CHAPTER 3

Conformal Mappings in Euclidean Space

Rigidity will prevent a wide and flexible class of conformal transformations in dimensions $n \geq 3$, and as we have said, this is a prime motivation for the general theory of quasiconformal mappings.

However the theory of conformal geometry in higher dimensions is interesting and important and connects with many other fields of mathematics. It is the purpose of this chapter to collect together many of these ideas and connections, not only for their interest but also to motivate and describe potential tools and results for quasiconformal mappings. These things include higher-dimensional cross ratios, the extension theorem of Poincaré, hyperbolic geometry, and the compactness and normal families properties of conformal transformations.

The contents of a significant part of the latter portions of this chapter are a synthesis of material found in Lars Ahlfors' Ordway Lectures [**9**] or in Alan Beardon's book *The Geometry of Discrete Groups*, [**16**]. Each of these we warmly recommend to the reader if they wish to discover more of the subject for themselves.

3.1. Linear conformal transformations

Let $n \geq 2$. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is *conformal* if T is nonsingular and preserves Euclidean angles, in the sense that

$$\theta[T(x), T(y)] = \theta(x, y)$$

for all nonzero vectors x and y in $\mathbb{R}^n$. It follows from statement (ii) in Theorem 3.1.1 that conformal linear transformations are actually "shape preserving", which is more faithful than "angle- preserving" to the literal meaning of "conformal". The foregoing definition permits a conformal linear transformation T to be either sense-preserving, the situation when $\det(T) > 0$, or sense-reversing, which occurs when $\det(T) < 0$.

The experienced reader will realize that this is at variance with the standard usage of "conformal" in classical complex analysis. In that setting, the term is traditionally reserved for transformations, linear or otherwise, that preserve angles not merely in size but also in orientation, while the expression "anti-conformal" is used when they are sense-reversing.

We note that the conformal linear transformations of $\mathbb{R}^n$ constitute a subgroup of $\mathrm{GL}(n)$, as do the sense-preserving conformal linear transformations. If a conformal linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ leaves $\mathbb{R}^m$ $(2 \leq m < n)$ invariant, then the restriction of T to $\mathbb{R}^m$ is a conformal linear transformation of the lower-dimensional space.

It will prove convenient to have several ways of characterizing conformal linear transformations, and we now set about trying to determine these criteria.

THEOREM 3.1.1. *Let $n \geq 2$. The following statements concerning a linear transformation $T \in \mathrm{GL}(n)$ are equivalent:*

(i) *T is conformal,*
(ii) *$T = \lambda U$, where λ is a positive number and U belongs to $\mathrm{O}(n)$,*
(iii) *$H(T) = 1$,*
(iv) *$\|T\|^n = |\det(T)|$,*
(v) *$T^*T = |\det(T)|^{2/n} I$,*

where I denotes the identity matrix.

PROOF. The implications (ii) $\Rightarrow$ (i), (ii) $\Rightarrow$ (iii), (ii) $\Rightarrow$ (iv), and (ii) $\Rightarrow$ (v) are clear. To demonstrate that (i) $\Rightarrow$ (ii), let $v_i = T(e_i)$ for $i = 1, 2, \ldots, n$ and set $\lambda = |v_1| > 0$. Because T is conformal, the vectors $v_1, v_2, \ldots, v_n$ are pairwise orthogonal. For $i \geq 2$, let Δ_i be the isosceles triangle in $\mathbb{R}^n$ with vertices 0, e_1, and e_i. Then T transforms Δ_i to a triangle Δ_i' with vertices $0, v_1$, and v_i. Conformality dictates that the angle of Δ_i' at any of its vertices be equal to the angle of Δ_i at the corresponding vertex. As a consequence, Δ_i' is also isosceles. It follows that $|v_i| = \lambda$ for $1 \leq i \leq n$. Thus $U = \lambda^{-1}T$ transforms $e_1, e_2, \ldots, e_n$ to another orthonormal basis for $\mathbb{R}^n$, which forces U into $\mathrm{O}(n)$ and gives T the stated form.

We next show that (iii) $\Rightarrow$ (ii). Writing $\lambda = L(T) = \ell(T) > 0$, we conclude that $|T(x)| = \lambda$ for every x in $\mathbb{S}^{n-1}$, so $|T(x)| = \lambda|x|$ for every x in $\mathbb{R}^n$. Accordingly, $U = \lambda^{-1}T$ has $|U(x)| = |x|$ for all x, i.e., U belongs to $\mathrm{O}(n)$. Again T has the required structure.

Assuming (iv), we let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n > 0$ denote the eigenvalues of TT^*. The observation used to verify 2.8 leads to

$$\lambda_1^n = L(T)^{2n} = \|T\|^{2n} = |\det(T)|^2 = |\det(TT^*)| = \lambda_1\lambda_2\cdots\lambda_n\ ,$$

which implies that $\lambda_1 = \lambda_2 = \cdots = \lambda_n$. In particular, $L(T) = \sqrt{\lambda_1} = \sqrt{\lambda_n} = \ell(T)$. Therefore $H(T) = 1$, so (iv) $\Rightarrow$ (iii).

Finally, given that (v) holds, we compute for any unit vector x in $\mathbb{R}^n$:

$$|T(x)|^2 = \langle T(x), T(x)\rangle = \langle T^*T(x), x\rangle = |\det(T)|^{2/n}\langle x, x\rangle = |\det(T)|^{2/n}\,.$$

From this we easily conclude that $\|T\|^n = |\det(T)|$. In other words, (v) $\Rightarrow$ (iv). This completes the proof of the theorem. $\square$

A straightforward consequence of Theorem 3.1.1 is that

$$\|ST\| = \|S\| \cdot \|T\| \tag{3.1}$$

whenever S and T are conformal linear transformations of $\mathbb{R}^n$.

Let D be a domain in $\mathbb{R}^n$ with $n \geq 2$—by a *domain* in a topological space X we understand a nonempty, open, connected subset of X—and let $f : D \to \mathbb{R}^n$ be a continuous injection (hence, a homeomorphism of D onto the domain $D' = f(D)$). We say that f is *conformal at a point* x of D under the condition that f is differentiable at x and its derivative $f'(x)$ is a conformal linear transformation.

Couched in geometric terms, this definition requires that f be infinitesimally angle-preserving (or, more generally, shape-preserving) at x. It demands, for instance, that f transform any pair of curves (or hypersurfaces) in D that intersect at x in an angle θ to curves (or hypersurfaces) whose angle of intersection at the point $f(x)$ is likewise θ.

We say that f is a *conformal mapping* of D onto D' provided that f is conformal at each point of D.

As a homeomorphism between domains in $\mathbb{R}^n$, such a mapping f must be either sense-preserving or sense-reversing. This fact finds its analytic expression in the statement that either the Jacobian determinant $J_f(x) > 0$ for all x in D, or $J_f(x) < 0$ for all x in D.

The preceding would be an elementary observation were f a diffeomorphism, but is not quite so obvious in the situation at hand, where no continuity conditions are imposed upon f'.

If f maps D conformally onto D' and g maps D' conformally onto D'', then $g \circ f$ is a conformal mapping of D onto D''. This follows from the chain rule and the fact that the conformal linear transformations of $\mathbb{R}^n$ form a group under composition. If f maps D conformally onto D', then Theorem 2.8.2 certifies that f^{-1} maps D' conformally onto D. If D is a domain in $\mathbb{R}^n$ with the property that $G = D \cap \mathbb{R}^m$ is a domain in $\mathbb{R}^m$ ($2 \leq m < n$) and if f, a conformal mapping of D onto some domain in $\mathbb{R}^n$, has the feature that $f(G)$ is a subset of $\mathbb{R}^m$, then $g = f|G$ is a conformal mapping of G onto $f(G)$. In fact, we deduce from (2.17) that for each x in G the linear transformation $f'(x)$ has $\mathbb{R}^m$ as an invariant subspace and infer that $g'(x) = f'(x)|\mathbb{R}^m$ is a conformal linear transformation of $\mathbb{R}^m$.

Consider a sense-preserving conformal mapping $f : D \to D'$ between domains D and D' in $\mathbb{R}^2$, which we now think of as the complex plane. A simple calculation confirms that, when viewed as a complex-valued function of a complex variable, any differentiable function $g : D \to \mathbb{C}$ has

$$J_g(z) = |\partial g(z)|^2 - |\,\overline{\partial} g(z)|^2\,, \tag{3.2}$$

which gives us a particularly nice formula for the Jacobian determinant.

In the present case we have $J_f > 0$ in D, so the conformality of f in combination with Theorem 3.1.1(iv), (2.22), (2.23), and (3.2) provides the information that

$$\big(|\partial f(z)| + |\,\overline{\partial} f(z)|\big)^2 = L_f(z)^2 = J_f(z) = |\partial f(z)|^2 - |\,\overline{\partial} f(z)|^2 \tag{3.3}$$

for every z in D. Thanks to the positivity of J_f, we can be certain that $|\partial f(z)| + |\,\overline{\partial} f(z)| > 0$ throughout D. Therefore, (3.3) reduces to

$$|\partial f(z)| + |\,\overline{\partial} f(z)| = |\partial f(z)| - |\,\overline{\partial} f(z)|$$

for each z in D.

The last equation demands that $\overline{\partial} f$ vanish identically in D, hence it forces f to be analytic there. The consequence of this discussion is that if $f : D \to \mathbb{C}$ is a sense-preserving conformal mapping of a plane domain D onto $D' = f(D)$, then f is an injective analytic function in D. The converse of that statement is also true—and not hard to prove. Similar reasoning reveals that the sense-reversing conformal mappings in the plane are the functions of the type $\bar{f}$, where $f : D \to \mathbb{C}$ is an injective analytic function.

The family of conformal mappings from $\mathbb{R}^n$ into itself plainly includes all mappings of the type $f = \lambda U + b$, in which λ is a positive number, U belongs to $\mathrm{O}(n)$, and b is an element of $\mathbb{R}^n$. The class of such mappings, which we refer to as *similarity transformations*, is a subgroup of the affine group $\mathrm{A}(n)$. We call it the (*general*) *similarity group* of $\mathbb{R}^n$ and denote it by $\mathrm{GS}(n)$. Again, we point out the divergence of this terminology from more common usage, which would exclude a sense-reversing mapping from inclusion under the "similarity transformation" heading. The sense-preserving similarities form a subgroup $\mathrm{GS}^+(n)$ of $\mathrm{GS}(n)$. Certain members of $\mathrm{GS}^+(n)$ bear distinguishing titles: a mapping of the type $f(x) = x + b$, with $b \neq 0$, is called a *translation*; a *dilation* (or *homothety*) is a transformation of the form $f(x) = \lambda x$ with $\lambda > 0$ but $\lambda \neq 1$.

We make the following relevant remark. Let f be an affine transformation of $\mathbb{R}^n$, say $f = T + b$ with $T \in \mathrm{GL}(n)$ and $b \in \mathbb{R}^n$. Since

$$|f(x)| \geq |T(x)| - |b| \geq \ell(T)|x| - |b| \to \infty$$

as $|x| \to \infty$, we see that by defining $f(\infty) = \infty$ we may extend f to a homeomorphism of $\hat{\mathbb{R}}^n$ onto itself. By extending each of its members in this way, we are free to regard the group $\mathrm{A}(n)$ as a group of homeomorphisms of $\hat{\mathbb{R}}^n$, which we shall tacitly do from now on. We note especially that, under this convention, any similarity transformation of $\mathbb{R}^n$ is a homeomorphism of $\hat{\mathbb{R}}^n$.

3.2. Reflections

Let P be a hyperplane in $\mathbb{R}^n$; that is, P is an $(n-1)$-dimensional affine subspace of $\mathbb{R}^n$. We can describe P by means of a linear equation, say

$$P : \langle \nu, x \rangle = d,$$

where ν is a unit vector in $\mathbb{R}^n$ and $d \geq 0$. As a matter of fact, subject to the stated normalizations, ν and d are uniquely determined by P except when $d = 0$, in which case there are two choices for ν, one the negative of the other.

The function $R : \hat{\mathbb{R}}^n \to \hat{\mathbb{R}}^n$ defined by

$$R(x) = x - 2\big(\langle \nu, x \rangle - d\big)\nu \tag{3.4}$$

for x in $\mathbb{R}^n$ and $R(\infty) = \infty$ is called the *reflection in P*. This term is geometrically motivated: R fixes every point of P and moves any point x of $\mathbb{R}^n \setminus P$ to its mirror image relative to P, which means to the point $R(x)$ such that P is the perpendicular bisector of the line segment with endpoints x and $R(x)$. In particular, R interchanges the two half-spaces into which P partitions $\mathbb{R}^n$.

We deduce from (3.4) by an elementary calculation that R is a Euclidean isometry. This makes R an affine transformation, one that is conformal when $n \geq 2$. Furthermore, if $n \geq 2$ and if the vectors $v_1, v_2, \ldots, v_{n-1}$ form a basis for the linear subspace P_0 of $\mathbb{R}^n$ defined by $P_0 : \langle \nu, x \rangle = 0$, then we see from (2.17) that the matrix of $T = R'(x)$ (in this instance $R'(x)$ is the same for all points x of $\mathbb{R}^n$) with respect to the basis $\nu, v_1, v_2, \ldots, v_{n-1}$ for $\mathbb{R}^n$ is

$$\begin{bmatrix} -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

It follows that $\det(T) = -1$, hence that R is sense-reversing. As a member of $\mathrm{A}(n)$, R is a homeomorphism of $\hat{\mathbb{R}}^n$ onto itself. On top of this, R is an involution, meaning that $R^{-1} = R$, or $R \circ R = I$, the identity transformation.

Next, let $S = \mathbb{S}^{n-1}(x_0, r)$ be a Euclidean sphere in $\mathbb{R}^n$. The *reflection in S* (also known as the *inversion in S*) is the function $R : \hat{\mathbb{R}}^n \to \hat{\mathbb{R}}^n$ defined by the rule

$$R(x) = x_0 + \frac{r^2(x - x_0)}{|x - x_0|^2} \tag{3.5}$$

for x in $\mathbb{R}^n \setminus \{x_0\}$, while $R(x_0) = \infty$ and $R(\infty) = x_0$.

Formula (3.5) again has a simple geometric interpretation: R transforms a point x of $\mathbb{R}^n \setminus \{x_0\}$ to the unique point $R(x)$ that lies on the Euclidean ray issuing from x_0 and passing through x and that satisfies the condition

$$|R(x) - x_0| \cdot |x - x_0| = r^2.$$

Just like the reflection in a hyperplane, the reflection in the sphere $S = \mathbb{S}(x_0, r)$ is an involutive homeomorphism of $\hat{\mathbb{R}}^n$. It leaves fixed every point of S and interchanges the two components of S^c, $B^n(x_0, r)$, and $\hat{\mathbb{R}}^n \setminus \bar{B}^n(x_0, r)$. This function is also a C^∞-diffeomorphism of $\mathbb{R}^n \setminus \{x_0\}$ onto itself, as (3.5) makes readily apparent.

We shall examine R_0, the reflection in the unit sphere $\mathbb{S}^{n-1}$, more closely. In this situation we have $R_0(x) = x/|x|^2$ for x different from 0 and ∞. As a function of class C^∞ in $\mathbb{R}^n \setminus \{0\}$, R_0 is differentiable at every point x of this set, and its derivative is not difficult to compute: the derivative matrix of R_0 at x is

$$R_0'(x) = |x|^{-2}[I - 2Q(x)], \tag{3.6}$$

where $Q(x)$ is the matrix whose $(i, j)^{th}$-entry is

$$(R_0'(x))_{i,j} = x_i x_j / |x|^2.$$

Now $Q = Q(x)$ is the matrix of a symmetric linear transformation, also labeled Q. It is an easy exercise to see that Q satisfies

$$Q^2 = Q.$$

As a consequence, we compute

$$(I - 2Q)(I - 2Q)^* = (I - 2Q)^2 = I - 4Q + 4Q^2 = I - 4Q + 4Q = I$$

and therefore discover that $|x|^2 R_0'(x) = I - 2Q(x)$ belongs to the orthogonal group $\mathrm{O}(n)$. This has the immediate consequence that

$$\det(R_0'(x)) = -|x|^{-2n}. \tag{3.7}$$

Assume now that $n \geq 2$. On the basis of Theorem 3.1.1(ii) we can assert that, for each x in $\mathbb{R}^n \setminus \{0\}$, $R_0'(x)$ is a conformal linear transformation of $\mathbb{R}^n$. In other words, R_0 is a conformal self-mapping of $\mathbb{R}^n \setminus \{0\}$. Moreover,

$$\|R_0'(x)\| = \frac{1}{|x|^2}\,. \tag{3.8}$$

Since the matrix corresponding to $R_0'(e_1)$, which once again is

$$\begin{bmatrix} -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix},$$

has determinant -1, we conclude that R_0 is sense-reversing.

The reflection R in an arbitrary sphere $\mathbb{S}^{n-1}(x_0, r)$ is expressible as the composition $R = f^{-1} \circ g \circ R_0 \circ f$, where $f(x) = x - x_0$ and $g(x) = r^2 x$. Because $f'(x) = I$ and $g'(x) = r^2 I$ for all x in $\mathbb{R}^n$, the chain rule yields

$$R'(x) = r^2 R_0'(x - x_0) \tag{3.9}$$

for every x in $\mathbb{R}^n \setminus \{x_0\}$. In case $n \geq 2$, we conclude that R is a sense-reversing conformal mapping of $\mathbb{R}^n \setminus \{x_0\}$ onto itself and, in view of (3.8), that

$$\|R'(x)\| = \frac{r^2}{|x - x_0|^2} \tag{3.10}$$

for x other than x_0 or ∞. We can take the determinant of both sides of (3.9) to obtain the following lemma.

LEMMA 3.2.1. *Let R be the reflection in an arbitrary sphere $\mathbb{S}^{n-1}(x_0, r)$. Then the Jacobian determinant of the mapping R is*

$$J_R(x) = -\frac{r^{2n}}{|x - x_0|^{2n}}. \tag{3.11}$$

One special case covered by the preceding discussion is worth highlighting. Suppose that $R : \hat{\mathbb{R}}^{n+1} \to \hat{\mathbb{R}}^{n+1}$ is the inversion in the sphere $\mathbb{S}^n(e_{n+1}, \sqrt{2}\,)$. For x in $\mathbb{R}^n$ we have $|x - e_{n+1}|^2 = |x|^2 + 1$, so for such x we obtain

$$R(x) = e_{n+1} + \frac{2(x - e_{n+1})}{|x|^2 + 1} = \left(\frac{2x_1}{|x|^2 + 1}, \ldots, \frac{2x_n}{|x|^2 + 1}, \frac{|x|^2 - 1}{|x|^2 + 1} \right) = \pi(x) \ ,$$

where $\pi : \hat{\mathbb{R}}^n \to \mathbb{S}^n$ is the stereographic projection. Therefore, π is the restriction to $\hat{\mathbb{R}}^n$ of a reflection operating in $\hat{\mathbb{R}}^{n+1}$!

This explains the well-known conformality of π as a mapping from $\mathbb{R}^n$ to $\mathbb{S}^n$, the latter endowed with the Riemannian structure it acquires as a smooth submanifold of $\mathbb{R}^{n+1}$.

Several consequences of this observation will be of value later on. One is that the function $f = R \circ S\big|\mathbb{H}^{n+1}$, in which

$$\mathbb{H}^{n+1} = \{x \in \mathbb{R}^{n+1} : x_{n+1} > 0\}$$

is the open "upper" half-space in $\mathbb{R}^{n+1}$ and S is reflection in $\mathbb{R}^n$ regarded as a hyperplane in $\mathbb{R}^{n+1}$, provides a sense-preserving conformal mapping of $\mathbb{H}^{n+1}$ onto B^{n+1}.

This is simply because R maps the domain $D = \{x \in \mathbb{R}^{n+1} : x_{n+1} < 0\}$ in a sense-reversing, conformal fashion into $\mathbb{R}^{n+1}$; being a homeomorphism of $\hat{\mathbb{R}}^{n+1}$ with $R(\partial D) = \mathbb{S}^n$, R maps D to one of the two components of $\hat{\mathbb{R}}^{n+1} \setminus \mathbb{S}^n$; as $R(-e_{n+1}) = 0$, said component has to be B^{n+1}; the restriction of S to $\mathbb{H}^{n+1}$ is a sense-reversing conformal mapping of $\mathbb{H}^{n+1}$ onto D; thus, f has all the stated properties.

Another is that $\pi : \mathbb{R}^n \to \mathbb{R}^{n+1}$ has

$$\|\pi'(x)\| = \frac{2}{1 + |x|^2} \ , \tag{3.12}$$

which is true because $\pi'(x) = R'(x)|\mathbb{R}^n$ once we recall (3.10).

In order to consolidate the notion of reflection in a sphere with that of reflection in a hyperplane, we introduce some convenient terminology. We call a subset Σ of

$\hat{\mathbb{R}}^n$ a *chordal sphere* if Σ is either a Euclidean sphere in $\mathbb{R}^n$ or a set of the type $\Sigma = P \cup \{\infty\}$, where P is a hyperplane in $\mathbb{R}^n$.

We will soon see that all the chordal spheres in $\hat{\mathbb{R}}^n$ are precisely the spheres of radius $\sqrt{2}$ or less for the chordal metric on $\hat{\mathbb{R}}^n$.

