Linear Fractional Maps of the Ball and Their Composition Operators

Carl C. Cowen*
Purdue University, West Lafayette, Indiana 47907
cowen@math.purdue.edu

Barbara D. MacCluer†
University of Virginia, Charlottesville, Virginia 22903
maccluer@virginia.edu

Acta Sci. Math. (Szeged) **66**(2000), 351–376.

AMS Classification: Primary 47B38, 32A30; Secondary 30C45

Abstract

In this paper, we describe a class of maps of the unit ball in $\mathbf{C}^N$ into itself that generalize the automorphisms and deserve to be called linear fractional maps. They are special cases or generalizations of the linear fractional maps studied by Kreĭn and Smul'jan, Harris and others. As in the complex plane, a linear fractional map on $\mathbf{C}^N$ is represented by an $(N+1)\times(N+1)$ matrix. Basic connections between the properties of the map and the properties of this matrix viewed as a linear transformation on an associated Kreĭn space are established. These maps are shown to induce bounded composition operators on the Hardy spaces $H^p(B_N)$ and some weighted Bergman spaces and we compute the adjoints of these composition operators on these spaces. Finally, we solve Schroeder's equation $f \circ \varphi = \varphi'(0)f$ when φ is a linear fractional self-map of the ball fixing 0.

*Supported by National Science Foundation grants DMS-9206965 and DMS-9500870.
†Supported by National Science Foundation grants DMS-9300525 and GER-9350040.

1 Introduction

Linear fractional maps of the unit disk into itself and the automorphisms of the disk play a basic role in the study of composition operators on spaces of analytic functions on the unit disk. We would expect the analogues of these maps in higher dimensions to play a similar role in the study of composition operators on spaces of analytic functions on the unit ball in $\mathbf{C}^N$. The goal of this paper is to introduce a class of maps that we hope will play this role, to establish some of their properties, and to prove basic facts about their composition operators. At the very least, because there are few concrete examples of composition operators in spaces of several variable functions, we hope that these examples will begin to outline the differences between the one variable and the several variable theory.

The unit ball B_N in $\mathbf{C}^N$ is the set $\{z : |z| < 1\}$ and the unit sphere is the set $\{z : |z| = 1\}$. If $\mathcal{H}$ is a Hilbert space of analytic functions on the unit ball and φ is an analytic map of the ball into itself, the composition operator C_φ is the operator given by $C_\varphi f = f \circ \varphi$ for f in $\mathcal{H}$. We will be most interested in composition operators induced by linear fractional maps acting on the Hardy space $H^2(B_N)$ and the Bergman space $A^2(B_N)$, but much of what we do will apply to other spaces as well.

Definition A map φ will be called a *linear fractional map* if

$$\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$$

where A is an $N \times N$ matrix, B and C are (column) vectors in $\mathbf{C}^N$, and D is a complex number. We will regard z as a column vector also and $\langle \cdot, \cdot \rangle$ denotes the usual Euclidean inner product on $\mathbf{C}^N$.

For example, a linear transformation is a linear fractional map ($Tz = (Tz + 0)(\langle z, 0\rangle + 1)^{-1}$) and a translation is a linear fractional map ($z + p = (Iz + p)(\langle z, 0\rangle + 1)^{-1}$); thus an *affine map* that is a linear transformation followed by a translation is a linear fractional map $((Tz + p)/(\langle z, 0\rangle + 1)$.

Clearly, the domain of a linear fractional map is the set of z in $\mathbf{C}^N$ for which $\langle z, C\rangle + D \neq 0$. For our purposes, we want the domain of φ to include the closed ball; since $z = -DC/|C|^2$ is a zero of $\langle z, C\rangle + D$, this means we will require $| - DC/|C|^2| > 1$ or, equivalently, $|D| > |C|$. Conversely, if $|D| > |C|$, then by the Cauchy-Schwarz inequality we will have $\langle z, C\rangle + D \neq 0$ for z in the closed ball. In particular, D is non-zero for the linear fractional maps we consider.

Identifying a 1×1 matrix with its entry, we occasionally write $\langle z, C\rangle = C^* z$. For example, using this identification we can see that a linear fractional map is constant if (and only if) $A = BC^*/D$. We will usually avoid the case of constant maps.

In order to use tools from the theory of Kreĭn spaces, we will sometimes identify $\mathbf{C}^N$ with equivalence classes of points in $\mathbf{C}^{N+1}$. If $v = (v_1, v_2)$ where v_1 is in $\mathbf{C}^N$ and $v_2 \neq 0$ is in $\mathbf{C}$, identify v with v_1/v_2; in particular, $z \leftrightarrow (z, 1)$. We introduce a Kreĭn space structure on $\mathbf{C}^{N+1}$ by letting $[v, w] = \langle Jv, w\rangle$ where $\langle \cdot, \cdot \rangle$ is the usual (Euclidean) inner product on $\mathbf{C}^{N+1}$ and

$$J = \begin{pmatrix} I & 0 \\ 0 & -1 \end{pmatrix}$$

In this setting, v represents a point of the unit sphere if and only if $|v_1| = |v_2|$ which occurs if and only if $[v, v] = |v_1|^2 - |v_2|^2 = 0$ and v represents a point of the unit ball if and only if $[v, v] < 0$.

Definition If $\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ is a linear fractional map, the matrix

$$m_\varphi = \begin{pmatrix} A & B \\ C^* & D \end{pmatrix}$$

will be called a *matrix associated with* φ.

Notice that if φ is a linear fractional map with $\varphi(z) = w$ and v is a point of $\mathbf{C}^{N+1}$ associated with z, then $m_\varphi v$ is associated with the point w and vice versa. If φ_1 and φ_2 are linear fractional maps, direct computation of $\varphi_1 \circ \varphi_2$ and $m_{\varphi_1} m_{\varphi_2}$ shows that $\varphi_1 \circ \varphi_2$ is a linear fractional map with associated matrix $m_{\varphi_1 \circ \varphi_2} = m_{\varphi_1} m_{\varphi_2}$. In particular, if φ has a linear fractional inverse, $m_{\varphi^{-1}} = (m_\varphi)^{-1}$ and if m_φ is invertible, φ has a linear fractional inverse.

Before we can begin to exploit the representation of the linear fractional maps of $\mathbf{C}^N$ as linear maps on $\mathbf{C}^{N+1}$, we should settle the question of when two linear fractional maps are the same, that is when does

$$(Az + B)(\langle z, C\rangle + D)^{-1} = (A'z + B')(\langle z, C'\rangle + D')^{-1}$$

The two maps are trivially the same if there is $\lambda \neq 0$ for which $A' = \lambda A$, $B' = \lambda B$, $C' = \overline{\lambda} C$, and $D' = \lambda D$. A somewhat tedious computation, which we omit, shows that the converse holds for non-constant maps.

In the second section of the paper, we review the connection between the Kreĭn space structure we have introduced on $\mathbf{C}^{N+1}$ and properties of linear fractional maps. Specifically, we show that φ maps the unit ball into itself if and only if m_φ is a multiple of a Kreĭn contraction, and φ maps the ball onto itself if and only if m_φ is a multiple of a Kreĭn isometry. In the latter case we have precisely the automorphisms of the ball, and we discuss various classes of automorphisms from the point of view of their matrix forms.

In Section 3, we establish several geometric facts about linear fractional maps. For example, we show that linear fractional maps take balls to ellipsoids and affine sets to affine sets, and we establish a correspondence between fixed points and fixed sets of linear fractional maps and invariant subspaces of the associated linear transformation. Examples are given to illustrate the results of Sections 2 and 3.

Section 4 contains the applications of these ideas to the study of composition operators with linear fractional symbol. For a variety of spaces, these operators are always bounded and they are compact if and only if they map the closed unit ball into the open unit ball. We introduce the adjoint mapping of a linear fractional map and use it to find the adjoint of composition operators on a class of Hilbert spaces of analytic functions including the classical Hardy and Bergman Hilbert spaces. The adjoint map also plays a role in the final section, where the several variable analogue of Schroeder's functional equation, for linear fractional maps, is discussed.

A number of authors have developed the theory of automorphisms of the unit ball, or more generally, of bounded symmetric domains in $\mathbf{C}^N$, for example, Arazy [1], Harris [9, 10], Rudin [22], and Stein [23]. Some of these have used an indefinite form, that is, a Kreĭn space approach, in their development. The linear fractional maps of the ball considered here are special cases of those considered by Kreĭn and Smul'jan in "On linear-fractional transformations with operator coefficients" ([15]) and by Harris in his work "Linear fractional transformations of circular domains in operator spaces," [9], because the unit ball in $\mathbf{C}^N$ is an operator space, and our approach is same as in these works. Our goal in studying linear fractional maps is to provide a foundation for the study of linear fractional composition operators on spaces of analytic functions in the ball, and this influences the issues we consider in Sections 2 and 3. It is clear that the techniques here will extend to maps of a variety of domains in $\mathbf{C}^N$ into other domains or themselves and the composition operators associated with those maps.

