(\mathfrak{A})+(r_1/r)(\mathfrak{A})$ is expressible as $(u+v, x)$ where $(u, x)\in q_1(\mathfrak{A})$, and $v=\sum c_jw_j$, $x=\sum b_jw_j$, for some $w_0, \cdots, w_m$, such that $(w_j, w_{j-1})\in\mathfrak{A}$. From $(u, x)\in q_1(\mathfrak{A})$ and $(x, w_0)\in r(\mathfrak{A})$ follows that $(u, w_0)\in(q_1r)(\mathfrak{A})$, by one of ARENS's results just cited. Indeed, by reference to the proof of that result [1, 2. 3] we see that we can even use

$$u=\sum_{j=0}^{n}(a_j-c_j)w_j \qquad (w_j, w_{j-1})\in\mathfrak{A},$$

with the same w_j as before as far as $j=m'$. But then $(u+v, x)\in(p/r)(\mathfrak{A})$. This proves „$\supseteq$" in the conclusion.

In the other direction no subtleties are involved: we are given $y=\sum a_jw_j$, $x=\sum b_jw_j$ as in Def. 6. 1, and we define $u=\sum(a_j-c_j)w_j$, $v=\sum c_jw_j$. By definition $(v, x)\in(r_1/r)(\mathfrak{A})$, it remains to prove that $(u, x)\in q_1(\mathfrak{A})$. Let $q_1(z)=\sum d_jz^j$, then for each i, $a_i-c_i=\sum_j d_{i-j}b_j$. Define $x_k=\sum_j b_jw_{j+k}$ $(k=0, 1, \cdots, \text{degree}\,(q_1))$, so that $x_0=x$, $(x_k, x_{k-1})\in\mathfrak{A}$, and $\sum d_kx_k=u$. Then by definition $(u, x)\in q_1(\mathfrak{A})$.

In the following propositions, Möbius transformations $\mu(z)=\dfrac{az+b}{cz+d}$ again play a special role.

Proposition 6. 2. $\mu(\mathfrak{A})$ *is closed.*

(We are assuming $\mathfrak{A}$ closed throughout.)

Proof. To say that $(y_\nu, x_\nu) \in \mu(\mathfrak{A})$ is to say that $y_\nu = aw_{1\nu} + bw_{0\nu}$ and $x_\nu = cw_{1\nu} + dw_{0\nu}$, with $(w_{1\nu}, w_{0\nu}) \in \mathfrak{A}$. We assume in addition that $y_\nu \to y$, $x_\nu \to x$. Then solving for $w_{1\nu}$ and $w_{0\nu}$ in terms of y_ν and x_ν, we deduce that $w_{1\nu} \to w_1$ and $w_{0\nu} \to w_0$ for w_i such that $y = aw_1 + bw_0$, $x = cw_1 + dw_0$. Since $\mathfrak{A}$ is closed, $(w_1, w_0) \in \mathfrak{A}$. This completes the proof.

I don't know how far this remains true when μ is replaced by the more general f considered earlier; cf. [13, § 2. 16], [1, 3. 8].

Proposition 6. 3. $(f \circ \mu)(\mathfrak{A}) = f(\mu(\mathfrak{A}))$.

Proof. This is clear when μ is just multiplication by a constant. When μ is a translation, $\mu(z) = z + b$, a calculation is needed which I will only summarize: in $y = \sum a_j w_j$ change to an expression $y = \sum a'_k w'_k$ by the substitution $w'_k = \sum_j \binom{k}{j} b^{k-j} w_j$ (and similarly for x, of course), and make the verification that $(w'_k, w'_{k-1}) \in \mathfrak{A} + b\mathfrak{I}$.

When μ is reciprocation, $\mu(z) = z^{-1}$, there is again a little calculation. It is convenient to depart from previous practice and write $p(z) = \sum_{j=0}^{n} a_j z^j$, $r(z) = \sum_{j=0}^{n} b_j z^j$, where not both a_n and b_n are 0 and not both a_0 and b_0 are 0. With this convention there is notational symmetry between f and $f \circ \mu$, and reference to definitions will verify the conclusion.

The observation that any Möbius transformation is obtained from these types by composition, completes the proof.

Proposition 6. 4. (Detailed spectral mapping theorem for Möbius transformations.) $\sigma_p(\mu(\mathfrak{A})) = \mu(\sigma_p(\mathfrak{A}))$, $\sigma_c(\mu(\mathfrak{A})) = \mu(\sigma_c(\mathfrak{A}))$, $\sigma_r(\mu(\mathfrak{A})) = \mu(\sigma_r(\mathfrak{A}))$.