Every chordal sphere Σ in $\hat{\mathbb{R}}^n$ admits descriptions of the type

$$\Sigma : a|x|^2 - 2\langle b, x\rangle + c = 0$$

with $a, c \in \mathbb{R}$, $b \in \mathbb{R}^n$, and $ac < |b|^2$, provided we establish the convention that ∞ is to be counted as a solution of such an equation if and only if $a = 0$. With the notation interpreted in the obvious way, the vector (a, b, c) in $\mathbb{R}^{n+2}$ is known as a *coefficient vector* for Σ. Any two coefficient vectors for Σ are nonzero scalar multiples of each another. Conversely, if the stated convention is respected, the solution locus in $\hat{\mathbb{R}}^n$ of an equation fitting the above description is a chordal sphere. With each chordal sphere Σ in $\hat{\mathbb{R}}^n$ is associated the reflection in Σ, a mapping we denote by R_Σ.

3.3. The Möbius group

The (*general*) *Möbius group* $\text{Möb}(n)$ is the group of homeomorphisms of $\hat{\mathbb{R}}^n$ generated by the reflections R_Σ, where Σ ranges over all chordal spheres in $\hat{\mathbb{R}}^n$. Its members are known as the *Möbius transformations* of $\hat{\mathbb{R}}^n$. The *special Möbius group* $\text{Möb}^+(n)$ is the subgroup of $\text{Möb}(n)$ consisting of those Möbius transformations that can be obtained by composing an even number of reflections.

Since any reflection R is a sense-reversing homeomorphism of $\hat{\mathbb{R}}^n$, the mappings in $\text{Möb}^+(n)$ are in fact the sense-preserving members of $\text{Möb}(n)$. From the fact that each reflection R is a diffeomorphism of the set $\hat{\mathbb{R}}^n \setminus \{\infty, R^{-1}(\infty)\}$ onto itself, we make the following inference: for any Möbius transformation f there is a finite subset E of $\hat{\mathbb{R}}^n$ such that the restriction of f to the set $D = \hat{\mathbb{R}}^n \setminus E$ is a diffeomorphism—a conformal diffeomorphism if $n \geq 2$—of D onto $f(D)$.

We shall discover later that we can always take $E = \{\infty, f^{-1}(\infty)\}$, but this is not yet entirely obvious from our definition of a Möbius transformation.

If $\pi : \hat{\mathbb{R}}^n \to \mathbb{S}^n$ is a stereographic projection, then

$$\text{Con}(n) = \pi \circ \text{Möb}(n) \circ \pi^{-1}$$

is an important and much studied group of homeomorphisms of $\mathbb{S}^n$, the *conformal group* of $\mathbb{S}^n$.

The Möbius group—indeed, the special Möbius group—contains all dilations and translations of $\mathbb{R}^n$: if $f(x) = \lambda x$ with $\lambda > 0$, we have $f = R \circ R_0$, where R_0 is the inversion in $\mathbb{S}^{n-1}$ and R is the inversion in $\mathbb{S}^{n-1}(\sqrt{\lambda})$; if $f(x) = x + b$ with $b \neq 0$, we can express f in the form $f = R_2 \circ R_1$, where R_1 is the reflection in the hyperplane $P_1 : \langle b, x\rangle = 0$ and R_2 is the reflection in $P_2 : \langle b, x\rangle = |b|^2/2$. The former assertion follows directly from (3.5), while the latter is established by a straightforward calculation, once it is observed that

$$R_1(x) = x - 2\langle \nu, x\rangle\nu, \qquad R_2(x) = x - 2\left(\langle \nu, x\rangle - \frac{|b|}{2}\right)\nu$$

with $\nu = b/|b|$. Less apparent is the fact that every orthogonal linear transformation of $\mathbb{R}^n$ belongs to $\text{Möb}(n)$.

LEMMA 3.3.1. *If U is a member of the orthogonal group $\mathrm{O}(n)$, then U can be represented as a composition of n or fewer reflections in hyperplanes that pass through the origin of $\mathbb{R}^n$.*

PROOF. If U is the identity, then $U = R \circ R$ for any such reflection. We can therefore assume $U \neq I$. We construct in a step-by-step fashion transformations $V_1, V_2, \ldots, V_n$ in $\mathrm{O}(n)$, each of them either the identity transformation I or a reflection in a hyperplane through the origin, so that for $m = 1, 2, \ldots, n$ the orthogonal linear transformation $U_m = V_m V_{m-1} \cdots V_1 U$ leaves fixed all of the vectors $e_1, e_2, \ldots, e_m$. If we can accomplish this, then U_n will fix $e_1, e_2, \ldots, e_n$, compelling the conclusion that $U_n = I$—hence, that $U = V_1 V_2 \cdots V_n$ (remember: $V_j^{-1} = V_j$) is a composition of n or fewer reflections in hyperplanes that contain the origin. To start the construction, write $b_1 = U(e_1) - e_1$. If $b_1 = 0$, take $V_1 = I$; if $b_1 \neq 0$, let V_1 be the reflection in the hyperplane $P_1 : \langle b_1, x\rangle = 0$. In the latter instance, we note that the vector $U(e_1) + e_1$ lies in P_1 for

$$\big\langle b_1, U(e_1) + e_1 \big\rangle = \big\langle U(e_1) - e_1, U(e_1) + e_1 \big\rangle = |U(e_1)|^2 - |e_1|^2 = 1 - 1 = 0\,.$$

In either case, the transformation $U_1 = V_1 U$ fixes e_1. This is trivial if $b_1 = 0$, since then $U_1(e_1) = U(e_1) = e_1$; if $b_1 \neq 0$, we compute

$$\begin{aligned} U_1(e_1) &= V_1\big(U(e_1)\big) = V_1\left(\frac{U(e_1) + e_1 + b_1}{2}\right) \\ &= \frac{1}{2}V_1\big(U(e_1) + e_1\big) + \frac{1}{2}V_1(b_1) = \frac{1}{2}\big(U(e_1) + e_1\big) - \frac{1}{2}b_1 = e_1\,. \end{aligned}$$

Assuming that $m < n$ and that we have been successful in producing transformations $V_1, V_2, \ldots, V_m$ of the type indicated such that U_m fixes e_j when $1 \leq j \leq m$, we proceed to construct V_{m+1}. Let $b_{m+1} = U_m(e_{m+1}) - e_{m+1}$. Mimicking what we did to begin the construction, we set $V_{m+1} = I$ if $b_{m+1} = 0$ and let V_{m+1} be the reflection in $P_{m+1} : \langle b_{m+1}, x\rangle = 0$ if $b_{m+1} \neq 0$. Exactly as before, we see that $U_{m+1} = V_{m+1} U_m$ fixes the vector e_{m+1}. It also fixes e_j for $j = 1, 2, \ldots, m$. Again this is clear when $b_{m+1} = 0$ (then $U_{m+1} = U_m$), while for $b_{m+1} \neq 0$ it is a consequence of the fact that U_m fixes e_j and that

$$\begin{aligned} \langle b_{m+1}, e_j\rangle &= \big\langle U_m(e_{m+1}) - e_{m+1}, e_j \big\rangle = \big\langle U_m(e_{m+1}), e_j \big\rangle \\ &= \big\langle U_m(e_{m+1}), U_m(e_j) \big\rangle = \langle e_{m+1}, e_j\rangle = 0\ , \end{aligned}$$

which places e_j in the fixed point locus of V_{m+1}. The procedure can thus be continued until the desired end result is achieved. □

We have shown that $\text{Möb}(n)$ includes $\mathrm{GS}(n)$, the group of similarity transformations of $\mathbb{R}^n$. In fact, it is not difficult to see that $\text{Möb}(n)$ is generated by the dilations, translations, and orthogonal linear transformations of $\mathbb{R}^n$, together with the inversion in $\mathbb{S}^{n-1}$. We also note that both $\text{Möb}(n)$ and $\text{Möb}^+(n)$ act transitively on $\hat{\mathbb{R}}^n$. This means that given $x, y \in \hat{\mathbb{R}}^n$ there is $f \in \text{Möb}^+(n)$ such that $f(x) = y$.

It is a simple exercise to show that $\text{Möb}(1)$ consists of all functions $f : \hat{\mathbb{R}} \to \hat{\mathbb{R}}$ that admit representations of the type $f(x) = (ax + b)/(cx + d)$, where $a, b, c,$ and d are real numbers with $ad - bc \neq 0$. Those f for which $ad - bc > 0$ constitute the group $\text{Möb}^+(1)$. In the classical case $n = 2$, members of the Möbius group also assume familiar forms.

THEOREM 3.3.2. *The group* Möb(2) *consists of the functions* $f : \hat{\mathbb{R}}^2 \to \hat{\mathbb{R}}^2$ *that can be represented with the aid of complex notation in either the form*

$$f(z) = \frac{az + b}{cz + d} \tag{3.13}$$

or the form

$$f(z) = \frac{a\bar{z} + b}{c\bar{z} + d} \ , \tag{3.14}$$

where $a, b, c,$ *and* d *are complex numbers and* $ad - bc \neq 0$*. The group* $\text{Möb}^+(2)$ *is made up of functions of the first type.*

PROOF. In order for the above formulas to define functions from $\hat{\mathbb{R}}^2 = \mathbb{C}$ to itself, they must be properly interpreted at certain troublesome points: if $c = 0$, $f(\infty) = \infty$; if $c \neq 0$, $f(\infty) = a/c$, $f(-d/c) = \infty$ in (3.13), and $f(-\bar{d}/\bar{c}) = \infty$ in (3.14). Modulo these conventions, a function given by (3.13) or (3.14) is a homeomorphism of $\mathbb{C}$ onto itself.

Under composition the collection of such functions forms a group, call it G. The functions of character (3.13) constitute a subgroup H of G.

Any line L in $\mathbb{C}$ admits complex equations of the kind $L : Az + \bar{A}\bar{z} + B = 0$, where $A \neq 0$ and B is real. The reflection R in L is then expressible as

$$R(z) = -(\bar{A}/A)\bar{z} - (B/A) \ ,$$

which has the form (3.14). The inversion R in $S^1(z_0, r)$ can be written as

$$R(z) = z_0 + \frac{r^2}{\bar{z} - \bar{z}_0} = \frac{z_0\bar{z} + r^2 - |z_0|^2}{\bar{z} - \bar{z}_0} \ ,$$

which is once more of type (3.14). Since the generators of Möb(2) lie in G, Möb(2) is definitely a subgroup of G. Furthermore, the composition of an even number of reflections clearly takes the form (3.13), making $\text{Möb}^+(2)$ a subgroup of H.

In complex notation the members of $\text{O}(2) \setminus \{identity\}$ have the appearance $f(z) = e^{i\theta}z$ or $f(z) = e^{i\theta}\bar{z}$, with θ real and not an integral multiple of 2π.

The function $f(z) = e^{i\theta}z$ is not itself a reflection, so in view of Lemma 3.3.1 it must be a composition of two reflections; that is, f belongs to $\text{Möb}^+(2)$.

It follows that $\text{Möb}^+(2)$ includes all functions having the structure $f(z) = az+b$ with $a \neq 0$. Indeed, we have $f = k \circ h \circ g$, where $g(z) = e^{i\theta}z$ for $\theta = \text{Arg}\,a$, $h(z) = |a|z$ is a dilation, and $k(z) = z + b$ is a translation. The group $\text{Möb}^+(2)$ also contains the function $f(z) = z^{-1}$ for $f = R_1 \circ R_0$, in which R_0 is reflection in S^1 and R_1 is reflection in the real axis. When $c \neq 0$, formula (3.13) can be rewritten

$$f(z) = \frac{a}{c} + \frac{bc - ad}{c^2} \ \frac{1}{z + (d/c)} \ ,$$

which implies that any such f belongs to $\text{Möb}^+(2)$. We have thus shown that H is a subgroup of $\text{Möb}^+(2)$. As a result, $\text{Möb}^+(2) = H$. Finally, since G is generated by H and the reflection R_1, itself a member of Möb(2), we conclude that $\text{Möb}(2) = G$. □

It is customary to normalize representations (3.13) and (3.14) by imposing the requirement that the "determinant" $ad - bc = 1$.

Such normalization is possible because the functions defined via these formulas are not affected when a, b, c, and d are replaced by $\lambda a, \lambda b, \lambda c$, and λd, where λ is any nonzero complex number.

One of the most important geometric features of Möbius transformations is the subject of the next theorem.

THEOREM 3.3.3. *If f is a Möbius transformation of $\hat{\mathbb{R}}^n$ and Σ is a chordal sphere in $\hat{\mathbb{R}}^n$, then $f(\Sigma)$ is also a chordal sphere.*

PROOF. If functions f and g map all chordal spheres to chordal spheres, and so does their composition. That similarity transformations of $\mathbb{R}^n$ transform chordal spheres to chordal spheres is evident. Because $\text{Möb}(n)$ is generated by $\text{GS}(n)$ and R_0, the inversion in $\mathbb{S}^{n-1}$, the proof boils down to confirming the assertion of the theorem for $f = R_0$.

Fix an equation for the chordal sphere Σ, say

$$\Sigma : a|x|^2 - 2\langle b, x\rangle + c = 0.$$

Then for x different from 0 and ∞ the point $y = R_0(x) = x/|x|^2$ satisfies

$$c|y|^2 - 2\langle b, y\rangle + a = \frac{c}{|x|^2} - 2\Big\langle b, \frac{x}{|x|^2}\Big\rangle + a = \frac{c - 2\langle b, x\rangle + a|x|^2}{|x|^2}\ ,$$

so x lies on Σ precisely when y lies on the chordal sphere $\Sigma' : c|y|^2 - 2\langle b, y\rangle + a = 0$. Furthermore, ∞ belongs to Σ if and only if $a = 0$, which occurs if and only if $0 = R_0(\infty)$ is on Σ'. Similarly, 0 is a point of Σ if and only if $\infty = R_0(0)$ lies on Σ'. Therefore $R_0(\Sigma) = \Sigma'$ is a chordal sphere, as claimed. □

When $n \geq 2$ the property of mapping chordal spheres to chordal spheres actually characterizes Möbius transformations:

LEMMA 3.3.4. *If $n \geq 2$ and $f : \hat{\mathbb{R}}^n \to \hat{\mathbb{R}}^n$ is an injective function with the property that $f(\Sigma)$ is a chordal sphere for every chordal sphere Σ, then f is a Möbius transformation of $\hat{\mathbb{R}}^n$.*

Since we do not intend to make any use of this observation, we leave its proof as an amusing and instructive exercise for the reader.

Theorem 3.3.3 has a generalization covering lower-dimensional chordal spheres:

LEMMA 3.3.5. *If f is a Möbius transformation of $\hat{\mathbb{R}}^n$ and Γ is a chordal p-sphere in $\hat{\mathbb{R}}^n$, then $f(\Gamma)$ is a chordal p-sphere.*

By a chordal p-sphere in $\hat{\mathbb{R}}^n$, $1 \leq p \leq n-1$, we mean either a p-dimensional Euclidean sphere contained in some $(p+1)$-dimensional affine subspace of $\mathbb{R}^n$ or a set of the type $V \cup \{\infty\}$, where V is a p-dimensional affine subspace of $\mathbb{R}^n$.

Thus "chordal sphere" is short for "chordal $(n-1)$-sphere".

As was the situation in Theorem 3.3.3, difficulties in the proof of its generalized version surface only for $f = R_0$, the inversion in $\mathbb{S}^{n-1}$. They are easily handled by doing an induction on n, using Theorem 3.3.3 and the fact that $U \circ R_0 = R_0 \circ U$ for every U from $\text{O}(n)$. The details are again left to the reader.

Let Σ be a chordal sphere in $\hat{\mathbb{R}}^n$. We can represent Σ uniquely in the manner $\Sigma = \tilde{\Sigma} \cap \hat{\mathbb{R}}^n$, in which $\tilde{\Sigma}$ is a chordal sphere in $\hat{\mathbb{R}}^{n+1}$ that is orthogonal to $\mathbb{R}^n$.

If $\pi : \hat{\mathbb{R}}^n \to \mathbb{S}^n$ denotes stereographic projection, then we are able to profit from an earlier discussion by remarking that

$$\pi(\Sigma) = \pi\big(\tilde{\Sigma} \cap \hat{\mathbb{R}}^n\big) = R\big(\tilde{\Sigma} \cap \hat{\mathbb{R}}^n\big) = R(\tilde{\Sigma}) \cap \mathbb{S}^n \ ,$$

where R denotes the inversion in $\mathbb{S}^n(e_{n+1}, \sqrt{2}\,)$. As the nondegenerate intersection of $\mathbb{S}^n$ with either another Euclidean sphere or a hyperplane, the set $\pi(\Sigma)$ is a Euclidean sphere of dimension $n-1$ sitting in $\mathbb{S}^n$. A moment's thought reveals that there must exist a point p_0 of $\mathbb{S}^n$ (there may be two choices for p_0) and a number r in $(0, \sqrt{2}\,]$ such that $\pi(\Sigma) = \{p \in \mathbb{S}^n : |p - p_0| = r\}$. From this we infer that $\Sigma = \{x \in \hat{\mathbb{R}}^n : q(x, x_0) = r\}$, in which $x_0 = \pi^{-1}(p_0)$. We conclude that Σ really is a sphere in the chordal metric on $\hat{\mathbb{R}}^n$.

The group $\text{Möb}(n)$ acts transitively on the family of chordal spheres in $\hat{\mathbb{R}}^n$: *if Σ and Σ' are chordal spheres in $\hat{\mathbb{R}}^n$, then there exists a Möbius transformation f of $\hat{\mathbb{R}}^n$ with the property that $f(\Sigma) = \Sigma'$.* This readily verified observation plays a role in the proof of the ensuing theorem, which would be false were $n = 1$.

THEOREM 3.3.6. *Let Σ be a chordal sphere in $\hat{\mathbb{R}}^n$ with $n \geq 2$, and let f be a member of* $\text{Möb}(n)$ *that fixes every point of Σ. Then f is either the identity transformation of $\hat{\mathbb{R}}^n$ or the reflection in Σ.*

PROOF. We deal initially with the special case $\Sigma = \mathbb{R}^{n-1} \cup \{\infty\}$. Then, in particular, $f(\infty) = \infty$. Consider a sphere $\Sigma' = \mathbb{S}^{n-1}(x_0, r)$ with x_0 in $\mathbb{R}^{n-1}$. Since ∞ is not a point of Σ', the set $f(\Sigma')$ must also be a Euclidean sphere, one for which

$$f(\Sigma') \cap \mathbb{R}^{n-1} = f(\Sigma') \cap f(\mathbb{R}^{n-1}) = f(\Sigma' \cap \mathbb{R}^{n-1}) = \Sigma' \cap \mathbb{R}^{n-1} \, .$$

Moreover, because Σ' is orthogonal to $\mathbb{R}^{n-1}$, the conformality of f in the complement of some finite subset of $\hat{\mathbb{R}}^n$ requires that $f(\Sigma')$, too, be orthogonal to $\mathbb{R}^{n-1}$. This information is enough to pin down the image of Σ' under f; namely, $f(\Sigma') = \Sigma'$.

Let x be an element of $\mathbb{R}^n \setminus \mathbb{R}^{n-1}$. We have seen that for each x_0 in $\mathbb{R}^{n-1}$ the sphere $S^{n-1}(x_0, |x - x_0|)$ remains invariant under f. Therefore $y = f(x)$ satisfies the condition $|y - x_0| = |x - x_0|$ for every x_0 in $\mathbb{R}^{n-1}$. The choice $x_0 = 0$ exposes the fact that $|y| = |x|$. We deduce that

$$\langle y, x_0 \rangle = \frac{|y|^2 + |x_0|^2 - |y - x_0|^2}{2} = \frac{|x|^2 + |x_0|^2 - |x - x_0|^2}{2} = \langle x, x_0 \rangle$$

for all x_0 in $\mathbb{R}^{n-1}$. Taking $x_0 = e_i$ yields $y_i = \langle y, e_i \rangle = \langle x, e_i \rangle = x_i$ for $1 \leq i \leq n-1$. Because we also know that $|y| = |x|$, either $y_n = x_n$ or $y_n = -x_n$. Now either the homeomorphism f leaves each component of $\mathbb{R}^n \setminus \mathbb{R}^{n-1}$ invariant or it interchanges these components. It follows that either $f(x) = x$ for every x in $\hat{\mathbb{R}}^n$ or $f(x) = (x_1, x_2, \ldots, -x_n)$ for all such x. In the latter case f is the reflection in $\mathbb{R}^{n-1}$.

Suppose finally that Σ is an arbitrary chordal sphere in $\hat{\mathbb{R}}^n$. We choose g in $\text{Möb}(n)$ for which $g(\Sigma_0) = \Sigma$, where $\Sigma_0 = \mathbb{R}^{n-1} \cup \{\infty\}$. Write $R = R_\Sigma$, $S = R_{\Sigma_0}$. The function $g^{-1} \circ R \circ g$ belongs to $\text{Möb}(n)$, is not the identity mapping of $\hat{\mathbb{R}}^n$, yet fixes Σ_0 pointwise. By the first part of the proof, $g^{-1} \circ R \circ g = S$. The Möbius transformation $g^{-1} \circ f \circ g$ also fixes every point of Σ_0. Again appealing to the special case treated first, we conclude that either $g^{-1} \circ f \circ g = I$, making $f = I$, or $g^{-1} \circ f \circ g = S$, in which case $f = g \circ S \circ g^{-1} = R$. □

The proof of Theorem 3.3.6 demonstrates that the reflection R_Σ associated with a chordal sphere Σ in $\hat{\mathbb{R}}^n$ is conjugate in the group Möb(n) to R_{Σ_0}, where $\Sigma_0 = \mathbb{R}^{n-1} \cup \{\infty\}$. Theorem 3.3.6 has the further consequence that

$$R_{f(\Sigma)} \circ f = f \circ R_\Sigma \tag{3.15}$$

for every f in Möb(n). To verify this, simply observe that the Möbius transformation $f^{-1} \circ R_{f(\Sigma)} \circ f$ fixes each point of Σ but is not the identity transformation. Accordingly, $f^{-1} \circ R_{f(\Sigma)} \circ f = R_\Sigma$, which is a rephrasing of (3.15). Recalling that points x and y of Σ^c are *symmetric with respect to* Σ provided $R_\Sigma(x) = y$ (hence, also, $R_\Sigma(y) = x$), we derive directly from (3.15) a useful symmetry principle (the case $n = 1$ requires a separate proof, which we omit).