The authors would like to thank Mike Dritschel for several valuable suggestions about this work.

2 Maps on the Ball: Isometries and Contractions

In this section we exploit the representation of linear fractional maps as $(N+1)\times(N+1)$ matrices acting on a Kreĭn space to investigate linear fractional maps of the ball B_N into and onto itself. Recall that a linear transformation T of a Kreĭn space is called a *contraction* (or *Kreĭn contraction* or *J–contraction*) if $[Tv,Tv]\le[v,v]$ for all vectors v, and is an isometry if $[Tv,Tv]=[v,v]$ for all v. We will see that the non-constant linear fractional transformations φ which map B_N into B_N are precisely those for which the matrix m_φ is a non-zero multiple of a contraction on our Kreĭn space, and φ maps B_N onto B_N if and only if m_φ is a multiple of an isometry. Since all automorphisms of the unit ball are linear fractional maps (see [7, page 99]), the latter result gives a characterization of the automorphism group of B_N . These results are not new; indeed they are special cases of work of Kreĭn and Smul'jan ([15]) done in the more general setting of linear fractional maps on operator spaces. However, to keep our exposition reasonably self-contained and specific to the concrete setting of maps of B_N into itself we will give proofs appropriate to our setting.

We begin with two preparatory lemmas that have been modified from the presentation of "plus" operators found in Bognar's book [2, Section II.8] . In Lemmas 1 and 2, the phrase "indefinite scalar product" in the hypothesis means that the scalar products should *not* be positive or negative semi-definite.

Lemma 1 *Suppose* $[\cdot,\cdot]_1$ *and* $[\cdot,\cdot]_2$ *are indefinite scalar products on the vector space* $\mathcal{V}$ *such that* $[x,x]_1=0$ *implies* $[x,x]_2\le 0$. *If* $[y,y]_1>0$ *and* $[z,z]_1<0$ *then*

$$\frac{[y,y]_2}{[y,y]_1}\le\frac{[z,z]_2}{[z,z]_1}$$

Proof. Suppose the conclusion is false, that is, suppose $[y,y]_1=1$ and $[z,z]_1=-1$ and $[y,y]_2+[z,z]_2>0$. Let $x=\epsilon y+z$ for $|\epsilon|=1$. Then

$$[x,x]_1=2\mathrm{Re}\,\epsilon[y,z]_1$$

and

$$[x,x]_2=[y,y]_2+2\mathrm{Re}\,\epsilon[y,z]_2+[z,z]_2>2\mathrm{Re}\,\epsilon[y,z]_2$$

Now there are at least two choices of ϵ for which $2\mathrm{Re}\,\epsilon[y,z]_1=0$ and for at least one of these, $2\mathrm{Re}\,\epsilon[y,z]_2\ge 0$. For this ϵ, we have $[x,x]_1=0$ and $[x,x]_2>0$ contrary to the hypothesis. This contradiction shows that

$[y,y]_1 = 1$ and $[z,z]_1 = -1$ implies $[y,y]_2 \le -[z,z]_2$ and the conclusion follows by multiplying y and z by appropriate constants. ■

Lemma 2 *If $[\cdot,\cdot]_1$ and $[\cdot,\cdot]_2$ are indefinite scalar products on the vector space $\mathcal{V}$ such that $[x,x]_1 = 0$ implies $[x,x]_2 \le 0$, then*

$$\gamma = \sup_{[y,y]_1=1} [y,y]_2 < \infty$$

and

$$[x,x]_2 \le \gamma [x,x]_1$$

for all x in $\mathcal{V}$.

Proof. Let $[y,y,]_1 = 1$ and $[z,z]_1 = -1$. Then Lemma 1 implies

$$[y,y]_2 = \frac{[y,y]_2}{[y,y]_1} \le \frac{[z,z]_2}{[z,z]_1} = -[z,z]_2$$

It follows that

$$\gamma = \sup_{[y,y]_1=1} [y,y]_2 \le -[z,z]_2 < \infty$$

and that

$$\gamma \le \inf_{[z,z]_1=-1} -[z,z]_2$$

which means

$$\sup_{[z,z]_1=-1} [z,z]_2 \le -\gamma$$

Now, if $[x,x]_1 > 0$, letting $y_0 = x/\sqrt{[x,x]_1}$, we have

$$[y_0,y_0]_1 = \frac{1}{[x,x]_1}[x,x]_1 = 1$$

which implies

$$\frac{[x,x]_2}{[x,x]_1} = [y_0,y_0]_2 \le \sup_{[y,y]_1=1} [y,y]_2 = \gamma$$

so it follows that $[x,x]_2 \le \gamma[x,x]_1$.

For $[x,x]_1 = 0$, the hypothesis says $[x,x]_2 \le 0 = \gamma[x,x]_1$.

Finally, if $[x,x]_1 < 0$, letting $z_0 = x/\sqrt{-[x,x]_1}$, we have

$$[z_0,z_0]_1 = \frac{1}{-[x,x]_1}[x,x]_1 = -1$$

so

$$\frac{[x,x]_2}{-[x,x]_1} = [z_0,z_0]_2 \le \sup_{[z,z]_1=-1} [z,z]_2 \le -\gamma$$

and

$$[x,x]_2 \le (-\gamma)(-[x,x]_1) = \gamma[x,x]_1$$

Thus, the desired inequality holds no matter what the sign of $[x,x]_1$.

■

The next two results characterize the linear fractional maps which take B_N into, and onto, B_N.

Theorem 3 *Let $\varphi(z) = (Az+B)(\langle z,C\rangle + D)^{-1}$ be a non-constant linear fractional map. If a non-zero multiple of the matrix*

$$m_\varphi = \begin{pmatrix} A & B \\ C^* & D \end{pmatrix}$$

is a contraction on the Kreĭn space with

$$J = \begin{pmatrix} I & 0 \\ 0 & -1 \end{pmatrix}$$

then φ maps the unit ball B_N into itself. Conversely, if φ maps the unit ball into itself, then m_φ is a non-zero multiple of a contraction on this Kreĭn space.

Proof. If v represents the point z in B_N, then for $\lambda \ne 0$, the vector $\lambda m_\varphi v$ represents the point $\varphi(z)$. If λm_φ is a contraction, then $[\lambda m_\varphi v, \lambda m_\varphi v] \le [v,v] < 0$. Since $|\lambda|^2[m_\varphi v, m_\varphi v] = [\lambda m_\varphi v, \lambda m_\varphi v] < 0$, it follows that $[m_\varphi v, m_\varphi v] < 0$ also and $\varphi(z)$ is a point of B_N.

Conversely, if φ maps the unit ball into itself, let $[x,y]_1 = [x,y]$ and let $[x,y]_2 = [m_\varphi x, m_\varphi y]$. The hypothesis that φ maps the ball into itself says that $[v,v]_1 = [v,v] < 0$ implies $[v,v]_2 = [m_\varphi v, m_\varphi v] < 0$. By continuity, this means that $[v,v]_1 \le 0$ implies $[v,v]_2 \le 0$. Since $\varphi(0)$ is in the unit ball, $[\cdot,\cdot]_2$ takes negative values. Since φ is non-constant, $A \ne BC^*/D$ and we may find $z_0 \in \mathbf{C}^N$ with z_0 in the domain of φ and $\varphi(z_0) \notin \overline{B_N}$: if there exists $w \perp C$ with $Aw \ne 0$, let $z_0 = \lambda w$ for $|\lambda|$ suitably large, while if $A = 0$ on $C^\perp$ and $A \not\equiv 0$, let $z_0 = \lambda C$ for λ sufficiently close to $-D/|C|^2$; the remaining case $A \equiv 0$ is trivial. Thus $[\cdot,\cdot]_2$ takes positive values as well as negative, and the hypotheses of Lemma 2 are satisfied.

The conclusion of Lemma 2 is that there is a number γ so that

$$[m_\varphi v, m_\varphi v] = [v,v]_2 \leq \gamma[v,v]_1 = \gamma[v,v]$$

Since φ is non-constant, the left side of this inequality takes both positive and negative values and $\gamma \neq 0$. Since we have a point z_0 as described above such that z_0 and $\varphi(z_0)$ both lie outside $\overline{B}_N$, we see that if v represents z_0, both $[m_\varphi v, m_\varphi v]$ and $[v,v]$ are positive. Thus $\gamma > 0$ and it follows that $\gamma^{-1/2} m_\varphi$ is a Kreĭn contraction as desired. ■

Now if T is a linear transformation on $\mathbf{C}^{N+1}$, then $[Tv,Tv] \leq [v,v]$ if and only if $\langle JTv, Tv\rangle \leq \langle Jv, v\rangle$ if and only if $0 \leq \langle (J - T^*JT)v, v\rangle$. This happens for all v in $\mathbf{C}^{N+1}$ if and only if the operator $J - T^*JT$ is positive semidefinite: this gives a quite concrete condition to check to see if a linear transformation is a Kreĭn contraction. On the other hand, the condition of the theorem, that a *multiple* of m_φ is a contraction, is much more subtle. While the definition of Hilbert space contraction is formally the same, the positivity of the inner product on a Hilbert space means *every* linear transformation is a multiple of a contraction! In an indefinite space, the condition treats positive vectors (that is, vectors with $[v,v] > 0$) differently than negative vectors (vectors with $[v,v] < 0$): to get a contraction, we want small multiples to satisfy the condition for positive vectors and large multiples to satisfy the condition for negative vectors. In many cases for which φ maps B_N into itself, there is precisely one positive multiple of m_φ that is a Kreĭn contraction! Examples will be given in the next section.