Again, we are at liberty to consider simple special Möbius transformations and then compose them to give the general μ.

Under linear transformation, σ_p, σ_c, and σ_r all behave as desired by Def. 2. 4, with the exception of ∞. Assume $\infty \in \sigma_p(\mathfrak{A})$ and let us prove $\infty \in \sigma_p(\mathfrak{A} + b\mathfrak{I})$. The assumption is equivalent to the existence of non-zero x such that $(x, 0) \in \mathfrak{A}$; but then $(x, 0) \in \mathfrak{A} + b\mathfrak{I}$ as well, and this gives the conclusion. Similarly for $\infty \in \sigma_c(\mathfrak{A})$. That $\infty \in \sigma_r(\mathfrak{A})$ entails $\infty \in \sigma_r(\mathfrak{A} + b\mathfrak{I})$ is a consequence of the remark following Def. 2. 4.

If we consider reciprocation, it is the λ other than 0 and ∞ which require checking. Assume $\lambda \in \sigma_c(\mathfrak{A})$, so that there are $(y_\nu, x_\nu) \in \mathfrak{A}$ such that $\|x_\nu\| = 1$ but $\|y_\nu - \lambda x_\nu\| \to 0$ (so that $\|y_\nu\| \to 0$; without loss of generality, $y_\nu \neq 0$). Let $x'_\nu = x_\nu / \|y_\nu\|$, $y'_\nu = y_\nu / \|y_\nu\|$. Clearly $(x'_\nu, y'_\nu) \in \mathfrak{A}^{-1}$, $\|y'_\nu\| = 1$, and $\|x'_\nu - \lambda^{-1} y'_\nu\| \to 0$, so $\lambda^{-1} \in \sigma'_c(\mathfrak{A}^{-1})$. For σ_p, it is even easier. Finally, take $\lambda \in \sigma_r(\mathfrak{A})$. This means $\mathfrak{R}(\mathfrak{A} - \lambda\mathfrak{I})$ is not dense. Choosen non-zero z such that, for all $(y, x) \in \mathfrak{A}$, $z \perp y - \lambda z$. Then also for all $(x, y) \in \mathfrak{A}^{-1}$, $z \perp x - \lambda^{-1} y$, so $\mathfrak{R}(\mathfrak{A}^{-1} - \lambda^{-1}\mathfrak{I})$ is not dense either. This completes the proof.

Again, I don't know how much of this result holds for more general f; I have partial results. Of course it can't hold in toto. For example, it is easy to construct an operator A whose square is O and hence has void continuous spectrum, while A itself has non-void continuous spectrum. (For the case of operators see [13, Theorem 5.12. 2].)

7. The shell and spectral sets

Definition 7. 1. The 'norm' of a relation $\mathfrak{A}$ is

$$\|\mathfrak{A}\| = \sup \{\|y\|/\|x\| : (0, 0) \neq (y, x) \in \mathfrak{A}\}.$$

Definition 7. 2. Let X be a closed subset of $\overline{C}$. X 'is a spectral set for' $\mathfrak{A}$ or 'is s.s. for' $\mathfrak{A}$) in case every rational function f having modulus $\leqq 1$ on X has also the property $\|f(\mathfrak{A})\| \leqq 1$.

It is clear by Prop. 2. 1 that $\|\mathfrak{A}\|$ is finite if and only if $\infty \notin \sigma_\pi(\mathfrak{A})$, that is, $\mathfrak{A}$ corresponds to an operator A which is bounded, and in this case $\|\mathfrak{A}\| = \|A\|$. Furthermore, as will appear in the course of developing the basic properties, statements about spectral sets for relations can be reduced to statements involving operators without much trouble. However there are two closely related virtues in the present definition: it applies to not-everywhere-defined operators; and, secondly, it brings in σ_π instead of σ (see in particular Prop. 7. 2 below).

Proposition 7. 1. *X is s.s. for $\mathfrak{A}$ if and only if $\mu(X)$ is s.s. for $\mu(\mathfrak{A})$.*

This follows easily from Prop. 6. 3.

Proposition 7. 2. *If X is s.s. for $\mathfrak{A}$ then $\sigma_\pi(\mathfrak{A}) \subseteqq X$.*

Proof. By Prop. 7. 1, it is enough to consider a special point of $\overline{C}$. Assume, then, $\infty \in \sigma_\pi(\mathfrak{A}) \setminus X$, so that X is a compact subset of C and $\mathfrak{A}$ is not a bounded operator. To show Def. 7. 2 is not satisfied, choose $f(z) = az$, for a sufficiently small constant $a > 0$.