Theorem 3.3.7. *If points x and y are symmetric with respect to a chordal sphere Σ in $\hat{\mathbb{R}}^n$ and if f belongs to* Möb(n)*, then $f(x)$ and $f(y)$ are symmetric with respect to $f(\Sigma)$.*

Symmetry with respect to a chordal sphere admits the following characterisation.

Corollary 3.3.8. *Let Σ be a chordal sphere in $\hat{\mathbb{R}}^n$ with $n \geq 2$, and let x and y be points of Σ^c, the complement of Σ in $\hat{\mathbb{R}}^n$. Then x and y are symmetric with respect to Σ if and only if every chordal sphere that passes through both x and y intersects Σ orthogonally.*

Proof. The assertions of the theorem are evident in the case of $\Sigma = \mathbb{R}^{n-1} \cup \{\infty\}$. Because of Theorem 3.3.7 and the conformality of Möbius transformations, both the symmetry property and the orthogonality property are Möbius invariant. This fact enables us to reduce the general situation to the special one, merely by selecting a transformation f in Möb(n) that maps a given chordal sphere Σ to $\mathbb{R}^{n-1} \cup \{\infty\}$. □

3.3.1. Cross ratios. If x, y, u, and v are distinct points of $\hat{\mathbb{R}}^n$, then the *chordal cross-ratio* $[x, y, u, v]$ is the quantity defined by

$$[x, y, u, v] = \frac{q(x,u) \cdot q(y,v)}{q(x,y) \cdot q(u,v)}\,. \tag{3.16}$$

When all four points are finite, equation (3.16) reduces to

$$[x, y, u, v] = \frac{|x-u| \cdot |y-v|}{|x-y| \cdot |u-v|}\,. \tag{3.17}$$

Euclidean expressions for cross-ratios in which one of the four points involved is ∞ are:

$$\begin{cases} [\infty, y, u, v] = \dfrac{|y-v|}{|u-v|}, \qquad [x, \infty, u, v] = \dfrac{|x-u|}{|u-v|}\,, \\[2ex] [x, y, \infty, v] = \dfrac{|y-v|}{|x-y|}, \qquad [x, y, u, \infty] = \dfrac{|x-u|}{|x-y|}\,. \end{cases} \tag{3.18}$$

Möbius transformations are characterized by their preservation of chordal cross-ratios or Euclidean cross-ratios.

Theorem 3.3.9. *Suppose that $f : \hat{\mathbb{R}}^n \to \hat{\mathbb{R}}^n$ is an injective function. Then f belongs to* Möb(n) *if and only if*

$$[x, y, u, v] = \big[f(x), f(y), f(u), f(v)\big] \tag{3.19}$$

whenever $x, y, u,$ and v are distinct points of $\hat{\mathbb{R}}^n$.

PROOF. Assume first that f is a member of Möb(n). If f is a dilation or a translation or an orthogonal linear transformation, then $f(\infty) = \infty$ and there is a constant $\lambda > 0$ with the property that $|f(z) - f(w)| = \lambda |z - w|$ for all points z and w of $\mathbb{R}^n$. It follows almost immediately from (3.17) and (3.18) that (3.19) holds for such f.

Consider next the inversion R_0 in $\mathbb{S}^{n-1}$. An elementary computation leads to the relation

$$|R_0(z) - R_0(w)| = \frac{|z - w|}{|z| \cdot |w|} \tag{3.20}$$

for z and w in $\mathbb{R}^n \setminus \{0\}$. With (3.17), identity (3.20) confirms (3.19) for $f = R_0$, provided that none of the points in question is 0 or ∞.

The continuity of cross-ratios is obvious from definition (3.16). From this continuity, (3.19) remains valid for this particular transformation when 0 and ∞ are included. Since (3.19) holds for $g \circ f$ whenever it holds for both f and g and since Möb(n) is generated by GS(n) and R_0, (3.19) obtains an arbitrary Möbius transformation f.

In proving the converse we are free to assume that $f(\infty) = \infty$. Should $f(\infty) \neq \infty$, we would simply consider $h = g \circ f$, where g is a Möbius transformation that takes $f(\infty)$ to ∞. Then h obeys (3.19), h fixes ∞, and $f = g^{-1} \circ h$ belongs to Möb(n) if h does.

Let x, y, u, and v be distinct points in $\mathbb{R}^n$. From (3.18) and the invariance assumption (3.19) we learn that

$$\frac{|y - v|}{|u - v|} = [\infty, y, u, v] = \big[\infty, f(y), f(u), f(v)\big] = \frac{|f(y) - f(v)|}{|f(u) - f(v)|}$$

and

$$\frac{|y - v|}{|x - y|} = [x, y, \infty, v] = \big[f(x), f(y), \infty, f(v)\big] = \frac{|f(y) - f(v)|}{|f(x) - f(y)|}.$$

These equations enable us to conclude that

$$\frac{|f(x) - f(y)|}{|x - y|} = \frac{|f(y) - f(v)|}{|y - v|} = \frac{|f(u) - f(v)|}{|u - v|},$$

the inference being that the ratio $|f(z) - f(w)|/|z - w|$ is the same for all pairs of distinct points z and w in $\mathbb{R}^n$. If we call the ratio in question r, then the function $r^{-1}f$ is a Euclidean isometry. This makes f a member of GS(n), a subgroup of Möb(n). □

Picking through the details of the preceding proof, we glean some additional information on the structure of Möbius transformations.

First, we draw attention to the analogue of (3.20) for the inversion $R = R_\Sigma$ in an arbitrary Euclidean sphere $\Sigma = \mathbb{S}^{n-1}(x_0, r)$:

$$|R(x) - R(y)| = \frac{r^2 |x - y|}{|x - x_0| \cdot |y - x_0|} \tag{3.21}$$

for x and y in $\mathbb{R}^n \setminus \{x_0\}$. This relation is derived from (3.20) by using the fact that $R = g^{-1} \circ f \circ R_0 \circ g$ with $g(x) = x - x_0$ and $f(x) = r^2 x$.

Next we point out:

THEOREM 3.3.10. *Let f be a member of* Möb(n). *If $f(\infty)=\infty$, then f is a similarity transformation. If $f(\infty)\neq\infty$, then $\Sigma=\{x\in\mathbb{R}^n : \|f'(x)\|=1\}$ is a Euclidean sphere centered at $x_0=f^{-1}(\infty)$, and f can be written as the composition $f=g\circ R_\Sigma$, in which g is a Euclidean isometry.*

PROOF. That a Möbius transformation fixing ∞ necessarily belongs to $\mathrm{GS}(n)$ is knowledge that emerged in the last part of the proof of Theorem 3.3.9. Assuming now that $f(\infty)\neq\infty$, let $x_0=f^{-1}(\infty)$ and let R be the inversion in the sphere $\mathbb{S}^{n-1}(x_0,1)$. Then $h=f\circ R$ is a Möbius transformation that fixes ∞. Thus h is in $\mathrm{GS}(n)$, say $h=\lambda T$ with $\lambda>0$ and T in the group $\boldsymbol{E}(n)$. In particular, $\|h'(x)\|=\lambda$ for every x in $\mathbb{R}^n$. Because $f=h\circ R^{-1}=h\circ R$, we conclude that f is everywhere differentiable in the domain $\mathbb{R}^n\setminus\{x_0\}$, where

$$\|f'(x)\|=\|h'\big[R(x)\big]\|\cdot\|R'(x)\|=\frac{\lambda}{|x-x_0|^2},$$

where we have recalled (3.1) and (3.10).

From this statement we extract the information that

$$\Sigma=\{x\in\mathbb{R}^n : \|f'(x)\|=1\}=\mathbb{S}^{n-1}(x_0,\sqrt{\lambda}).$$

Moreover, (3.21) gives

$$|f(x)-f(y)|=\lambda|R(x)-R(y)|=\frac{\lambda|x-y|}{|x-x_0|\cdot|y-x_0|}=|x-y| \tag{3.22}$$

whenever x and y lie on Σ. The function $g=f\circ R_\Sigma$ is yet another similarity transformation, one whose restriction to Σ is shown by (3.22) to be a Euclidean isometry. We infer that g itself is a Euclidean isometry. And, of course, $f=g\circ R_\Sigma$. □

If f is a Möbius transformation that does not fix the point ∞, then the set $\Sigma=\{x\in\mathbb{R}^n : \|f'(x)\|=1\}$ is called the *isometric sphere* of f. Theorem 3.3.10 ties up a tiny loose end left hanging at the beginning of this section.

We are now in a position to state unequivocally: *a Möbius transformation f of $\hat{\mathbb{R}}^n$ restricts to a C^∞-diffeomorphism of the domain $\mathbb{R}^n\setminus\{f^{-1}(\infty)\}$ onto $\mathbb{R}^n$. This diffeomorphism is conformal when $n\geq 2$.*

As a second corollary of Theorem 3.3.10 we obtain a criterion for deciding whether two potentially different Möbius transformations actually are the same.

COROLLARY 3.3.11. *Suppose that f and g are both Möbius transformations of $\hat{\mathbb{R}}^n$. If the set $E=\{x\in\hat{\mathbb{R}}^n : f(x)=g(x)\}$ does not lie on any chordal sphere, then $f\equiv g$.*

PROOF. We demonstrate that the function $h=g^{-1}\circ f$, which fixes E pointwise, must be the identity transformation of $\hat{\mathbb{R}}^n$. Through any set of $n+1$ or fewer points of $\hat{\mathbb{R}}^n$ there always passes at least one chordal sphere. Thus the set E here can contain no fewer than $n+2$ points. By performing a preliminary conjugation we are therefore at liberty to assume that both 0 and ∞ belong to E.

The nonzero finite points of E must include a basis for $\mathbb{R}^n$. Otherwise the linear subspace of $\mathbb{R}^n$ spanned by such points would have dimension no greater than $n-1$, so E would definitely be contained in some chordal sphere. Since $h(0)=0$

and $h(\infty) = \infty$, we infer from Theorem 3.3.10 that h is a linear transformation of $\mathbb{R}^n$; since h also fixes the members of a basis for $\mathbb{R}^n$, $h = I$, the identity. Thus $f \equiv g$. □

It is a direct consequence of Theorem 3.3.9 that every isometry of the metric space $(\hat{\mathbb{R}}^n, q)$ is a Möbius transformation. The chordal isometry group of $\hat{\mathbb{R}}^n$ is therefore a subgroup of $\text{Möb}(n)$. As a matter of fact, the chordal isometries of $\hat{\mathbb{R}}^n$ are quite easy to characterize.

THEOREM 3.3.12. *The chordal isometries of* $\hat{\mathbb{R}}^n$ *are the functions* $f = \pi^{-1} \circ U \circ \pi$, *where* $\pi : \hat{\mathbb{R}}^n \to \mathbb{S}^n$ *is a stereographic projection and* U *is an arbitrary member of* $\mathrm{O}(n+1)$. *In particular, the chordal isometry group acts transitively on* $\hat{\mathbb{R}}^n$.

PROOF. Every function f of the type described is obviously a chordal isometry. Because $\mathrm{O}(n+1)$ acts transitively on $\mathbb{S}^n$, the last assertion in the theorem then becomes clear as well. It remains to show that an arbitrary chordal isometry f has the structure $f = \pi^{-1} \circ U \circ \pi = g_U$ for some U in $\mathrm{O}(n+1)$. To demonstrate this, we first choose V from $\mathrm{O}(n+1)$ in such a way that $h = g_V \circ f$ maps ∞ to ∞. Then h is also a chordal isometry and, because

$$q[h(0), \infty] = q[h(0), h(\infty)] = q(0, \infty) = 2 \ ,$$

we find that $h(0) = 0$. We conclude with the help of Theorem 3.3.10 that h takes the form $h = \lambda T$, where $\lambda > 0$ and T belongs to $\mathrm{O}(n)$. Since $|h(e_1)| = \lambda$, we see that

$$\frac{\lambda}{\sqrt{1+\lambda^2}} = q\big[h(e_1), 0\big] = q(e_1, 0) = \frac{1}{\sqrt{2}} \ ,$$

which yields $\lambda = 1$ and places h in $\mathrm{O}(n)$. There is a unique transformation U in $\mathrm{O}(n+1)$ that satisfies $U(e_{n+1}) = e_{n+1}$ and $U|\mathbb{R}^n = h$. The transformation U maps the ray that emanates from e_{n+1} and passes through a point x of $\mathbb{R}^n$ to the analogous ray determined by e_{n+1} and $h(x)$. This ensures that $\pi \circ h = U \circ \pi$, so $h = g_U$. Finally, $f = g_V^{-1} \circ h = g_V^{-1} \circ g_U = g_{V^{-1}U}$, giving f the prescribed structure. □

In a number of upcoming discussions it will be important to know the exact nature of the Möbius transformations that leave invariant certain special domains in $\hat{\mathbb{R}}^n$. When D is a domain in $\hat{\mathbb{R}}^n$, we use the notation $\text{Möb}(D)$ to represent $\{f \in \text{Möb}(n) : f(D) = D\}$, while $\text{Möb}^+(D)$ stands for $\text{Möb}(D) \cap \text{Möb}^+(n)$. Both $\text{Möb}(D)$ and $\text{Möb}^+(D)$ are subgroups of $\text{Möb}(n)$.

We shall sometimes ignore the fact that the members of $\text{Möb}(D)$ are global mappings of $\hat{\mathbb{R}}^n$ and think of them simply as homeomorphisms of D onto itself, blurring in the process the distinction between a transformation from $\text{Möb}(D)$ and its restriction to D.

We now proceed to determine the Möbius transformations that belong to $\text{Möb}(\mathbb{H}^n)$ and $\text{Möb}(B^n)$ for $n \geq 2$. In describing the first class we write $\mathrm{GS}(\mathbb{H}^n)$ for $\text{Möb}(\mathbb{H}^n) \cap \mathrm{GS}(n)$.

THEOREM 3.3.13. *Let* $n \geq 2$. *A similarity transformation* f *is in* $\mathrm{GS}(\mathbb{H}^n)$ *if and only if* f *has the form*

$$f = \lambda U + b,$$

in which $\lambda > 0$*,* b *is an element of* $\mathbb{R}^{n-1} = \partial\mathbb{H}^n$*, and* U *is an orthogonal transformation in* $\mathrm{O}(n)$ *that fixes* e_n*.*

A Möbius transformation f *belongs to* $\text{Möb}(\mathbb{H}^n)$ *if and only if* f *is in* $\mathrm{GS}(\mathbb{H}^n)$ *or has the structure*

$$f = g \circ R_\Sigma,$$

where g *is in* $\mathrm{GS}(\mathbb{H}^n)$ *and* Σ *is a Euclidean sphere that is orthogonal to* $\mathbb{R}^{n-1}$*. In the latter case one can take* Σ *to be the isometric sphere of* f*, in which event* g *is a Euclidean isometry.*

PROOF. Let f belong to $\mathrm{GS}(\mathbb{H}^n)$, say $f = \lambda U + b$ with $\lambda > 0$, b in $\mathbb{R}^n$, and U in $\mathrm{O}(n)$. Since $f(\mathbb{H}^n) = \mathbb{H}^n$, it follows that $f(\hat{\mathbb{R}}^{n-1}) = \hat{\mathbb{R}}^{n-1}$. Then $b = f(0)$ lies in $\mathbb{R}^{n-1}$, so $U(e_n) = \lambda^{-1}(f(e_n) - b)$ is a point of $\mathbb{H}^n$. On the other hand, $U(e_n)$ is a unit vector that is orthogonal to $U(\mathbb{R}^{n-1}) = f(\mathbb{R}^{n-1}) = \mathbb{R}^{n-1}$ and conformal mappings preserve angles. These conditions mandate that $U(e_n) = e_n$ and that f has the stated form.

Conversely, every similarity transformation of the type under consideration is clearly an element of $\mathrm{GS}(\mathbb{H}^n)$.

Next, an arbitrary transformation f in $\text{Möb}(\mathbb{H}^n)$ either fixes ∞, in which case Theorem 3.3.10 places f in $\mathrm{GS}(\mathbb{H}^n)$, or has the structure $f = g \circ R_\Sigma$, where Σ is the isometric sphere of f and g is a Euclidean isometry. In the latter case, the center $x_0 = f^{-1}(\infty)$ of Σ must lie in $\mathbb{R}^{n-1}$, which lets us know that Σ is orthogonal to $\mathbb{R}^{n-1}$. The inversion R_Σ leaves invariant every ray emanating from x_0, a fact which reveals that R_Σ preserves $\mathbb{H}^n$. This causes R_Σ and, consequently, $g = f \circ R_\Sigma$ to be members of $\text{Möb}(\mathbb{H}^n)$. Because g fixes ∞, it thus belongs to $\mathrm{GS}(\mathbb{H}^n)$. Lastly, a Euclidean sphere that meets $\mathbb{R}^{n-1}$ orthogonally has its center in $\mathbb{R}^{n-1}$, implying as earlier that the associated reflection R transforms $\mathbb{H}^n$ to itself, hence that R and $g \circ R$ for any g from $\mathrm{GS}(\mathbb{H}^n)$ are transformations in $\text{Möb}(\mathbb{H}^n)$. □

The group $\text{Möb}(\mathbb{H}^n)$ for $n \geq 2$ is generated by the dilations of $\mathbb{R}^n$, the translations of $\mathbb{R}^n$ in directions parallel to $\mathbb{R}^{n-1}$, the transformations from $\mathrm{O}(n)$ that fix the vector e_n, and the inversion R_0 in $\mathbb{S}^{n-1}$. The group $\text{Möb}^+(\mathbb{H}^n)$ is generated by the sense-preserving members of $\mathrm{GS}(\mathbb{H}^n)$ together with the transformation $f_0 = g_0 \circ R_0$, with g_0 being the reflection in the hyperplane $P : x_1 = 0$. Notice that the group $\mathrm{GS}^+(\mathbb{H}^n)$—hence, each of the groups $\text{Möb}(\mathbb{H}^n)$ and $\text{Möb}^+(\mathbb{H}^n)$—acts transitively on $\mathbb{H}^n$. In the classical setting of the complex plane, the mappings in $\text{Möb}(\mathbb{H}^2)$ take one of the normalized forms

$$f(z) = \frac{az + b}{cz + d}\ ,$$

with a, b, c, and d real and $ad - bc = 1$, or

$$f(z) = \frac{a\bar{z} + b}{c\bar{z} + d}\ ,$$

in which a, b, c, and d are purely imaginary and $ad - bc = 1$. The transformations of the first kind make up $\text{Möb}^+(\mathbb{H}^2)$.

We note in passing that when $n \geq 2$ the quantity $|y - x|^2/(y_n x_n)$ is conserved under the action of $\text{Möb}(\mathbb{H}^n)$, meaning that

$$\frac{|f(y) - f(x)|^2}{f_n(y) f_n(x)} = \frac{|y - x|^2}{y_n x_n} \tag{3.23}$$

whenever f is a member of $\text{Möb}(\mathbb{H}^n)$ and the points x and y belong to $\mathbb{H}^n$. That this holds for any function f in $\text{GS}(\mathbb{H}^n)$ is all but trivial: writing $f = \lambda U + b$ as described in Theorem 3.3.13, we see that $|f(y) - f(x)| = \lambda|y - x|$, $f_n(y) = \lambda y_n$, and $f_n(x) = \lambda x_n$ from which (3.23) falls out easily. If $f = R_\Sigma$ for a Euclidean sphere $\Sigma = \mathbb{S}^{n-1}(x_0, r)$ orthogonal to $\mathbb{R}^{n-1}$, then

$$f(x) = x_0 + \frac{r^2(x - x_0)}{|x - x_0|^2}\,. \tag{3.24}$$

Thus, remembering that x_0 lies in $\mathbb{R}^{n-1}$, we get

$$f_n(y) = \frac{r^2 y_n}{|y - x_0|^2}, \qquad f_n(x) = \frac{r^2 x_n}{|x - x_0|^2}\,. \tag{3.25}$$

Also, by (3.21),

$$|f(y) - f(x)|^2 = \frac{r^4|y - x|^2}{|y - x_0|^2|x - x_0|^2}\,.$$

These observations combine to give (3.23) for $f = R_\Sigma$. Because relation (3.23) is preserved under composition, we can certify its correctness for an arbitrary f in $\text{Möb}(\mathbb{H}^n)$ by an appeal to Theorem 3.3.13.

An extremely significant consequence of (3.23), in tandem with the conformality of Möbius transformations, is the identity

$$\|f'(x)\| = \lim_{y \to x} \frac{|f(y) - f(x)|}{|y - x|} = \frac{f_n(x)}{x_n} \tag{3.26}$$

which holds for all x in $\mathbb{H}^n$ once we recall (2.21), (2.22), and Theorem 3.1.1, provided f comes from the group $\text{Möb}(\mathbb{H}^n)$.

