SPHERE-PRESERVING MAPS IN INVERSIVE GEOMETRY

A. F. BEARDON AND D. MINDA

(Communicated by Juha M. Heinonen)

ABSTRACT. We give an extensive discussion of sphere-preserving maps defined on subdomains of Euclidean n-space, and their relationship to Möbius maps and to the preservation of cross-ratios. In the case $n = 2$ (the complex plane) we also relate these ideas to the solutions of certain functional equations.

1. INTRODUCTION

In this paper we consider various connections that exist between Möbius transformations acting on $\mathbb{R}^n$, absolute cross-ratios of points in $\mathbb{R}^n$, sphere-preserving maps of $\mathbb{R}^n$ into itself and, in the case of the complex plane (where $n = 2$), the solutions of the simultaneous functional equations

$$(1.1) \qquad\qquad f(z + w) = f(z) + f(w),$$

$$(1.2) \qquad\qquad f(zw) = f(z)f(w).$$

Given four distinct points a, b, c and d in the extended complex plane $\mathbb{C}_\infty$, the cross-ratio and the absolute cross-ratio of these points are defined by

$$(1.3) \qquad [a, b, c, d] = \frac{(a - c)(b - d)}{(a - d)(b - c)}, \qquad |a, b, c, d| = \frac{|a - c| \cdot |b - d|}{|a - d| \cdot |b - c|},$$

respectively. Of course, $|a, b, c, d|$ is the absolute value of $[a, b, c, d]$, but the absolute cross-ratio is defined for points in $\mathbb{R}^n$ whereas the cross-ratio is not. A Möbius map

$$(1.4) \qquad\qquad g(z) = \frac{az + b}{cz + d}, \quad ad - bc \neq 0,$$

of $\mathbb{C}_\infty$ onto itself preserves cross-ratios (and absolute cross-ratios); a Möbius map of $\mathbb{R}_\infty^n$ onto itself preserves absolute cross-ratios (see Section 2).

There are well-known elementary proofs that if f is a continuous injective map of the extended complex plane $\mathbb{C}_\infty$ into itself that preserves cross-ratios, or that maps circles into circles, then f is a Möbius map. Similarly, the continuous simultaneous solutions of (1.1) and (1.2) are well known. In fact, in both cases the same conclusions can be drawn from significantly weaker hypotheses, and we are concerned with such results here. We shall also investigate (where appropriate) local versions of these problems for maps of $\mathbb{R}_\infty^n$ into itself; this case has largely been ignored in the literature.

The connection between Möbius maps, cross-ratios and the preservation of circles is well known, so we shall begin by showing the relevance of the functional equations

(1.1) and (1.2) to these topics; this has already been explored in [1]. We say that an injective map f of $\mathbb{C}_\infty$ into itself *preserves cross-ratios with value α* if and only if $[a, b, c, d] = \alpha$ implies that $[f(a), f(b), f(c), f(d)] = \alpha$.

Theorem 1.1. *Suppose that f is an injective map of $\mathbb{R}^1_\infty$ into itself, or of $\mathbb{C}_\infty$ into itself, and that f fixes 0, 1 and ∞. Then f preserves cross-ratios with value -1 if and only if f satisfies the functional equations* (1.1) *and* (1.2) *on $\mathbb{R}^1$ or $\mathbb{C}$, respectively.*

Theorem 1.1 (whose proof will be given later) provides us with the link between Möbius maps and the functional equations. Suppose first that an injective map $g : \mathbb{R}^1_\infty \to \mathbb{R}^1_\infty$ preserves cross-ratios with value -1. Then there is a Möbius map g_1 such that $g_1 g$ fixes 0, 1 and ∞, so that if we put $f = g_1 g$, then, by Theorem 1.1, f satisfies (1.1) and (1.2). As the only real solution of (1.1) and (1.2) that fixes 0, 1 is the identity map, we find that *if $g : \mathbb{R}^1_\infty \to \mathbb{R}^1_\infty$ preserves cross-ratios with value -1, then g is Möbius.* By contrast, there exist injective complex solutions of (1.1) and (1.2) that are not continuous, and hence not Möbius (see [10], Exercise 6, p. 317), so we see that the condition that $f : \mathbb{C}_\infty \to \mathbb{C}_\infty$ preserves cross-ratios with value -1 does *not* imply that f is Möbius. Of course, the functional equations have no meaning in higher dimensions.

We now know that if $f : \mathbb{C}_\infty \to \mathbb{C}_\infty$ preserves all cross-ratios, then it is Möbius, but that if it only preserves cross-ratios with value -1, then it need not be Möbius. What can be said of the intermediate situations? More precisely, is it possible to characterize subsets E of $\mathbb{C}$ with the property that if $[f(a), f(b), f(c), f(d)] \in E$ whenever $[a, b, c, d] \in E$, then f is Möbius? In fact, there is a dense subset E of $\mathbb{C}$ for which this is not true because any injective solution f of (1.1) and (1.2) is such that either $f(z)$ fixes every complex rational number or $\overline{f(z)}$ does. On the other hand, it was shown recently in [9] (although not stated in this form) that if f is meromorphic in $\mathbb{C}$, and if f preserves cross-ratios of absolute value 1, then f is Möbius (this is the case when E is the unit circle). We shall show that the assumption that f is meromorphic here is irrelevant (just as smoothness conditions in the classical results are irrelevant), so that the preservation of cross-ratios of absolute value 1 is, by itself, enough to guarantee that the map is Möbius. This shows, incidentally, that if $f : \mathbb{C} \to \mathbb{C}$ satisfies (1.1) and (1.2), fixes 0, 1 and ∞, and maps the unit circle into itself (that is, it preserves numbers of norm 1), then f is the identity.