Theorem 4 *If the matrix*

$$m_\varphi = \begin{pmatrix} A & B \\ C^* & D \end{pmatrix}$$

is a multiple of an isometry on the Kreĭn space with

$$J = \begin{pmatrix} I & 0 \\ 0 & -1 \end{pmatrix}$$

then $\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ *maps the unit ball* B_N *onto itself. Conversely, if* $\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ *is a linear fractional map of the unit ball onto itself, then* m_φ *is a multiple of an isometry.*

Proof. Since multiples of m_φ give the same linear fractional map, we will assume m_φ is an isometry. Since $\mathbf{C}^{N+1}$ is a finite dimensional Kreĭn space, all isometries are invertible. Now if z is a point of the ball with $\varphi(z) = w$ and if v is a vector in $\mathbf{C}^{N+1}$ that represents z, then $u = m_\varphi v$ represents w. Since m_φ is an isometry, $[v, v] < 0$ if and only if $[u, u] < 0$ which says z is in the ball if and only if $\varphi(z)$ is in the ball.

To prove the converse, first note that the linear transformation m_φ must be invertible. Indeed, writing e_j for the j^{th} standard basis vector in $\mathbf{C}^N$, since φ maps onto the ball, the points 0 and $e_j/2$ for $j = 1, 2, \cdots, N$, are in $\varphi(B_N)$. Since the range of m_φ is a subspace, this means that $(0, \cdots, 0, 1)$, and $(e_j/2, 1)$ are in the range of m_φ. But these $N + 1$ vectors are linearly independent, so the range of m_φ is $\mathbf{C}^{N+1}$.

Since m_φ is invertible, φ has a linear fractional inverse $\varphi^{-1} : B_N \to B_N$, and $(m_\varphi)^{-1} = m_{\varphi^{-1}}$. By Theorem 3 there are positive numbers k_1 and k_2 so that $k_1 m_\varphi$ and $k_2 m_{\varphi^{-1}}$ are contractions on the given Kreĭn space. In particular,

$$k_2^2[v, v] = [k_2 v, k_2 v] \leq [m_\varphi v, m_\varphi v] \leq \frac{1}{k_1^2}[v, v]$$

Since this holds for both positive and negative vectors v, we must have $k_1 k_2 = 1$, and consequently

$$[v, v] \leq k_1^2[m_\varphi v, m_\varphi v] \leq [v, v]$$

which implies $[k_1 m_\varphi v, k_1 m_\varphi v] = [v, v]$ for all v, as desired. ∎

For U to be an invertible isometry on a Kreĭn space means that $U^{-1} = U^\times$, the Kreĭn adjoint. In general, the Kreĭn adjoint is the operator that satisfies $[Tv, w] = [v, T^\times w]$ and it is not difficult to see that $T^\times = JT^*J$ where T^* is the usual Hilbert space adjoint. In the case we are interested in,

$$\begin{pmatrix} A & B \\ C^* & D \end{pmatrix}^\times = \begin{pmatrix} A^* & -C \\ -B^* & D^* \end{pmatrix} \qquad (1)$$

Writing out $UU^\times = I$ and $U^\times U = I$ gives eight equations that characterize isometries on our Kreĭn space, that is, the automorphisms of B_N.

Examples We wish to identify the expressions for certain automorphisms. First, note that the automorphism given by a unitary operator U on C^N, $\varphi(z) = Uz$, can be represented by

$$m_\varphi = \begin{pmatrix} U & 0 \\ 0 & 1 \end{pmatrix}$$

The automorphisms φ fixing $e_1 = (1, 0, \cdots, 0)$ and $-e_1$ can be written as $\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ where $B = C = ce_1$, c real, $D = d > 1$ with $c = \pm\sqrt{d^2 - 1}$, and A has the block form

$$\begin{pmatrix} d & 0 \\ 0 & V \end{pmatrix}$$

where V is an $(N - 1) \times (N - 1)$ (Euclidean) unitary. When $V = I$ the corresponding automorphism is called a *non-isotropic dilation* of the ball.

Finally, if p is a point of B_N, we would like to find an automorphism φ so that $\varphi(\varphi(z)) = z$ and $\varphi(0) = p$. Such automorphisms will be called *involutions*. The transitivity of the automorphism group follows from the existence of involutions. An involution satisfies $\varphi^{-1} = \varphi$ so, choosing a representation m_φ of φ for which m_φ is an isometry and $D \geq 1$, there is λ so that $\lambda {m_\varphi}^\times = m_\varphi$. Equation (1) gives $A = \lambda A^*$, $B = -\lambda C$, and $D = \lambda D^* = \lambda D$. The final equality implies $\lambda = 1$, so we get $A = A^*$ and $B = -C$. From the relations $m_\varphi {m_\varphi}^\times = {m_\varphi}^\times m_\varphi = I$ we obtain $D = (1 - |p|^2)^{-1/2}$, $Ap = -Dp$, and $A^2 = I$ on the orthogonal complement of the subspace spanned by p. Conversely, if $D = (1 - |p|^2)^{-1/2}$ and A is a selfadjoint matrix with $Ap = -Dp$ and $A^2 = I$ on the orthogonal complement of the subspace spanned by p, letting $\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ where $B = Dp$ and $C = -Dp$ gives an automorphism of B_N with $\varphi(0) = p$ and $\varphi(\varphi(z)) = z$. ∎

Proposition 5 *The unitary maps and the non-isotropic dilations generate the full group of automorphisms of the ball. Specifically, if φ is a (non-unitary) automorphism of the ball, then there are unitaries W_1 and W_2 and a non-isotropic dilation δ so that $\varphi = W_1 \delta W_2$.*

Proof. First, for $0 < b < 1$, consider the involution automorphism φ_b that takes be_1 to 0 given by $\varphi_b(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ where A is diagonal with diagonal entries $(-s^{-1}, 1, \cdots, 1)$, $B = -C = bs^{-1}e_1$, and $D = s^{-1}$ for $s = \sqrt{1 - |b|^2}$. Let δ_b be the non-isotropic dilation given by $\delta_b(z) = (A'z + B')(\langle C', z\rangle + D')^{-1}$ where A' is diagonal with diagonal entries $(s^{-1}, 1, \cdots, 1)$, $B' = C' = bs^{-1}e_1$, and $D' = s^{-1}$. If U is the diagonal unitary with diagonal entries $(-1, 1, \cdots, 1)$, then a computation with the associated linear transformations shows that $\varphi_b = \delta_b U$.

Now for any point p of the ball, $p \neq 0$, if V is a unitary that such that $Vp = |p|e_1$ then $V^* \varphi_{|p|} V$ is an involution automorphism that takes p to 0.

Moreover, for any automorphism φ, if $p = \varphi(0)$ and ψ is an involution that takes p to 0, then $\psi \circ \varphi$ maps 0 to 0 and, by Cartan's Theorem [22, Theorem 2.1.3], is therefore a unitary W. Since $\psi \circ \varphi = W$ and ψ is an involution, $\varphi = \psi \circ \psi \circ \varphi = \psi W$. Putting these maps together, we see that we have $\varphi = V^* \delta_{|p|}(UVW)$ for any automorphism φ. ∎

The automorphisms allow us to "change variables" via conjugation: we say φ and ψ are *conjugate* if there is an automorphism of the ball τ such that $\psi = \tau^{-1} \circ \varphi \circ \tau$. Conjugation allows convenient normalizations. For example, every linear fractional map φ of the ball into itself with a fixed point in the ball is conjugate to one fixing 0, via an involution automorphism interchanging 0 and a fixed point of φ. Similarly, every linear fractional self-map $\varphi(z) = (Az+B)(\langle z, C\rangle + D)^{-1}$ of the ball is conjugate to a map $\psi(z) = (A'z + B')(cz_1 + 1)^{-1}$ for some $0 \leq c < 1$, via a unitary automorphism U chosen so that $U^*C = \overline{D}|D|^{-1}|C|e_1$, with $A' = D^{-1}U^*AU$, $B' = D^{-1}U^*B$ and $c = |D|^{-1}|C|$. The condition $|C| < |D|$ implies $c < 1$.