If Σ is a chordal sphere in $\hat{\mathbb{R}}^n$, then as remarked earlier there is a unique chordal sphere $\tilde{\Sigma}$ in $\hat{\mathbb{R}}^{n+1}$ that is orthogonal to $\mathbb{R}^n$ and has $\tilde{\Sigma} \cap \hat{\mathbb{R}}^n = \Sigma$. It is apparent from the formulas defining the reflections $R = R_\Sigma$ and $\tilde{R} = R_{\tilde{\Sigma}}$ that $R = \tilde{R}|\hat{\mathbb{R}}^n$. It is equally apparent that $\tilde{R}$ leaves $\mathbb{H}^{n+1}$ invariant. On the strength of these comments, we are able to conclude that each member f of $\text{Möb}(n)$ can be extended to a Möbius transformation $\tilde{f}$ belonging to $\text{Möb}(\mathbb{H}^{n+1})$, a transformation known as the *Poincaré extension* of f. There can be only one such extension. If $\tilde{f}_1$ were a second extension, then $g = \tilde{f}^{-1} \circ \tilde{f}_1$ would be a member of $\text{Möb}(n+1)$ that fixed $\hat{\mathbb{R}}^n$ pointwise, yet was neither the identity ($\tilde{f} \neq \tilde{f}_1$) nor the reflection in $\hat{\mathbb{R}}^n$ (g preserves $\mathbb{H}^{n+1}$) contrary to Theorem 3.3.6. By Theorem 3.3.13 each function in $\text{Möb}(\mathbb{H}^{n+1})$ arises as the Poincaré extension of a unique member of $\text{Möb}(n)$. The members of $\text{Möb}^+(\mathbb{H}^{n+1})$ are the mappings $\tilde{f}$ with f in $\text{Möb}^+(n)$.

One can directly compute a formula for the Poincaré extension of an element $f \in \text{Möb}^+(n)$ to an element of $\text{Möb}^+(\mathbb{H}^{n+1})$.

3.3.2. The Poincaré extension. Suppose that $f \in \text{Möb}^+(n-1)$. There are two cases to consider. If $f(\infty) = \infty$, then f is a similarity of the form $f : \lambda U + b$ with $U \in O(n-1)$ and $b \in \mathbb{R}^{n-1}$. The obvious thing to do is set

$$\tilde{U} = \begin{pmatrix} U & 0^t \\ 0 & 1 \end{pmatrix}, \qquad \tilde{b} = (b_1, b_2, \ldots, b_{n-1}, 0)$$

and set $\tilde{f} = \lambda\tilde{U} + \tilde{b}$. Here $0 \in \mathbb{R}^{n-1}$. Thus $(x,t) \mapsto (f(x), \lambda t)$, so to write the extension we need to identify only the scale factor λ which we must choose positive to preserve $\mathbb{H}^n$. Then $\lambda = |f(e_1) - f(0)|$, and if we define

$$\tilde{f}(x,t) = (f(x), |f(e_1) - f(0)|\, t),$$

it is easy to check that this is a Möbius transformation of $\mathbb{H}^n$ and $\tilde{f}\big|\mathbb{R}^{n-1} = f$.

Next, if $f(\infty) = y_0 \neq \infty$, we put

$$g(x) = y_0 + \frac{f(x) - y_0}{|f(x) - y_0|^2}$$

to see that g is a Möbius transformation for which $g(\infty) = \infty$. Then g has an extension $\tilde{g}$ as above, and once we note that $g = \Phi \circ f$, where Φ is inversion in the unit sphere about y_0 as per (3.24), then the extension we seek will be $\tilde{\Phi}^{-1} \circ \tilde{g}$. The necessary calculation is simplified by observing that all we need to do is calculate the height (that is, the n^{th} coordinate) of $(\tilde{\Phi}^{-1} \circ \tilde{g})(x,t) = (\tilde{\Phi} \circ \tilde{g})(x,t)$, as we already know that this map is f on $\mathbb{R}^{n-1}$. We can therefore use (3.25) to obtain

$$\begin{aligned}
s &= \tilde{g}(x,t)_n = |g(e_1) - g(0)|t \\
&= \Big|\frac{f(e_1) - f(\infty)}{|f(e_1) - f(\infty)|^2} - \frac{f(0) - f(\infty)}{|f(0) - f(\infty)|^2}\Big|\, t, \\
\tilde{\Psi}(\tilde{g}(x,t))_n &= \frac{s}{|\tilde{g}(x,t) - (y_0, 0)|^2} = \frac{s}{|g(x) - y_0|^2 + \tilde{g}(x,t)_n^2} \\
&= \frac{s}{|f(x) - f(\infty)|^{-2} + s^2}.
\end{aligned}$$

This gives a simple procedure for writing the Poincaré extension.

Lemma 3.3.14. *Let $f \in \text{Möb}^+(n-1)$ with $f(\infty) \neq \infty$ and set*

$$\alpha = \Big|\frac{f(e_1) - f(\infty)}{|f(e_1) - f(\infty)|^2} - \frac{f(0) - f(\infty)}{|f(0) - f(\infty)|^2}\Big|.$$

Then the Poincaré extension $\tilde{f} \in \text{Möb}^+(\mathbb{H}^n)$ is given by the formula

$$\tilde{f}(x,t) = \Big(f(x), \frac{\alpha t}{|f(x) - f(\infty)|^{-2} + \alpha^2 t^2}\Big). \tag{3.27}$$

Of course an alternative approach to the proof of this lemma is via the cross-ratio. Here we tacitly use the inclusion $\mathbb{R}^{n-1} \hookrightarrow \mathbb{R}^n$. Then we must have from Theorem 3.3.9 and (3.18),

$$\begin{aligned}
[0, e_1, \infty, (x,t)] &= [\tilde{f}(0), \tilde{f}(e_1), \tilde{f}(\infty), \tilde{f}(x,t)] = [f(0), f(e_1), f(\infty), (f(x), \tilde{t})], \\
|x - e_1|^2 + t^2 &= \frac{|f(0) - f(\infty)|^2 |f(e_1) - (f(x), \tilde{t})|^2}{|f(0) - f(e_1)|^2 |f(\infty) - (f(x), \tilde{t})|^2} \\
&= \frac{|f(0) - f(\infty)|^2\, (|f(e_1) - f(x)|^2 + \tilde{t}^2)}{|f(0) - f(e_1)|^2\, (|f(\infty) - f(x)|^2 + \tilde{t}^2)}. \qquad (3.28)
\end{aligned}$$

From this one can solve for a positive $\tilde{t}$, and the result will follow. The appearance of the variable x in the formula for $\tilde{t}$ obtained seems not to coincide with (3.27).

This is accounted for in the invariance of another cross-ratio which relates x and $f(x)$. For instance if $f(\infty)=\infty$, $f:\lambda U+b$, $\lambda>0$, (3.28) becomes

$$\begin{aligned} |x-e_1|^2+t^2 &= \frac{|f(e_1)-f(x)|^2+\tilde{t}^{\,2}}{|f(0)-f(e_1)|^2}=\frac{\lambda^2\,|x-e_1|^2+\tilde{t}^{\,2}}{\lambda^2}, \\ \lambda^2 t^2 &= \tilde{t}^{\,2}, \end{aligned}$$

as before. Indeed there are even further approaches using the hyperbolic geometry of lines and their perpendiculars which the reader may care to explore.

3.3.3. The group Möb(B^n). Before attempting to characterize the transformations in the group Möb(B^n) for $n\geq 2$, we make some preparatory remarks. Recall that the function $\Phi=R\circ S$, where S is reflection in $\mathbb{R}^{n-1}$ and R is reflection in $\mathbb{S}^{n-1}(e_n,\sqrt{2}\,)$, defines a Möbius transformation that maps $\mathbb{H}^n$ to B^n. Let Σ be a chordal sphere that is orthogonal to $\mathbb{S}^{n-1}$. The chordal sphere $\Sigma'=\Phi^{-1}(\Sigma)$ is orthogonal to $\mathbb{R}^{n-1}$, and $R_\Sigma=\Phi\circ R_{\Sigma'}\circ\Phi^{-1}$. We already know that $R_{\Sigma'}$ preserves $\mathbb{H}^n$, allowing us to conclude: *the reflection R_Σ in any chordal sphere Σ that meets $\mathbb{S}^{n-1}$ orthogonally leaves B^n invariant.* A hyperplane P in $\mathbb{R}^n$ is orthogonal to $\mathbb{S}^{n-1}$ if and only if P passes through the origin.

The question then is to determine under what conditions does the Euclidean sphere Σ with center x_0 and radius r intersect $\mathbb{S}^{n-1}$ orthogonally? We maintain that this happens precisely when

$$|x_0|^2=1+r^2\,. \tag{3.29}$$

Given that $\Sigma\cap S^{n-1}\neq\emptyset$, the law of cosines determines the angle θ formed by radii of the two spheres at a point of intersection:

$$\cos\theta=\frac{1+r^2-|x_0|^2}{2r}\,.$$

To get $\theta=\pi/2$, (3.29) is necessary. Conversely, if (3.29) holds, then the center of Σ lies in $\mathbb{R}^n\setminus\bar{B}^n$. The nearest point of Σ to the origin has norm $|x_0|-r$, which is a number in $(0,1)$ —a glance at (3.29) shows that $|x_0|>r$ and $(|x_0|-r)(|x_0|+r)=1$— so Σ intersects B^n. Therefore Σ intersects $\mathbb{S}^{n-1}$, and the law of cosines affirms that the angle of intersection is a right angle. With these comments behind us, we may prove the following theorem.

THEOREM 3.3.15. *Let $n\geq 2$. A Möbius transformation f belongs to* Möb(B^n) *if and only if f is an orthogonal transformation in* O(n) *or has the structure*

$$f=g\circ R_\Sigma,$$

where $g\in$ O(n) and Σ is a Euclidean sphere that is orthogonal to $\mathbb{S}^{n-1}$.

PROOF. The transformations of the two types mentioned are definitely included in Möb(B^n). To establish the converse, consider an arbitrary member f of Möb(B^n). Out of necessity f transforms $\mathbb{S}^{n-1}$ to itself. If $f(0)=0$, f also fixes ∞ (Theorem 3.3.7). Theorem 3.3.10 then implies that f is a linear transformation, and a linear transformation that leaves B^n invariant must be an orthogonal transformation. Assuming next that $x_0=f^{-1}(0)$ is nonzero, let $r=(|x_0^*|^2-1)^{1/2}$ and $\Sigma=\mathbb{S}^{n-1}(x_0^*,r)$, where x_0^* is the point that is symmetric to x_0 with respect to $\mathbb{S}^{n-1}$. By reason of (3.29) the sphere Σ is orthogonal to $\mathbb{S}^{n-1}$, so R_Σ maps B^n to itself. Since $R_\Sigma(\infty)=x_0^*$, Theorem 3.3.7 guarantees that $R_\Sigma(0)=x_0$.

We infer that $g = f \circ R_\Sigma$ is a transformation from $\text{Möb}(B^n)$ and fixes the origin. The first part of our discussion implies that g belongs to $\text{O}(n)$ and $f = g \circ R_\Sigma$. $\square$

In light of Lemma 3.3.1, Theorem 3.3.15 shows that $\text{Möb}(B^n)$ is generated by the family of reflections R_Σ in which Σ is a chordal sphere orthogonal to $\mathbb{S}^{n-1}$. Further, an analogous statement is true for $\text{Möb}(\mathbb{H}^n)$: it is generated by the reflections R_Σ with Σ orthogonal to $\mathbb{R}^{n-1}$.

Of course it follows that the members of the group $\text{Möb}^+(B^n)$ are the transformations from $\text{SO}(n)$ and the mappings of the type $f = g \circ R_\Sigma$, where g is a sense-reversing orthogonal transformation and Σ is a Euclidean sphere that meets $\mathbb{S}^{n-1}$ at a right angle.

We draw particular attention to the transformations T_b in $\text{Möb}^+(B^n)$ defined for b in B^n as follows: $T_0 = I$, and for $b \neq 0$

$$T_b = g_b \circ R_b \ , \tag{3.30}$$

where R_b is the inversion in the Euclidean sphere of radius $r = (|b^*|^2 - 1)^{1/2}$ centered at b^* (notation as earlier), and g_b is the reflection in the hyperplane P: $\langle b, x \rangle = 0$.

The transformation T_b maps b to the origin. If $b \neq 0$, T_b fixes the points $b/|b|$ and $-b/|b|$. It thus leaves invariant the diameter of B^n through b. The presence of the transformations T_b in $\text{Möb}^+(B^n)$ shows directly that the action of this group on B^n is transitive. The mappings in $\text{Möb}(B^2)$ admit complex representations of the character

$$f(z) = \frac{az + b}{\bar{b}z + \bar{a}} \quad \text{or} \quad f(z) = \frac{a\bar{z} + b}{\bar{b}\bar{z} + \bar{a}} \ ,$$

in which a and b are complex numbers satisfying $|a|^2 - |b|^2 = 1$. Those of the first type are the mappings in $\text{Möb}^+(B^2)$.

The identities (3.23) and (3.26) have counterparts for the group $\text{Möb}(B^n)$ with $n \geq 2$:

$$\frac{|f(y) - f(x)|^2}{(1 - |f(y)|^2)(1 - |f(x)|^2)} = \frac{|y - x|^2}{(1 - |y|^2)(1 - |x|^2)} \tag{3.31}$$

whenever f belongs to $\text{Möb}(B^n)$ and the points x and y lie in B^n; if f is in $\text{Möb}(B^n)$, then

$$\|f'(x)\| = \frac{1 - |f(x)|^2}{1 - |x|^2} \tag{3.32}$$

for all x in B^n. The relation (3.31) is preserved under composition and holds trivially if f is in $\text{O}(n)$. To complete the verification of this fact, we must check that it holds if $f = R_\Sigma$ for a Euclidean sphere Σ perpendicular to $\mathbb{S}^{n-1}$. Let Σ have center x_0 and radius r, and write $R = R_\Sigma$. Then $|x_0|^2 = 1 + r^2$, $R(x_0^*) = 0$, and by (3.21)

$$\begin{aligned} |R(x)| &= |R(x) - R(x_0^*)| = \frac{r^2 |x - x_0^*|}{|x - x_0| \cdot |x_0^* - x_0|} = \frac{r^2 |x_0| \cdot |x - x_0^*|}{|x - x_0|(|x_0|^2 - 1)} \\ &= \frac{|x_0| \cdot |x - x_0^*|}{|x - x_0|} \end{aligned}$$

for every $x \in B^n$. Consequently,

$$\begin{aligned} 1 - |R(x)|^2 &= 1 - \frac{|x_0|^2 |x - x_0^*|^2}{|x - x_0|^2} \\ &= \frac{|x|^2 - 2\langle x, x_0 \rangle + |x_0|^2 - |x_0|^2|x|^2 + 2\langle x, x_0\rangle - 1}{|x - x_0|^2}, \end{aligned}$$

which simplifies to

$$1 - |R(x)|^2 = \frac{r^2(1 - |x|^2)}{|x - x_0|^2}.$$

When coupled with (3.21), this yields (3.31) for $f = R_\Sigma$. Identity (3.32) is derived from (3.31) in the same manner that (3.26) was derived from (3.23):

$$\|f'(x)\| = \lim_{y \to x} \frac{|f(y) - f(x)|}{|y - x|} = \frac{1 - |f(x)|^2}{1 - |x|^2}.$$

Just as every f in Möb(n) has its Poincaré extension $\tilde{f}$ in Möb($\mathbb{H}^{n+1}$), so each f in the conformal group Con(n) can be extended in a unique way to a Möbius transformation $\tilde{f}$ (also called the Poincaré extension of f) belonging to Möb(B^{n+1}). This follows from the fact that we know a transformation Φ in Möb(n+1) that maps $\mathbb{H}^{n+1}$ to B^{n+1} and agrees with stereographic projection on $\hat{\mathbb{R}}^n$. Thus Möb(B^{n+1}) = $\Phi \circ$ Möb($\mathbb{H}^{n+1}$) $\circ\, \Phi^{-1}$ and Con(n) = $\Phi \circ$ Möb(n) $\circ\, \Phi^{-1}$.

3.4. Hyperbolic geometry

In this section we derive some elementary aspects of hyperbolic geometry which are not only interesting in and of themselves, but which will prove useful for later applications. Further, the famous Mostow rigidity theorem we shall establish in Chapter 9 concerns itself with deformations of certain groups of isometries of hyperbolic space, and so we will need to know not only what these are, but will need to have some familiarity with hyperbolic geometry. There are many books which explore this subject in great depth, and we would recommend the monograph of J. G. Ratcliffe [**138**] at least as a starting point.

Here we start with a fairly general construction of metrics on domains in Euclidean space.

3.4.1. Conformally Euclidean metrics. Let D be a domain in $\mathbb{R}^n$. There is a standard procedure for modifying the Euclidean geometry of D in such a way that the interpretations of length, area, volume, and similar quantities undergo modification, while the measurement of angles is unaffected. These are the conformally Euclidean metrics.

We start by choosing from the class $C^\infty(D)$ a positive function ρ, termed in this context as a *metric density* in D.

Given a piecewise smooth path $\gamma : [a, b] \to D$ (γ is piecewise C^1, say), one defines its *ρ-length* $\ell_\rho(\gamma)$ by

$$\ell_\rho(\gamma) = \int_\gamma \rho(x)\,|dx| = \int_a^b \rho\big(\gamma(t)\big) \cdot |\dot{\gamma}(t)|\,dt. \tag{3.33}$$

The *ρ-distance* $d_\rho(x, y)$ between points x and y of D is then defined by the rule

$$d_\rho(x, y) = \inf_\gamma \ell_\rho(\gamma), \tag{3.34}$$

where the infimum is taken over the class of piecewise smooth paths γ in D with initial point x and endpoint y.

Again we point out that, at least in the locally Euclidean setting, we need only γ to be locally rectifiable so that (3.33) is defined (possibly equal to $+\infty$) and we could take the infimum over such paths. It is easily seen that these routes all lead to the same distance function for a given metric density ρ.

When $d_\rho : D \times D \to [0, \infty)$ is defined in this way, the pair (D, d_ρ) becomes a metric space. The only possible issue is with the triangle inequality, but this follows easily by joining paths in the obvious manner.

In fact, d_ρ is the distance function associated with the Riemannian metric in D whose fundamental form is

$$ds^2 = \rho^2(x)(dx_1^2 + dx_2^2 + \cdots + dx_n^2). \tag{3.35}$$

For instance, the choice $\rho \equiv 1$ gives rise to the *relative Euclidean metric* in D, which may differ radically from the restriction to D of the ordinary Euclidean metric unless the domain D is convex, since our infimum is taken only over paths lying in D, and the Euclidean line segment between two points of D may not.

In general, metrics of the type under discussion here are said to be *conformally Euclidean metrics* since if u and v are nonzero vectors based at a point x of D, then the angle between u and v in the Riemannian geometry of D corresponding to such a metric happens to be the same as the Euclidean angle between u and v. This follows from the form of the fundamental form at (3.35)—the inner product on the tangent space at x, where u and v lie, is a scalar multiple (in fact $\rho(x)$) times the Euclidean inner product. Thus the angles are identical.

The reader should note that in fact smoothness of the metric density is not a key feature in much of our discussion and the definitions of ℓ_ρ and d_ρ continue to make perfectly good sense if our insistence that the density ρ belong to $C^\infty(D)$ is relaxed to the demand that ρ be merely a positive continuous function in D. The pair (D, d_ρ) is still a metric space, albeit one possibly without the full array of geometric structure that would normally accompany a Riemannian metric in D. However many of these geometric structures—such as the existence of geodesics—remain with mild growth assumptions on ρ near ∂D.

The continuity of ρ allows the conclusion that

$$\lim_{y \to x} \frac{d_\rho(x, y)}{|x - y|} = \rho(x) \tag{3.36}$$

for every x in D. From (3.36) it is a short step to the observation that a sequence $\langle x_\nu \rangle$ in D converges to a point x of D in the Euclidean metric if and only if $d_\rho(x_\nu, x) \to 0$ as $\nu \to \infty$. In other words, the metric topology in D associated with d_ρ agrees with the topology induced in D by the standard topology of $\mathbb{R}^n$, independent of ρ.

Assume that D is a proper subdomain of $\mathbb{R}^n$. We give an example of the foregoing construction which will later play a very important role in the theory. We take as a metric density in D the function

$$\rho(x) = \frac{1}{\mathrm{dist}(x, \partial D)},$$

which is continuous but not usually differentiable in D. The metric d_ρ that arises from this choice of density is known as the *quasi-hyperbolic metric in D* and is usually denoted by k_D. We will next see the motivation for the terminology "quasi-hyperbolic". It can be shown that k_D is a complete metric—every quasi-hyperbolic Cauchy sequence in D converges to some point of D—a property definitely not shared by the Euclidean metric in D, be it the restricted metric or the relative one. Metrics such as this, and which reflect the geometry of the domain in question, are very useful tools for studies in conformal geometry. The quasi-hyperbolic metric was introduced in [**52**] where they further showed the existence of geodesics for this metric. It has since proven to be a useful tool.