We now consider some of these ideas in a little more detail. First, we consider the main theorem in [9] which we choose to write in terms of cross-ratios rather than Apollonius quadrilaterals. In [9], four distinct points a, b, c and d in $\mathbb{C}$ are defined to be the vertices of an *Apollonius quadrilateral* if and only if $|a, b, c, d| = 1$ (although cross-ratios are never mentioned explicitly in [9]). A Möbius map sends one Apollonius quadrilateral to another, and the main result in [9] is the converse to this which, when stated in terms of absolute cross-ratios, reads as follows.

Theorem A. *Suppose that f is meromorphic in some domain in $\mathbb{C}$, and that for every a, b, c and d, $|a, b, c, d| = 1$ implies $|f(a), f(b), f(c), f(d)| = 1$. Then f is a Möbius map.*

In the case of a meromorphic function, much more is true and Theorem A can be greatly strengthened without any difficulty. It is well known that an analytic

bijection of one disk onto another is a Möbius map, and this fact leads to the following result.

Theorem B. *Suppose that f is meromorphic in a region Ω in $\mathbb{C}_\infty$. Suppose also that there is an open disc D, whose closure $\overline{D}$ lies in Ω, such that f is injective in $\overline{D}$ and maps the boundary ∂D of D into a circle C. Then f is the restriction of some Möbius map to Ω.*

The proof is easy. As f is meromorphic and injective on a slightly larger disc than $\overline{D}$, its restriction to $\overline{D}$ is a homeomorphism. This implies that $f(\partial D) = C$, and also that f maps D bijectively onto one of the two components of $\mathbb{C}_\infty \backslash C$. This in turn implies that the restriction of f to D is Möbius and analytic continuation implies that f is Möbius throughout $\mathbb{C}_\infty$. To show that Theorem B includes Theorem A we have only to observe first that as f preserves absolute cross-ratios equal to 1, it maps circles into circles (this is elementary and will be discussed later), and second that f is locally injective at all but a countable set of isolated points.

Our aim is to give an extensive discussion of sphere-preserving maps in all dimensions, and to give results of this type that follow from local, rather than global, hypotheses. In Section 2 we introduce some terminology and notation for higher-dimensional spaces, and in Section 3 we discuss cross-ratios. Section 4 is devoted to a brief discussion of maps of the extended real line $\mathbb{R}^1_\infty$ onto itself, and this is the first step of a later proof by induction on the dimension. In Section 5 we consider maps that preserve absolute cross-ratios equal to 1, and in Section 6 we consider sphere-preserving maps.

We mention that related results can be found in [5], [11] and [12].

2. Higher-dimensional spaces

We write $x = (x_1, \ldots, x_n)$ for a point x in $\mathbb{R}^n$, and the Euclidean distance in $\mathbb{R}^n$ is denoted by $|x - y|$. The space $\mathbb{R}^n \cup \{\infty\}$ is denoted by $\mathbb{R}^n_\infty$, and this supports the usual chordal distance (see [2], p. 22, and [13], p. 113). The absolute cross-ratio is defined in $\mathbb{R}^n_\infty$ in terms of the chordal distance, and this agrees with the definition (1.3) in terms of the Euclidean distance when the four points are in $\mathbb{R}^n$.

If a and b are in $\mathbb{R}^n$, $[a, b]$ denotes the closed line segment with a and b as endpoints. A *hyperplane* in $\mathbb{R}^n_\infty$ is the set of x (including ∞) satisfying $|x - a| = |x - b|$ for some distinct a and b in $\mathbb{R}^n$; a *Euclidean sphere* is the set of x satisfying $|x - a| = r$ for some a in $\mathbb{R}^n$ and some positive r. By analogy with the terminology used in the complex case, a *sphere* is either a hyperplane or a Euclidean sphere. Note that a sphere in $\mathbb{R}^n_\infty$ has dimension $n - 1$.

A *Möbius map* g acting on $\mathbb{R}^n_\infty$ is a composition of a finite number of reflections (or inversions) in spheres, and the group of Möbius maps acting on $\mathbb{R}^n_\infty$ is denoted by $\mathcal{M}_n$. Briefly, we review the main properties of Möbius maps. First, Möbius maps preserve spheres, and symmetry with respect to spheres. It is known (and easy to prove) that a map $g : \mathbb{R}^n_\infty \to \mathbb{R}^n_\infty$ is Möbius if and only if it preserves absolute cross-ratios ([2], p. 32). Finally, Möbius maps satisfy a unique continuation property: if f is a map of a region Ω in $\mathbb{R}^n_\infty$ into $\mathbb{R}^n_\infty$, and if for each point a in Ω there is an open ball B_a containing a such that $f|B_a$ is the restriction to B_a of an element of $\mathcal{M}_n$, then f is the restriction to Ω of an element of $\mathcal{M}_n$ (as usual, the restriction of a map g to a set E is denoted by $g|E$). This follows immediately from the fact that if two Möbius maps agree on an open set, then they are identical (because they agree on and inside a sphere, and they preserve symmetry).

For each n we embed $\mathbb{R}_\infty^n$ in $\mathbb{R}_\infty^{n+1}$ by identifying $(x_1, \ldots, x_n)$ and $(x_1, \ldots, x_n, 0)$ and with this identification every element of $\mathcal{M}_n$ extends (in exactly two ways) to an element of $\mathcal{M}_{n+1}$ (see [2], p. 33). To be definite, we take this extension to preserve the upper half-space (given by $x_{n+1} > 0$) of $\mathbb{R}_\infty^{n+1}$. We also identify the extended complex plane $\mathbb{C}_\infty$ with $\mathbb{R}_\infty^2$, and the Möbius maps acting on $\mathbb{R}_\infty^2$ are then of the form $g(z)$ or $g(\bar{z})$, where g is given in (1.4). Henceforth, each of these types of maps will be considered to be a Möbius map. For more details on Möbius maps we refer the reader to [2] and [13].