Theorem 7. 1. *The unit disk $D = \{z \in C: |z| \leqq 1\}$ is s.s. for $\mathfrak{A}$ if and only if $\|\mathfrak{A}\| \leqq 1$.*

Proof. Either condition implies we are dealing with a bounded operator A. If $\mathfrak{D}(A) = \mathfrak{H}$, the theorem is just VON NEUMANN's basic result, as in [17, § 154]. More generally, define $\bar{A}$ as the extension of A which is zero on $\mathfrak{D}(A)^\perp$. By applying VON NEUMANN's theorem to $\bar{A}$, it is easy to deduce (the non-trivial half of) the present theorem for A.

This, theorem, together with Prop. 7. 1, tells which disks are s.s. for $\mathfrak{A}$. The result is familiar in general outlines: it resembles that in [17] except that no special

exemptions need to be made for functions with poles in $\sigma(\mathfrak{A})$ — even in $\sigma_\pi(\mathfrak{A})$. Namely,

(i) the finite disk $\{z\in\boldsymbol{C}\colon |z-z_0|\leqq r\}$ is s.s. for $\mathfrak{A}$ if and only if $\|\mathfrak{A}-z_0\mathfrak{I}\|\leqq r$;

(ii) the complementary disk $\{z\in\boldsymbol{C}\colon |z-z_0|\geqq r\}\cup\{\infty\}$ is s.s. for $\mathfrak{A}$ if and only if $\|(\mathfrak{A}-z_0\mathfrak{I})^{-1}\|\leqq r^{-1}$;

(iii) the half-plane $\{z\in\boldsymbol{C}\colon \mathrm{Re}\,(\bar{\zeta}z)\geqq a\}\cup\{\infty\}$ is s.s. for $\mathfrak{A}$ if and only if it contains $w(\mathfrak{A})$.

The following theorem contains all three of these, in a geometric form, invariant under Möbius transformations of $\overline{\boldsymbol{C}}$.

Theorem 7.2. *Let X be a disk. Then X is s.s. for $\mathfrak{A}$ if and only if the shell $s(\mathfrak{A})$ is contained in the convex hull of the stereographic projection $\tau(X)$.*

Proof. Let μ be the Möbius transformation such that $\mu(X)=D$, the unit disk. We have just seen that X is s.s. for $\mathfrak{A}$ if and only if $\|\mu(\mathfrak{A})\|\leqq 1$. Comparing Defs. 7.1 and 1.1, this is seen to be equivalent to saying that $s(\mu(\mathfrak{A}))$ lies in the half-space $\{(\zeta,h)\in\boldsymbol{R}^3\colon h\leqq 0\}$; and this half-space is the convex hull of $\tau(D)=\tau(\mu(X))$. To return from $\mu(\mathfrak{A})$ and $\mu(X)$ to $\mathfrak{A}$ and X, we want to consider the transformation μ' inverse to μ, and the projective transformation $\varrho(\mu')$ of $\boldsymbol{R}^3$ to which it gives rise. This was discussed in § 5. $\varrho(\mu')$ preserves B, and preserves planes; consequently it takes convex hulls to convex hulls. It takes $\tau(\mu(X))$ to $\tau(X)$. Finally, by Thm. 5.1, it takes $s(\mu(\mathfrak{A}))$ to $s(\mathfrak{A})$. The theorem is proved.

Alternatively, I could have confined consideration to rigid rotations of B (cf. § 5) with only slight modification in the proof.

The set of disks which are s.s. for fixed $\mathfrak{A}$ is hereby represented as a closed convex set $s(\mathfrak{A})^\dagger$, the dual of $s(\mathfrak{A})$. The convexity of this set is inherent in the following easily proved fact, a generalization of the Lemma of [8, § 3.1]: if two disks are s.s. for $\mathfrak{A}$, then so is any disk containing their intersection. The closedness of $s(A)^\dagger$ is also easy to prove directly.

However, if it was a question only of representing all disks s.s. for $\mathfrak{A}$ in a simple geometric way, the closed convex hull of $s(\mathfrak{A})$ would do exactly as well as $s(\mathfrak{A})$, for it has the same dual. It is like the situation for the numerical range $w(\mathfrak{A})$. Only $\overline{w(\mathfrak{A})}$ is needed in criterion (iii) above, telling which half-planes are spectral sets. Still for some purposes the richer structure of $w(\mathfrak{A})$ itself is interesting. The shell is like the numerical range with one dimension added (cf. Thm. 3.1), and its structure is very much richer than that of its closed convex hull.