3 Geometric Properties

Next, we consider some geometric facts about linear fractional maps. In the complex plane, the principal geometric fact about linear fractional maps is that they map circles to circles. In higher dimensions, there is much more flexibility, but the appropriate analogies are not obvious. By an *ellipsoid*, we mean a translate of the image of the unit ball under an invertible (complex) linear transformation. Because translations and linear transformations are linear fractional maps, every ellipsoid is the image of the unit ball under a linear fractional map. The next result gives a converse.

Theorem 6 *If φ is a one-to-one linear fractional map defined on a ball, $\overline{\mathbf{B}}$, in $\mathbf{C}^N$, then $\varphi(\mathbf{B})$ is an ellipsoid.*

Proof. Without loss of generality, we may assume that the ball $\mathbf{B}$ is B_N, and we may write $\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ where $|D|^2 = 1 + |C|^2$. Then $\varphi(B_N)$ is an ellipsoid if and only if $\psi(\varphi(B_N))$ is an ellipsoid for an invertible affine map ψ. Following φ by the translation $z \mapsto z - BD^{-1}$ gives an invertible linear fractional map taking 0 to 0:

$$\begin{pmatrix} I & -BD^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A & B \\ C^* & D \end{pmatrix} = \begin{pmatrix} A - BD^{-1}C^* & 0 \\ C^* & D \end{pmatrix}$$

Moreover, since φ is one-to-one, m_φ is invertible and this implies $A - BD^{-1}C^*$ is invertible; if not find a non-zero vector x with $(A-BD^{-1}C^*)x = 0$ and note that the non-zero vector $(x, -D^{-1}C^*x)$ is mapped by m_φ to 0. Now if ψ is the affine map with

$$m_\psi = \begin{pmatrix} (A - BD^{-1}C^*)^{-1} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I & -BD^{-1} \\ 0 & 1 \end{pmatrix}$$

then $\psi(\varphi(z)) = z(\langle z, C\rangle + D)^{-1}$.

Now consider the affine map $\Psi(z) = Xz + Y$ where $X = \sqrt{(I + CC^*)^{-1}}$ and $Y = XC$. Note that the matrix CC^* is positive semidefinite, so $I + CC^*$ is invertible and positive definite. The matrix X is the unique positive definite square root of this inverse and by the spectral theorem, X commutes with CC^*. The linear fractional map $\Psi(\psi(\varphi(z)))$ then is associated with the linear transformation

$$\begin{pmatrix} X & Y \\ 0 & 1 \end{pmatrix} \begin{pmatrix} I & 0 \\ C^* & D \end{pmatrix} = \begin{pmatrix} X + YC^* & YD \\ C^* & D \end{pmatrix}$$

and

$$\begin{pmatrix} X + YC^* & YD \\ C^* & D \end{pmatrix} \begin{pmatrix} X + CY^* & -C \\ -D^*Y^* & D^* \end{pmatrix}$$

has upper left entry

$$\begin{aligned} &(X + YC^*)(X + CY^*) - YDD^*Y^* \\ &\quad = (X + XCC^*)(X + CC^*X) - XCDD^*C^*X \\ &\quad = X^2\left((I + CC^*)^2 - |D|^2CC^*\right) \\ &\quad = X^2(I + CC^*) = I \end{aligned}$$

(where we used the relationships $|D|^2 = 1 + |C|^2$ and $C^*C = |C|^2$), has upper right entry

$$\begin{aligned} (X + YC^*)(-C) + YDD^* &= -XC - YC^*C + YDD^* \\ &= -XC - XC|C|^2 + XC(1 + |C|^2) = 0 \end{aligned}$$

has lower left entry

$$C^*(X + CY^*) - DD^*Y^* = 0$$

and has lower right entry

$$-C^*C + DD^* = 1$$

That is, the matrix

$$\begin{pmatrix} X + YC^* & YD \\ C^* & D \end{pmatrix}$$

is a Kreĭn isometry. This means that the linear fractional map $\Psi \circ \psi \circ \varphi$ is an automorphism of the ball! In particular, the image $\varphi(B_N)$ is the same as the image of the affine map $(\Psi \circ \psi)^{-1}$ which is an ellipsoid. ■

If φ is a linear fractional map defined on a closed ball $\overline{\mathbf{B}}$ that is not one-to-one, the image $\varphi(\mathbf{B})$ will be the intersection of an ellipsoid with a translate of a k dimensional subspace of $\mathbf{C}^N$, for some $k < N$. To see this, we again assume without loss of generality that $\varphi(0) = 0$ and note that

$$m_\varphi = \begin{pmatrix} A & 0 \\ C^* & D \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & D \end{pmatrix} \begin{pmatrix} I & 0 \\ D^{-1}C^* & 1 \end{pmatrix}$$

The first factor on the right-hand side corresponds to the linear transformation $D^{-1}Az$ and the second factor corresponds to a one-to-one linear fractional map also defined on $\overline{\mathbf{B}}$. If φ is not one-to-one, then $\operatorname{rank} A = k < N$, and by Theorem 6 $\varphi(\mathbf{B})$ is a translate of the image of a ball under some linear transformation of rank k.

Recall that if $z_1, z_2, \cdots, z_n$ are points in $\mathbf{C}^N$, the *affine set* determined by these points is the set

$$[z_1, z_2, \cdots, z_n] = \left\{ \sum_{j=1}^{n} \alpha_j z_j : \sum_{j=1}^{n} \alpha_j = 1;\ \alpha_j \in \mathbf{C} \right\}$$

Affine sets are just translates of (complex) subspaces and the dimension of an affine set is the dimension of this subspace. The following theorem says that linear fractional maps take affine sets to affine sets.

Theorem 7 *If φ is a linear fractional transformation and $z_1, z_2, \cdots, z_n$ are points in the domain Δ of φ, then the closure of $\varphi([z_1, z_2, \cdots, z_n] \cap \Delta)$ is $[\varphi(z_1), \varphi(z_2), \cdots, \varphi(z_n)]$.*

Note that the closure in the statement of the theorem is necessary: if $\varphi(z) = 1/z$ in the plane, and $z_1 = -1$ and $z_2 = 1$, then the affine set

$[z_1, z_2]$ is the complex plane, but $\varphi([z_1, z_2] \cap \Delta)$ does not contain 0.

Proof. Let $\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ and z_1, z_2, $\cdots$, z_n be given. If z is in the domain of φ where $z = \sum \alpha_j z_j$ for $\sum \alpha_j = 1$, let

$$\beta_j = \alpha_j \frac{\langle z_j, C\rangle + D}{\langle z, C\rangle + D}$$

Now a calculation shows that $\sum \beta_j = 1$ and $\sum \beta_j \varphi(z_j) = \varphi(z)$. Thus if z in Δ is an affine combination of z_1, z_2, $\cdots$, z_n, then $\varphi(z)$ is an affine combination of $\varphi(z_1)$, $\varphi(z_2)$, $\cdots$, $\varphi(z_n)$.

Conversely, suppose u_1 and u_2 are points in $[z_1, z_2, \cdots, z_n] \cap \Delta$ and $w = \beta_1\varphi(u_1)+\beta_2\varphi(u_2)$ where $\beta_1+\beta_2 = 1$. If $\beta_1(\langle u_2, C\rangle+D)+\beta_2(\langle u_1, C\rangle+D) \neq 0$, let

$$\alpha_1 = \frac{\beta_1(\langle u_2, C\rangle + D)}{\beta_1(\langle u_2, C\rangle + D) + \beta_2(\langle u_1, C\rangle + D)}$$

and

$$\alpha_2 = \frac{\beta_2(\langle u_1, C\rangle + D)}{\beta_1(\langle u_2, C\rangle + D) + \beta_2(\langle u_1, C\rangle + D)}$$

so that $\alpha_1 + \alpha_2 = 1$. A computation shows that

$$\langle(\alpha_1 u_1 + \alpha_2 u_2), C\rangle + D = \frac{(\langle u_1, C\rangle + D)(\langle u_2, C\rangle + D)}{\beta_1(\langle u_2, C\rangle + D) + \beta_2(\langle u_1, C\rangle + D)}$$

which is non-zero so that $\alpha_1 u_1 + \alpha_2 u_2$ is in the domain of φ. Further computation verifies $\varphi(\alpha_1 u_1 + \alpha_2 u_2) = w$ so that w is in $\varphi([z_1, z_2, \cdots, z_n] \cap \Delta)$.