3. Cross-ratios

In this section we take a closer look at cross-ratios. First consider points in $\mathbb{C}_\infty$. Then the cross-ratio $[a, b, c, d]$ is defined by (1.3) whenever a, b, c and d are distinct. If these points are not distinct, then the value of the cross-ratio is defined (by continuity) *if and only if no three of the points a, b, c and d coincide.* Explicitly,

(1) $[a, b, c, d] = 0$ if and only if $a = c$ or $b = d$ (or both);
(2) $[a, b, c, d] = 1$ if and only if $a = b$ or $c = d$ (or both);
(3) $[a, b, c, d] = \infty$ if and only if $a = d$ or $b = c$ (or both).

Likewise, in all dimensions the absolute cross-ratio is defined whenever a, b, c and d are distinct. When they are not distinct, similar statements to (1) and (3) hold, but there is no statement corresponding to (2) (because, for example, $|i, 1, 0, \infty| = 1$).

We shall now describe some geometric properties of cross-ratios. First, in Lemma 3.1 we give a geometric characterization of quadruples of points in $\mathbb{R}_\infty^n$ with absolute cross-ratio equal to one. In Lemma 3.2, we give a geometric characterization of quadruples of points in $\mathbb{R}_\infty^1$ with cross-ratio equal to -1, and in Lemma 3.3 we give a condition for three geodesics in the hyperbolic plane to be concurrent. Finally, in Lemma 3.4, we give a sufficient condition (in terms of cross-ratios) for a function to be injective.

Given two points a and b in $\mathbb{R}_\infty^n$, there is a unique pencil of spheres each of which has a and b as inverse points. These are called the *Apollonius spheres* determined by a and b, and we denote this pencil of spheres by $\mathcal{A}(a, b)$. Of course, $\mathcal{A}(a, b) = \mathcal{A}(b, a)$, and $\mathcal{A}(a, \infty)$ consists of all Euclidean spheres with center a. If a, $b \in \mathbb{R}^n$, then $\mathcal{A}(a, b)$ consists of the spheres given by

$$\frac{|x - a|}{|x - b|} = k,$$

where $0 < k < +\infty$. For $k \neq 1$ this is a Euclidean sphere, while for $k = 1$ it is the hyperplane, which we denote by $\perp(a, b)$, that is, the perpendicular bisector of the segment $[a, b]$. Each point x (other than a and b) of $\mathbb{R}_\infty^n$ lies on exactly one of the (non-degenerate) Apollonius spheres in $\mathcal{A}(a, b)$, and we denote this sphere by $S_{a,b}[x]$. Note that for a, $b \in \mathbb{R}^n$, $S_{a,b}[\infty] = \perp(a, b)$. Since $|a, b, c, d| = 1$ if and only if

$$(3.1) \qquad\qquad \frac{|a - c|}{|a - d|} = \frac{|b - c|}{|b - d|},$$

the following lemma is clear.

Lemma 3.1. *For distinct a, b, c and d in $\mathbb{R}_\infty^n$, $|a, b, c, d| = 1$ if and only if $S_{c,d}[a] = S_{c,d}[b]$.*

Lemma 3.1 says that $|a, b, c, d| = 1$ if and only if there is a (necessarily unique) sphere through a and b that has c and d as inverse points. If we rearrange (3.1) we see that $S_{c,d}[a] = S_{c,d}[b]$ if and only if $S_{a,b}[c] = S_{a,b}[d]$; thus there exists a sphere S' through a and b with inverse points c and d if and only if there exists a sphere S'' through c and d with inverse points a and b. Note that in these circumstances, S' and S'' are orthogonal to each other (this follows easily when $a = 0$ and $b = \infty$, and it is true in general as Möbius maps are angle-preserving).

Lemma 3.2. *Given distinct points a, b, c and d in $\mathbb{R}^1_\infty$, let $C_{a,b}$ and $C_{c,d}$ be the circles in $\mathbb{C}_\infty$ with diameters $[a, b]$ and $[c, d]$, respectively. Then $C_{a,b}$ crosses $C_{c,d}$ if and only $[a, b, c, d] < 0$. Further, $[a, b, c, d] = -1$ if and only if $C_{a,b}$ and $C_{c,d}$ are orthogonal.*

By invariance under Möbius maps, this result need only be verified when $a = 0$, $b = \infty$ and $c = 1$, and in this case it is trivially true. The geometry of geodesics in the hyperbolic plane $\mathbb{H} = \{x + iy : y > 0\}$ is intimately connected to the cross-ratios of their endpoints, and we now identify each geodesic γ in $\mathbb{H}$ with its unordered pair of endpoints, say $\{a, b\}$, and write $\langle a, b \rangle$ for γ. Lemma 3.2 implies that the two geodesics $\langle a, b \rangle$ and $\langle c, d \rangle$ intersect in $\mathbb{H}$ if and only if $[a, b, c, d] < 0$, and are orthogonal if and only if $[a, b, c, d] = -1$. The same result holds for four points in $\mathbb{C}_\infty$ if we construct the geodesics in hyperbolic 3-space. The next lemma exploits this connection still further.