In spite of these remarks, Thm. 7.2 throws emphasis on $s(\mathfrak{A})^\dagger$, and this will persist in the final section of the paper. Let me therefore say a few more words about this set.

It has already been pointed out in § 5 that half-spaces in $\boldsymbol{R}^3$ may be for present purposes conveniently represented by quadruples $\Delta=(\Delta_1,\Delta_2,\Delta_3,\Delta_4)$; here

$\Delta_1, \Delta_4 \in \boldsymbol{R}$, $\Delta_2 = \bar{\Delta}_3 \in \boldsymbol{C}$, and if $\lambda > 0$ then $\lambda\Delta$ represents the same half-space as Δ. Namely, it is the half-space of all $(\zeta, h) \in \boldsymbol{R}^3$ such that $\Delta_1(1+h) + \bar{\Delta}_2\zeta + \bar{\Delta}_3\bar{\zeta} + \Delta_4(1-h) \geqq 0$.

Let us regard the dual $C^\dagger$ of an arbitrary set $C \subseteqq \boldsymbol{R}^3$ as the convex cone of all $\Delta \in \boldsymbol{R}^4$ representing half-spaces $\supseteqq C$. This sort of dual has been used often in the theory of convex sets [23].

Proposition 7.3. *$\Delta \in B^\dagger$ if and only if Δ gives homogeneous co-ordinates for a point of B in the δ-representation of* § 5, *i.e., if and only if $\Delta_1 + \Delta_4 > 0$ and $\Delta_2\Delta_3 \leqq \Delta_1\Delta_4$.*

This is just self-duality of the unit ball; I will not bother translating the familiar proof into this notation.

8. Generalization of Berger's theorem

The most striking results relating disk spectral sets to dilation theory are the well-known theorem of Sz.-Nagy [20] on dilations of contractions, and C. Berger's theorem [11], [21] on dilations of operators with $w(A)$ in the unit disk. Sz.-Nagy and Foiaş [22] have recently given a common generalization of the two, which I will relate to the ideas of this paper. Their theorem may be stated as follows. [2])

Theorem 8.1 (Sz.-Nagy—Foiaş). *For a bounded, everywhere-defined operator A, and for any $\varrho > 0$, consider these conditions:*

(i) *$A \in C_\varrho$, that is, there exists a unitary U on a Hilbert space $\mathfrak{K} \supseteqq \mathfrak{H}$ such that $A^n = \varrho \cdot \mathrm{pr}\, U^n$ $(n = 1, 2, \cdots)$, where* pr *denotes compression to $\mathfrak{H}$;*

(ii) *for every z with $|z| < 1$, $\|zA((\varrho-1)zA - \varrho)^{-1}\| \leqq 1$;*

(iii) *for every $x \in \mathfrak{H}$ and every $t \in [0, 1]$,*

$$2|\varrho - 1|t \cdot |x^*Ax| \leqq \varrho\|x\|^2 - (2-\varrho)t^2\|Ax\|^2.$$

Conclusions: Conditions (ii) *and* (iii) *are equivalent. Condition* (i) *is equivalent to* (iii) *together with*

(iv) *$\sigma(A) \subseteqq D$, the closed unit disk.*

If $\varrho \leqq 2$, (iv) *follows from* (iii).

This formulation is not quite that of Sz.-Nagy and Foiaş; let me bridge the short gap. My (ii) is equivalent to their (5). My (iii) is obtained from their (I_ϱ), an inequality which must be asserted for every $z \in D$, by rewriting it in such a way that the phase of z need no longer be kept in view.

