The Numerical Ranges of Automorphic Composition Operators*

Paul S. Bourdon

Washington and Lee University, Lexington, VA 24450
E-mail: pbourdon@wlu.edu

Joel H. Shapiro

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

We investigate the shape of the numerical range for composition operators induced on the Hardy space H^2 by conformal automorphisms of the unit disc. We show that usually, but not always, such operators have numerical ranges whose closures are discs centered at the origin. Surprising open questions arise from our investigation.

Key Words: Numerical range, composition operator

1. INTRODUCTION

We work on the Hardy space H^2 of functions f holomorphic on the open unit disc $\mathbb{U}$ with square-summable MacLaurin series coefficients. More precisely, a function $f(z) = \sum_{n=0}^{\infty} \hat{f}(n) z^n$, holomorphic on $\mathbb{U}$, belongs to H^2 if and only if $\|f\|^2 := \sum_{n=0}^{\infty} |\hat{f}(n)|^2 < \infty$. The functional $\|\cdot\|$ so defined is a norm on H^2 that makes it into a Hilbert space that is isometrically isomorphic, via the map $f \to \hat{f}$, to the sequence space ℓ^2.

A consequence of a famous theorem of J. E. Littlewood [15, 1925] asserts that each holomorphic selfmap φ of $\mathbb{U}$ induces on H^2 a bounded *composition operator* C_φ defined by the equation $C_\varphi f = f \circ \varphi \quad (f \in H^2)$. That C_φ maps the space of *all* holomorphic functions on $\mathbb{U}$ linearly into itself is elementary, but Littlewood's result that C_φ also preserves H^2 is quite remarkable (see [8, Chapter 1] or [21, Chapters 1 and 9] for more on Littlewood's theorem).

* The research of both authors was supported in part by the National Science Foundation

Littlewood's theorem raises the possibility of connecting operator theory with the exquisite function theory of holomorphic selfmaps of the unit disc that developed roughly between 1870 and 1930, driven by the work of H. A. Schwarz, Koenigs, Julia, Wolff, Denjoy and Carathéodory (see, e.g., [21, Chapters 4–6]). Beginning with seminal papers of Ryff [19], Nordgren [17] and H. J. Schwartz [20] in the late 1960s, this subject has been pursued enthusiastically by a growing cadre of researchers. To get an idea of the current flavor of the subject the reader may consult the monographs [5] and [21], and the recent conference proceedings [12].

The driving force behind this paper is the operator theoretic concept of *numerical range.* For a bounded operator T on a Hilbert space $\mathcal{H}$ this is the subset $W(T)$ of the complex plane that is the image of the unit sphere of $\mathcal{H}$ under the quadratic form associated with T, that is,

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

It is clear that every eigenvalue of T belongs to $W(T)$, and a little more effort shows that the *spectrum* of T belongs to the closure $\overline{W}(T)$ of $W(T)$ (we will observe shortly that $W(T)$ need not always be closed). The most remarkable fact about the numerical range, however, is that it is always *convex*. This is the famous theorem of Toeplitz [23] and Hausdorff [14], for which many additional proofs have been given since their time. None, however, reduces the result to a triviality; see [11, Chapter 1], [6], and [7] (most conveniently found [13, Problem 210]) for a few of these.

Because of the Toeplitz-Hausdorff theorem and the spectral containment mentioned above, $\overline{W}(T)$ contains the convex hull of the spectrum of T. A striking difference between spectrum and numerical range is that while the former is similarity invariant, the latter is not; it is precisely this lack of similarity invariance that dominates the work of this paper.

Here we study the numerical ranges of composition operators C_φ induced by *conformal automorphisms* φ of $\mathbb{U}$. It is well known that these are all linear fractional transformations, and that they come in three flavors (see, e.g., [21, Chapter 0]):

Hyperbolic Conformally conjugate to a positive dilation; two fixed points, both on $\partial\mathbb{U}$.

Parabolic Conformally conjugate to a translation; fixed point on $\partial\mathbb{U}$, no other fixed point on the Riemann sphere.

Elliptic Conformally conjugate to a rotation about the origin; two fixed points in the sphere, one in $\mathbb{U}$, one outside its closure.

A quick look at the elliptic case suffices to capture the tenor of our investigation. If φ is an elliptic automorphism that fixes the origin, then it is a rotation through some angle α: $\varphi(z) = e^{i\alpha}z$ ($z \in \mathbb{U}$), in which case C_φ is a unitary operator on H^2. If α is a rational multiple of π, or equivalently if φ (and therefore C_φ) is periodic with some period n, then it is elementary to check that $W(C_\varphi)$ is the closed polygon whose vertices are the n-th roots of unity, while if α is an *irrational* multiple of π then $W(C_\varphi)$ consists of the open unit disc along with the dense subset $\{e^{in\alpha} : n \in \mathbb{Z}\}$ of the unit circle (note that in this case $W(C_\varphi)$ is not closed); see [16, Prop. 2.1] for the details.

However the fun begins when φ is an elliptic automorphism that does *not* fix the origin. We show that if such a map φ is not periodic then the closure of $W(C_\varphi)$ is a disc centered at the origin (Theorem 4.1.). However we have only crude estimates of the radius of this disc, and we do not know what points of the boundary, if any, belong to $W(C_\varphi)$. If φ is periodic then, surprisingly, the situation seems even murkier: For period 2 we can show that the closure of $W(C_\varphi)$ is an elliptical disc with foci at ± 1 (Corollary 4.4.), but for period $n > 2$ then all we can say is that the numerical range of C_φ has n-fold symmetry—while we strongly suspect that in this case the closure is not a disc, we are not yet able to prove this.