Lemma 3.3. *Let a, b, p, q, u and v be six distinct points in $\mathbb{R}^1_\infty$, and suppose that $\langle p, q \rangle$ crosses $\langle u, v \rangle$ at the point ζ in $\mathbb{H}$. Then a necessary and sufficient condition for $\langle a, b \rangle$ to pass through ζ (so that the three geodesics are concurrent) is that $[u, v, p, a] \, [u, v, q, b] = 1$.*

Proof. We may assume that $u = 0$ and $v = \infty$. Then (by calculation)

$$[u, v, p, a] \, [u, v, q, b] = 1 \quad \text{if and only if} \quad pq = ab.$$

Suppose now that $\langle u, v \rangle$ and $\langle p, q \rangle$ cross at the point iy, where $y > 0$. Then, by the Intersecting Chord Theorem for the Euclidean circle with diameter $[p, q]$, we have $pq = -y^2$. Equally, $\langle a, b \rangle$ meets $\langle u, v \rangle$ at iy if and only if $ab = -y^2$ and the result follows. $\square$

Next, we need to address the issue of injectivity of functions. For most maps $f : \mathbb{R}^n_\infty \to \mathbb{R}^n_\infty$, there are points a, b, c and d for which $|a, b, c, d|$ is defined but $|f(a), f(b), f(c), f(d)|$ is not (for example, if y is such that the equation $f(x) = y$ has three distinct solutions a, b and c, and if d is distinct from these). However, a mild assumption about cross-ratios does imply injectivity.

Lemma 3.4. *Let Ω be a region in $\mathbb{R}^n_\infty$, and suppose that $f : \Omega \to \mathbb{R}^n_\infty$ assumes at least three values, and is such that if $|a, b, c, d| \neq 0$, then $|f(a), f(b), f(c), f(d)|$ is defined and non-zero. Then f is injective. A similar conclusion holds if 0 is replaced by ∞.*

Proof. Suppose that $f(a) = f(c)$ for some a and c in Ω. As f assumes at least three values we can choose b and d in Ω such that the values $f(a), f(b), f(d)$ are distinct. This implies that the points a, b, c, d are distinct except possibly that $a = c$. Now if $a \neq c$, then $|a, b, c, d| \neq 0$ but $|f(a), f(b), f(c), f(d)| = 0$ which is a contradiction. We conclude that f is injective. The corresponding statement concerning ∞ is proved in a similar way. $\square$

Finally, we introduce some terminology that we will use when discussing functions that preserve absolute cross-ratios that are equal to 1.

Definition. We say that a function f mapping a region Ω in $\mathbb{R}_\infty^n$ into $\mathbb{R}_\infty^n$ *has the property $\mathcal{A}$ in Ω* (the Apollonius property) if f is injective in Ω, and if for each distinct a, b, c, d in Ω with $|a, b, c, d| = 1$ we have $|f(a), f(b), f(c), f(d)| = 1$. We say that *$f$ has property $\mathcal{A}$ locally in Ω* if each point of Ω has a neighborhood N such that the restriction of f to N has the property $\mathcal{A}$, and we emphasize that this implies f is injective in some neighborhood of each point of Ω.

4. Maps of intervals into $\mathbb{R}_\infty^1$

We begin by sketching the proof of Theorem 1.1 and the argument is valid whether f acts on $\mathbb{R}_\infty^1$ or on $\mathbb{C}_\infty$. First, if f satisfies the two functional equations (1.1) and (1.2), then $f(-1) = f(0) - f(1) = -1$, so that if $[a, b, c, d] = -1$, then

$$[f(a), f(b), f(c), f(d)] = f\big([a, b, c, d]\big) = f(-1) = -1.$$

An obvious modification holds when one of the points is ∞ (because the cross-ratios are still rational expressions), so if f satisfies (1.1) and (1.2), then it preserves cross-ratios with value -1.

Suppose now that f is injective, fixes 0, 1 and ∞, and preserves cross-ratios with value -1. Now $[c, \infty, a, b] = -1$ if and only if $c = (a + b)/2$. Thus, given $c = (a + b)/2$, we have $[f(c), \infty, f(a), f(b)] = -1$ and hence

$$f\left(\frac{a+b}{2}\right) = \frac{f(a) + f(b)}{2}.$$

If we put $b = 0$ we obtain $f(a/2) = f(a)/2$ for all a, and hence f satisfies the functional equation (1.1). As $[x^2, 1, x, -x] = -1$ we see that $[f(x^2), 1, f(x), -f(x)] = -1$. Writing $v = f(x^2)$, $u = f(x)$ and noting that $f(-x) = -u$, we find $[v, 1, u, -u] = -1$ which simplifies to $v = u^2$. Thus for any x, $f(x^2) = f(x)^2$. If we now let $x = a + b$ we easily see that f also satisfies the functional equation (1.2). This completes the proof of Theorem 1.1.

Note that in the case when f acts on $\mathbb{R}_\infty^1$, this proof gives more. As in this case, $f(x^2) = f(x)^2 \geq 0$, so $f(y) \geq 0$ if $y \geq 0$. This and (1.1) imply that f is increasing on $\mathbb{R}^1$, and as it can be shown (in the usual way) that $f(r) = rf(1) = r$ for all rational numbers r, it follows that $f(x) = x$ on $\mathbb{R}^1$. For a similar global result and proof, in projective geometry, see [3], pp. 32–33, where it is credited to Darboux.

The rest of this section contains a proof of the following result.

Theorem 4.1. *Let Ω be an open interval in $\mathbb{R}_\infty^1$, and let $f : \Omega \to \mathbb{R}_\infty^1$ be any map. Then the following are equivalent:*
    (1)   *f is the restriction of some Möbius map to Ω;*
    (2)   *f preserves cross-ratios of points in Ω;*
    (3)   *f assumes at least three values and preserves absolute cross-ratios;*
    (4)   *f has the property $\mathcal{A}$ locally in Ω;*
    (5)   *$[f(a), f(b), f(c), f(d)] = -1$ when a, b, c and d are distinct points in Ω with $[a, b, c, d] = -1$.*

Proof. It is clear that (1) implies (5). Next, a simple modification of the argument given above in the proof of Theorem 1.1 shows that if (5) holds, then, given any x_0 in Ω, f is the restriction of some Möbius map to some neighborhood of the point x_0; we omit the details. The unique continuation property of Möbius maps now

guarantees that (1) follows; thus we have now shown that (1) and (5) are equivalent. The reader should note that there is no assumption in (5) that f is injective, or continuous, or even that $[f(a), f(b), f(c), f(d)]$ is defined when $[a, b, c, d] \neq -1$.