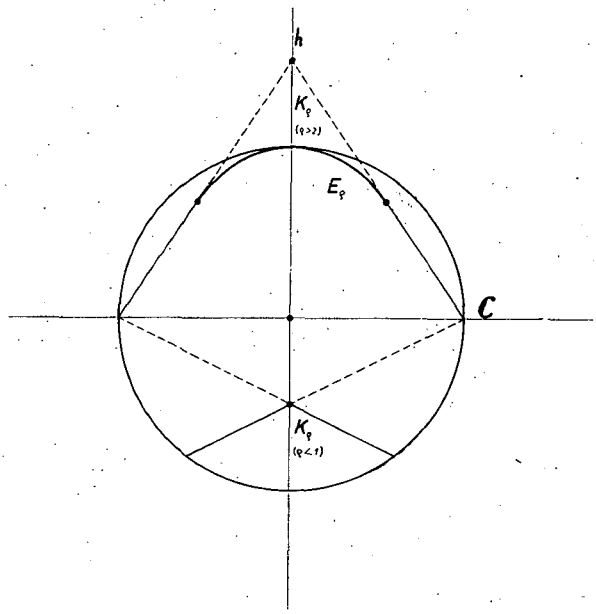
[2]) The real parameter ϱ is not to be confused with the symbol $\varrho(\mu)$ already introduced in § 5.

Now for each t, the inequality (iii) is a homogeneous linear inequality relating the moduli of the homogeneous co-ordinates of points of $s(A)$. As such, it is readily interpreted geometrically: it restricts $s(A)$ to lie in a certain solid half-cone (from now on I will say simply 'cone'). The cone may degenerate into a half-space, in particular it does so for $t=0$. To complete the account, we must examine the consequences for $s(A)$ of imposing the restrictions (iii) for all t simultaneously.

For this purpose, here are a few special notations. Let

$$f_x(t) = (2-\varrho)\|Ax\|^2 t^2 + 2|\varrho-1|\cdot|x^*Ax|t - \varrho\|x\|^2.$$

For each non-zero x, f_x is a real quadratic polynomial, considered as a function on [0, 1]; (iii) asserts that every f_x is $\leqq 0$ on the whole interval.



Secondly, let K_ϱ denote the cone $\{(\zeta, h)\in \boldsymbol{R}^3\colon |\varrho-1|\cdot|\zeta| \leqq \varrho-1-h\}$. The assertion $f_x(1)\leqq 0$ is readily transformed into the assertion $\varphi(Ax, x)\in K_\varrho$. This can be done directly from (1. 3) (normalizing by assuming $\|x\|^2+\|Ax\|^2=1$).

The first part of the picture can now be completed.

Proposition 8. 1. *For $\varrho\leqq 2$, A satisfies condition* (iii) *above if and only if $s(A)\subseteqq K_\varrho$.*

Proof. Refer to the definition of f_x. We have just noted that $s(A)\subseteqq K_\varrho$ if and only if $f_x(1)\leqq 0$ for all x; and $f_x(0)\leqq 0$ in any case. But f_x is a quadratic polynomial

with leading coefficient $\geqq 0$; therefore it is non-positive at the endpoints of the interval [0, 1] if and only if it is non-positive throughout. This gives the equivalence.

K_ϱ is of course symmetric about the h-axis, and has apex at $h=\varrho-1$. Its generators pass through the equatorial points $(e^{i\theta}, 0)$; but for $\varrho<1$, the equator lies in the other nappe of the cone, and K_ϱ itself lies entirely below the equatorial plane. For $\varrho=1$ (SZ.-NAGY's case), K_ϱ is the half-space $h\leqq 0$. For $\varrho=2$ (BERGER's case), the apex is at the north pole, and $s(A)\subseteqq K_\varrho$ is equivalent to $w(A)\subseteqq D$ by Thm. 3. 1.

For $\varrho>2$, the apex of K_ϱ is above the north pole. But then further restrictions on $s(A)$ result from (iii).

Let E_ϱ be obtained by removing from K_ϱ all points lying between the apex and the upper half of the ellipsoid $(\varrho-1)^2|\zeta|^2=\varrho(\varrho-2)(1-h^2)$. This ellipsoid, in addition to being evidently symmetric with respect to the h-axis and with respect to the ζ-plane, is tangent to ∂K_ϱ. The definition of E_ϱ requires the following interpretation. The "upper half" of the ellipsoid is the portion above the circle of tangency.

The ellipsoid lies entirely in B, and is tangent to S at the poles. Hence in $h>0$, E_ϱ has no points in common with S except (0, 1).

Proposition 8. 2. *For $\varrho>2$, A satisfies condition* (iii) *above if and only if $s(A)\subseteqq E_\varrho$.*

Proof. The cones to which $s(A)$ is restricted by (iii) are as follows: for $t=0$, the half-space $h\leqq 1$; for $t=1$, K_ϱ; and for intermediate t, the intermediate cones tangent to the ellipsoid. To see this, one reduces by symmetry to consideration of $\zeta>0$, and then makes an elementary computation which will not be reproduced here. Now the equivalence of $s(A)\subseteqq E_\varrho$ becomes clear.