On the other hand, if $\beta_1(\langle u_2, C\rangle+D)+\beta_2(\langle u_1, C\rangle+D) = 0$, then $\langle u_1, C\rangle+D = \beta_1\langle(u_1 - u_2), C\rangle$. Since u_1 is in the domain of φ, $\langle u_1, C\rangle + D \neq 0$ and this means $\langle(u_1 - u_2), C\rangle \neq 0$. Therefore, we can apply the above argument to $\beta_1+\epsilon$ and $\beta_2-\epsilon$ for all non-zero ϵ to show that $(\beta_1+\epsilon)\varphi(u_1)+(\beta_2-\epsilon)\varphi(u_2)$ is in $\varphi([z_1, z_2, \cdots, z_n] \cap \Delta)$ for all non-zero ϵ. Since

$$w = \lim_{\epsilon\to 0}(\beta_1 + \epsilon)\varphi(u_1) + (\beta_2 - \epsilon)\varphi(u_2)$$

this shows w is in the closure of $\varphi([z_1, z_2, \cdots, z_n] \cap \Delta)$.

Thus, the closure of $\varphi([z_1, z_2, \cdots, z_n] \cap \Delta)$ is an affine set that contains $\varphi(z_1)$, $\varphi(z_2)$, $\cdots$, $\varphi(z_n)$ so it contains $[\varphi(z_1), \varphi(z_2), \cdots, \varphi(z_n)]$. Together with the containment proved in the first half of the argument, this shows that the conclusion is correct. ∎

A *slice* of the ball is the intersection of the ball with a subspace of $\mathbf{C}^N$. A slice is copy of a unit ball of lower dimension and if a map φ fixes a slice as a set, there is a unitary transformation of this ball of lower dimension onto the slice that allows one to pull the restriction of φ to the slice back to a map of the ball of lower dimension.

The next result follows immediately from Theorem 7 via conjugation by an automorphim sending a point of the fixed affine set to 0.

Corollary 8 *If φ is a linear fractional map of the unit ball into itself that fixes an affine set (as a set) then φ is conjugate to a linear fractional map ψ that fixes a slice (as a set).*

We wish to determine when a linear fractional map of the ball into itself fixes an affine set (as a set) or has a fixed point.

Theorem 9 *Let φ be a linear fractional map of the ball into itself with associated linear transformation m_φ. If $\mathcal{S}$ is an affine set whose intersection with the ball is fixed as a set by φ, then the closure of the set of points in $\mathbf{C}^{N+1}$ whose equivalence classes are points of $\mathcal{S}$ is a subspace $M_{\mathcal{S}}$ of $\mathbf{C}^{N+1}$ that is invariant for m_φ. Conversely, if M is a subspace of $\mathbf{C}^{N+1}$ that is invariant for m_φ, then the set of equivalence classes in the ball of the vectors in M is (either empty or) the intersection $\mathcal{S}_M$ of the ball with an affine set that is fixed as a set by φ. In particular, z is a fixed point of φ, that is, $\varphi(z) = z$, if and only if $(z, 1)$ is an eigenvector of m_φ.*

Proof. Subspaces of $\mathbf{C}^{N+1}$ correspond, under the quotient mapping $(v_1, v_2) \leftrightarrow v_1/v_2$ (where v_1 is in $\mathbf{C}^N$ and $v_2 \neq 0$ is a number), to affine sets in $\mathbf{C}^N$. Indeed, if (v_1, v_2) is in the subspace, then so is $(v_1/v_2, 1)$. Thus, if $z = v_1/v_2$ and $w = u_1/u_2$ are points of $\mathbf{C}^N$ that are quotients of vectors in the subspace, then $\alpha z + (1-\alpha)w$ is also a quotient of a point in the subspace, namely

$$\alpha(v_1/v_2, 1) + (1-\alpha)(u_1/u_2, 1) = (\alpha v_1/v_2 + (1-\alpha)u_1/u_2, 1)$$

Conversely, if $\mathcal{S}$ is an affine subset of $\mathbf{C}^N$, then the set of points of $\mathbf{C}^{N+1}$ whose quotients are in $\mathcal{S}$ is a subspace: clearly every multiple such a vector is in the set because the multiple has the same quotient. Similarly, if (v_1, v_2) and (u_1, u_2) have quotients z and w in $\mathcal{S}$ and $v_2 + u_2 \neq 0$ then

$$v_2/(v_2+u_2)z + u_2/(v_2+u_2)w$$

is an affine combination of z and w that is a quotient of $(v_1 + u_1, v_2 + u_2)$. (If $v_2 + u_2 = 0$, then take the limit as t approaches 1 of $(v_1, v_2) + (tu_1, tu_2)$.)

The conclusions follow from the observation that $\varphi(z) = w$ if and only if $m_\varphi(z, 1) = (\lambda w, \lambda)$ for some $\lambda \neq 0$. ■

There are intuitive difficulties in dealing with $\mathbf{C}^N$ for large N (that is, $N > 1$) because $\mathbf{C}^N$ is hard to visualize. The goal of the following theorem is to give a foundation for expanding our intuition by restricting to the real case — we can imagine that the unit balls of $\mathbf{C}^2$ and $\mathbf{C}^3$ look "just like" the unit balls of $\mathbf{R}^2$ and $\mathbf{R}^3$.

Theorem 10 *Let $\varphi(z) = (Az+B)(\langle z, C\rangle + D)^{-1}$ be a linear fractional map for which the matrices A, B, C, and D are real. Then φ maps the unit ball in $\mathbf{R}^N$ into itself if and only if φ maps the unit ball in $\mathbf{C}^N$ into itself.*

Proof. Since φ maps $\mathbf{R}^N$ into itself, if φ maps the unit ball in $\mathbf{C}^N$ into itself, then φ will map the unit ball of $\mathbf{R}^N$ into itself since it is the intersection of B_N with $\mathbf{R}^N$.

Conversely, suppose φ maps the unit ball in $\mathbf{R}^N$ into itself. This means that if $v = (x, c)$ represents the point x/c in the unit ball of $\mathbf{R}^N$, then for $(p, a) = m_\varphi v$, we get $p/a = \varphi(x/c)$ in the unit ball as well. We wish to show that m_φ as a linear transformation on $\mathbf{C}^{N+1}$ has the same property.

Suppose (z, ζ) satisfies $|\zeta| = 1$ and $|z|^2 < 1$ so that z/ζ is in B_N. For each real number θ, the vector $(e^{i\theta}z, e^{i\theta}\zeta)$ represents the same point of B_N as (z, ζ). Write $e^{i\theta}z = x + iy$ (x, y in $\mathbf{R}^N$) and $e^{i\theta}\zeta = c + is$ (c, s in $\mathbf{R}$). Since $c^2 + s^2 = 1$, $|z|^2 = |x|^2 + |y|^2 < 1$, and c varies from -1 to 1 as θ varies, $|x|^2 - c^2$ takes on both negative and non-negative values as θ ranges over $\mathbf{R}$. By continuity we may choose θ so that $|x|^2 - c^2 = 0$.

Suppose $c \neq 0$. Since $|x|^2 - c^2 = 0$ and

$$|y|^2 - s^2 = |x|^2 - c^2 + |y|^2 - s^2 = |z|^2 - |\zeta|^2 < 0$$

both (x, c) and (y, s) represent points of the closed unit ball in $\mathbf{R}^N$. The hypothesis guarantees that $m_\varphi(x, c) = (p, a)$ represents a point of the closed unit ball and $m_\varphi(y, s) = (q, b)$ represents a point of the unit ball in $\mathbf{R}^N$. Now,

$$\begin{pmatrix} e^{i\theta}z \\ e^{i\theta}\zeta \end{pmatrix} = \begin{pmatrix} x \\ c \end{pmatrix} + i \begin{pmatrix} y \\ s \end{pmatrix}$$

and

$$\begin{aligned} m_\varphi \begin{pmatrix} e^{i\theta} z \\ e^{i\theta}\zeta \end{pmatrix} &= m_\varphi \begin{pmatrix} x \\ c \end{pmatrix} + i m_\varphi \begin{pmatrix} y \\ s \end{pmatrix} \\ &= \begin{pmatrix} p \\ a \end{pmatrix} + i \begin{pmatrix} q \\ b \end{pmatrix} \equiv \begin{pmatrix} w \\ \eta \end{pmatrix} \end{aligned}$$

We see that

$$|w|^2 - |\eta|^2 = |p|^2 - a^2 + |q|^2 - b^2 < 0$$

so (w, η) represents a point of the unit ball in $\mathbf{C}^N$.