We have just seen that (1) and (5) are equivalent. It is clear that (1) implies (2), and that (2) implies (3). Further, the implication (3) implies (4) is trivial if we use (3) in conjunction with Lemma 3.4 which shows that f is injective on Ω. It remains to show that (4) implies (1), so we assume now that (4) holds. Take any point x_0 in $\mathbb{R}^1_\infty$, and let N be an open interval about x_0 on which f is injective. Suppose now that a, b, c and d are distinct points in N with $[a, b, c, d] = -1$. As (4) holds, we find that $|f(a), f(b), f(c), f(d)| = 1$, so that $[f(a), f(b), f(c), f(d)] = \pm 1$. Now as f is injective on N, the points $f(a)$, $f(b)$, $f(c)$ and $f(d)$ are distinct and so this cross-ratio cannot be 1 (see the remark in Section 3). It follows that the cross-ratio is -1 so that the condition (5) holds with Ω replaced by N. As (5) implies (1) (for any choice of Ω), we now see that f agrees with some Möbius map, say g, on N. As this holds for any x_0 in Ω we see (as before) that (1) holds. The proof is now complete. $\square$

Theorem 4.1 shows that if $f : \mathbb{R}^1_\infty \to \mathbb{R}^1_\infty$ preserves cross-ratios, then f extends naturally to a Möbius map that maps the upper half-plane $\mathbb{H}$ onto itself. Now $\mathbb{H}$ is the hyperbolic plane, and $\mathbb{R}^1_\infty$ is its boundary, so it is of interest to see how one can give a geometric construction of the action of f as a hyperbolic isometry of $\mathbb{H}$ starting from the action of f on $\mathbb{R}^1_\infty$ alone.

Given that $f : \mathbb{R}^1_\infty \to \mathbb{R}^1_\infty$ preserves cross-ratios, it is immediate that f is injective. As f is injective it induces a map from the space of all hyperbolic geodesics to itself by the rule

$$f^* : \langle a, b \rangle \mapsto \langle f(a), f(b) \rangle.$$

Now take any z_0 in $\mathbb{H}$ and choose two geodesics, say $\langle p, q \rangle$ and $\langle u, v \rangle$ that cross at z_0 in $\mathbb{H}$. It follows that $[u, v, p, q] < 0$ (see Lemma 3.2) and so, as f preserves cross-ratios, the geodesics $\langle f(u), f(v) \rangle$ and $\langle f(p), f(q) \rangle$ cross at some point w_0. We want to define $f(z_0) = w_0$ and to do this we need to know that w_0 is independent of the original choice of the two geodesics through z_0. This, however, is an immediate consequence of Lemma 3.3 for this implies that f^* maps the pencil of all geodesics through z_0 to the pencil of all geodesics through w_0. We have now extended f so as to be a map of $\mathbb{H}$ into itself.

We now show that f is an isometry of $\mathbb{H}$. Take two distinct points z_1 and z_2 in $\mathbb{H}$, let ℓ be the geodesic through z_1 and z_2, and let ℓ_1 and ℓ_2 be the geodesics through z_1 and z_2, respectively, that are orthogonal to ℓ. Write $\ell = \langle u, v \rangle$, $\ell_1 = \langle p, q \rangle$ and $\ell_2 = \langle a, b \rangle$. It follows that $[u, v, p, q] = -1$ and $[u, v, a, b] = -1$, and also

$$[p, q, a, b] = \tanh^2 \tfrac{1}{2} \rho(z_1, z_2),$$

where ρ is the hyperbolic metric in $\mathbb{H}$ (one only has to check this when $u = 0$, $v = \infty$ and $p = 1$, in which case $q = -1$, $a = R$ and $b = -R$, say, where $R > 0$). The invariance of cross-ratios shows immediately that $\rho\big(f(z_1), f(z_2)\big) = \rho(z_1, z_2)$.

5. Maps with the property $\mathcal{A}$ in $\mathbb{R}^n_\infty$

We will characterize maps that locally have property $\mathcal{A}$, but first we need a lemma.

Lemma 5.1. *Suppose that Ω is a region in $\mathbb{R}^n_\infty$ and that $f : \Omega \to \mathbb{R}^n_\infty$ has property $\mathcal{A}$. Then, for any sphere S with $S \cap \Omega \neq \emptyset$, there is a sphere S' with $f(S \cap \Omega) \subset S'$. Also, if c and d in Ω are symmetric about S, then $f(c)$ and $f(d)$ are symmetric about S'. In particular, if S_0 is a sphere that is orthogonal to S at a point in Ω, and if $f(S_0) \subset S'_0$, then S'_0 is orthogonal to S'.*

Proof. Suppose that S is a sphere in $\mathbb{R}^n_\infty$ with $S \cap \Omega \neq \emptyset$, and take any a in $S \cap \Omega$. Suppose that c and d in Ω are symmetric about S (such symmetric points exist because Ω is open). Then $S = S_{c,d}[a]$, and we will show that $f(S) \subset S_{f(c),f(d)}[f(a)]$. For any point b in S we have $|a, b, c, d| = 1$. Then $|f(a), f(b), f(c), f(d)| = 1$, so by Lemma 3.1, $f(b) \in S_{f(c),f(d)}[f(a)]$. It follows that f maps S into S', where $S' = S_{f(c),f(d)}[f(a)]$, and that $f(c), f(d)$ are symmetric about S'. Finally, suppose that the sphere S_0 is orthogonal to S, and select points c and d on S_0 that are symmetric with respect to S. As $f(c)$ and $f(d)$ are symmetric with respect to S' every sphere through these two points, and in particular S'_0, is orthogonal to S'. $\square$

Theorem 5.2. *Suppose that Ω is a region in $\mathbb{R}^n_\infty$, and that $f : \Omega \to \mathbb{R}^n_\infty$ has property $\mathcal{A}$ locally in Ω. Then f is the restriction to Ω of some element in $\mathcal{M}_n$.*

First, we recall that f is Möbius if it is locally Möbius; thus, it suffices to establish the conclusion of Theorem 5.2 in some neighborhood of each point of Ω. The proof of Theorem 5.2 is by induction on n, and we already know that the conclusion is true when $n = 1$ for this is Theorem 4.1. We suppose now that it is true for $n = k$, *and for all regions Ω in $\mathbb{R}^k_\infty$*, and consider a region Ω in $\mathbb{R}^{k+1}_\infty$ and a function $f : \Omega \to \mathbb{R}^{k+1}_\infty$ that has property $\mathcal{A}$ locally in Ω.