3.4.2. Hyperbolic spaces. Let $n \geq 2$. The function

$$\rho(x) = \frac{1}{x_n} \tag{3.37}$$

is a metric density of class C^∞ in the upper half-space $\mathbb{H}^n$. With ρ is associated a conformally Euclidean metric d_ρ in $\mathbb{H}^n$, a metric known as the *hyperbolic metric* or *Poincaré metric*. The metric space $(\mathbb{H}^n, d_\rho)$ is called the *Poincaré half-space model for n-dimensional hyperbolic space* $\mathbb{H}^n$, the name deriving from the fact that the Riemannian geometry of $\mathbb{H}^n$ given by ρ is an n-dimensional non-Euclidean hyperbolic geometry and, for instance, the sum of the vertex angles of any geodesic triangle in this geometry is less than π.

In fact, to phrase things in the standard terminology of differential geometry, hyperbolic n-space $\mathbb{H}^n$ is the unique—up to isometry—complete, simply connected, n-dimensional Riemannian manifold having constant sectional curvature equal to -1. This abstract space has a number of concrete realizations, of which $\mathbb{H}^n$ endowed with the Poincaré metric is one such.

In this setting we switch to the notation $\ell_\mathbb{H}$ and $d_\mathbb{H}$ in place of the generic ℓ_ρ and d_ρ—we do not include the dimension $n \geq 2$, which will always be understood from the circumstances, to avoid clumsiness. The set $\hat{\mathbb{R}}^{n-1} = \partial\mathbb{H}^n$ also acquires a special name here; it is designated the *sphere at infinity* as one easily sees that the distance from a finite point x (an element of $\mathbb{H}^n$) to the boundary $\partial\mathbb{H}^n$ is infinite, by which we mean

$$\lim_{\mathbb{H}^n \ni y \to \partial\mathbb{H}^n} d_\mathbb{H}(x, y) = +\infty.$$

This fact is an easy consequence of the distance formula (3.39) established below.

Suppose f belongs to the group Möb$(\mathbb{H}^n)$. If $\gamma : [a, b] \to \mathbb{H}^n$ is a piecewise smooth path, then the path $\beta = f \circ \gamma$ is piecewise smooth as well and, because of the conformality of f, satisfies

$$|\dot\beta(t)| = |f'\big[\gamma(t)\big]\dot\gamma(t)| = \|f'\big[\gamma(t)\big]\| \cdot |\dot\gamma(t)|$$

whenever t is a point at which γ is differentiable. Recalling equation (3.26), we compute that

$$\begin{aligned} \ell_\mathbb{H}(\beta) &= \int_\beta \frac{|dx|}{x_n} = \int_a^b \frac{|\dot\beta(t)|\,dt}{\beta_n(t)} = \int_a^b \frac{\|f'[\gamma(t)]\| \cdot |\dot\gamma(t)|\,dt}{f_n[\gamma(t)]} \\ &= \int_a^b \frac{|\dot\gamma(t)|\,dt}{\gamma_n(t)} = \int_\gamma \frac{|dx|}{x_n} = \ell_\mathbb{H}(\gamma)\,. \end{aligned}$$

We thus recognize that the hyperbolic path length is invariant under f. This, in turn, makes certain that

$$d_{\mathbb{H}}\big[f(x), f(y)\big] = d_{\mathbb{H}}(x, y)$$

for all points x and y of $\mathbb{H}^n$ whenever f belongs to Möb($\mathbb{H}^n$). Stated differently, the mappings in Möb($\mathbb{H}^n$) are hyperbolic isometries. It is a fact that the group Möb($\mathbb{H}^n$) includes every hyperbolic isometry of $\mathbb{H}^n$, as we shall subsequently demonstrate.

Let r and s satisfy $0 < r < s < \infty$ and let $\gamma : [a, b] \to \mathbb{H}^n$ be a piecewise smooth path with $\gamma(a) = re_n$ and $\gamma(b) = se_n$. Then

$$\begin{aligned} \ell_{\mathbb{H}}(\gamma) &= \int_a^b \frac{|\dot\gamma(t)|\,dt}{\gamma_n(t)} \geq \int_a^b \frac{|\dot\gamma_n(t)|\,dt}{\gamma_n(t)} \geq \left|\int_a^b \frac{\dot\gamma_n(t)\,dt}{\gamma_n(t)}\right| \\ &= |\log\gamma_n(b) - \log\gamma_n(a)| = \log\frac{s}{r}\,. \end{aligned}$$

Moreover, $\ell_{\mathbb{H}}(\gamma) > \log(s/r)$ will definitely prevail unless two conditions are met. The first is that $\dot\gamma_1 = \dot\gamma_2 = \cdots = \dot\gamma_{n-1} = 0$ on $[a, b]$, which implies that $\gamma_1 = \gamma_2 = \cdots = \gamma_{n-1} = 0$ (remember that $\gamma_1(a) = \gamma_2(a) = \cdots = \gamma_{n-1}(a) = 0$). The second is that, once allowances are made for a finite number of points t at which $\dot\gamma_n(t)$ might fail to exist, $\dot\gamma_n$ is nonnegative on $[a, b]$, making γ_n a nondecreasing function on this interval. On the other hand, taking $\gamma(t) = te_n$ for $r \leq t \leq s$ we find that $\ell_{\mathbb{H}}(\gamma) = \log(s/r)$. The upshot of this discussion is that

$$d_{\mathbb{H}}(re_n, se_n) = \left|\log\frac{s}{r}\right| \tag{3.38}$$

for r and s in $(0, \infty)$ and that any piecewise smooth path γ between re_n and se_n with a chance of minimizing hyperbolic length (i.e., satisfying $\ell_{\mathbb{H}}(\gamma) = d_{\mathbb{H}}(re_n, se_n)$) must have the Euclidean line segment with endpoints re_n and se_n as its trajectory.

Given points x and y in $\mathbb{H}^n$, one can easily produce a transformation f in Möb($\mathbb{H}^n$) such that $f(x) = e_n$ and $f(y) = re_n$ with $r \geq 1$. In conjunction with (3.38), this comment leads to a nice formula for $d_{\mathbb{H}}(x, y)$:

$$\cosh d_{\mathbb{H}}(x, y) = 1 + \frac{|x-y|^2}{2x_n y_n}\,. \tag{3.39}$$

Indeed, (3.38) proves that (3.39) holds when e_n and re_n are substituted for x and y, respectively. However, both sides of (3.39) are invariant under Möb($\mathbb{H}^n$)—don't forget (3.23)—so the validity of (3.39) persists for $x = f^{-1}(e_n)$ and $y = f^{-1}(re_n)$.

Let us employ (3.39), for example, to compute $d_{\mathbb{H}}(x, |x|e_n)$ for x in $\mathbb{H}^n$. We claim that

$$d_{\mathbb{H}}(x, |x|e_n) = \log\big[\cot(\varphi/2)\big]\,, \tag{3.40}$$

in which $\varphi = (\pi/2) - \theta(x, e_n)$. We assume that x is not a scalar multiple of e_n, for (3.40) holds trivially otherwise. Since both sides of (3.40) go unchanged when x is replaced by λx with $\lambda > 0$, we may further suppose that $|x| = 1$. Write $x = ae_n + bu$ with $a > 0$, $b > 0$, and u a unit vector in $\mathbb{R}^{n-1}$. Considering that $a^2 + b^2 = |x|^2 = 1$, we deduce from (3.39) that

$$\begin{aligned}\cosh d_{\mathbb{H}}(x,e_n) &= 1+\frac{|(a-1)e_n+bu|^2}{2a}=1+\frac{(a-1)^2+b^2}{2a}\\ &= \frac{1+a^2+b^2}{2a}=\frac{1}{a}=\csc\varphi\,.\end{aligned}$$

Solving for $d_{\mathbb{H}}(x,e_n)$ gives (3.40) when $|x|=1$.

Formula (3.39) has a number of other noteworthy implications. One is that the hyperbolic sphere of radius r centered at a point x of $\mathbb{H}^n$, by which we mean the set

$$S^{n-1}_{\mathbb{H}}(x,r)=\{y\in H^n : d_{\mathbb{H}}(y,x)=r\},$$

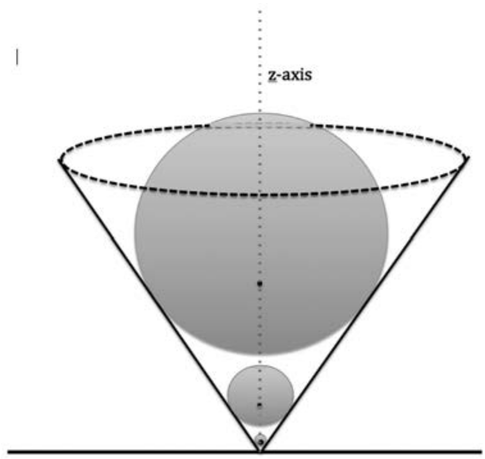
is actually a Euclidean sphere. In fact, a brief calculation shows us that

$$S^{n-1}_{\mathbb{H}}(x,r)=\mathbb{S}^{n-1}(\tilde{x},s)$$

with

$$\tilde{x}=(x_1,x_2,\ldots,x_{n-1},x_n\cosh r)\quad\text{and}\quad s=x_n\sinh r.$$

Thus the families of closed hyperbolic balls in $\mathbb{H}^n$ and closed Euclidean balls in $\mathbb{H}^n$ coincide, although the hyperbolic center and hyperbolic radius of any such ball may differ from their Euclidean counterparts.



Hyperbolic balls of the same radius in $\mathbb{H}^3$. The cone consists of points a fixed distance from the z-axis.

A subset L of $\mathbb{H}^n$ is called a *hyperbolic line* (an $\mathbb{H}$-*line*, for short) if L can be written as $L=\Gamma\cap\mathbb{H}^n$, where Γ is a chordal circle in $\hat{\mathbb{R}}^n$ that is orthogonal to $\mathbb{R}^{n-1}$.

Through any pair of distinct points x and y of $\mathbb{H}^n$ there passes one and only one hyperbolic line L; the subarc of L with endpoints x and y is called the *hyperbolic segment* (or just the $\mathbb{H}$-*segment*) between x and y. A hyperbolic line will have two endpoints on the sphere at infinity; these are not, however, points of hyperbolic space.

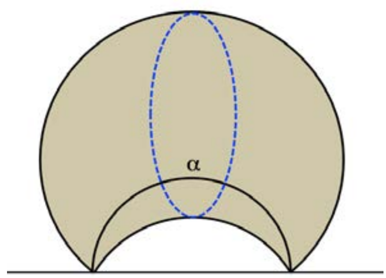
It is an easy exercise to verify that the group Möb($\mathbb{H}^n$) acts transitively on the family of all hyperbolic lines in $\mathbb{H}^n$—although this is more easily seen in other models of hyperbolic space. Thus remarks in the preceding paragraphs quickly bring one to the realization that any (piecewise smooth) path from x to y whose hyperbolic length is minimal among all paths joining x to y must have the $\mathbb{H}$-segment between these points as its trajectory. For this reason we often speak of $\mathbb{H}$-segments as *hyperbolic geodesic segments* and of $\mathbb{H}$-lines as *hyperbolic geodesics*. The *standard* (or *hyperbolic arclength*) *parametrization* γ_0 of the $\mathbb{H}$-segment between x and y, described in the direction from x to y, is given as follows:

$$\gamma_0(t) = f^{-1}(e^t e_n), \qquad 0 \le t \le \log r,$$

where f is any member of $\text{Möb}(\mathbb{H}^n)$ that satisfies $f(x) = e_n$ and $f(y) = re_n$ with $r > 1$. Notice that a point z of $\mathbb{H}^n$ lies on this $\mathbb{H}$-segment if and only if $d_{\mathbb{H}}(x, z) + d_{\mathbb{H}}(z, y) = d_{\mathbb{H}}(x, y)$.

With regard to angles in hyperbolic geometry, we reiterate something that was pointed out earlier in reference to angles in arbitrary conformally Euclidean geometries; namely, that the hyperbolic angle determined at a point x of $\mathbb{H}^n$ by two hyperbolic rays issuing from x is just the Euclidean angle between these rays, which is to say the Euclidean angle between the vectors tangent to these rays at x.

Let $\bar{B}$ be a closed ball in $\mathbb{H}^n$. The set $\bar{B}$ is *hyperbolically convex* (abbreviated $\mathbb{H}$-*convex*), which means that for every pair of points x and y in B the entire $\mathbb{H}$-segment between x and y is contained in $\bar{B}$.



The collection of points a fixed distance from a hyperbolic line $\alpha \in \mathbb{H}^3$ is hyperbolically convex but not convex unless α is also a Euclidean line.

To recognize this it is enough to notice that the intersection of $\bar{B}$ with any hyperbolic line L, if nonempty, is an $\mathbb{H}$-segment or a single point. The truth of this observation is clear for $L = L_0$, the positive x_n-axis; its validity for arbitrary L is established by transforming L to L_0 with a member of $\text{Möb}(\mathbb{H}^n)$ and remembering that the mappings in this group preserve the classes of $\mathbb{H}$-lines, $\mathbb{H}$-segments, and closed balls in $\mathbb{H}^n$.

We may exploit the $\mathbb{H}$-convexity of $\bar{B}$ in deriving the bounds

$$a|x - y| \le d_{\mathbb{H}}(x, y) \le b|x - y| \tag{3.41}$$

for all x and y in $\bar{B}$, where

$$a = \min\{x_n^{-1} : x \in \bar{B}\}, \qquad b = \frac{\pi}{2}\ \max\{x_n^{-1} : x \in \bar{B}\}\,.$$

Consider a pair of points x and y of $\bar{B}$, which we assume to be distinct, and let γ_0 denote the standard parametrization of the $\mathbb{H}$-segment A between x and y.

From elementary geometry we obtain estimates for the Euclidean length $\ell(A)$ of A,

$$|x-y| \le \ell(A) \le \frac{\pi|x-y|}{2}\,.$$

Since A lies completely in the ball $\bar{B}$, we discover that

$$\begin{aligned} a|x-y| \le a\ell(A) &= a\int_{\gamma_0}|dx| \le \int_{\gamma_0}\frac{|dx|}{x_n} = d_{\mathbb{H}}(x,y) \le \frac{2b}{\pi}\int_{\gamma_0}|dx| \\ &= \frac{2b\ell(A)}{\pi} \le b|x-y|\,. \end{aligned}$$

The estimates in (3.41) are instrumental in the proof of an important fact.

THEOREM 3.4.1. *The metric space* $(\mathbb{H}^n, d_{\mathbb{H}})$ *is complete.*

PROOF. Let $\langle x_\nu\rangle$ be a hyperbolic Cauchy sequence. Then $r = \sup_\nu d_{\mathbb{H}}(x_\nu, x_1) < \infty$, so $\langle x_\nu\rangle$ is a sequence in the closed ball $\bar{B} = \{x \in \mathbb{H}^n : d_{\mathbb{H}}(x, x_1) \le r\}$. The first inequality in (3.41) implies that $\langle x_\nu\rangle$ is a Euclidean Cauchy sequence, hence that it converges in the Euclidean metric to some point x_0 of the closed set $\bar{B}$. The second half of (3.41) then informs us that $d_{\mathbb{H}}(x_\nu, x_0) \to 0$. Thus $\langle x_\nu\rangle$ converges to a finite point x_0 in the hyperbolic metric. □

We make note of the fact that, for fixed x in $\mathbb{H}^n$, $d_{\mathbb{H}}(y,x) \to \infty$ as $y \to \partial\mathbb{H}^n$ (i.e., as $q(y, \partial\mathbb{H}^n) \to 0$). Otherwise there would be a sequence $\langle y_\nu\rangle$ in $\mathbb{H}^n$ such that $y_\nu \to \partial\mathbb{H}^n$, yet $r = \sup_\nu d_{\mathbb{H}}(y_\nu, x) < \infty$. But the closed ball $\bar{B} = \{y \in \mathbb{H}^n : d_{\mathbb{H}}(y,x) \le r\}$ is a compact set and is disjoint from $\partial\mathbb{H}^n$, implying that $\inf_\nu q(y_\nu, \partial\mathbb{H}^n) \ge q(\bar{B}, \partial\mathbb{H}^n) > 0$ and thereby contradicting the assumption that $q(y_\nu, \partial\mathbb{H}^n) \to 0$.

As indicated earlier, there are a number of alternative models for the hyperbolic space $\mathbb{H}^n$, one of which is the *Poincaré ball model*. It is provided by the unit ball B^n when it is equipped with the conformally Euclidean metric that corresponds to the density

$$\rho(x) = \frac{2}{1-|x|^2}. \tag{3.42}$$

We continue to write $\ell_{\mathbb{H}}$ and $d_{\mathbb{H}}$ for ρ-length and ρ-distance here, letting the context determine whether we are working in $\mathbb{H}^n$ or B^n.

As a matter of fact, most of what has been said about $(\mathbb{H}^n, d_{\mathbb{H}})$ can be transferred without effort to the setting of $(B^n, d_{\mathbb{H}})$ simply by using the mapping Φ that has aided us on several previous occasions: $\Phi = R \circ S$, where S is the reflection in $\mathbb{R}^{n-1}$ and R is the inversion in $\mathbb{S}^{n-1}(e_n, \sqrt{2})$. We know that Φ belongs to $\text{Möb}^+(n)$ and has $\Phi(\mathbb{H}^n) = B^n$. Taking heed of the fact that $R(-e_n) = 0$ and making reference to (3.21) and (3.10), we realize that for x different from e_n and ∞ it is true that

$$|R(x)|^2 = |R(x) - R(-e_n)|^2 = \frac{|x+e_n|^2}{|x-e_n|^2} = 1 + \frac{4x_n}{|x-e_n|^2}$$

and

$$\|R'(x)\| = \frac{2}{|x-e_n|^2}\,.$$

Accordingly, for any x in $\mathbb{H}^n$ we get

$$1-|\Phi(x)|^2 = 1-|R\big[S(x)\big]|^2 = \frac{4x_n}{|S(x)-e_n|^2} \tag{3.43}$$

and, by an appeal to the conformality of S in $\mathbb{H}^n$ and of R in $S(\mathbb{H}^n)$,

$$\|\Phi'(x)\| = \|R'\big[S(x)\big]\|\cdot\|S'(x)\| = \frac{2}{|S(x)-e_n|^2}\,. \tag{3.44}$$

From (3.43) and (3.44) it follows that

$$\frac{2\|\Phi'(x)\|}{1-|\Phi(x)|^2} = \frac{1}{x_n} \tag{3.45}$$

whenever x lies in $\mathbb{H}^n$. If $\gamma:[a,b]\to\mathbb{H}^n$ is a piecewise smooth path and $\beta=\Phi\circ\gamma$, then (3.45) now yields

$$\begin{aligned}\ell_{\mathbb{H}}(\beta) &= \int_\beta \frac{2|dx|}{1-|x|^2} = \int_a^b \frac{2|\dot\beta(t)|\,dt}{1-|\beta(t)|^2} = \int_a^b \frac{2\|\Phi'[\gamma(t)]\|\cdot|\dot\gamma(t)|\,dt}{1-|\Phi[\gamma(t)]|^2}\\ &= \int_a^b \frac{|\dot\gamma(t)|\,dt}{\gamma_n(t)} = \int_\gamma \frac{|dx|}{x_n} = \ell_{\mathbb{H}}(\gamma)\ ,\end{aligned}$$

which implies that

$$d_{\mathbb{H}}\big(\Phi(x),\Phi(y)\big) = d_{\mathbb{H}}(x,y) \tag{3.46}$$

for all points x and y of $\mathbb{H}^n$. Equation (3.46) therefore expresses the fact that Φ is an isometry between the metric spaces $(\mathbb{H}^n, d_{\mathbb{H}})$ and $(B^n, d_{\mathbb{H}})$.

Once in possession of this knowledge we can painlessly translate statements about the Poincaré half-space to statements about the Poincaré ball, and vice versa. For instance the transformations in $\text{Möb}(B^n) = \Phi\circ\text{Möb}(\mathbb{H}^n)\circ\Phi^{-1}$ are hyperbolic isometries of B^n; the hyperbolic geodesics in B^n are the sets L of the form $L=\Gamma\cap B^n$, where Γ is a chordal circle in $\hat{\mathbb{R}}^n$ that meets $\mathbb{S}^{n-1}$ at right angles, and so forth.

It will later be important to know that

$$d_{\mathbb{H}}(0,x) = \log\frac{1+|x|}{1-|x|} \tag{3.47}$$

for every x in B^n. Since for any x in B^n there is an orthogonal linear transformation U with $U(x)=|x|e_n$, since the members of $\mathrm{O}(n)$ are simultaneously hyperbolic and Euclidean isometries of B^n, and since Φ maps the ray $\{re_n : r\geq 1\}$ to the radial segment of B^n from 0 to e_n, it suffices to prove (3.47) for $x=\Phi(re_n)$ with $r>1$.

Formula (3.38) gives

$$d_{\mathbb{H}}(0,x) = \log r\ ,$$

while (3.43) tells us that

$$1-|x|^2 = \frac{4r}{(r+1)^2}\,.$$

This equation can be solved for r to get

$$r = \frac{1+|x|}{1-|x|}\ ,$$

from which (3.47) directly follows.

We next show that $\mathrm{M\ddot{o}b}(B^n)$ and $\mathrm{M\ddot{o}b}(\mathbb{H}^n)$ are the full isometry groups of the metric spaces $(B^n, d_{\mathbb{H}})$ and $(\mathbb{H}^n, d_{\mathbb{H}})$. We prepare the way with two lemmas. The first characterizes local Euclidean isometries.