The hyperbolic and parabolic cases offer their own challenges. For these we are able to show that $W(C_\varphi)$ is a disc centered at the origin (Theorem 3.1.), but we do not know if this disc is open or closed, and we do not know its radius. It is intriguing, however, that in the "canonical hyperbolic" case (*antipodal* fixed points on $\partial\mathbb{U}$) we *do* know all of this: here $W(C_\varphi)$ is the open disc centered at the origin whose radius is the spectral radius of C_φ (Theorem 3.2.). Unfortunately our method of proof does not survive a conformal conjugation.

It is this juxtaposition of the straightforward and the mysterious that makes numerical ranges of composition operators interesting to study; we hope readers of this paper will agree, and be motivated to attack the problems we leave open here. In a related paper [1] we consider the numerical ranges of composition operators induced by non-automorphic selfmaps of $\mathbb{U}$, focusing this time on the problem of when $0 \in W(C_\varphi)$—an apparently straightforward problem that evolves, like the present one, into an intriguing adventure. In [16] Valentin Matache determines (among other things) $W(C_\varphi)$ where φ is a constant multiple of a monomial, and a section of [22] relates numerical ranges and composition operators induced by inner functions.

2. PRELIMINARIES

We gather here some well known results about Hardy spaces and convex sets that will figure in the sequel.

2.1. MORE ON H^2. For each function $f(z) = \sum_{n=0}^{\infty} \hat{f}(n) z^n$ in H^2 there is a naturally associated trigonometric series $\sum_0^{\infty} \hat{f}(n) e^{in\theta}$ which, by the Riesz-Fisher theorem, is the Fourier series of some function $f^* \in L^2 = L^2(m, \partial\mathbb{U})$ (here m denotes arclength measure on $\partial\mathbb{U}$, normalized to have total mass one). Clearly the map $f \to f^*$ takes H^2 isometrically onto the closed subspace of L^2 consisting of functions whose Fourier coefficients of negative index vanish. The boundary function f^* turns out to be the natural extension of f to $\partial\mathbb{U}$, namely the radial limit function $f^*(\zeta) := \lim_{r\to 1-} f(r\zeta)$, where the limit exists for m almost every $\zeta \in \partial\mathbb{U}$ (see, e.g., [8, Chapter 1] or [18, Chapter 17] for the details). From now on we simply write $f(\zeta)$ instead of $f^*(\zeta)$ for each $\zeta \in \partial\mathbb{U}$ at which this radial limit exists, letting the context determine the meaning of the symbol f.

With these observations the norm and inner product in H^2 can be computed on the boundary of the unit circle as:

$$\|f\|^2 = \int_{\partial\mathbb{U}} |f|^2 \, dm \quad \text{and} \quad \langle f, g \rangle = \int_{\partial\mathbb{U}} f\overline{g} \, dm \qquad (f, g \in H^2). \tag{1}$$

2.2. THE SUPPORT FUNCTION OF A CONVEX SET. The following material is needed for the proof of Theorem 4.2.. Suppose E is a bounded convex subset of the plane. For $0 \le \theta < 2\pi$ define

$$p_{_E}(\theta) = \sup\{\mathrm{Re}\,(e^{-i\theta} z) : z \in E\}. \tag{2}$$

Note that for $z \in \mathbb{C}$ the number $\mathrm{Re}\,(e^{-i\theta} z)$ is the real dot product of the plane vectors $e^{i\theta}$ and z, i.e., the signed length of the projection of z in the direction of $e^{i\theta}$. Thus the set

$$\Pi_\theta := \{z \in \mathbb{C} : \mathrm{Re}\,(e^{-i\theta} z) \le p_{_E}(\theta)\}$$

is a closed half-plane that contains E and intersects ∂E. The boundary L_θ of Π_θ is called the *support line* of E perpendicular to $e^{i\theta}$. The magnitude of $p_{_E}(\theta)$ is the orthogonal distance from the origin to L_θ. The function $p_{_E} : [0, 2\pi) :\to \mathbb{R}$ defined by (2) is called the *support function* of E. The Hahn-Banach theorem insures that the closure of E is the intersection of all the half-planes Π_θ as θ runs from 0 to 2π, hence two bounded convex sets with the same support function have the same closures.

In our applications the set E will always contain the origin in its closure, in which case $p_{_E} \ge 0$. We will be particularly interested in the support function of a standard ellipse.

2.3. PROPOSITION. *Suppose $a, b > 0$ and E is the elliptical disc determined by the inequality $\frac{x^2}{a^2} + \frac{y^2}{b^2} \le 1$. Then $p_{_E}(\theta) = \sqrt{a^2 \cos^2\theta + b^2 \sin^2\theta}$ $\quad (0 \le \theta < 2\pi)$.*

Proof. We parameterize the boundary of E by the complex equation $z(t) = a\cos t + i\,b\sin t$, with $0 \le t < 2\pi$, so that the derivative $z'(t) = -a\sin t + i\,b\cos t$ is tangent to the boundary at the point $z(t)$, and therefore determines a support line L_θ for E at $z(t)$, where θ is the angle between the horizontal axis and the vector $b\cos t + i\,a\sin t$ which is normal to L_θ, and points away from E. Now $p_E(\theta)$ is the length of the orthogonal projection of the vector $z(t)$ onto this normal vector, i.e., the magnitude of the (real) dot product of $z(t)$ with the unit normal

$$n(t) = \frac{b\cos t + i\,a\sin t}{\sqrt{b^2\cos^2 t + a^2\sin^2 t}} \quad \text{so} \quad p_E(\theta) = \frac{ab}{\sqrt{b^2\cos^2 t + a^2\sin^2 t}}\ . \quad (3)$$

All that remains is to determine t in terms of θ. Since $n(t) = e^{i\theta}$ this relation is coded into the equations

$$\cos\theta = \frac{b\cos t}{\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}}\ , \qquad \sin\theta = \frac{a\sin t}{\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}}\ ,$$

which imply that $\tan t = \frac{b}{a}\tan\theta$. After a little calculation with right triangles this yields the equations

$$\cos t = \frac{a\cos\theta}{\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}}\ , \qquad \sin t = \frac{b\sin\theta}{\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}}\ ,$$

which in turn exhibit the denominator of the expression on the right-hand side of the second equation of (3) as $ab/\sqrt{a^2\cos^2\theta + b^2\sin^2\theta}$. The desired equation for $p_E(\theta)$ follows immediately. ∎

We note in closing that this result persists in the limiting case $b = 0$. In this case E is the real segment $[-a, a]$, for which the definition (2) of support function yields $p_E(\theta) \equiv a|\cos\theta|$.