If $c = 0$, then $|x|^2 - c^2 = 0$ means $x = 0$ also. It follows that

$$\frac{z}{\zeta} = \frac{iy}{is} = \frac{y}{s}$$

so z/ζ is point of the unit ball of $\mathbf{R}^N$ and $\varphi(z/\zeta)$ is a point of the unit ball of $\mathbf{R}^N$, hence of B_N, by hypothesis. ∎

We can illustrate Theorems 3, 6, and 10 with some examples. If

$$A = \begin{pmatrix} 1 & 0 \\ 0 & t \end{pmatrix} \quad B = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad C = \begin{pmatrix} -1 \\ 0 \end{pmatrix} \quad \text{and} \quad D = 3$$

then $\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ maps B_2 into itself for $-2 \le t \le 2$. Indeed, if $T = \frac{1}{2} m_\varphi$, then $J - T^* JT$ is positive semidefinite so that $\frac{1}{2} m_\varphi$ is a J–contraction; moreover, this is the only positive multiple of m_φ with this property. The more conventional way to write these maps, say in a class on multivariate calculus, would be

$$\varphi((x, y)) = (\frac{1 + x}{3 - x}, \frac{ty}{3 - x})$$

Since all the matrices in m_φ are real, φ maps the unit ball in $\mathbf{R}^2$ into itself as well. Figure 1 shows the intersections of the ellipsoids that are the images of $\varphi(B_2)$ with the unit ball in $\mathbf{R}^2$, that is, the images of the restriction of φ to the unit ball of $\mathbf{R}^2$, for $t = 0$ (the line segment), .25, .50, $\cdots$, 1.75, and 2 (the largest ellipse).

Figure 1: Images of some linear fractional maps.

4 The Adjoint Mapping and Composition Operators with Linear Fractional Symbol

Our goal was to introduce linear fractional maps of the ball to be able to better understand composition operators on spaces of analytic functions in several variables. While composition operators induced by automorphisms have received some previous attention (e.g. [17]), using general linear fractional maps as the symbol has not been specifically addressed. In this section, we prove some basic facts about the composition operators induced by linear fractional maps. This study provides further justification for the introduction of the Kreĭn space structure because the map in the following definition plays a role in the construction of the adjoints of composition operators and the proof of boundedness. In the one variable case, this definition leads to a notion of duality between Hilbert spaces of analytic functions (see Hurst [12]).

Definition If $\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ is a linear fractional map of the ball B_N into itself, define the *adjoint map* $\sigma = \sigma_\varphi$ by

$$\sigma(z) = \frac{A^*z - C}{\langle z, -B\rangle + D^*}$$

Notice that $\varphi(B_N) \subset B_N$ implies $\varphi(0) = |BD^{-1}| < 1$ so $|-B| < |D|$ and the domain of σ contains the closed unit ball.

Proposition 11 *If $\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ is a linear fractional transformation mapping B_N into itself, then the adjoint map $\sigma(z) = (A^*z - C)(\langle z, -B\rangle + D^*)^{-1}$ maps B_N into itself.*

Proof. If φ is a map of B_N into itself, we have seen (Theorem 3) that some non-zero multiple of m_φ is a Kreĭn contraction. Thus, suppose $\lambda \neq 0$ is a scalar such that m_φ satisfies $[\lambda m_\varphi v, \lambda m_\varphi v] \leq [v, v]$. Since we are working in a finite dimensional space, this means $\overline{\lambda} m_\varphi{}^\times = (\lambda m_\varphi)^\times$ is also a Kreĭn contraction [14, page 106]. Thus, by Theorem 3 and the observation that $m_\sigma = m_\varphi{}^\times$, we see that σ maps the ball into itself. ∎

Lemma 12 *If φ, ψ are linear fractional maps of B_N into B_N, then $\sigma_{\varphi\circ\psi} = \sigma_\psi \circ \sigma_\varphi$.*

Proof. By Proposition 11 we know $\sigma_\psi \circ \sigma_\varphi$ is a linear fractional map of B_N into itself. The conclusion follows from

$$m_{\sigma_{\varphi\circ\psi}} = m^{\times}_{\varphi\circ\psi} = (m_\varphi m_\psi)^{\times} = m^{\times}_{\psi} m^{\times}_{\varphi} = m_{\sigma_\psi} m_{\sigma_\varphi}$$

■

Proposition 13 *Let φ be a linear fractional map of B_N into B_N and let σ be its adjoint map. Then σ is an automorphism of B_N if and only if φ is an automorphism, σ is one-to-one if and only if φ is one-to-one, and $\sigma(B_N)$ is contained in an affine set of dimension $k < N$ if and only if $\varphi(B_N)$ is contained in an affine set of dimension k.*

Proof. Since the adjoint map of φ is σ if and only if the adjoint map for σ is φ, it is enough to only prove each of the "if" statements.

If φ is an automorphism of B_N, we may assume without loss of generality that m_φ is an isometry. As in the proof of Proposition 11, we see that $m_\sigma = m_\varphi{}^{\times}$ is also an isometry, and hence that σ is an automorphism. An alternate proof, based on Proposition 5 and Lemma 12 can be given, since the adjoint map of a unitary is unitary, and the adjoint of a non-isotropic dilation is again a non-isotropic dilation.

If φ is a linear fractional map of B_N into B_N with $\varphi(0) = a$, then $\varphi_a \circ \varphi$ is a linear fractional map of B_N into B_N fixing 0, where φ_a is an involution automorphism interchanging a and 0 (see the discussion preceding Proposition 5). Since $\varphi = \varphi_a \circ (\varphi_a \circ \varphi)$, Lemma 12 shows that $\sigma_\varphi = \sigma_{\varphi_a \circ \varphi} \circ \sigma_{\varphi_a}$, where by the first part of the proof, σ_{φ_a} is an automorphism. Thus it suffices to prove the remaining parts of the proposition under the additional hypothesis that $\varphi(0) = 0$ so that $\varphi(z) = Az(\langle z, C\rangle + D)^{-1}$. If φ is one-to-one, A is non-singular, and $\sigma_\varphi(z) = (A^* z - C)/D^*$ is clearly also one-to-one. Similarly, if $\varphi(B_N)$ is contained in an affine set of dimension $k < N$, then the rank of A is at most k, and the expression $\sigma_\varphi(z) = (A^* z - C)/D^*$ shows that $\sigma_\varphi(B_N)$ is contained in an affine set of dimension at most k. ■

We use Proposition 11 in the next result, which shows that all linear fractional maps of B_N into B_N induce bounded composition operators on the Hardy spaces $H^p(B_N)$.

Theorem 14 *If φ is a linear fractional map of B_N into B_N then C_φ is bounded on $H^p(B_N)$ for all $p > 0$.*

Proof. This is trivially true for $p = \infty$, so we consider $0 < p < \infty$. Since the automorphisms of B_N give bounded composition operators [7, page 172], it is enough to prove the theorem under the assumption that $\varphi(0) = 0$. Thus we may assume that $\varphi(z) = Az/(\langle z, C\rangle + 1)$ where $|C| < 1$.

Since φ is smooth on the closed ball, we may appeal to a theorem of W. R. Wogen [24] to see that if C_φ fails to be bounded on $H^p(B_N)$ then there must exist ζ and η in ∂B_N with $\varphi(\zeta) = \eta$ and τ in ∂B_N with $\langle \zeta, \tau\rangle = 0$ so that

$$D_\zeta \varphi_\eta(\zeta) = |D_{\tau\tau}\varphi_\eta(\zeta)| \tag{2}$$

where $\varphi_\eta(z) = \langle \varphi(z), \eta\rangle$ is the coordinate of φ in the $\eta-$ direction and D_ζ is the derivative in the ζ direction. By pre– and post–composing φ with unitary maps we may assume without loss of generality that $\zeta = \eta = e_1 = (1, 0')$; that is, if there is a linear fractional map of B_N fixing 0 for which Equation (2) holds, then there is a linear fractional map of B_N fixing 0 for which the normalized version of Equation (2) holds: $\varphi(e_1) = e_1$ and

$$D_1\varphi_1(e_1) = |D_{\tau\tau}\varphi_1(e_1)| \tag{3}$$

for some τ in ∂B_N with $\langle e_1, \tau\rangle = 0$.

Writing $A = (a_{ij})$ and $C = (c_1, \dots, c_N)^t$ the condition $\varphi(e_1) = e_1$ says that

$$a_{11} = 1 + \overline{c_1} \quad \text{and} \quad a_{j1} = 0 \text{ for } j = 2, 3, \dots, N \tag{4}$$

A computation then shows that $D_1\varphi_1(e_1) = a_{11}(1+\overline{c_1})^{-2} = (1+\overline{c_1})^{-1}$. Now choose $e^{i\theta}$ so that $e^{i\theta}\overline{c_1} = -|c_1|$. Since $\varphi(e^{i\theta}e_1)$ must be in the closed ball, we have

$$\frac{|a_{11}|}{1 - |c_1|} = \frac{|1 + \overline{c_1}|}{1 - |c_1|} \le 1$$

so that $\overline{c_1} = -|c_1|$. In particular, this shows that $D_1\varphi_1(e_1) = (1-|c_1|)^{-1} > 0$.