LEMMA 3.4.2. *Let B be an open Euclidean ball in $\mathbb{R}^n$, and let $f : B \to \mathbb{R}^n$ be an isometric embedding, meaning that $|f(x) - f(y)| = |x - y|$ for all x and y in B. Then f is the restriction to B of a unique transformation g from the group $\boldsymbol{E}(n)$ of Euclidean isometries.*

PROOF. Suppose $B = B^n(x_0, r)$. We may assume that $x_0 = f(x_0) = 0$. (If not, consider in place of f the mapping f_0 defined in $B^n(r)$ by $f_0(x) = f(x+x_0) - f(x_0)$. If true for f_0, the lemma is also true for f.) For any x and y in B we have $\langle f(x), f(y)\rangle = \langle x, y\rangle$. Indeed,

$$\begin{aligned} \langle f(x), f(y)\rangle &= \frac{|f(x)|^2 + |f(y)|^2 - |f(x) - f(y)|^2}{2} = \frac{|x|^2 + |y|^2 - |x-y|^2}{2} \\ &= \langle x, y\rangle \,. \end{aligned}$$

Next, fixing a number s in $(0, r)$, we observe that the vectors $u_1 = s^{-1}f(se_1)$, $u_2 = s^{-1}f(se_2), \ldots, u_n = s^{-1}f(se_n)$ form an orthonormal basis for $\mathbb{R}^n$. Thus we discover that

$$f(x) = \sum_{i=1}^{n} \langle f(x), u_i\rangle u_i = s^{-1}\sum_{i=1}^{n}\langle f(x), f(se_i)\rangle u_i = s^{-1}\sum \langle x, se_i\rangle u_i = \sum_{i=1}^{n} x_i u_i$$

for every x in B. Now the function $g : \mathbb{R}^n \to \mathbb{R}^n$ defined by $g(x) = \sum_{i=1}^n x_i u_i$ is an orthogonal linear transformation, so $f = g|B$ with g a member of $\boldsymbol{E}(n)$. The uniqueness of g is a consequence of Corollary 3.3.11. ☐

The second lemma recasts Lemma 3.4.2 in its "infinitesimal" formulation.

LEMMA 3.4.3. *Let B be an open Euclidean ball in $\mathbb{R}^n$, and let $f : B \to \mathbb{R}^n$ be an injective mapping with the property that*

$$\lim_{y\to x} \frac{|f(y) - f(x)|}{|y - x|} = 1 \tag{3.48}$$

for every point x of B. Then f is an isometric embedding of B into $\mathbb{R}^n$—hence, is the restriction to B of a unique Euclidean isometry $g \in \boldsymbol{E}(n)$. In particular, $B' = f(B)$ is an open Euclidean ball. If both B and B' are centered at the origin, then $B = B'$ and g is an orthogonal linear transformation.

PROOF. As in the proof of Lemma 3.4.2, it suffices to consider the situation where $B = B^n(0, r)$ and $f(0) = 0$, in which case the claim is that $f(B) = B$ and that $f = g|B$, with g an orthogonal transformation of $\mathrm{O}(n)$. Condition (3.48) plainly implies that f is a continuous function. Thus f maps B homeomorphically onto a domain that contains the origin, a domain we label D. Also, it is an easy matter to check that (3.48) is true with f^{-1} in place of f at each point of D. The first step in the proof of the lemma is to demonstrate that $|f(y) - f(x)| \le |y - x|$ holds for all x and y in B. Fix x and y in B, with $x \ne y$, and $m > 1$. We proceed to verify that

$$|f(y) - f(x)| \le m|y - x| \,. \tag{3.49}$$

However, as $m > 1$ is arbitrary, this enables us to conclude that $|f(y) - f(x)| \le |y - x|$. Write $u = y - x$ and let $A = \{t \in [0, 1] : |f(x + tu) - f(x)| \le mt|u|\}$.

Then $0 \in A$, so this set is nonempty, $A \neq \emptyset$. By the continuity of f, $t_0 = \sup A$ belongs to A. Suppose that t_0 were smaller than 1. On the basis of 3.48 we would have

$$\lim_{t \to t_0^+} \frac{|f(x+tu) - f(x+t_0u)|}{(t-t_0)|u|} = 1 < m \,,$$

permitting us to pick t in the interval $(t_0, 1]$ for which

$$|f(x+tu) - f(x+t_0u)| < m(t-t_0)|u| \,.$$

The triangle equality would then give

$$|f(x+tu) - f(x)| \leq mt|u|$$

and would thus place t in A, in contradiction with the definition of t_0. To avoid a contradiction we must have $t_0 = 1$. Therefore $|f(y) - f(x)| \leq m|y-x|$, which proves (3.49). We can therefore be certain that $|f(y) - f(x)| \leq |y-x|$ is obtained for all x and y in B. One implication of this inequality is that B contains the domain D.

Now let $B' = B^n(0,d)$, where $d = d(0, \partial D)$. We can apply the foregoing argument to the function f^{-1} in the ball B'. It leads to the conclusion that $|f^{-1}(y) - f^{-1}(x)| \leq |y-x|$ whenever x and y lie in B'. On the other hand, the statement

$$|y-x| = \left|f\left[f^{-1}(y)\right] - f\left[f^{-1}(x)\right]\right| \leq |f^{-1}(y) - f^{-1}(x)|$$

is also valid for such x and y, showing that the restriction of f^{-1} to B' is an isometric imbedding. Lemma 3.4.2 tells us that f^{-1} coincides in B' with a unique mapping h from $\boldsymbol{E}(n)$. Indeed, since $h(0) = f^{-1}(0) = 0$, h is an orthogonal linear transformation. If y is a point of ∂D with $|y| = d$, then $f^{-1}(x)$ tends to ∂B as $x \to y$ through B', whence

$$d = \lim_{x \to y} |x| = \lim_{x \to y} |f^{-1}(x)| = r \,.$$

From this we infer that $B' = B = D$ and that f is the restriction to B of $g = h^{-1}$, as maintained. ☐

We are now in a position to establish the result we have really been after.

THEOREM 3.4.4. *The isometry group of* $(B^n, d_{\mathbb{H}})$ *is* $\text{Möb}(B^n)$*; the isometry group of* $(\mathbb{H}^n, d_{\mathbb{H}})$ *is* $\text{Möb}(\mathbb{H}^n)$.

PROOF. We have already shown that the members of $\text{Möb}(B^n)$ and $\text{Möb}(\mathbb{H}^n)$ are hyperbolic isometries. Let $f : B^n \to B^n$ be an arbitrary isometry of $(B^n, d_{\mathbb{H}})$. We maintain that $f = g|B^n$ for some g in $\text{Möb}(B^n)$. In proving this we are free to assume that $f(0) = 0$. (Otherwise choose h in $\text{Möb}(B^n)$ that sends $f(0)$ to 0 and consider $h \circ f$ instead of f.) Consider a point x of B^n. Then $d_{\mathbb{H}}[0, f(x)] = d_{\mathbb{H}}(0,x)$, so (3.47) implies that $|f(x)| = |x|$. Recalling (3.36), we find that

$$\begin{aligned} \lim_{y \to x} \frac{|f(y) - f(x)|}{|y-x|} &= \lim_{y \to x} \frac{|f(y) - f(x)|/d_{\mathbb{H}}[f(y), f(x)]}{|y-x|/d_{\mathbb{H}}(y,x)} \\ &= \frac{(1-|f(x)|^2)/2}{(1-|x|^2)/2} = \frac{1-|x|^2}{1-|x|^2} = 1 \end{aligned}$$

for every x in B^n. Lemma 3.4.3 informs us that $f = g|B^n$ for some g in $\text{O}(n)$. Finally, if f is a general hyperbolic isometry of $\mathbb{H}^n$, then $\Phi \circ f \circ \Phi^{-1}$ is a hyperbolic isometry of B^n. It follows that there is a transformation g in Möb(B^n) such that $\Phi \circ f \circ \Phi^{-1} = g|B^n$. This makes f the restriction to $\mathbb{H}^n$ of $\Phi^{-1} \circ g \circ \Phi$, which belongs to Möb($\mathbb{H}^n$). □

3.5. Classification of hyperbolic isometries

Let $n \geq 2$ and let $f \in$ Möb(B^n) be a Möbius transformation different from the identity. The Brouwer fixed point theorem implies the existence of at least one fixed point for f in $\bar{B}^n$, though f need not fix any point of the open ball B^n.

3.5.1. Elliptic transformations. If f does have a fixed point in B^n, then we say that f is an *elliptic transformation*. In this case we can select an arbitrary point x_0 of B^n fixed by f and choose a transformation g in Möb(B^n) with $g(0) = x_0$. The mapping $f_0 = g^{-1} \circ f \circ g$ belongs to Möb(B^n) and fixes the origin, so f_0 is in $\mathrm{O}(n)$. The set $\mathrm{fix}(f_0)$ defined by

$$\mathrm{fix}(f_0) = \{x \in \hat{\mathbb{R}}^n : f_0(x) = x\}$$

is then of the form $T \cup \{\infty\}$, where T is a linear subspace of $\mathbb{R}^n$. Moreover, $\mathrm{fix}(f) = g[\mathrm{fix}(f_0)]$. On the strength of these comments we can assert the following about f, should it be elliptic: f is conjugate in Möb(B^n) to an orthogonal transformation; $\mathrm{fix}(f)$ is either a two-point set or a chordal p-sphere ($1 \leq p \leq n-1$) that is orthogonal to $\mathbb{S}^{n-1}$; f leaves invariant any hyperbolic sphere or ball in B^n whose hyperbolic center is fixed by f.

3.5.2. Loxodromic and parabolic transformations. Looking next at a transformation f from Möb(B^n) that is free of fixed points in B^n, we remark that $\mathrm{fix}(f)$ must be a subset of $\mathbb{S}^{n-1}$. If f were to fix a point x of $\hat{\mathbb{R}}^n \setminus \bar{B}^n$, then by Theorem 3.3.7 it would also fix x^*, the point symmetric with x relative to $\mathbb{S}^{n-1}$, which is a point of B^n. Moreover, we shall see in a moment that $\mathrm{fix}(f)$ can contain no more than two elements. A transformation from Möb(B^n) with exactly two fixed points on $\mathbb{S}^{n-1}$ and none in B^n is termed *loxodromic*; a member of Möb(B^n) whose one and only fixed point lies on $\mathbb{S}^{n-1}$ is said to be *parabolic*.

Our proposed classification of the nonidentity members of Möb(B^n) with $n \geq 2$ into elliptic, loxodromic, and parabolic transformations has an obvious parallel in Möb($\mathbb{H}^n$). The following fact makes certain that in both cases the classification is exhaustive: if f in Möb($\mathbb{H}^n$) fixes three points of ∂H^n, then f must have a fixed point in $\mathbb{H}^n$. Indeed, by passing from f to one of its conjugates in Möb($\mathbb{H}^n$) we may suppose that f fixes 0, e_1, and ∞.

According to Theorem 3.3.13, we can write $f = \lambda U + b$, where $\lambda > 0$, b is in $\mathbb{R}^{n-1}$, and U is a member of $\mathrm{O}(n)$ that fixes e_n. Then $b = f(0) = 0$, whence $\lambda = |\lambda U(e_1)| = |f(e_1)| = |e_1| = 1$. We conclude that $f(e_n) = U(e_n) = e_n$, thereby exhibiting a fixed point for f in $\mathbb{H}^n$. Notice that in both B^n and $\mathbb{H}^n$ the classification of hyperbolic isometries is conjugacy invariant: if f and g are members of Möb(B^n) (respectively, Möb($\mathbb{H}^n$)) and if f is not the identity transformation, then f and $g^{-1} \circ f \circ g$ fall into the same category in the classification.

Elliptic transformations are easiest to picture in the ball model of hyperbolic space, where we have already observed that they are conjugate to orthogonal linear transformations. The structures, up to conjugacy, of parabolic and loxodromic transformations become more transparent in the Poincaré half-space. To set the stage for an analysis of these transformations, we make a remark concerning the nonexistence of finite fixed points for certain Euclidean isometries.

Let f in $\boldsymbol{E}(n)$ be of the form $f = U + b$, with b a nonzero vector in $\mathbb{R}^n$ and U in $\mathrm{O}(n)$. We claim that the fixed point set $\mathrm{fix}(f) = \{\infty\}$ if and only if b does not belong to $T^\perp$, where $T = \mathrm{fix}(U)$ and $T^\perp$ indicates the orthogonal complement of the linear space T. Note that under either condition, $T \neq \{0\}$. This fact is obvious if b is not in $T^\perp$, for then it is immediate that $T^\perp \neq \mathbb{R}^n$ and hence that $T \neq \{0\}$. If $T = \{0\}$, then 1 is not an eigenvalue of U, the linear transformation $U - I$ is nonsingular, and the equation $U(x) - x = -b$, which is equivalent to $f(x) = x$, has a finite solution.

Consequently, $\mathrm{fix}(f) \neq \{\infty\}$ when $T = \{0\}$. In substantiating this claim we make use of the direct sum decomposition $\mathbb{R}^n = T \oplus T^\perp$. Thus each vector x in $\mathbb{R}^n$ has a unique representation $x = x' + x''$ with x' in T and x'' in $T^\perp$. In particular, $b = b' + b''$. As $U(x') = x'$, for x in $\mathbb{R}^n$ we can express $f(x)$ as

$$\begin{aligned} f(x) &= f(x' + x'') = U(x') + U(x'') + b' + b'' \\ &= (x' + b') + \big[U(x'') + b''\big]. \end{aligned} \tag{3.50}$$

Because U is an orthogonal transformation, $T^\perp$ is invariant under U; hence $U(x'') \in T^\perp$ and also $U(x'') + b'' \in T^\perp$. We infer from (3.50) and the uniqueness of the orthogonal decomposition that $f(x) = x$ if and only if $x' + b' = x'$ and $U(x'') + b'' = x''$. By construction, 1 is not an eigenvalue of $U\big|T^\perp$, so the equation $U(x'') + b'' = x''$ definitely has a solution x'' in $T^\perp$. In order for f to be free of fixed points in $\mathbb{R}^n$ it is therefore both necessary and sufficient that $b' \neq 0$, which is equivalent to the requirement that b not be an element of $T^\perp$.

Formula (3.50) suggests another interesting observation about a Euclidean isometry f with $\mathrm{fix}(f) = \{\infty\}$. Writing $f^0 = I$, $f^k = f \circ f \circ \cdots \circ f$ (k factors) for $k = 1, 2, \ldots$ and $f^{-k} = (f^{-1})^k$ for $k = 1, 2, \ldots$, we learn from (3.50) that $f^k(x) = x' + kb' + z_k$, where z_k is a vector from $T^\perp$. Since $b' \neq 0$, we infer that

$$|f^k(x)| = \big(|x' + kb'|^2 + |z_k|^2\big)^{1/2} \geq |x' + kb'| \geq k|b'| - |x'| \to \infty$$

as $k \to \infty$. That is,

$$f^k(x) \to \infty \qquad \text{as } k \to \infty \text{ for every } x \text{ in } \hat{\mathbb{R}}^n.$$

Because f^{-1} has the same structure as f, we see also that $f^k(x) \to \infty$ as $k \to -\infty$ for all x in $\hat{\mathbb{R}}^n$.

Consider now a parabolic transformation f from $\mathrm{Möb}(\mathbb{H}^n)$ whose sole fixed point is ∞. We know from Theorem 3.3.13 that $f = \lambda U + b$ for some $\lambda > 0$, b in $\mathbb{R}^{n-1}$, and U in $\mathrm{O}(n)$ fixing e_n. We must have $b = f(0) \neq 0$. Also, $\lambda = 1$ is required, for $\lambda \neq 1$ would mean that λ^{-1} is certainly not an eigenvalue of U and hence would make $U - \lambda^{-1}I$ an invertible linear transformation. As a result, the equation $\lambda U(x) + b = x$ would have a solution in $\mathbb{R}^n$, a situation incompatible with our assumptions about f having the unique fixed point $\{\infty\}$. We have thus shown that f has the form $f = U + b$, where U is a transformation in $\mathrm{O}(n)$ that fixes e_n and b is a vector in $\mathbb{R}^{n-1}$ that, by our preliminary remarks, does not lie in $\mathrm{fix}(U)^\perp$. Of course, every transformation of the type just described is a parabolic transformation in $\mathrm{Möb}(\mathbb{H}^n)$ that fixes ∞, while an arbitrary parabolic isometry of $(\mathbb{H}^n, d_{\mathbb{H}})$ is conjugate in $\mathrm{Möb}(\mathbb{H}^n)$ to such a transformation.

These facts bring to light two key features of a general parabolic transformation f from Möb($\mathbb{H}^n$), say one with fix(f) $= \{p\}$. First, f leaves invariant each horosphere to ∂H^n at p. Here, by a *horosphere* to a chordal sphere Σ at a point p we mean a chordal sphere Σ' such that $\Sigma \cap \Sigma' = \{p\}$. Second, $f^k(x) \to p$ as $|k| \to \infty$ for every x in $\hat{\mathbb{R}}^n$. Both of these facts are easily seen when $p = \infty$—the horospheres are the hyperplanes $\{x \in \mathbb{H}^n : x_n = c > 0\}$; the general case then follows by means of conjugation. The obvious analogues of these two properties are enjoyed by parabolic transformations belonging to Möb(B^n).

Suppose next that f is a loxodromic transformation in Möb($\mathbb{H}^n$) for which fix(f) $= \{0, \infty\}$. It follows from Theorem 3.3.13 that f can be represented in the manner $f = \lambda U$, where $\lambda > 0$, $\lambda \neq 1$, and U is an orthogonal transformation that fixes e_n. If x lies on the hyperbolic line L with endpoints 0 and ∞, we have $f(x) = \lambda x$, so L is invariant under f and the geometric effect of f on L is to translate each point a hyperbolic distance $|\log \lambda|$ along L, in the direction toward the origin if $\lambda < 1$ and away from it if $\lambda > 1$. Furthermore, for the iterate f^k we get $|f^k(x)| = \lambda^k |x|$ for all x in $\mathbb{R}^n$, which yields the information that

$$f^k(x) \to 0 \text{ and } f^{-k}(x) \to \infty \quad \text{as } k \to \infty \text{ for every } x \in \hat{\mathbb{R}}^n \setminus \{0, \infty\}$$

when $\lambda < 1$, whereas

$$f^k(x) \to \infty \text{ and } f^{-k}(x) \to 0 \quad \text{as } k \to \infty \text{ for every } x \in \hat{\mathbb{R}}^n \setminus \{0, \infty\}$$

when $\lambda > 1$.

Then through the process of conjugation we draw the following conclusions about a general loxodromic transformation f in Möb($\mathbb{H}^n$) or Möb(B^n), assuming that fix(f) $= \{p, q\}$: f leaves invariant the hyperbolic line L whose terminal points are p and q and L is called the *axis* of f; on L the transformation f acts as a hyperbolic translation; the labeling of the fixed points of f can be done so that p is its *attracting fixed point* and q its *repelling fixed point* (i.e., $f^k(x) \to p$ and $f^{-k}(x) \to q$ as $k \to \infty$ for every x in $\hat{\mathbb{R}}^n \setminus \{p, q\}$).

We remark that a transformation f from the Möbius group Möb(n) or from the conformal group Con(n) is classified as parabolic, loxodromic, or elliptic under the condition that its Poincaré extension $\tilde{f}$, which is a member of Möb($\mathbb{H}^{n+1}$) when f belongs to Möb(n) and of Möb(B^{n+1}) when f is in Con(n), is so classified. We can summarise the above discussions in a theorem

THEOREM 3.5.1. *Assume that* $f \in$ Möb(n) *is not the identity mapping of* $\hat{\mathbb{R}}^n$. *Then*

- f *is parabolic if it has a unique fixed point;*
- f *is loxodromic if its Poincaré extension has exactly two fixed points;*

in all other cases—in particular, when f *has no fixed points—*f *is elliptic.*

Of course a similar result is true for $f \in$ Con(n).

3.6. The distortion, compactness and convergence properties of Möbius transformations

In this section we will discuss the convergence properties of sequences of Möbius transformations of hyperbolic n-space $\mathbb{H}^n$, the unit ball B^n and the n-sphere $\mathbb{S}^n$. Given the close relationships between these spaces, the compactness properties will also be similar. Further, it is to be expected that the results we achieve should be very similar to those well known in the theory of complex analysis and in particular the classical theory of normal families—a term coined by P. Montel in 1912 [**139**, p. 154]—and the key idea of equicontinuity. The theory of normal families—well covered in J. Schiff's book [**145**] on this topic—is important because it is closely connected with a standard proof of the Riemann mapping theorem, and with the circle of ideas surrounding basic theorems of Picard, Schottky, Landau and Bloch.

We will see a little later that the compactness results we achieve for families of Möbius transformations hold in much greater generality—in fact for families of quasiconformal mappings—and so the results obtained here foreshadow those.

The situation for families of Möbius transformations of $\mathbb{H}^n$ or B^n is a bit easier since we have already shown that these mappings act as isometries of these spaces when given their respective hyperbolic metrics. From this one can in fact easily deduce compactness results for sequences in $\mathrm{Con}(n)$, however it is possible to achieve interesting and useful distortion estimates which frame more general discussions later.

3.6.1. Distortion of chordal distances. With the aid of hyperbolic geometry one can give sharp bounds for the distortion of chordal distances under a Möbius transformation. These bounds stem from the following lemma relating spherical distortion and hyperbolic geometry.