3. PARABOLIC AND HYPERBOLIC AUTOMORPHISMS

3.1. THEOREM. *Suppose φ is a conformal automorphism of $\mathbb{U}$ that is either parabolic or hyperbolic. Then $W(C_\varphi)$ is a disc centered at the origin.*

Proof. (a) THE PARABOLIC CASE. Let $\zeta \in \partial\mathbb{U}$ be the unique fixed point that φ has in the Riemann sphere. The Möbius transformation $\tau(z) = i(\zeta + z)/(\zeta - z)$ maps $\mathbb{U}$ onto the upper halfplane Π^+, and takes ζ to ∞. The resulting Möbius transformation $\Phi = \tau \circ \varphi \circ \tau^{-1}$ maps Π^+ onto itself and fixes only ∞, so it must be a real translation: $\Phi(w) = w + a$ ($w \in \mathbb{C} \cup \{\infty\}$), with $a \in \mathbb{R}$. Now the composition operator C_Φ operating on the space $H(\Pi^+)$ of all functions holomorphic on Π^+ has, for each $\gamma \in \mathbb{C}$, an eigenfunction E_γ defined by: $E_\gamma(w) = e^{\gamma w}$ $(w \in \Pi^+)$, where the corresponding eigenvalue is $e^{\gamma a}$.

We are particularly interested in the eigenfunctions E_{it} for $t \geq 0$. Each one is a bounded analytic function on Π^+, and pulls back to a bounded eigenfunction $e_t = E_{it} \circ \tau$ of C_φ on H^2, corresponding to the eigenvalue e^{iat}. In fact this eigenfunction is *inner* because E_{it} itself has modulus one at every point of the real axis.

Suppose $w \in W(C_\varphi)$. We are going to show that $e^{it}w \in W(C_\varphi)$ for every $t > 0$, which will establish the desired circularity. We know that $w = \langle C_\varphi f, f\rangle$ for some unit vector $f \in H^2$. For $t > 0$ let $f_t = e_t \cdot f$ (pointwise product) and set $w_t = \langle C_\varphi f_t, f_t\rangle$. Because e_t is inner, f_t is, by (1), also a unit vector in H^2; hence $w_t \in W(C_\varphi)$. Now

$$C_\varphi f_t = (C_\varphi e_t) \cdot (C_\varphi f) = e^{iat} e_t \cdot C_\varphi f,$$

hence

$$w_t = e^{iat}\langle e_t \cdot C_\varphi f, \, e_t \cdot f\rangle = e^{iat} \int_{\partial\mathbb{U}} e_t \cdot C_\varphi f \cdot \overline{e_t} \cdot \overline{f} \, dm = e^{iat} \int_{\partial\mathbb{U}} C_\varphi f \cdot \overline{f} \, dm,$$

with the last line following from the fact that $|e_t| \equiv 1$ on $\partial\mathbb{U}\backslash\{\zeta\}$. Thus for each $t > 0$,

$$w_t = e^{iat}\langle C_\varphi f, \, f\rangle = e^{iat} w,$$

so w belongs to an origin-centered circle of points that lie in $W(C_\varphi)$. This completes the proof that $W(C_\varphi)$ is a disc centered at the origin.

(b) The hyperbolic case. The argument follows the same lines as the parabolic case, but is complicated by the fact that the eigenfunctions of C_φ are no longer inner.

We are given that φ is a Möbius transformation that maps $\mathbb{U}$ onto itself and fixes two distinct points ζ_0 and ζ_∞ of $\partial\mathbb{U}$. Let us choose our notation so that ζ_0 is the attracting fixed point of φ. The Möbius transformation $\tau_0(z) \equiv (\zeta_0 - z)/(\zeta_\infty - z)$ maps the unit circle to an extended line, taking ζ_0 to 0 and ζ_∞ to ∞. Thus for an appropriately chosen unimodular complex number ω the Möbius transformation $\tau = \omega \cdot \tau_0$ maps the unit disc to the upper half-plane and the unit circle to the extended real line. This time the mapping $\Phi = \tau \circ \varphi \circ \tau^{-1}$ fixes both the origin and ∞, and since it maps Π^+ onto itself it must be a positive dilation, say $\Phi(w) \equiv rw$ for some positive number $r \neq 1$. Because we have chosen ζ_0 to be attracting for φ, the origin must be attracting for Φ, so $0 < r < 1$. Let us call r the *dilation parameter* of both Φ and φ.

As in part (a) we begin with a convenient collection of eigenfunctions for $C_\Phi : H(\Pi^+) \to H(\Pi^+)$. For $\gamma \in \mathbb{C}$ let $E_\gamma(w) = w^\gamma = \exp\{\gamma \operatorname{Log} w\}$ $(w \in$

Π^+), where Log denotes the principal branch of the logarithm function. One checks easily that E_γ is an eigenfunction of C_Φ for the eigenvalue r^γ. Just as in the parabolic case, we will be particularly interested in the case of unimodular eigenvalues, i.e., γ pure imaginary.

We claim that each of the eigenfunctions E_{it} is bounded ($t \in \mathbb{R}$). In fact for $0 \neq w$ in the closed upper halfplane $\overline{\Pi^+}$ we have $|E_{it}(w)| = \exp\{-t \operatorname{Arg} w\}$, where $\operatorname{Arg} w$ is the principal branch of the argument of w, defined to take values in the interval $[0, \pi]$ on $\overline{\Pi^+}\backslash\{0\}$. Note that, in particular, $|E_{it}|$ is bounded on Π^+ (< 1 if $t > 0$ and $< e^{\pi|t|}$ if $t < 0$), $\equiv 1$ on the positive real axis, and $\equiv e^{-\pi t}$ on the negative real axis.