Lemma 5.3. *Given that the induction hypothesis holds when $n = k$, suppose that Ω is a region in $\mathbb{R}^{k+1}_\infty$, and that $f : \Omega \to \mathbb{R}^{k+1}_\infty$ has property $\mathcal{A}$ locally in Ω. Then, for every sphere S contained in Ω, there is a Möbius g (depending on S) in $\mathcal{M}_{n+1}$ such that $f = g$ on S. In particular, the restriction of f to S is continuous on S.*

The proof of Lemma 5.3. Suppose that S is a sphere contained in Ω, and that $f(S) \subset S'$ (as in Lemma 5.1). Select Möbius maps γ and η in $\mathcal{M}_{k+1}$ such that $\gamma(\mathbb{R}^k_\infty) = S$ and $\eta(S') = \mathbb{R}^k_\infty$. It follows that the map $F = \eta \circ f \circ \gamma$ is an injective map of $\mathbb{R}^k_\infty$ into itself that also has property $\mathcal{A}$. We deduce (from the induction hypothesis) that there is some σ in $\mathcal{M}_k$ such that $F|\mathbb{R}^k_\infty = \sigma$. Now σ extends to a Möbius map of $\mathbb{R}^{k+1}_\infty$ onto itself (we continue to use σ for this map), and we find that $f|S = g|S$, where $g = \eta^{-1} \circ \sigma \circ \gamma^{-1}$ and $g \in \mathcal{M}_{k+1}$. $\square$

The proof of Theorem 5.2. Select a point a in Ω, and an open ball B containing a whose closure $\overline{B}$ lies in Ω such that f has property $\mathcal{A}$ and is injective on $\overline{B}$. Let $S = \partial B$, and let g_S be as in Lemma 5.3. Now define $f_1 = g_S^{-1} \circ f$; this satisfies property $\mathcal{A}$ in Ω, is injective in $\overline{B}$, and is the identity map on S. Moreover, by composing this if necessary with the reflection in S, we may assume that $f_1(a) \in B$. We shall now show that $f_1(a) = a$.

Consider the image $f_1(a)$ in B. First, we recognize that the point a is the intersection of all hyperplanes that are orthogonal to S (and all such hyperplanes meet Ω as $B \subset \Omega$). The only other point in this intersection is, of course, ∞. If H is one of these hyperplanes, then its image $f_1(H)$ lies on a sphere which, by Lemma 5.1, is orthogonal to S. However, as f_1 is the identity on S, we see that $f_1(H \cap S) = H \cap S$, and this means that $f_1(H) \subset H$. We deduce that $f_1(a)$, which

lies in B and also in the intersection of all such $f_1(H)$, must be fixed by f_1; that is, $f_1(a) = a$.

A similar argument shows that for each x in B, $f_1(x)$ is either x, or the reflection x^* of x in S. Now let Σ_x be the unique sphere with diameter $[x, a]$, and note that $\Sigma_x \subset B$. Then $f(\Sigma_x)$ is a connected set (from Lemma 5.3) that lies in the union of Σ_x and its image under reflection in S. As $f_1(a) = a$ and $a \in \Sigma_x$, we deduce that $f_1(\Sigma_x) \subset \Sigma_x$, so that $f_1(x) \in B$. It follows that $f_1(x) = x$ for all x in B, and hence that $f = g_S$ throughout B. This completes the proof. $\square$

Corollary 5.4. *For any $f : \mathbb{R}_\infty^n \to \mathbb{R}_\infty^n$, where $n \geq 2$, the following are equivalent:*
  (1) *f is Möbius;*
  (2) *f assumes at least three values and preserves absolute cross-ratio;*
  (3) *f has property $\mathcal{A}$;*
  (4) *f locally has property $\mathcal{A}$.*

Proof. Trivially, (1) implies (2), and (2) implies (3) because the injectivity of f follows from Lemma 3.4. Next, (3) implies (4) is trivial, while (4) implies (1) follows from Theorem 5.2. $\square$

6. Sphere-preserving maps in $\mathbb{R}_\infty^n$

Circle-preserving maps have a long history that goes back to Möbius himself (see [1], [7], and also [8], p. 109), and sometimes the function is not even required to be measurable. For example, we have the following result (see [4], and an earlier, but slightly weaker, result in [6], p. 309).

Theorem C. *Suppose that $f : D \to \mathbb{C}_\infty$ is injective in a plane domain D, and that f maps each circle in D onto some circle in $\mathbb{C}_\infty$. Then there is some Möbius map g such that f is the restriction to D of one of the maps $g(z)$ or $g(\bar{z})$.*

Of course, the hypothesis in Theorem C is that $[f(a), f(b), f(c), f(d)]$ is real whenever $[a, b, c, d]$ is real. We shall now show a similar result holds in all higher dimensions.