LEMMA 3.6.1. *If f belongs to* $\mathrm{M\ddot{o}b}(B^n)$ *with $n \geq 2$, then*

$$\sup\left\{\frac{|f(y)-f(x)|}{|y-x|} : x, y \in \mathbb{S}^{n-1}\ ,\ x \neq y\right\} = e^{d_{\mathbb{H}}[0,f(0)]}\,. \tag{3.51}$$

PROOF. If f is in $\mathrm{O}(n)$, then (3.51) is a trivial statement as $|f(y)-f(x)| = |y-x|$ and $f(0) = 0$. Consider next the inversion R in a Euclidean sphere $\mathbb{S}^{n-1}(x_0, r)$ that is orthogonal to $\mathbb{S}^{n-1}$. Once we recall (3.29) and (3.21) we have $1 + r^2 = |x_0|^2$ and

$$\frac{|R(x)-R(y)|}{|x-y|} = \frac{r^2}{|x-x_0|\cdot|y-x_0|}$$

for x and y on $\mathbb{S}^{n-1}$ with $x \neq y$. The right-hand side of this equation attains its maximum on $\mathbb{S}^{n-1}$ when $x = y = x_0/|x_0|$, in which case $|x-x_0| = |y-x_0| = |x_0| - 1$. Remembering that $R(0) = x_0^*$, the image of x_0 under reflection in $\mathbb{S}^{n-1}$, and appealing to formula (3.47), we compute that

$$\begin{aligned}\sup\left\{\frac{|R(x)-R(y)|}{|x-y|} : x, y \in S^{n-1}\ ,\ x \neq y\right\} &= \frac{r^2}{(|x_0|-1)^2} = \frac{|x_0|^2-1}{(|x_0|-1)^2}\\ &= \frac{|x_0|+1}{|x_0|-1} = \frac{1+|x_0^*|}{1-|x_0^*|} = \frac{1+|R(0)|}{1-|R(0)|} = e^{d_{\mathbb{H}}[0,R(0)]}\,.\end{aligned}$$

Thus when $f = R$, relation (3.51) again holds true. Finally, from Theorem 3.3.15 we see that (3.51) is valid for an arbitrary f from $\mathrm{M\ddot{o}b}(B^n)$. □

As suggested, Lemma 3.6.1 is the stepping stone to a basic distortion theorem showing that Möbius transformations are bilipschitz in the spherical metric and giving the best possible estimate on the bilipschitz constant.

THEOREM 3.6.2. *Let f be a Möbius transformation of $\hat{\mathbb{R}}^n$, and let $\tilde{f}$ be its Poincaré extension. Let*

$$b = \exp\{d_{\mathbb{H}}[e_{n+1}, \tilde{f}(e_{n+1})]\}.$$

Then the bounds

$$b^{-1} q(x,y) \le q\big[f(x), f(y)\big] \le b\, q(x,y) \tag{3.52}$$

hold for all x and y in $\hat{\mathbb{R}}^n$. Moreover, the constant b is sharp.

PROOF. We establish the right-hand inequality in (3.52); this can subsequently be applied to f^{-1} to produce the lower bound once we observe that if $g = f^{-1}$, then $\tilde{g} = \tilde{f}^{-1}$ and $d_{\mathbb{H}}[e_{n+1}, \tilde{f}(e_{n+1})] = d_{\mathbb{H}}[e_{n+1}, \tilde{g}(e_{n+1})]$, which shows that f and f^{-1} give rise to the same constant b.

We once again call upon the transformation Φ in Möb$(n+1)$ that maps $\mathbb{H}^{n+1}$ to B^{n+1} and agrees with stereographic projection on $\hat{\mathbb{R}}^n$. Recall that Φ is an isometry between $(\mathbb{H}^{n+1}, d_{\mathbb{H}})$ and $(B^{n+1}, d_{\mathbb{H}})$, as well as an isometry between $(\hat{\mathbb{R}}^n, q)$ and $(\mathbb{S}^n, d)$, where d denotes the Euclidean metric.

Now $g = \Phi \circ \tilde{f} \circ \Phi^{-1}$ is a member of Möb(B^{n+1}) and, since $\Phi(e_{n+1}) = 0$,

$$d_{\mathbb{H}}\big[e_{n+1}, \tilde{f}(e_{n+1})\big] = d_{\mathbb{H}}\big[0, g(0)\big]\,.$$

Lemma 3.6.1 thus yields

$$\begin{aligned} &\sup\left\{\frac{q[f(x), f(y)]}{q(x,y)} : x, y \in \hat{\mathbb{R}}^n\ ,\ x \ne y\right\} \\ &= \quad \sup\left\{\frac{|g(z) - g(w)|}{|z - w|} : z, w \in \mathbb{S}^n\ ,\ z \ne w\right\} = b\ , \end{aligned}$$

which confirms the upper estimate in (3.52) and at the same time demonstrates that the constant b cannot be replaced by any smaller constant. □

The next result furnishes a more localized method for gauging the chordal distortion of a Möbius transformation.

THEOREM 3.6.3. *If D is a domain in $\hat{\mathbb{R}}^n$ with at least two boundary points and if f belongs to* Möb(n)*, then*

$$q\big[f(x), f(y)\big]\, q\big[f(D)^c\big] \le \frac{8\, q(x,y)}{\sqrt{q(x, \partial D)\, q(y, \partial D)}} \tag{3.53}$$

for all points x and y of D.

PROOF. Let x and y be arbitrary distinct points of D, and let u and v be points of the complement of D, that is, D^c, such that $q[f(u), f(v)] = q[f(D)^c]$. Put $c = 8/q[f(D)^c]$. Invoking Theorem 3.3.9, we calculate that

$$\begin{aligned} &\frac{q(x,y)\, q(u,v)}{q(x,u)\, q(y,v)} \cdot \frac{q(x,y)\, q(u,v)}{q(x,v)\, q(y,u)} \\ &\quad = \quad [x,u,y,v][x,v,y,u] \\ &\quad = \quad \big[f(x), f(u), f(y), f(v)\big]\big[f(x), f(v), f(y), f(u)\big] \\ &\quad = \quad \frac{q[f(x), f(y)]\, q[f(u), f(v)]}{q[f(x), f(u)]\, q[f(y), f(v)]} \cdot \frac{q[f(x), f(y)]\, q[f(v), f(u)]}{q[f(x), f(v)]\, q[f(y), f(u)]}\,. \end{aligned}$$

This therefore leads us to the estimates

$$\begin{aligned}
&\left\{\frac{q[f(x),f(y)]}{q(x,y)}\right\}^2\\
&= \left\{\frac{q(u,v)}{q[f(u),f(v)]}\right\}^2 \left\{\frac{q[f(x),f(u)]\,q[f(x),f(v)]\,q[f(y),f(u)]\,q[f(y),f(v)]}{q(x,u)\,q(x,v)\,q(y,u)\,q(y,v)}\right\}\\
&\leq \left\{\frac{q(u,v)}{q[f(u),f(v)]}\right\}^2 \left\{\frac{16}{q(x,u)\,q(x,v)\,q(y,u)\,q(y,v)}\right\}\\
&= \left\{\frac{4}{q[f(u),f(v)]}\right\}^2 \left\{\frac{q(u,v)}{q(x,u)\,q(x,v)}\right\}\left\{\frac{q(u,v)}{q(y,u)\,q(y,v)}\right\}\\
&\leq \frac{c^2}{4}\left\{\frac{q(u,x)+q(x,v)}{q(x,u)\,q(x,v)}\right\}\left\{\frac{q(u,y)+q(y,v)}{q(y,u)\,q(y,v)}\right\}\\
&\leq \frac{c^2}{4}\left\{\frac{1}{q(x,v)}+\frac{1}{q(x,u)}\right\}\left\{\frac{1}{q(y,u)}+\frac{1}{q(y,v)}\right\}\\
&\leq \frac{c^2}{4}\left\{\frac{1}{q(x,D^c)}+\frac{1}{q(x,D^c)}\right\}\left\{\frac{1}{q(y,\partial D)}+\frac{1}{q(y,\partial D)}\right\}\\
&= \frac{c^2}{q(x,\partial D)\,q(y,\partial D)}.
\end{aligned}$$

This then results in the estimate

$$\left\{\frac{q[f(x),f(y)]}{q(x,y)}\right\}^2 \leq \frac{c^2}{q(x,\partial D)\,q(y,\partial D)}$$

that is the same as (3.53) which we were seeking. □

3.6.2. Normal families and convergence. Let $f_\nu : X \to Y$ $(\nu = 1, 2, \ldots)$ be a sequence of functions from a metric space (X, d) to a metric space (Y, d'). There are numerous ways in which to think of the sequence $\langle f_\nu\rangle_{\nu=1}^{\infty}$ (which for simplicity we denote by $\langle f_\nu\rangle$) converging to a function $f : X \to Y$. For starters, $\langle f_\nu\rangle$ could converge to f *pointwise in* X, meaning simply that $f_\nu(p) \to f(p)$ for each point p of X. A significantly stronger type of convergence is *uniform convergence on* X. This form of convergence demands that for each $\varepsilon > 0$ there should be an index N such that $d'[f_\nu(p), f(p)] < \varepsilon$ is true for every p in X as soon as $\nu \geq N$.

Between these extremes lies the mode of convergence appropriate to most of the convergence questions that come up in this book. It is called a *locally uniform convergence in* X and requires that each point p of X have a neighbourhood $U = U_p$ with the property that $f_\nu \to f$ uniformly on U. This is easily seen to be equivalent to the statement that for each compact subset F of X we have $f_\nu \to f$ uniformly on F. Another common name for this type of convergence is "uniform convergence on compact subsets".

Given a sequence $\langle f_\nu\rangle$ of mappings from X to Y, we would like to have a criterion telling us whether it is possible to extract from $\langle f_\nu\rangle$ a subsequence that converges in the locally uniform sense to some function $f : X \to Y$. More generally, we would like to characterize the families $\mathcal{F}$ of functions from X to Y that enjoy the following property: each sequence $\langle f_\nu\rangle$ from $\mathcal{F}$ has a locally uniformly convergent subsequence. Note that at this point we do not insist that the limit of such a subsequence be a member of $\mathcal{F}$. However this is clearly desirable and an intriguing separate problem.

In a classic 1907 paper [**118**], families fitting this description were dubbed *normal families* by the French analyst Paul Montel. The property of normality in a family $\mathcal{F}$ is closely allied with the notion of equicontinuity. We say that a family $\mathcal{F}$ of functions from X to Y is *equicontinuous at the point* p if for each $\varepsilon > 0$ there is a $\delta = \delta(p, \varepsilon) > 0$ such that $d'[f(q), f(p)] < \varepsilon$ holds for every f in $\mathcal{F}$ whenever $d(q, p) < \delta$.

A family $\mathcal{F}$ that is equicontinuous at each point of X is known as an *equicontinuous family*. The link between normality and equicontinuity is established in the following version of the Arzelà-Ascoli theorem that is general enough to meet all our needs in this book.

THEOREM 3.6.4. *Let $\mathcal{F}$ be a nonempty family of continuous functions from a separable and locally compact metric space X to a complete metric space Y. Then $\mathcal{F}$ is a normal family if and only if $\mathcal{F}$ is an equicontinuous family and $\mathcal{F}(p) = \{f(p) : f \in \mathcal{F}\}$ is a relatively compact subset of Y for each p in X.*

Without any explicit statement to the contrary, it will be a standing assumption that whenever the terms "normal" and "equicontinuous" are used in this book in reference to a family of mappings $f : A \to \hat{\mathbb{R}}^n$ with A a subset of $\hat{\mathbb{R}}^n$, the metric involved in both A and $\hat{\mathbb{R}}^n$ is the chordal metric.

Theorem 3.6.3 now leads to a very convenient method for detecting normality in a family of Möbius transformations.

THEOREM 3.6.5. *Suppose that D is a domain in $\hat{\mathbb{R}}^n$ and that $\mathcal{F}$ is a nonempty subfamily of* $\text{Möb}(n)$. *If each point x of D has an open neighbourhood $U = U_x$ with the property that* $\inf\{q[f(U)^c] : f \in \mathcal{F}\} > 0$, *then* $\mathcal{F}|D = \{f|D : f \in \mathcal{F}\}$ *is a normal family. In particular, $\mathcal{F}|D$ is a normal family whenever*

$$\inf\{q[f(D)^c] : f \in \mathcal{F}\} > 0. \tag{3.54}$$

PROOF. The metric spaces $X = D$, $Y = \hat{\mathbb{R}}^n$, and the family $\mathcal{F}|D$ satisfy all the conditions laid down in Theorem 3.6.4. Moreover, since $\hat{\mathbb{R}}^n$ is compact, the set $\mathcal{F}(x)$ is relatively compact for every x in $\hat{\mathbb{R}}^n$. To establish the normality of $\mathcal{F}|D$ we have to check only that this family is equicontinuous. Given x_0 in D, we invoke the hypothesis of the theorem to choose $r > 0$ for which the chordal ball $B = \{x : q(x, x_0) < 2r\}$ is contained in D and has $d = \inf\{q[f(B)^c] : f \in \mathcal{F}\} > 0$. We elicit from Theorem 3.6.3 (applied in B rather than D) the information that

$$q\big[f(x), f(x_0)\big] \leq \frac{8q(x, x_0)}{d\sqrt{q(x, B^c)\, q(x_0, B^c)}} \leq \frac{8q(x, x_0)}{dr}$$

whenever f comes from $\mathcal{F}$ and $q(x, x_0) < r$. This inequality is more than adequate to guarantee the equicontinuity of $\mathcal{F}|D$ at x_0, an arbitrary point of D. □

It is worthwhile to cite a corollary of Theorem 3.6.5 that is frequently easier to use than the theorem itself.

COROLLARY 3.6.6. *Suppose that a nonempty subfamily $\mathcal{F}$ of* $\text{Möb}(n)$ *enjoys the following property: there exist points a, b, and c of $\hat{\mathbb{R}}^n$ and a constant $m > 0$ such that*

$$q\big[f(a), f(b)\big] \geq m, \qquad q\big[f(a), f(c)\big] \geq m, \qquad q\big[f(b), f(c)\big] \geq m$$

for every f in $\mathcal{F}$. Then $\mathcal{F}$ is a normal family.

PROOF. If $D = \hat{\mathbb{R}}^n \setminus \{a, b\}$, then $q[f(D)^c] \geq q[f(a), f(b)] \geq m$ for every f in $\mathcal{F}$, so Theorem 3.6.5 ensures the normality of $\mathcal{F}|D$ and thus the equicontinuity of $\mathcal{F}$ at each point of D. The same argument applies to $\hat{\mathbb{R}}^n \setminus \{a, c\}$ and $\hat{\mathbb{R}}^n \setminus \{b, c\}$, revealing that $\mathcal{F}$ is equicontinuous at every point of $\hat{\mathbb{R}}^n$. An appeal to Theorem 3.6.4 finishes the proof. □

Next, having established a good general criterion to determine whether a limit function of a sequence $\langle f_\nu \rangle$ exists, we should decide what properties this limit has. The next result supplies the answer to this question.

THEOREM 3.6.7. *Let $\langle f_\nu \rangle$ be a sequence from* $\text{Möb}(n)$ *that converges pointwise in $\hat{\mathbb{R}}^n$ to a function f. Then there are exactly three possibilities for the limit function:*

(1) *f is constant;*
(2) *f assumes exactly two values;*
(3) *f belongs to* $\text{Möb}(n)$.

In the third case both $f_\nu \to f$ and $f_\nu^{-1} \to f^{-1}$ uniformly on $\mathbb{R}^n$.

PROOF. Assuming that neither (1) nor (2) above describes the limit f, we prove that (3) offers the only alternative. We take as an assumption, therefore, the existence of points a, b, and c in $\hat{\mathbb{R}}^n$ such that the values $f(a)$, $f(b)$, and $f(c)$ are all different. We first observe that f is injective. If $f(x) = f(y)$ were to hold with $x \neq y$, we could choose distinct elements u and v from the set $\{a, b, c\}$ so that $f(u)$ and $f(v)$ both differ from $f(x)$ and compute

$$\begin{aligned} 0 &= \frac{q[f(x), f(y)]\, q[f(u), f(v)]}{q[f(x), f(u)]\, q[f(y), f(v)]} = \lim_{\nu\to\infty} \big[f_\nu(x), f_\nu(u), f_\nu(y), f_\nu(v)\big] \\ &= [x, u, y, v] > 0 \ , \end{aligned}$$

a contradiction. Thus f is injective. Furthermore,

$$\big[f(x), f(y), f(u), f(v)\big] = \lim_{\nu\to\infty} \big[f_\nu(x), f_\nu(y), f_\nu(u), f_\nu(v)\big] = [x, y, u, v]$$

whenever x, y, u, and v are distinct elements of $\hat{\mathbb{R}}^n$. Theorem 3.3.9 now certifies that f is a Möbius transformation of $\hat{\mathbb{R}}^n$.

Because $f_\nu(a) \to f(a)$, $f_\nu(b) \to f(b)$, and $f_\nu(c) \to f(c)$, it is apparent that there is a constant $m > 0$ such that

$$q\big[f_\nu(a), f_\nu(b)\big] \geq m, \qquad q\big[f_\nu(a), f_\nu(c)\big] \geq m, \qquad q\big[f_\nu(b), f_\nu(c)\big] \geq m$$

for all ν. Corollary 3.6.6 tells us that $\{f_\nu : \nu = 1, 2, \ldots\}$ is an equicontinuous family. Given $\varepsilon > 0$, we can choose for each x in $\hat{\mathbb{R}}^n$ a $\delta_x > 0$ with the feature that $q[f_\nu(y), f_\nu(x)] < \varepsilon/3$ for every $\nu \geq 1$ whenever y lies in $B_x = \{y : q(y, x) < \delta_x\}$. Next, we can select a finite number of these chordal balls—$B_{x_1}, B_{x_2}, \ldots, B_{x_r}$, say—to cover $\hat{\mathbb{R}}^n$ and then fix N large enough that $q[f_\nu(x_j), f_\mu(x_j)] < \varepsilon/3$ for $j = 1, 2, \ldots, r$ whenever $\mu \geq \nu \geq N$. Any y in $\hat{\mathbb{R}}^n$ belongs to B_{x_j} for some j, so

$$\begin{aligned} q\big[f_\nu(y), f_\mu(y)\big] &\leq q\big[f_\nu(y), f_\nu(x_j)\big] + q\big[f_\nu(x_j), f_\mu(x_j)\big] + q\big[f_\mu(x_j), f_\mu(y)\big] \\ &< \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} + \tfrac{\varepsilon}{3} = \varepsilon \end{aligned}$$

whenever $\mu \geq \nu \geq N$. Since y was an arbitrary point of $\mathbb{R}^n$, we have just demonstrated that in case (3) the sequence $\langle f_\nu \rangle$ is a uniform Cauchy sequence with respect to the chordal metric on $\hat{\mathbb{R}}^n$. It therefore follows that f_ν converges uniformly to f on $\hat{\mathbb{R}}^n$.

To establish that $f_\nu^{-1} \to f^{-1}$ uniformly on $\hat{\mathbb{R}}^n$ in case (3), it suffices by what we have just shown to check that $f_\nu^{-1} \to f^{-1}$ pointwise in $\hat{\mathbb{R}}^n$. Given x in $\hat{\mathbb{R}}^n$, we consider the sequence $\langle y_\nu \rangle$ defined by $y_\nu = f_\nu^{-1}(x)$. Let y be an arbitrary accumulation point of $\langle y_\nu \rangle$. We can select a subsequence $\langle y_{\nu_k} \rangle$ of $\langle y_\nu \rangle$ such that $y_{\nu_k} \to y$. Because $f_\nu \to f$ uniformly on $\hat{\mathbb{R}}^n$, it is easy to see that $f_{\nu_k}(y_{\nu_k}) \to f(y)$. Consequently,

$$f(y) = \lim_{k\to\infty} f_{\nu_k}(y_{\nu_k}) = \lim_{k\to\infty} x = x\,.$$

We conclude that $y = f^{-1}(x)$ is the sole accumulation point of the sequence of points $\langle f_\nu^{-1}(x) \rangle$. In other words, $f_\nu^{-1}(x) \to f^{-1}(x)$ for each x in $\hat{\mathbb{R}}^n$, as desired. □

As examples, suppose that $f_\nu(x) = x + \nu e_1$. Then $f_\nu(x) \to \infty$ for each x in $\hat{\mathbb{R}}^n$; if $f_\nu(x) = \nu x$, then $f_\nu(x) \to \infty$ for $x \neq 0$ and $f_\nu(0) \to 0$. These examples illustrate that situations (1) and (2) in Theorem 3.6.7 really do arise.

In neither of these cases is convergence uniform on $\hat{\mathbb{R}}^n$: in case (1) uniform convergence is prohibited by the fact that $q[f_\nu(\hat{\mathbb{R}}^n)] = 2$ for every ν; in case (2) it is ruled out by the discontinuity of the limit mapping.

Theorem 3.6.7 implies the following lemma.

LEMMA 3.6.8. *If $\langle f_\nu \rangle$ and $\langle g_\nu \rangle$ are sequences in* Möb(n) *and if these sequences converge pointwise in $\hat{\mathbb{R}}^n$ (and hence, uniformly on $\hat{\mathbb{R}}^n$) to Möbius transformations f and g, respectively, then $g_\nu \circ f_\nu \to g \circ f$ uniformly on $\hat{\mathbb{R}}^n$.*

Next we establish the following useful result which we will use to identify the equivalence of various topologies on the Möbius group.