To pull these observations back to the unit disc, for $t \in \mathbb{R}$ set $e_t = E_{it} \circ \tau$, a bounded analytic function on $\mathbb{U}$ with $C_\varphi e_t = r^{it} e_t$. Let Γ_+ be the arc of the unit circle that gets taken by τ to the positive real axis and Γ_- that arc taken to the negative real axis. Then the work of the last paragraph shows that $|e_t|$ is bounded on $\mathbb{U}$, $\equiv 1$ on Γ_+, and $\equiv e^{-\pi t}$ on Γ_-.

Now down to business! Fix $0 \neq w \in W(C_\varphi)$; our goal is to show that $W(C_\varphi)$ contains the entire circle through w centered at the origin. Initially the argument will proceed exactly as the one for the parabolic case. We have $w = \langle C_\varphi f, f\rangle$ for some unit vector $f \in H^2$, and we set $f_t = e_t \cdot f$ for each $t \in \mathbb{R}$. Since e_t is bounded on $\mathbb{U}$ we have $f_t \in H^2$ and may proceed as before to calculate

$$C_\varphi f_t = (C_\varphi e_t) \cdot (C_\varphi f) = (r^{it} e_t) \cdot C_\varphi f,$$

so that

$$\langle C_\varphi f_t, f_t\rangle = r^{it} \langle e_t \cdot C_\varphi f, e_t f\rangle = r^{it} \int_{\partial \mathbb{U}} e_t \cdot C_\varphi f \cdot \overline{e_t} \cdot \overline{f} \, dm$$

Upon recalling the constant values that e_t assumes on Γ_+ and Γ_-, we have

$$\langle C_\varphi f_t, f_t\rangle = r^{it}(A + e^{-2\pi t} B), \tag{4}$$

where

$$A = \int_{\Gamma_+} C_\varphi f \cdot \overline{f} \, dm \qquad \text{and} \qquad B = \int_{\Gamma_-} C_\varphi f \cdot \overline{f} \, dm. \tag{5}$$

In order to produce a point of $W(C_\varphi)$ we have to divide the result of this calculation by

$$\|f_t\|^2 = C + e^{-2\pi t} D, \tag{6}$$

where

$$C = \int_{\Gamma_+} |f|^2 \, dm \qquad \text{and} \qquad D = \int_{\Gamma_-} |f|^2 \, dm \tag{7}$$

(note in particular that $C + D = \|f\|^2 = 1$), thus for each $t \in \mathbb{R}$,

$$w_t := \frac{\langle C_\varphi f_t,\, f_t\rangle}{\|f_t\|^2} = r^{it}\,\frac{A + e^{-2\pi t}B}{C + e^{-2\pi t}D}\ , \tag{8}$$

is a point of $W(C_\varphi)$. Because $w_0 = w$ we can view $\Omega_w := \{w_t : t \in \mathbb{R}\}$ as a curve in $\mathbb{C}$ that passes through w and lies entirely in the numerical range of C_φ.

Now observe that $|w_t|$ converges to $\rho_+ := |A|/C$ as $t \to +\infty$ and to $\rho_- := |B|/D$ as $t \to -\infty$. Because of the way Ω_w spirals around the origin, each of the circles $\{|z| = \rho_+\}$ and $\{|z| = \rho_-\}$ lies in its closure, hence by convexity the closures of the discs $\Delta_+ := \{|z| < \rho_+\}$ and $\Delta_- := \{|z| < \rho_-\}$ lie in $\overline{W}(C_\varphi)$, so by convexity the discs themselves lie in $W(C_\varphi)$.

Where is w relative to these discs? Because neither f nor $C_\varphi f$ is the zero-function, neither boundary function can vanish a.e. on a subset of $\partial\mathbb{U}$ having positive measure. In particular, neither C nor D is zero. There are three possibilities:

$$\frac{|B|}{D} < \frac{|A|}{C}, \quad \frac{|B|}{D} > \frac{|A|}{C}, \quad \text{or} \quad \frac{|B|}{D} = \frac{|A|}{C}.$$

In the first case

$$|w| \le \frac{|A| + |B|}{C + D} < \frac{|A|}{C} = \rho_+, \tag{9}$$

so $w \in \Delta_+$. In the second case the same kind of estimate shows that $|w| < \rho_-$, hence $w \in \Delta_-$. Thus in both cases the circle of radius $|w|$ centered at the origin lies entirely in $W(C_\varphi)$.

For the third case we have

$$\frac{|A| + |B|}{C + D} = \frac{|B|}{D} = \frac{|A|}{C} := \rho > 0,$$

the strict positivity of ρ coming from the fact that $A + B = w$, where we have assumed that $w \neq 0$. Thus $\Delta_+ = \Delta_- := \Delta$, and the inequality (9) is replaced by $|w| \le \rho$, so w lies in the closure of Δ. Now we have to consider two subcases. If $|w| < \rho$ then, just as in the previous cases, $w \in \Delta \subset W(C_\varphi)$. If, on the other hand, $|w| = \rho$, then there is equality in *both* inequalities of (9), and the first dictates that the complex numbers A and B point in the same direction; $A = \omega|A|$ and $B = \omega|B|$ for some unimodular ω. It follows that $w_t = r^{it}\omega\rho$ for every real t, hence Ω_w, which we recall lies entirely in $W(C_\varphi)$, is the circle of radius $\rho = |w|$. This completes the proof that $W(C_\varphi)$ is a disc centered at the origin. We remark that were it not for this last subcase, we could conclude that $W(C_\varphi)$ is actually open (see Theorem 3.2. below for more on “openness”). ▮

For φ a hyperbolic or a parabolic automorphism the result above shows that $W(C_\varphi)$ is either open or it is closed, but—with one exception—we do not know which, and we have little information about the radius of $W(C_\varphi)$, other than the general fact [11, Theorem 1.3–1, page 9] that the numerical radius of a Hilbert space operator T must lie between $\|T\|/2$ and $\|T\|$. Thus the radius of $\overline{W}(C_\varphi)$ is bounded above and below by constant multiplies of $(1-|\varphi(0)|)^{-1}$ (see the discussion centered on (11) below). Additional comments about the radius of $W(C_\varphi)$ may be found in Section 5.2. below.