Suppose that Ω is a region in $\mathbb{R}_\infty^n$. A mapping $f : \Omega \to \mathbb{R}_\infty^n$ is called *locally sphere preserving* in Ω if each point a in Ω has a neighborhood N such that for each sphere S in N, $f|S$ is a bijection of S onto a sphere $f(S)$. Note that a locally sphere-preserving map must be locally injective. We say that f is *sphere-preserving* in Ω if for each sphere S in Ω, $f|S$ is a bijection of S onto a sphere $f(S)$. Our main result is as follows.

Theorem 6.1. *Suppose that Ω is a region in $\mathbb{R}_\infty^n$, where $n \geq 2$, and that $f : \Omega \to \mathbb{R}_\infty^n$ is locally sphere-preserving. Then $f \in \mathcal{M}_n$.*

We begin with a preliminary lemma.

Lemma 6.2. *Suppose that $f : \Omega \to \mathbb{R}_\infty^n$ is a sphere-preserving injection, and that S is a sphere in Ω. If a component E of $\mathbb{R}_\infty^n \backslash S$ is contained in Ω, then $f(E)$ is a component of $\mathbb{R}_\infty^n \backslash f(S)$.*

Proof. Take any a in E. There are elements γ_1 and γ_2 in $\mathcal{M}_n$ such that $\gamma_1(\infty) = a$ and $\gamma_2(f(a)) = \infty$. Then $F = \gamma_2 \circ f \circ \gamma_1$ is a sphere-preserving injection that fixes ∞. Also, $\gamma_1^{-1}(S) = T$ and $\gamma_2(f(S)) = F(T)$ are Euclidean spheres in $\mathbb{R}^n$. Note that the exterior of T lies in the domain of F, and we will now show that F maps the exterior of T onto the exterior of $F(T)$. As F is an injection, it preserves the

cardinality of the intersection of two sets. For a in T, let H_a denote the hyperplane tangent to T at a (hence F is defined on H_a). Then $F(H_a)$ is a hyperplane tangent to $F(T)$ at $F(a)$, so F maps $H_a \setminus \{a\}$ into the exterior of $F(T)$. Because every point in the exterior of T is contained in some hyperplane tangent to T, we conclude that F maps the exterior of T into the exterior of $F(T)$. Since F is a bijection of T onto $F(T)$, $\{F(H_a) : a \in T\}$ is the family of all hyperplanes tangent to the sphere $F(T)$. Therefore, the union of these hyperplanes is the union of $F(T)$ and its exterior. Thus, F maps the exterior of T onto the exterior of $F(T)$ and this completes the proof. $\square$

The proof of Theorem 6.1. It suffices to prove that f is locally Möbius. The proof will be by induction, but we first establish some general properties of sphere-preserving maps. The first step is to reduce consideration to a normalized situation, and with this in mind we let $\mathbb{E}_n$ be the exterior of the closed ball with center $(0, \dots, 0, -2)$ and radius 1.

Given $a \in \Omega$, there is an open ball B_a containing a such that the restriction of f to B_a is sphere-preserving and injective. Now, there is some γ_1 in $\mathcal{M}_n$ such that $\gamma_1(\mathbb{E}_n) = B_a$ and $\gamma_1(0) = a$. Then $f \circ \gamma_1$ is a sphere-preserving map of $\mathbb{E}_n$ into $\mathbb{R}^n_\infty$, so that $f \circ \gamma_1(\mathbb{R}^{n-1}_\infty)$ is a sphere in $\mathbb{R}^n_\infty$. Hence, there exists some γ_2 in $\mathcal{M}_n$ such that $F : \mathbb{E}_n \to \mathbb{R}^n_\infty$ defined by $F = \gamma_2 \circ f \circ \gamma_1$ is a sphere-preserving injection that leaves $\mathbb{R}^{n-1}$ invariant. If Σ is a sphere in the half-space $x_n > 0$, then $F(\Sigma)$ is connected (it is a sphere) and it does not meet $\mathbb{R}^n_\infty$ ($= F(\mathbb{R}^n_\infty)$) because of the injectivity of F. It follows that F maps this upper half-space into itself, or into the lower half-space, and by composing F with reflection in $\mathbb{R}^n_\infty$ if necessary, we may also assume that F leaves the half-space $x_n > 0$ invariant. The proof of Theorem 6.1 will be complete if we show F is Möbius in a neighborhood of the origin, and the proof is by induction starting with $n = 2$.

We now establish the theorem when $n = 2$. Of course, the case $n = 2$ is Theorem C, but our proof is shorter than the proof of Theorem C in [4] and in any case, our argument is needed for the general inductive case. We employ complex notation, so that $F : \mathbb{E}_2 \to \mathbb{C}_\infty$ is a circle preserving injection into $\mathbb{C}_\infty$ that leaves the upper half-plane invariant. By composing F with a real Möbius map, we may also assume that F fixes 0, 1 and ∞. Take any point $z_0 = x_0 + iy_0$ in the upper half-plane and construct the circle C with diameter $[x_0, z_0]$, and the horizontal tangent line L to C at z_0. Because of the injectivity, the sphere-preserving property, and the fixed point ∞ of F, we see that $F(L)$ is a horizontal line (tangent to $\mathbb{R}$ at ∞) that is tangent to $F(C)$ at $F(z_0)$. This implies that $F(x_0)$ and $F(z_0)$ have the same real part, and hence that we can write

(6.1) $$F(x_0 + iy_0) = F(x_0) + i\varphi(y_0),$$

for some function φ that maps $(0, +\infty)$ into itself. We should note that in this construction, φ may appear to depend on x_0 but it does not, for we can apply this same argument simultaneously to all circles that have L and $\mathbb{R}$ as horizontal tangents. Note also that exactly the same argument holds for circles tangent to, but below, $\mathbb{R}$ providing only that their diameter is less than 1.