Proposition 8. 3. *In Thm.* 8. 1, *condition* (iv) *follows from* (iii) *in case $\varrho>2$ also. Hence* (i) *and* (iii) *are equivalent for all values of ϱ.*

Proof. For a bounded, everywhere-defined operator, $\sigma(A)\subseteqq D$ is known to follow from $\sigma_\pi(A)\subseteqq D$. By Thm. 2. 3 and the above description of E_ϱ, any $\mathfrak{A}$ with $s(\mathfrak{A})\subseteqq E_\varrho$ must have $\sigma_\pi(\mathfrak{A})\subseteqq D\cup\{\infty\}$. Here ∞ is ruled out because we are in the bounded, single-valued case. The conclusion therefore follows from Prop. 8. 2.

These ideas lead to the following question: "If $s(\mathfrak{A})\subseteqq K_\varrho$ if $\varrho\leqq 2$ (or $\subseteqq E_\varrho$ if $\varrho>2$), what can we conclude about dilations of $\mathfrak{A}$?" They do not, however, lead to an answer. The difficulties arise even for $\varrho<2$; although we then know we have to deal with operators, we do not know that they are everywhere defined. Some new idea seems to be needed to cope with this dilation problem.

References

[1] R. ARENS, Operational calculus of linear relations, *Pacific J. Math.*, **11** (1961), 9—23.
[2] R. BALDUS, *Nichteuklidische Geometrie. Hyperbolische Geometrie der Ebene*, Sammlung Göschen 970 (Berlin—Leipzig, 1927).
[3] S. K. BERBERIAN, Approximate proper vectors, *Proc. Amer. Math. Soc.*, **13** (1962), 111—114.
[4] H. S. M. COXETER, *Introduction to geometry* (New York, 1961).
[5] ——— The inversive plane and hyperbolic space, *Abh. Math. Sem. Univ. Hamburg*, **29** (1966), 217—242.
[6] C. W. CURTIS and I. REINER, *Representation theory of finite groups and associative algebras* (New York, 1962).
[7] CH. DAVIS, Separation of two linear subspaces, *Acta Sci. Math.*, **19** (1958), 172—187.
[8] CH. DAVIS and D. G. RIDER, Spectral sets and numerical range, *Revue Roumaine de Math. Pures et Appl.*, **10** (1965), 125—131.
[9] W. F. DONOGHUE, On the numerical range of a bounded operator, *Michigan Math. J.*, **4** (1957), 261—263.
[10] P. R. HALMOS, *Introduction to Hilbert space and the theory of spectral multiplicity* (New York, 1951).
[11] ——— Positive definite sequences and the miracle of *w*. Mimeographed lecture notes, University of Michigan, 1965.
[12] S. HILDEBRANDT, Über den numerischen Wertebereich eines Operators, *Math. Ann.*, **163** (1966), 230—247.
[13] E. HILLE and R. S. PHILLIPS, *Functional analysis and semigroups*, American Mathematical Society Colloquium Publications, Vol. 31 (Providence, R. I., 1957).
[14] T. KATO, *Perturbation theory for linear operators* (Berlin, 1967).
[15] J. VON NEUMANN, Über adjungierte Funktionaloperatoren, *Annals of Math.*, **33** (1932), 294—310.
[16] ——— Eine Spektraltheorie für allgemeine Operatoren eines unitären Raumes, *Math. Nachr.*, **4** (1950/51), 258—281.
[17] F. RIESZ and B. SZ.-NAGY, *Leçons d'analyse fonctionnelle*, 2nd ed. (Budapest, 1953).
[18] M. SCHREIBER, Numerical range and spectral sets, *Michigan Math. J.*, **10** (1963), 283—288.
[19] H. SCHWERDTFEGER, *Geometry of complex numbers* (Toronto, 1962).
[20] B. SZ.-NAGY, Sur les contractions de l'espace de Hilbert, *Acta Sci. Math.*, **15** (1953), 87—92.
[21] ——— Positiv-definite, durch Operatoren erzeugte Funktionen, *Wiss. Zeitschrift Techn. Univ. Dresden*, **15** (1966), 219—222.
[22] B. SZ.-NAGY and C. FOIAŞ, On certain classes of power-bounded operators in Hilbert space, *Acta Sci. Math.*, **27** (1966), 17—25.
[23] F. A. VALENTINE, *Convex sets* (New York, 1964).

(Received May 9, 1967)