The exceptional case is that of a "canonical" hyperbolic automorphism, i.e., one with antipodal fixed points ($\zeta_\infty = -\zeta_0$ in part (b) of the proof above). Here we have a complete characterization of the numerical range.

3.2. Theorem. *If φ is a canonical hyperbolic automorphism of $\mathbb{U}$ with dilation parameter $0 < r < 1$, then $W(C_\varphi)$ is the open disc of radius $1/\sqrt{r}$ centered at the origin.*

Proof. We may, without loss of generality, take $+1$ to be the attractive fixed point of φ and -1 to be the repulsive one. Were this not the case initially, a rotational conjugation of φ would make it so, and this would induce a unitary equivalence (which does not alter the numerical range) of the respective composition operators. With this normalization $\varphi = \tau^{-1}\circ\Phi\circ\tau$ where $\Phi(w)\equiv rw$ and the map τ, this time defined by $\tau(z) = (1-z)/(1+z)$ for $z\in\mathbb{C}$, maps the unit disc to the open right half plane and the unit circle to the extended imaginary axis. Upon working out the arithmetic one sees quickly that for $z\in\overline{\mathbb{U}}$,

$$\varphi(z) = \frac{\rho+z}{1+\rho z} \quad \text{where} \quad \rho := \frac{1-r}{1+r}\ . \tag{10}$$

In [17] Nordgren showed that for any automorphism φ of the unit disc:

$$\|C_\varphi\| = \sqrt{\frac{1+|\varphi(0)|}{1-|\varphi(0)|}} \tag{11}$$

(and that furthermore the right hand side of this equation is an upper bound for the norm of *any* composition operator C_φ on H^2, see also [21, page 16]). In our canonical hyperbolic case $\varphi(0) = \rho$, so (10) and (11) yield $\|C_\varphi\| = 1/\sqrt{r}$. In particular, $W(C_\varphi)$ lies in the closed disc $\{|w| \le 1/\sqrt{r}\}$. On the other hand, referring back to the discussion of eigenvalues and eigenvectors that played an important role in part (b) of the proof of the previous theorem, we see that for each $\gamma\in\mathbb{C}$ the function $e_\gamma(z) := \tau(z)^\gamma$ $(z\in\mathbb{U})$ is an eigenvector of $C_\varphi : H(\mathbb{U})\to H(\mathbb{U})$ that corresponds to the eigenvalue r^γ. Now $e_\gamma\in H^2$ if and only if $|\mathrm{Re}\,\gamma| < 1/2$, so in particular all the corresponding eigenvalues, which form an open annulus $\{\sqrt{r} < |w| < 1/\sqrt{r}\}$, lie in $W(C_\varphi)$.

Before proceeding further we remark that to this point the argument simply repeats a portion of the one Nordgren used to find the spectrum of C_φ. Now we strike out on our own! By the Toeplitz-Hausdorff theorem $W(C_\varphi)$ must contain the entire open disc $\Delta := \{|w| < 1/\sqrt{r}\}$. But we have already seen that $\overline{W}(C_\varphi)$ lies in the closed disc $\overline{\Delta}$, so it coincides with this disc. Thus to complete the proof of our theorem (which in the current notation states that $W(C_\varphi) = \Delta$) we need only show that no point of $\partial\Delta$ belongs to $W(C_\varphi)$.

For this it will be enough to show that no point on $\partial\Delta$ is an eigenvalue of C_φ. Indeed, once this has been established, then if $|w| = 1/\sqrt{r}$ and $w = \langle C_\varphi f,\, f\rangle$ for some unit vector $f \in H^2$, we would have

$$\|C_\varphi\| = \frac{1}{\sqrt{r}} = |w| = |\langle C_\varphi f,\, f\rangle| \leq \|C_\varphi f\|\,\|f\| \leq \|C_\varphi\|.$$

Thus there would be equality throughout this display, and in particular in the first inequality, which is the Cauchy-Schwartz inequality. This could only happen if $C_\varphi f$ were a constant multiple of f. By its definition, w would then have to be this constant, making w an eigenvalue of C_φ—a contradiction.

So all depends on showing that the boundary of Δ contains no eigenvalue of C_φ. Fix $w \in \partial\Delta$, and suppose that $C_\varphi f = wf$ for some $f \in H^2$. Our goal is to show that $f \equiv 0$ on $\mathbb{U}$. To this end fix $z \in \mathbb{U}$ and observe that for each positive integer n,

$$w^n f(z) = C_\varphi^n f(z) = f(\varphi_n(z)) \tag{12}$$

where φ_n is the composition of φ with itself n times. Now on one hand there is this standard estimate for functions in H^2:

$$|f(p)| = \mathrm{o}\left(\frac{1}{\sqrt{1-|p|}}\right) \qquad \text{as } |p| \to 1- \tag{13}$$

(the corresponding "big-oh" estimate, $|f(p)| \leq \|f\|/\sqrt{1-|p|^2}$, follows via the Cauchy-Schwarz inequality from the power series representation of f, and the "little-oh" improvement is an easy consequence of this and the density of polynomials in H^2). Upon letting $p = \varphi_n(z)$ in (13) and substituting the result into (12) we obtain (because $\varphi_n(z) \to 1$ as $n \to \infty$),

$$r^{-n/2}|f(z)| = \mathrm{o}\left(\frac{1}{\sqrt{1-|\varphi_n(z)|}}\right) \qquad (n \to \infty). \tag{14}$$