Now consider a circle C of radius less than 1 whose center x_0 is in $\mathbb{R}$. Let the vertical diameter of C be $[z_0, \bar{z}_0]$, where $z_0 = x_0 + iy_0$. Next, construct the two circles C^+ and C^- with diameters $[x_0, z_0]$ and $[x_0, \bar{z}_0]$, respectively, and apply the arguments given above. As the injectivity implies that tangency is preserved

by F, we see immediately that F maps the center of C to the center of $F(C)$. Algebraically, this implies that if $a, b \in [-1, 1]$, then (by considering the circle with diameter $[a, b]$)

$$F\left(\frac{a+b}{2}\right) = \frac{F(a) + F(b)}{2}.$$

As in Section 4, this implies that $F(r) = r$ for all rational numbers r in $[-1, 1]$. For all rational r and s with $-1 \le r < s \le 1$, F maps the circle with diameter $[r, s]$ onto itself since F fixes both the center and the points r and s, and also the interior of this circle onto itself (by Lemma 6.2 and the fact that F fixes ∞). This implies that if $r < x < s$, then $r < F(x) < s$, and consequently that $F(x) = x$ for all $x \in [-1, 1]$.

To complete the proof when $n = 2$ we show that F is the identity function on the square $\{z = x + iy : |x| < 1, |y| < 1\}$. We know (from (6.1) and the previous paragraph) that

$$F(x + iy) = F(x) + i\varphi(y) = x + i\varphi(y),$$

and that $\phi(y)$ has the same sign as y. Consider the circle C with center at the origin and radius $|y|$. Since $F(C) = C$, we conclude $\varphi(y) = y$, and hence that $F(z) = z$ on the square. This completes the proof when $n = 2$.

Finally, we consider the inductive step and this is slightly easier. We suppose that the result is true when $n = k$ and consider the case when $n = k + 1$. We may restrict our attention to the normalized situation described before we began the proof for the case $n = 2$, but now, because of the induction hypothesis, we know that the restriction of F to $\mathbb{R}_\infty^k$ is some σ in $\mathcal{M}_k$. As σ extends to an element of $\mathcal{M}^{k+1}$ that preserves the half-space $x_{k+1} > 0$, we may combine F and σ^{-1} and so assume that the restriction of F to $\mathbb{R}_\infty^k$ is the identity map. The same arguments as those given above readily show that if $x_0 \in \mathbb{R}^k$ and e_{k+1} is the usual basis vector $(0, \ldots, 0, 1)$ (with $k + 1$ entries), then $F(x_0 + y_0 e_{k+1}) = x_0 + \psi(y_0)e_{k+1}$, for some suitable function ψ. The proof now continues exactly as above to show that F is now the identity in some neighborhood of the origin. This completes the proof. $\square$

Corollary 6.3. *Suppose that $n \ge 2$. For any $f : \mathbb{R}_\infty^n \to \mathbb{R}_\infty^n$, the following are equivalent.*

(1)   *f is Möbius;*
(2)   *the restriction of f to every sphere is Möbius;*
(3)   *f is sphere-preserving and bijective;*
(4)   *f is sphere-preserving and injective;*
(5)   *f is locally sphere-preserving.*

Proof. The implications (1) implies (2), (1) implies (3), (2) implies (4), (3) implies (4), and (4) implies (5) are all trivial. Finally, the implication (5) implies (1) follows from Theorem 6.1. $\square$

REFERENCES

[1]   Aczél, J. and McKiernan, M.A., On the characterization of plane projective and complex Moebius-transformations, *Math. Nach.* 33 (1967), 315-337 MR **36:**5806
[2]   Beardon, A.F., *The geometry of discrete groups*, Springer-Verlag, GTM 91, 1983. MR **85d:**22026
[3]   Busemann, H. and Kelly, P.L., *Projective geometry and projective metrics*, Academic Press, New York, 1953. MR **14:**1008e

[4] Carathéodory, C., The most general transformations of plane regions which transform circles into circles, *Bull. Amer. Math. Soc.* 43 (1937), 573-579.

[5] Chubarev, A. and Pinelis, I., Fundamental Theorem of geometry without the 1-to-1 assumption, *Proc. Amer. Math. Soc.* 127 (1999), 2735-2744. MR **99m:**51002

[6] Coolidge, J.L., *A treatise on the circle and the sphere*, Chelsea, 1971 (reprinted from Oxford, 1916). MR **52:**10346

[7] Coxeter, H.S.M., Similarities and conformal transformations, *Annali di Matematica pura ed applicata* 53 (1961), 165-172. MR **26:**648

[8] Coxeter, H.S.M., *Introduction to Geometry*, Wiley, 1969. MR **49:**11369

[9] Haruki, H. and Rassias, T.M., A new characteristic of Möbius transformations by use of Apollonius quadilaterals, *Proc. Amer. Math. Soc.* 126 (1998), 2857-2861. MR **99a:**30012

[10] Hungerford, T.W., *Algebra*, Holt, Rinehart and Winston, New York, 1974. MR **50:**6693

[11] Jeffers, J., Lost theorems of geometry, *American Math. Monthly* 107 (2000), 800-812. CMP 2001:03

[12] McKemie, M.J. and Väisälä, J., Spherical maps of Euclidean spaces, *Result. Math.* 35 (1999), 145-160. MR **2000a:**30044

[13] Radford, J.G., *Foundations of hyperbolic manifolds*, Springer-Verlag, GTM 149, 1994.

DEPARTMENT OF PURE MATHEMATICS AND MATHEMATICAL STATISTICS, CENTRE FOR MATHE-MATICAL SCIENCES, WILBERFORCE ROAD, CAMBRIDGE CB3 0WB, ENGLAND
  E-mail address: `A.F.Beardon@dpmms.cam.ac.uk`

DEPARTMENT OF MATHEMATICAL SCIENCES, UNIVERSITY OF CINCINNATI, CINCINNATI, OHIO 45221-0025
  E-mail address: `David.Minda@math.uc.edu`