Now by Proposition 11 we know that $\sigma(z) = A^*z - C$ also maps B_N into B_N. In particular, $|\sigma(e_1)| \le 1$. But

$$\sigma(e_1) = (\overline{a_{11}} - c_1, \overline{a_{12}} - c_2, \dots, \overline{a_{1N}} - c_N)$$

and we know $a_{11} = 1 + \overline{c_1}$ so we must have

$$\overline{a_{12}} = c_2,\ \overline{a_{13}} = c_3,\ \dots,\ \overline{a_{1N}} = c_N \tag{5}$$

Returning to φ we see by direct computation (using (4) and (5)) that for $j, k = 2, 3, \dots, N$

$$D_{jk}\varphi_1(e_1) = 0$$

From this it follows that whenever $\langle \tau, e_1 \rangle = 0$ the second order derivative $D_{\tau\tau}\varphi_1(e_1)$ must be 0. This contradicts Equation (3) and hence C_φ must in fact be bounded on $H^p(B_N)$. ■

Having established that all linear fractional maps induce bounded composition operators on the Hardy spaces, we can extend this result to Bergman or weighted Bergman spaces, defined for $0 < p < \infty$ and $\alpha > -1$ by

$$A^p_\alpha(B_N) = \{f \text{ analytic} : \int_{B_N} |f(z)|^p (1 - |z|^2)^\alpha \, d\nu(z) < \infty\}$$

where ν is normalized volume measure on B_N.

Theorem 15 *If φ is a linear fractional map of B_N into itself, then C_φ is bounded on $A^p_\alpha(B_N)$ for all $\alpha > -1$ and $p > 0$.*

Proof. Again we may restrict attention to those linear fractional maps φ fixing 0 since the automorphisms give bounded composition operators on these spaces as well (see, for example, Exercises 3.5.4 or 3.5.9 in [7]). For $0 < r < 1$ consider $\varphi_r(z) = \varphi(rz)$. By the theorem just proved, C_φ is bounded on $H^2(B_N)$ and it is easy to see that $\|C_{\varphi_r}\| \leq \|C_\varphi\|$ on $H^2(B_N)$. By this observation and the Carleson measure criterion for boundedness on $H^2(B_N)$ ([7, page 161])there is a finite constant K so that

$$\mu(\varphi_r^{-1} S(\zeta, h)) \leq K h^N$$

for all Carleson sets $S(\zeta, h) = \{z \in \overline{B_N} : |1 - \langle z, \zeta \rangle| < h\}$ in the ball, where μ is normalized Lebesgue measure on ∂B_N. Since $\|C_{\varphi_r}\| \leq \|C_\varphi\|$ the constant K may be chosen independent of r. We are assuming $\varphi(0) = 0$ so that by the Schwarz lemma we have $|\varphi(z)| \leq |z|$ for z in B_N. Changing to polar coordinates we see that

$$\int_{\varphi^{-1}S(\zeta,h)} (1 - |z|^2)^\alpha \, d\nu(z)$$

$$= 2N \int_{1-h}^{1} r^{2N-1} (1 - r^2)^\alpha \, dr \int_{\partial B_N} \chi_{\varphi^{-1}S(\zeta,h)}(r\eta) \, d\mu(\eta)$$

where the inner integral is bounded above by Kh^N. This yields

$$\nu_\alpha \varphi^{-1} S(\zeta, h) \leq K' h^{\alpha+N+1}$$

for some finite constant K', where $d\nu_\alpha(z) = (1 - |z|^2)^\alpha d\nu(z)$. This guarantees that C_φ is bounded on the weighted Bergman space $A^p_\alpha(B_N)$ (see [7, page 164]). ■

Linear fractional maps can induce compact composition operators on the Hardy or weighted Bergman spaces only if $\|\varphi\|_\infty < 1$. This follows immediately from the observation that if $|\varphi(\zeta)| = 1$ for some ζ in ∂B_N then the smoothness of φ on $\overline{B_N}$ implies that φ has finite angular derivative at ζ. This prevents C_φ from being compact on $H^p(B_N)$ or $A^p_\alpha(B_N)$ ([7, page 171]).

If $\mathcal{H}$ a Hilbert space of analytic functions on the ball and h is an analytic function on the ball, we say h is a multiplier of $\mathcal{H}$ if the operator T_h defined by $T_h(f) = hf$ for f in $\mathcal{H}$ is bounded on $\mathcal{H}$. If w is a point of the ball, the reproducing kernel function K_w is the function in $\mathcal{H}$ that gives the linear functional of evaluation at w, that is, for which $\langle f, K_w\rangle = f(w)$ for all f in $\mathcal{H}$. It is not difficult to check that $T_h^*(K_w) = \overline{h(w)}K_w$ when h is a multiplier of $\mathcal{H}$. In particular, $\|T_h\| \geq \sup_{|z|<1} |h(z)|$, so multipliers must be in $H^\infty(B_N)$. The following theorem identifies the adjoints of linear fractional composition operators on many spaces of analytic functions on the ball, including the usual Hardy and Bergman Hilbert spaces, since $K_w(z) = (1 - \langle z, w\rangle)^{-N}$ is the reproducing kernel on the Hardy space $H^2(B_N)$ and $K_w(z) = (1 - \langle z, w\rangle)^{-N-1}$ is the reproducing kernel on the Bergman space $A^2(B_N)$. The following theorem is a generalization of results of Cowen [4] and Hurst [12].

Theorem 16 *Let $\mathcal{H}$ be a Hilbert space of analytic functions on the unit ball for which all functions in $H^\infty(B_N)$ are multipliers and for which the reproducing kernel functions are given by*

$$K_w(z) = (1 - \langle z, w\rangle)^{-r}$$

for some positive number r. Suppose $\varphi(z) = (Az + B)(\langle z, C\rangle + D)^{-1}$ is a linear fractional map of B_N into itself for which C_φ is a bounded operator on $\mathcal{H}$. Let $\sigma(z) = (A^*z - C)(\langle z, -B\rangle + D^*)^{-1}$ be the adjoint mapping. Then C_σ is a bounded operator on $\mathcal{H}$, $g(z) = (\langle z, -B\rangle + D^*)^{-r}$ and $h(z) = (\langle z, C\rangle + D)^r$ are in $H^\infty(B_N)$, and

$$C_\varphi^* = T_g C_\sigma T_h^*$$

Proof. Because φ and therefore also σ map B_N into itself, we have $|C| < |D|$ and $|B| < |D|$. Thus the functions h, h^{-1}, g, and g^{-1} are in $H^\infty(B_N)$. We will show $C_\sigma = T_g^{-1} C_\varphi^* (T_h^*)^{-1}$.

Let w be in B_N. Since

$$(T_h^*)^{-1} K_w = (T_h^{-1})^* K_w = (\overline{h(w)})^{-1} K_w$$

and $C_\varphi^* K_w = K_{\varphi(w)}$ a calculation gives

$$T_g^{-1} C_\varphi^* (T_h^*)^{-1}(K_w)(z) = \overline{h(w)}^{-1} (g(z))^{-1} K_{\varphi(w)}(z)$$

$$\begin{aligned}
&= \overline{h(w)}^{-1}(g(z))^{-1}\left(1 - \langle z, (Aw+B)(C^*w+D)^{-1}\rangle\right)^{-r} \\
&= \overline{(C^*w+D)}^{-r}(-B^*z+D^*)^r\overline{(C^*w+D)}^r\left(\overline{(C^*w+D)} - \langle z, Aw+B\rangle\right)^{-r} \\
&= (-B^*z+D^*)^r\left(-B^*z+D^* - \langle A^*z, w\rangle + \langle C, w\rangle\right)^{-r} \\
&= \left(1 - \langle (A^*z - C)(-B^*z+D^*)^{-1}, w\rangle\right)^{-r} \\
&= C_\sigma(K_w)(z)
\end{aligned}$$

Since the K_w span a dense set of $\mathcal{H}$ and T_g^{-1}, C_φ^*, and $(T_h^*)^{-1}$ are each bounded operators, $C_\sigma = T_g^{-1} C_\varphi^* (T_h^*)^{-1}$ and C_σ is a bounded operator. The formula in conclusion follows easily from this relation. ∎

The hypotheses of Theorem 16 are more restrictive than necessary. In particular the positivity of r is inessential, and rather than all $H^\infty(B_N)$ functions being multipliers of $\mathcal{H}$ it suffices to only know that g, h, and their reciprocals are multipliers. Typically, though, when r is negative determining the multipliers of $\mathcal{H}$ becomes more difficult.

5 Schroeder's Equation

For an arbitrary analytic map φ of the disk D to itself, fixing 0 with $\varphi'(0) = \lambda$ satisfying $0 < |\lambda| < 1$, work of Koenigs in 1884 ([13]) gives an essentially unique analytic function f in D solving Schroeder's functional equation

$$f \circ \varphi = \lambda f. \tag{6}$$

By analogy, when φ is an analytic map of B_N into B_N with $\varphi(0) = 0$ we may seek a $\mathbf{C}^N$-valued analytic function f on B_N solving the several variable Schroeder equation

$$f \circ \varphi = Af \tag{7}$$

where $A = \varphi'(0)$ is an $N \times N$ matrix. To avoid certain exceptional cases we assume $\varphi'(0)$ has no eigenvalues of modulus 1. In [8] we use tools from the theory of compact composition operators to construct solutions to Schroeder's functional equation in several variables. Certain complications may arise is the case $N > 1$ which are never present when $N = 1$; in particular Equation (7) may fail to have a locally univalent solution f in B_N when certain algebraic relationships hold between the eigenvalues of $\varphi'(0)$ (see [8]). However, in this section we show that when φ is a linear fractional map fixing 0, Equation (7) always has an analytic solution which is univalent on B_N if $\varphi'(0)$ has no eigenvalue of modulus 1. The argument uses the Kreĭn adjoint of φ introduced in the last section.