To finish the proof we need to get serious about $|\varphi_n(z)|$. Because φ_n is a canonical hyperbolic automorphism with the same fixed points as φ and dilation parameter r^n (i.e., $\varphi(z) \equiv \tau^{-1}(r\tau(z))$, so $\varphi_n(z) \equiv \tau^{-1}(r^n\tau(z)))$, an explicit formula for φ_n is given by (10) with ρ replaced by

$$\rho_n := (1 - r^n)/(1 + r^n) \qquad (n = 1, 2, \ldots). \tag{15}$$

It follows that:

$$\begin{aligned} 2(1 - |\varphi_n(z)|) \ &\geq \ 1 - |\varphi_n(z)|^2 = \frac{(1 - \rho_n^2)(1 - |z|^2)}{|1 - \rho_n z|^2} \\ &\geq \ \frac{1 - |z|^2}{(1 + |z|)^2}\,(1 - \rho_n) \geq \frac{1 - |z|}{1 + |z|}\,r^n, \end{aligned}$$

where the equality on the first line is a standard identity that can be found, e.g., in [10, page 3] or [21, §4.3, pp. 59–60], and the last inequality follows from (15). This estimate and (14) imply that $r^{-n/2}|f(z)| = \mathrm{o}\left(r^{-n/2}\right)$ as $n \to \infty$, which can only happen if $f(z) = 0$. ▮

We conjecture that the numerical range of *any* composition operator induced by a hyperbolic automorphism is an open disc, but have, as yet, no proof of this. The problem is that for non-canonical hyperbolic automorphisms the norm of C_φ is no longer the same as the spectral radius. In fact an amusing calculation shows that if we keep the attractive fixed point at $+1$ and hold the dilation parameter r fixed, then as the repulsive fixed point tends to $+1$, the point $\varphi(0) \to 1$, and therefore by (11), $\|C_\varphi\| \to \infty$.

We can, however, show that if φ is a hyperbolic *non-automorphism* of $\mathbb{U}$ that is "canonical" then $W(C_\varphi)$ is an open disc centered at the origin. See §5.3. below for more details.

4. ELLIPTIC AUTOMORPHISMS

In the Introduction we discussed the numerical range of C_φ where φ is a "model" elliptic automorphism, i.e., an "ω-rotation" $z \to \omega z$ where ω is a complex number of modulus one. The general elliptic automorphism fixes a point $p \in \mathbb{U}$, and has the form $\varphi = \tau^{-1} \circ \Phi \circ \tau$ where Φ is an ω-rotation, and τ is a conformal automorphism that takes p to the origin. Let us call ω the *rotation parameter* of φ.

4.1. Theorem. *Suppose φ is an elliptic automorphism of $\mathbb{U}$ with rotation parameter ω. If ω is not a root of unity then $\overline{W}(C_\varphi)$ is a disc centered at the origin.*

Proof. We continue with the notation introduced just before the statement of the theorem. For each non-negative integer n the monomial $E_n(z) = z^n$ is a bounded eigenfunction of C_Φ with eigenvalue ω^n. The same is therefore true (with C_φ in place of C_Φ) of $e_n = E_n \circ \tau$, which is an inner function.

Fix $w \in W(C_\varphi)$, and a unit vector $f \in H^2$ for which $w = \langle C_\varphi f, f\rangle$. As is now our custom, let $f_n = e_n \cdot f$ for each non-negative integer n, noting that $f_0 = f$. Now just as in our parabolic argument, because e_n is an eigenfunction for ω^n that is *inner*,

$$\langle C_\varphi f_n, f_n\rangle = \omega^n \langle C_\varphi f, f\rangle = \omega^n w. \tag{16}$$

Because ω is not a root of unity, the points $\{\omega^n\}_0^\infty$ are dense in the unit circle, so w as a limit point of the set $\{\omega^n w\}_0^\infty$, each point of which, as (16) reveals, belongs to $W(C_\varphi)$. Thus the closure of $W(C_\varphi)$ contains the circle through w that is centered at the origin, hence this closure is itself a disc with center at the origin. ▮

In case ω is a root of unity we strongly suspect that $\overline{W}(C_\varphi)$ is not a disc, but up to now have a proof only for the case $\omega = -1$. In this case $C_\varphi^2 = I$, so our result follows from the more general one below.

4.2. Theorem. *Suppose $T \neq \pm I$ is an operator on the Hilbert space $\mathcal{H}$ with $T^2 = I$. Then $\overline{W}(T)$ is a (possibly degenerate) elliptical disc with foci at ± 1. In particular, it is not a circular disc.*

Proof. We compute the support function p_T of $W(T)$ in this standard fashion:

$$\begin{aligned} p_T(\theta) &:= \sup\{\mathrm{Re}\,(e^{-i\theta} z) : z \in W(T)\} \\ &= \sup\{\mathrm{Re}\,(e^{-i\theta}\langle Tf, f\rangle : f \in \mathcal{H}, \|f\| = 1\} \\ &= \sup\{\langle H_\theta f, f\rangle : f \in \mathcal{H}, \|f\| = 1\}, \end{aligned}$$

where $H_\theta := \mathrm{Re}\,(e^{-i\theta}T) = \frac{1}{2}(e^{-i\theta}T + e^{i\theta}T^*)$.
Claim: $W(T) = -W(T)$.

We defer the proof of this claim so as not to interrupt the flow of argument.

Since H_θ is a self-adjoint operator on $\mathcal{H}$ this claim and the last calculation show that for each $0 \le \theta < 2\pi$,

$$p_T(\theta) = \sup\{|\langle H_\theta f, f\rangle| : f \in \mathcal{H}, \|f\| = 1\} = \|H_\theta\|.$$

Now for every unit vector $f \in \mathcal{H}$ a routine computation shows that for each θ,

$$\|H_\theta f\|^2 = \frac{1}{4}\left\{\|Tf\|^2 + \|T^* f\|^2 + 2\,\mathrm{Re}\,(e^{-2i\theta}\langle T^2 f, f\rangle)\right\} \tag{17}$$

which, upon setting

$$\beta(t) = \frac{1}{2}\sup\{\|Tf\|^2 + \|T^* f\|^2 : f \in \mathcal{H}, \|f\| = 1\}$$

and noting our hypothesis that $T^2 = I$, yields

$$\begin{aligned} p_T(\theta) &= \sqrt{\frac{\beta(T)}{2} + \frac{\cos 2\theta}{2}} \\ &= \sqrt{\frac{\beta(T)+1}{2}\cos^2\theta + \frac{\beta(T)-1}{2}\sin^2\theta} \\ &= \sqrt{a^2\cos^2 t + b^2\sin^2 t}, \end{aligned}$$

where $a = \sqrt{(\beta(T)+1)/2}$ and $b = \sqrt{(\beta(T)-1)/2}$. Our definition of b depends on the fact that $\beta(T) \geq 1$. To see that this is so, note that $T^2 = I$ implies that T has 1 or -1 as an eigenvalue, so if f is a corresponding unit eigenvector then $\|Tf\| = 1$, hence

$$\|T^*f\| \geq |\langle T^*f,\, f\rangle| = \langle f,\, Tf\rangle = 1.$$

By the work of Proposition 2.3. and the remark following its proof, p_T is the support function of either the interval $[-1, 1]$ (if $\beta(T) = 1$) or the nondegenerate elliptical disc $\frac{x^2}{a^2} + \frac{y^2}{b^2} \leq 1$ (if $\beta > 1$). This latter disc has foci at $\pm\sqrt{a^2 - b^2} = \pm 1$, so we are done pending verification of the claim that $-W(T) = W(T)$.

For this it suffices to show that $-W(T) \subset W(T)$. To get started, note that 1 must be an eigenvalue for T: for any $g \in \mathcal{H}$, $T(Tg+g) = Tg+g$ and $Tg+g$ must be nonzero for some $g \in \mathcal{H}$ because $T \neq -I$. Similarly, -1 must also be an eigenvalue. Let $\mathcal{M}$ denote the eigenspace for T corresponding to its eigenvalue 1.

Suppose $w \in W(T)$ so that there is a vector $f \in \mathcal{H}$ of norm 1 such that $w = \langle Tf, f\rangle$. Decompose f as a sum of a vector $g \in \mathcal{M}$ and a vector $h \in \mathcal{M}^\perp$. If $h = 0$, then $w = 1$ and $-w = -1$ is in $W(T)$ since -1 is an eigenvalue. If $g = 0$, then

$$w = \langle Th, h\rangle = \langle (Th+h) - h, h\rangle = \langle (Th+h), h\rangle - \|h\|^2 = -1,$$

where the last equality follows from the fact that $(Th + h) \in \mathcal{M}$. Once again we have that $-w = 1$ is in $W(T)$. Thus, we may assume that both components g and h of f are nonzero elements of $\mathcal{H}$.

Let $s = \frac{\|h\|}{\|g\|}g - \frac{\|g\|}{\|h\|}h$ and observe that $\|s\|^2 = \|h\|^2 + \|g\|^2 = \|f\|^2 = 1$. Then because $g \in \mathcal{M}$ we have

$$\begin{aligned} \langle Ts, s\rangle &= \left\langle \frac{\|h\|}{\|g\|}g - \frac{\|g\|}{\|h\|}Th, \frac{\|h\|}{\|g\|}g - \frac{\|g\|}{\|h\|}h \right\rangle \\ &= \|h\|^2 - \langle Th, g\rangle + \frac{\|g\|^2}{\|h\|^2}\langle Th, h\rangle \end{aligned}$$

$$\begin{aligned} &= (1-\|g\|^2) - \langle Th, g\rangle + \frac{1-\|h\|^2}{\|h\|^2}\langle Th, h\rangle \\ &= -\left[\|g\|^2 + \langle Th, g\rangle + \langle Th, h\rangle\right] + (1 + \frac{1}{\|h\|^2}\langle Th, h\rangle) \\ &= -[\langle T(g+h), g+h\rangle] + \frac{\langle Th+h, h\rangle}{\|h\|^2}. \end{aligned}$$

Upon recalling that $f = g+h$ and $Th+h \in \mathcal{M}$, we see that the last line of the calculation above is just $-\langle Tf, f\rangle = -w$, as desired. ▮

4.3. Remark. In this last result: *The degenerate case* $\overline{W}(T) = [-1, 1]$ *occurs if and only if the involution* T *is self-adjoint*. To see why this is true, note that if T is self-adjoint then $\beta(T) = 1$, hence $b = 0$ in the proof above, so that $\overline{W}(T) = [-1, 1]$. Conversely, whenever $W(T) \subset \mathbb{R}$ then T is self-adjoint [11, Theorem 1.2-2].

4.4. Corollary. *If* φ *is an elliptic automorphism of* $\mathbb{U}$ *with multiplier* -1*, then* $\overline{W}(C_\varphi)$ *is a (possibly degenerate) ellipse with foci* ± 1*. The degenerate case occurs if and only if* $\varphi(0) = 0$*, in which case* $\varphi(z) \equiv -z$*.*

5. FINAL REMARKS

5.1. The essential numerical range. For an operator T on a Hilbert space this is the intersection $W_e(T)$ of all the sets $\overline{W}(T+K)$ where K ranges through all compact operators on the space. Clearly $W_e(T)$ is compact, convex, and invariant under compact perturbations of T. A result of Fillmore, Stampfli, and Williams [9] characterizes $W_e(T)$ as the set of points w for which there exists a weakly null sequence $\{x_n\}$ of unit vectors such that $\langle Tx_n, x_n\rangle \to w$. A careful analysis of the proofs given in Sections 3 and 4 (see [2] for the details) discloses that for each of the operators in question, each point of the numerical range is obtained as precisely such a limit, hence: $\overline{W}(C_\varphi) = W_e(C_\varphi)$ *for each conformal automorphism* φ *of* $\mathbb{U}$.