Since we are assuming $\varphi(0) = 0$ (this being a normalization of the requirement that φ has a fixed point in B_N) the maps we wish to consider may be written as

$$\varphi(z) = \frac{Az}{\langle z, C\rangle + 1}$$

Recall that $\varphi(B_N) \subset B_N$ implies $|C| < 1$. The case $C = 0$ is uninteresting, so we assume $C \neq 0$. Note that $\varphi'(0) = A$. We will produce a C^N- valued analytic solution $f(z)$ to Schroeder's equation $f \circ \varphi = Af$ which is initially defined in a neighborhood of 0 and univalent there. In the case that $\varphi'(0)$ has no eigenvalue of modulus 1, this solution will in fact be analytic and univalent on all of B_N.

Lemma 17 *If $\varphi(z) = Az/(\langle z, C\rangle + 1)$ maps B_N into B_N then C is orthogonal to the null-space of $A - I$.*

Proof. Suppose ζ is in $\ker(A - I)$ with $|\zeta| = 1$. If $\langle \zeta, C\rangle \neq 0$ choose the complex number λ with $|\lambda| = 1$ and $\lambda\langle \zeta, C\rangle = -|\langle \zeta, C\rangle| \neq 0$. Then $\lambda\zeta$ is a point of the unit sphere but

$$\varphi(\lambda\zeta) = \frac{A(\lambda\zeta)}{\langle \lambda\zeta, C\rangle + 1} = \frac{\lambda\zeta}{1 - |\langle \zeta, C\rangle|}$$

which is not in the closed unit ball. Thus φ doesn't map B_N into B_N, a contradiction. ∎

Theorem 18 *Suppose $\varphi : B_N \to B_N$ is a linear fractional map with $\varphi(0) = 0$. Then there is an invertible linear fractional map f defined in a neighborhood of 0, with $f \circ \varphi = \varphi'(0)f$.*

Proof. Write $\varphi(z) = Az/(\langle z, C\rangle + 1)$. Since C is orthogonal to the null-space of $A - I$ we may solve $(I - A^*)P = C$ for P. Define f by

$$f(z) = \frac{z}{\langle z, P\rangle + 1}$$

Then f is analytic in $\{z : \langle z, P\rangle \neq -1\}$, and this is a neighborhood of 0.

A computation shows that $f\circ\varphi = Af = \varphi'(0)f$. Moreover f is univalent, since if

$$\frac{z}{\langle z, P\rangle + 1} = \frac{w}{\langle w, P\rangle + 1} \tag{8}$$

we must have $w = \lambda z$ for some $\lambda \in C$. But then Equation (8) implies $\lambda = 1$ and $z = w$. ∎

The next result shows that if $\varphi'(0)$ has no eigenvalue of modulus 1, then the mapping f given by the previous theorem is actually analytic (and univalent) on B_N.

Theorem 19 *Let $\varphi(z) = Az/(\langle z, C\rangle+1)$ map B_N into B_N and assume that A has no eigenvalue of modulus 1. Then there is a univalent, $\mathbf{C}^N$-valued mapping f defined on B_N with $f \circ \varphi = \varphi'(0)f$.*

Proof. By hypothesis, 1 is not an eigenvalue of A^*. Let $P = (I - A^*)^{-1}C$ and set $f(z) = z/(\langle z, P\rangle + 1)$. As before, $f \circ \varphi = \varphi'(0)f$ and f is univalent. We need only check that f is analytic on B_N. For this it suffices to show $|P| \leq 1$, because it will then follow that $|\langle z, P\rangle| \leq |z||P| \leq |z| < 1$, so $\langle z, P\rangle + 1$ is non-zero in the open ball, B_N.

Since φ maps B_N into B_N, so does its Krein adjoint $\tau(z) = A^*z - C$ by Proposition 11. Since $|C| < 1$, this says $\tau(C)$, $\tau(\tau(C))$, $\tau(\tau(\tau(C)))$, ... are all in B_N. Computing these iterates, we find that for each positive integer n

$$(A^*)^nC - (A^*)^{n-1}C - \cdots - A^*C - C \in B_N \tag{9}$$

Since all eigenvalues of A have modulus less than 1, there is a positive integer m so that $|A^m| < 1$. Thus, given $\epsilon > 0$, there exists M so that for all $n \geq M$

$$|(A^*)^nC| < \epsilon \tag{10}$$

Since $(I - A^*)^{-1} = \sum (A^*)^k$ we have

$$C + A^*C + (A^*)^2C + \cdots + (A^*)^nC \to (I - A^*)^{-1}C = P$$

as $n \to \infty$. We may suppose that M is chosen large enough that

$$|P - (C + A^*C + (A^*)^2C + \cdots + (A^*)^{n-1}C)| < \epsilon \tag{11}$$

for all $n \geq M$. Properties (9), (10), and (11) say that $|P| < 1 + 2\epsilon$; since ϵ is arbitrary, $|P| \leq 1$.

$\blacksquare$

References

[1] J. Arazy, A survey of invariant Hilbert spaces of analytic functions on bounded symmetric domains, *in* "Multivariable Operator Theory," Contemporary Mathematics **185**, Amer. Math. Soc., Providence, 1995, 7–65.

[2] J. Bognar, "Indefinite Inner Product Spaces", Ergebnisse der Math., **78**, Springer-Verlag, New York, 1974.

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

[4] C. C. Cowen, Linear fractional composition operators on H^2, *J. Integral Equations Operator Theory* **11**(1988), 151–160.

[5] C. C. Cowen, An application of Hadamard multiplication to operators on weighted Hardy spaces, *Linear Alg. Appl.* **133**(1990), 21–32.

[6] C. C. Cowen, Composition operators on Hilbert spaces of analytic functions: A status report, *Proc. Symposia Pure Math.* **51**(part 1)(1990), 131–145.

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

[8] C. C. Cowen and B. D. MacCluer, Schroeder's equation in several variables, preprint, 1997.

[9] L. A. Harris, Linear fractional transformations of circular domains in operator spaces, *Indiana Univ. Math. J.* **41**(1992), 125–147.

[10] L. A. Harris, Unbounded symmetric homogeneous domains in spaces of operators, *Ann. Scuola Norm. Sup. Pisa* **22**(1995), 449–467.

[11] K. Hoffman, "Banach Spaces of Analytic Functions," Prentice–Hall, Englewood Cliffs, 1962.

[12] P. R. Hurst, Relating composition operators on different weighted Hardy spaces, *Arch. Math. (Basel)* **68**(1997), 503–513.

[13] G. Koenigs, Recherches sur les intégrales de certaines équations functionelles, *Annales Ecole Normale Superior (Sér. 3)* **1**(1884), supplément, 3–41.

[14] M. G. Kreĭn and J. L. Šmul'jan, Plus-operators in a space with indefinite metric, *Amer. Math. Soc. Transl. (2)* **85**(1969), 93–113.

[15] M. G. Kreĭn and J. L. Šmul'jan, On linear fractional transformations with operator coefficients, *Amer. Math. Soc. Transl. (2)* **103**(1974), 125–152.

[16] B. D. MacCluer, Spectra of compact composition operators on $H^p(B_N)$, *Analysis* **4**(1984), 87–103.

[17] B. D. MacCluer, Spectra of automorphism–induced composition operators on $H^p(B_N)$, *J. London Math. Soc. (2)* **30**(1984), 95–104.

[18] B. D. MacCluer and J. H. Shapiro, Angular derivatives and compact composition operators on Hardy and Bergman spaces, *Canadian J. Math.* **38**(1986), 878–906.

[19] R. Nevanlinna, "Analytic Functions," Springer–Verlag, Berlin, 1970.

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

[21] E. A. Nordgren, Composition operators on Hilbert space, *in* "Hilbert Space Operators," Lecture Notes in Math. **693**, Springer–Verlag, Berlin, 1978, 37–63.

[22] W. Rudin, "Function Theory in the Unit Ball of C^n," Grundlehren der math. Wiss. **241**, Springer–Verlag, New York, 1980.

[23] E. M. Stein, "Boundary Behavior of Holomorphic Functions of Several Complex Variables," Princeton University Press, Princeton, 1972.

[24] W. R. Wogen, The smooth mappings which preserve the Hardy space $H^2_{B_n}$, *Operator Theory: Advances Appl.* **35**(1988), 249–267.