5.2. The numerical radius. For an operator T on a Hilbert space this is

$$w(T) := \sup\{\langle Tx, x\rangle : \|x\| = 1\}.$$

We have already mentioned that $\|T\|/2 \le w(T) \le \|T\|$. Another lower bound for $w(T)$ is furnished by (17), which is valid for any Hilbert space operator. Upon setting $T = C_\varphi$, $\theta = 0$ and $f \equiv 1$ in (17) we obtain

$$w(C_\varphi) \geq \frac{1}{2}\sqrt{3 + \frac{1}{1 - |\varphi(0)|^2}}$$

Note that the quantity on the right exceeds 1 whenever $\varphi(0) \neq 0$, and by using (11) one sees readily that it exceeds $\|C_\varphi\|/2$ whenever $|\varphi(0)| < (\sqrt{13} - 1)/4$. Clearly this is just the beginning; more effort should produce better estimates of the numerical radius for the composition operators studied here.

5.3. More open discs. The proof of Theorem 3.2. also establishes that $W(C_\varphi)$ is an open disc for a certain class of hyperbolic linear fractional selfmaps of $\mathbb{U}$ that are *not automorphisms*. These are the hyperbolic non-automorphisms φ that have attractive fixed point (necessarily) on $\partial\mathbb{U}$ and repulsive one either at ∞ (e.g., $\varphi(z) \equiv (1+z)/2$) or on the ray through the origin emanating from the attractive fixed point (hence in $\mathbb{C}\backslash\overline{\mathbb{U}}$, with $\mathbb{U}$ lying between the two fixed points).

For such maps φ the crucial facts used in the proof of Theorem 3.2. continue to hold for C_φ, namely:

1. The spectrum of C_φ contains an annulus of eigenvalues centered at the origin having outer radius equal to the spectral radius of C_φ, and this equal to $1/\sqrt{r}$, where r is the derivative of φ at its attractive fixed point (see [3, Theorem 2.1 and Theorem 4.5]).
2. The norm and spectral radius of C_φ coincide (because, e.g., C_φ^* is subnormal; see [4, Theorem 7]).
3. No complex number of modulus $1/\sqrt{r} = \|C_\varphi\|$ can be an eigenvalue of C_φ (by estimates similar to those in the last paragraph of the proof of Theorem 3.2.).

Thus the proof of Theorem 3.2. actually yields this:

5.4. Theorem. *If φ is any hyperbolic linear fractional selfmap of $\mathbb{U}$ with repulsive fixed point either at ∞ or on the ray through the origin that emanates from the attractive fixed point, then $W(C_\varphi)$ is the open disc $\{|w| < 1/\sqrt{r}\}$, where r is the derivative of φ at its attractive fixed point.*

REFERENCES

1. P. S. Bourdon and J. H. Shapiro, Accretive composition operators, *In preparation*.
2. P. S. Bourdon and J. H. Shapiro, What is the numerical range of a composition operator?, unpublished preprint available from "Preprints" link at www.math.msu.edu/∼shapiro.

3. C. C. Cowen, Composition operators on H^2, *J. Operator Theory* **9** (1983), 77–106.

4. C. C. Cowen and T. L. Kriete, Subnormality and composition operators on H^2, *J. Funct. Anal.* **81** (1988), 298–319.

5. C. C. Cowen and B. D. MacCluer, "Composition Operators on Spaces of Analytic Functions," CRC Press, Boca Raton, 1995.

6. C. Davis, The Toeplitz-Hausdorff theorem explained, *Canad. Math. Bull.*, **14** (1971), 245–246.

7. N. P. Dekker, "Joint Numerical Range and Joint Spectrum of Hilbert Space Operators," Thesis, Free University of Amsterdam, 1969.

8. P. L. Duren, "Theory of H^p Spaces," Academic Press, New York, 1970.

9. P. A. Fillmore, J. G. Stampfli, and J. P. Williams, On the essential numerical range, the essential spectrum, and a problem of Halmos, *Acti Sci. Math. (Szeged)* **33** (1972), 179–192.

10. J. B. Garnett, "Bounded Analytic Functions," Academic Press, New York, 1981.

11. K.E. Gustafson and D.K.M. Rao, "Numerical Range, The Field of Values of Linear Operators and Matrices," Springer-Verlag, New York, 1997.

12. F. Jafari et al., editors, "Studies on Composition Operators," *Contemp. Math. Vol. 213*, American Math. Soc. 1998.

13. P. R. Halmos, "A Hilbert Space Problem Book," Second Edition, Springer-Verlag, New York 1982.

14. F. Hausdorff, Der Vervorrat einer Bilinearform, *Math. Z.* **3** (1919), 314–316.

15. J. E. Littlewood, On inequalities in the theory of functions, *Proc. London Math. Soc.* **23** (1925), 481–519.

16. V. Matache, Numerical ranges of composition operators, *preprint* 1999.

17. E. A. Nordgren, Composition operators, *Canadian J. Math.* **20** (1968), 442–449.

18. W. Rudin, "Real and Complex Analysis," Third edition, *McGraw-Hill, New York*, 1987.

19. J. V. Ryff, Subordinate H^p functions, *Duke Math. J.* **33** (1966), 347–354.

20. H. J. Schwartz, "Composition operators on H^p" Thesis, Univ. of Toledo, 1969.

21. J. H. Shapiro, "Composition Operators and Classical Function Theory," *Springer-Verlag*, 1993.

22. J. H. Shapiro, What do composition operators know about inner functions?, *Monatshefte für Mathematik*, to appear.

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