LEMMA 3.6.9. *Let $\langle f_\nu \rangle$ be a sequence from* Möb$(\mathbb{H}^n)$ *with $n \geq 2$, and let f belong to* Möb$(\mathbb{H}^n)$*. The following statements are equivalent:*

(i) *$f_\nu \to f$ uniformly on $\hat{\mathbb{R}}^n$;*
(ii) *$f_\nu \to f$ uniformly on $\hat{\mathbb{R}}^{n-1}$;*
(iii) *$f_\nu \to f$ pointwise in $\mathbb{H}^n$.*

PROOF. The implications (i) ⇒ (ii) and (i) ⇒ (iii) are trivial. Assume that either (ii) or (iii) holds. It is clear from Theorem 3.6.5 that, in either situation, $\mathcal{F} = \{f_\nu : \nu = 1, 2, \ldots\}$ is a normal family. Therefore, every subsequence of $\langle f_\nu \rangle$ has a further subsequence that converges uniformly on $\hat{\mathbb{R}}^n$, necessarily to a member of Möb$(\mathbb{H}^n)$. If $\langle f_\nu \rangle$ failed to converge to f uniformly on $\hat{\mathbb{R}}^n$, there would be a subsequence $\langle f_{\nu_k} \rangle$ of $\langle f_\nu \rangle$ such that $f_{\nu_k} \to g$ uniformly on $\hat{\mathbb{R}}^n$, where g belonged to Möb$(\mathbb{H}^n)$ but was different from f. Given that (ii) is true, we would have $g = f$ on $\hat{\mathbb{R}}^{n-1}$, so $g^{-1} \circ f$ would be a Möbius transformation that fixed $\hat{\mathbb{R}}^{n-1}$ pointwise, yet was neither the identity transformation of $\hat{\mathbb{R}}^n$ nor the reflection in $\mathbb{R}^{n-1}$, contrary to Theorem 3.3.6. Assuming that (iii) holds, we would have $g = f$ in $\mathbb{H}^n$; hence, by continuity, $g = f$ on $\hat{\mathbb{R}}^{n-1}$, leading to precisely the same contradiction as above. In order for a contradiction not to follow, it must be true that $f_\nu \to f$ uniformly on $\hat{\mathbb{R}}^n$ whenever condition (ii) or (iii) holds. □

3.7. The Möbius group as a matrix group

We have already seen indications of the relationship between Möbius transformations, or linear fractional transformations, of the Riemann sphere $\hat{\mathbb{C}}$ and matrix groups. In this section we intend to clarify these relationships. As a first example of this, suppose that f and g are linear fractional transformations,

$$f(z) = \frac{az+b}{cz+d}, \quad g(z) = \frac{a'z+b'}{c'z+d'}$$

normalised so that both

$$ad - bc = 1, \quad a'd' - b'c' = 1.$$

We can identify f and g with the matrices

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad B = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix}$$

respectively. The matrices A and B lie in the group $\mathrm{SL}(2,\mathbb{C})$ of 2×2 matrices with complex entries and determinant equal to 1. The group operation in the Möbius group is composition. We can compare the group operations with the following calculation:

$$(f\circ g)(z) = \frac{a\left(\frac{a'z+b'}{c'z+d'}\right)+b}{c\left(\frac{a'z+b'}{c'z+d'}\right)+d} = \frac{a(a'z+b')+b(c'z+d')}{c(a'z+b')+d(c'z+d')}$$

$$(3.55) \qquad = \frac{(aa'+bc')z+(ab'+bd')}{(ca'+dc')z+(cb'+dd')},$$

$$(3.56) \qquad A.B = \begin{pmatrix} aa'+bc' & ab'+bd' \\ ca'+dc' & cb'+dd' \end{pmatrix}.$$

Then notice that (3.55) and (3.56) effect the same correspondence between matrices and Möbius transformations that we used between f and A and also between g and B. That is, the natural matrix representative for $f\circ g$ is simply the product of the matrix representation for f and that for g using the usual matrix multiplication. The correspondence $f \leftrightarrow A$ as above therefore describes an isomorphism of the groups $\text{Möb}^+(2)$ and $\mathrm{PSL}(2,\mathbb{C}) = \mathrm{SL}(2,\mathbb{C})/\{\pm identity\}$—we have to projectivise as the matrices A and $-A$ give rise to the same linear fractional transformation.

We now consider to what extent this remarkable correspondence continues in higher dimensions. To do so we will have to develop a bit more general theory.

3.7.1. The Möbius group as a topological group. A *topological group* is a pair $\langle G, \mathcal{T}\rangle$, where G is a group and $\mathcal{T}$ is a Hausdorff topology on G with the property that the group operations of G are continuous relative to $\mathcal{T}$. Thus the function from $G\times G$ to G that sends (g_1, g_2) to $g_1 g_2$ and the function from G to itself that takes g to g^{-1} are continuous. Here the topology on $G\times G$ is, as expected, the product topology associated with $\mathcal{T}$. One usually speaks of "the topological group G", with it being implicit that the underlying topology $\mathcal{T}$ has somehow been prescribed and is to remain fixed for the duration of the discussion.

The chordal metric q on $\hat{\mathbb{R}}^n$ induces in a natural way a metric q^* on the group $\text{Möb}(n)$: for f and g in $\text{Möb}(n)$,

$$q^*(f,g) = \max_{x\in\hat{\mathbb{R}}^n} q\big[f(x), g(x)\big].$$

The associated metric topology on Möb(n) is nothing but the topology of uniform convergence on $\hat{\mathbb{R}}^n$, which is to say that a sequence $\langle f_\nu \rangle$ from Möb(n) converges to a member f of Möb(n) in the metric q^* if and only if $f_\nu \to f$ uniformly on $\hat{\mathbb{R}}^n$. It was observed in the previous section that if sequences $\langle f_\nu \rangle$ and $\langle g_\nu \rangle$ from Möb(n) converge uniformly on $\hat{\mathbb{R}}^n$ to mappings f and g, then Theorem 3.6.7 and Lemma 3.6.8 imply that these limits are of necessity Möbius transformations and also that $g_\nu \circ f_\nu \to g \circ f$ and $f_\nu^{-1} \to f^{-1}$ uniformly on $\hat{\mathbb{R}}^n$. In other words, when endowed with the metric topology that derives from q^*, Möb(n) becomes a topological group. The groups Möb($\mathbb{H}^n$) and Möb(B^n) are closed topological subgroups of Möb(n).

Topological groups G and G' are *isomorphic* if there is a bijection $\varphi : G \to G'$ that is simultaneously a group homomorphism and a homeomorphism. Any such φ is called a *topological isomorphism* from G onto G'. For example, we know of a bijective homomorphism φ : Möb(n) $\to$ Möb($\mathbb{H}^{n+1}$), the one given by $\varphi(f) = \tilde{f}$, where $\tilde{f}$ is the Poincaré extension of f. We assert that φ is a topological isomorphism. Indeed, that φ is also a homeomorphism follows instantly from Lemma 3.6.9.

Of course Lemma 3.6.9 has an obvious analogue for the group Möb(B^n), and this shows that the topological groups Möb($\mathbb{H}^n$) and Möb(B^n) are isomorphic: the function φ defined by $\varphi(f) = g^{-1} \circ f \circ g$, where g is any member of Möb(n) that maps B^n onto $\mathbb{H}^n$, delivers a topological isomorphism of Möb($\mathbb{H}^n$) onto Möb(B^n). We infer that Möb(n) and Möb(B^{n+1}) are isomorphic as topological groups.

3.7.2. Another model for hyperbolic space. To identify the matrix groups we need, we shall need another description of hyperbolic space and in particular its group of isometries.

We have already pointed out that a number of realizations exist for hyperbolic n-space $\mathbb{H}^n$. In the Poincaré half-space and ball models we have already met two of these. We now consider a third. The new model has the advantage that its group of isometries is a subgroup of GL($n+1$). Since this isometry group is also isomorphic as a topological group to Möb($\mathbb{H}^n$), we arrive at a means to represent the group Möb($\mathbb{H}^n$)—and, ultimately, Möb(n)—as a group of matrices.

For reasons that will soon become apparent we shall, in this subsection and the next, write vectors x in $\mathbb{R}^{n+1}$ in the manner $x = (x_0, x_1, \dots, x_n)$ and, with the obvious meaning, use $e_0, e_1, \dots, e_n$ for the standard basis of $\mathbb{R}^{n+1}$.

For $n \geq 2$, let B_n denote the nondegenerate, symmetric, bilinear form defined on $\mathbb{R}^{n+1}$ by

$$B_n(x, y) = x_0 y_0 - x_1 y_1 - x_2 y_2 - \cdots - x_n y_n$$

or, to express it differently,

$$B_n(x, y) = \langle J(x), y \rangle \ ,$$

where $J : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ is the symmetric linear involution given by

$$J(x) = (x_0, -x_1, -x_2, \dots, -x_n) \, .$$

With B_n is associated a quadratic form Q_n,

$$Q_n(x) = B_n(x, x) = x_0^2 - x_1^2 - x_2^2 - \cdots - x_n^2 \, .$$

We define a subset $\mathbb{H}^n$ of $\mathbb{R}^{n+1}$ as follows (we use the notation $\mathbb{H}^n$ since it will turn out to be a model for hyperbolic space when endowed with the correct metric of course):

$$\mathbb{H}^n = \big\{x \in \mathbb{R}^{n+1} : Q_n(x) = 1 \ , \ x_0 > 0\big\}.$$

Thus $\mathbb{H}^n$ is the "positive" sheet of the hyperboloid of two sheets in $\mathbb{R}^{n+1}$ determined by the equation

$$x_0^2 - x_1^2 - x_2^2 - \cdots - x_n^2 = 1\,.$$

Observe that $x_0 \geq 1$ for every point x of $\mathbb{H}^n$. If $\gamma : [a,b] \to \mathbb{R}^{n+1}$ is a smooth path whose trajectory lies on $\mathbb{H}^n$, say $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_n)$, then

$$\big[\gamma_0(t)\big]^2 = 1 + \big[\gamma_1(t)\big]^2 + \big[\gamma_2(t)\big]^2 + \cdots + \big[\gamma_n(t)\big]^2$$

for every t in $[a,b]$. It follows from differentiation of this identity that

$$\gamma_0(t)\dot\gamma_0(t) = \gamma_1(t)\dot\gamma_1(t) + \gamma_2(t)\dot\gamma_2(t) + \cdots + \gamma_n(t)\dot\gamma_n(t)$$

for all such t. From these relations we deduce that

$$\begin{aligned} Q_n\big[\dot\gamma(t)\big] &= \big[\dot\gamma_0(t)\big]^2 - \sum_{i=1}^n \big[\dot\gamma_i(t)\big]^2 = \frac{\Big[\sum_{i=1}^n \gamma_i(t)\dot\gamma_i(t)\Big]^2}{\big[\gamma_0(t)\big]^2} - \sum_{i=1}^n \big[\dot\gamma_i(t)\big]^2 \\ &\leq \frac{\Big\{\sum_{i=1}^n \big[\gamma_i(t)\big]^2\Big\}\Big\{\sum_{i=1}^n \big[\dot\gamma_i(t)\big]^2\Big\}}{\big[\gamma_0(t)\big]^2} - \sum_{i=1}^n \big[\dot\gamma_i(t)\big]^2 \\ &= -\frac{\sum_{i=1}^n \big[\dot\gamma_i(t)\big]^2}{\big[\gamma_0(t)\big]^2} \leq 0 \end{aligned}$$

for each t in $[a,b]$, with strict inequality unless $\dot\gamma_1(t) = \dot\gamma_2(t) = \cdots = \dot\gamma_n(t) = 0$ (hence, also $\dot\gamma_0(t) = 0$). As a result, we see that the form Q_n is negative definite on tangent vectors to $\mathbb{H}^n$.

Mimicking what we previously did to construct the hyperbolic distances, we can use Q_n to define a distance $d_{\mathbb{H}}$ in $\mathbb{H}^n$: for any piecewise smooth path $\gamma : [a,b] \to \mathbb{H}^n$, we set

$$\ell_{\mathbb{H}}(\gamma) = \int_a^b \sqrt{-Q_n[\dot\gamma(t)]}\, dt\ ;$$

for x and y in $\mathbb{H}^n$, we define

$$d_{\mathbb{H}}(x,y) = \inf_\gamma \ell_{\mathbb{H}}(\gamma)\ ,$$

where the infimum extends over all piecewise smooth paths γ on $\mathbb{H}^n$ with initial point x and terminal point y. In the language of differential geometry, $\ell_{\mathbb{H}}$ and $d_{\mathbb{H}}$ are the length and distance functions associated with the Riemannian metric on $\mathbb{H}^n$ given by the fundamental form $ds^2 = dx_1^2 + \cdots + dx_n^2 - dx_0^2$.

We now want to show that, equipped with the above distance, the metric space $\mathbb{H}^n$ is isometric to the Poincaré ball B^n.

THEOREM 3.7.1. *The mapping* $F : \mathbb{H}^n \to B^n$ *defined by*

$$F(x) = \left(\frac{x_1}{1+x_0}, \frac{x_2}{1+x_0}, \ldots, \frac{x_n}{1+x_0}\right)$$

is an isometry between $(\mathbb{H}^n, d_{\mathbb{H}})$ *and* $(B^n, d_{\mathbb{H}})$.

Proof. Write $y = F(x)$ for x in $\mathbb{H}^n$. A computation gives

$$|y|^2 = \frac{x_0 - 1}{x_0 + 1} < 1 \ ,$$

so $F(\mathbb{H}^n)$ is contained in B^n. An inverse for F is easily exhibited:

$$F^{-1}(y) = \left(\frac{1 + |y|^2}{1 - |y|^2}, \frac{2y_1}{1 - |y|^2}, \dots, \frac{2y_n}{1 - |y|^2} \right)$$

for y in B^n. This now establishes that F is a bijection between $\mathbb{H}^n$ and B^n. To prove that F is an isometry it suffices to show that $\ell_{\mathbb{H}}(F \circ \gamma) = \ell_{\mathbb{H}}(\gamma)$ whenever γ is a smooth path on $\mathbb{H}^n$. Fix γ and write $\beta = F \circ \gamma$. Then

$$\beta_i(t) = \frac{\gamma_i(t)}{1 + \gamma_0(t)}$$

and

$$\dot\beta_i(t) = \frac{\dot\gamma_i(t)}{1 + \gamma_0(t)} - \frac{\gamma_i(t)\dot\gamma_0(t)}{[1 + \gamma_0(t)]^2}$$

for $i = 1, 2, \dots, n$.

Remembering that $\gamma_0\dot\gamma_0 = \gamma_1\dot\gamma_1 + \gamma_2\dot\gamma_2 + \cdots + \gamma_n\dot\gamma_n$, we calculate that

$$\begin{aligned}
&\frac{4|\dot\beta(t)|^2}{(1 - |\beta(t)|^2)^2} \\
&= 4\left(\frac{1 + \gamma_0(t)}{2}\right)^2 \sum_{i=1}^{n} \left(\frac{\dot\gamma_i(t)}{1 + \gamma_0(t)} - \frac{\gamma_i(t)\dot\gamma_0(t)}{[1 + \gamma_0(t)]^2} \right)^2 \\
&= \sum_{i=1}^{n} \big[\dot\gamma_i(t)\big]^2 - \frac{2\dot\gamma_0(t)}{1 + \gamma_0(t)} \sum_{i=1}^{n} \gamma_i(t)\dot\gamma_i(t) + \frac{[\dot\gamma_0(t)]^2}{[1 + \gamma_0(t)]^2} \sum_{j=1}^{n} \big[\gamma_i(t)\big]^2 \\
&= \sum_{i=1}^{n} \big[\gamma_i(t)\big]^2 - \frac{2\gamma_0(t)[\dot\gamma_0(t)]^2}{1 + \gamma_0(t)} + \frac{[\dot\gamma_0(t)]^2[\gamma_0(t) - 1]}{1 + \gamma_0(t)} \\
&= \sum_{i=1}^{n} \big[\dot\gamma_i(t)\big]^2 - \big[\dot\gamma_0(t)\big]^2 = -Q_n\big[\dot\gamma(t)\big] \,.
\end{aligned}$$

If $\gamma : [a, b] \to \mathbb{H}^n$, we thus obtain

$$\ell_{\mathbb{H}}(\beta) = \int_a^b \frac{2|\dot\beta(t)|\,dt}{1 - |\beta(t)|^2} = \int_a^b \sqrt{-Q_n[\dot\gamma(t)]}\,dt = \ell_{\mathbb{H}}(\gamma) \ ,$$

as desired. □

The isometry F from $\mathbb{H}^n$ to B^n can be described in geometric terms as a stereographic projection with respect to the point $-e_0$: for x in $\mathbb{H}^n$, $F(x)$ is the point where B^n, here identified with $\{x \in B^{n+1} : x_0 = 0\}$, is intersected by the ray in $\mathbb{R}^{n+1}$ that issues from $-e_0$ and passes through x. On the basis of Theorem 3.7.1 we realize $(\mathbb{H}^n, d_{\mathbb{H}})$ as a third model for hyperbolic n-space—the hyperboloid model. The correspondence $f \mapsto F^{-1} \circ f \circ F$ sets up an algebraic isomorphism between Möb(B^n) and Isom$(\mathbb{H}^n)$, the group of isometries of $(\mathbb{H}^n, d_{\mathbb{H}})$. We shall exploit this isomorphism to identify the members of Isom$(\mathbb{H}^n)$. It is through this process that we shall establish a topological isomorphism between Möb(B^n) and one of the classical matrix groups.

3.7.3. The isometry group $\mathrm{Isom}^+(\mathbb{H}^n)$**.** For $n \geq 2$ we denote by $\mathrm{O}(1,n)$ the subgroup of $\mathrm{GL}(n+1)$ consisting of all transformations T that preserve the bilinear form B_n: a member T of $\mathrm{GL}(n+1)$ belongs to $\mathrm{O}(1,n)$ if and only if

$$B_n(x,y) = B_n[T(x), T(y)]$$

for all x and y in $\mathbb{R}^{n+1}$. Recalling that $B_n(x,y) = \langle J(x), y\rangle$, where $J(x) = (x_0, -x_1, -x_2, \ldots, -x_n)$, the last condition can be rewritten as

$$\big\langle J(x), y\big\rangle = \big\langle [JT](x), T(y)\big\rangle = \big\langle [T^* JT](x), y\big\rangle$$

for all x and y in $\mathbb{R}^{n+1}$. It follows that a linear transformation $T : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ is in $\mathrm{O}(1,n)$ if and only if $T^* JT = J$. This is equivalent to the requirement $T^{-1} = JT^* J$.

Any transformation T from $\mathrm{O}(1,n)$ preserves the quadratic form Q_n, so T leaves invariant the hyperboloid $S : Q_n(x) = 1$, although it may interchange the two sheets of S. We use $\mathrm{O}^+(1,n)$ to denote the set

$$\mathrm{O}^+(1,n) = \{T \in \mathrm{O}(1,n) : T(\mathbb{H}^n) = \mathbb{H}^n\}\ .$$

Then $\mathrm{O}^+(1,n)$ is a subgroup of $\mathrm{O}(1,n)$. In fact, $\mathrm{O}^+(1,n)$ has index two in $\mathrm{O}(1,n)$, the left cosets of $\mathrm{O}^+(1,n)$ in $\mathrm{O}(1,n)$ being $\mathrm{O}^+(1,n)$ and $R\mathrm{O}^+(1,n)$, where R is the reflection in the hyperplane $P : x_0 = 0$. It is evident that a transformation T from $\mathrm{O}(1,n)$ belongs to $\mathrm{O}^+(1,n)$ if and only if $T(e_0)$ lies in $\mathbb{H}^n$, which happens if and only if $\langle T(e_0), e_0\rangle > 0$.

The relation $T^* JT = J$ for T in $\mathrm{O}(1,n)$ implies that $\det(T) = \pm 1$. Those T in $\mathrm{O}(1,n)$ with $\det(T) = 1$ form an index-two subgroup of $\mathrm{O}(1,n)$, a subgroup that we designate $\mathrm{SO}(1,n)$. The notation $\mathrm{SO}^+(1,n)$ stands for the group $\mathrm{O}^+(1,n) \cap \mathrm{SO}(1,n)$, which is of index two in $\mathrm{O}^+(1,n)$. All of the above groups are closed subgroups of the topological group $\mathrm{GL}(n+1)$, whose topology we take to be the metric topology associated with the operator norm. Of course, a sequence of linear transformations $T_\nu : \mathbb{R}^n \to \mathbb{R}^m$ converges in operator norm to $T : \mathbb{R}^n \to \mathbb{R}^m$ if and only if $T_\nu(x) \to T(x)$ for each x in $\mathbb{R}^n$, or, for that matter, for each x in some basis for $\mathbb{R}^n$.

The groups just introduced can be viewed more concretely as groups of $(n+1) \times (n+1)$ invertible matrices through the customary device of identifying a linear transformation $T : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ with its matrix A_T relative to the basis $e_0, e_1, \ldots, e_n$:

$$A_T = \begin{bmatrix} a_{00} & a_{01} & \cdots & a_{0n} \\ a_{10} & a_{11} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n0} & a_{n1} & \cdots & a_{nn} \end{bmatrix},$$

where $a_{ij} = \langle T(e_j), e_i\rangle$ for $0 \leq i,j \leq n$. The correspondence $T \to A_T$ provides a topological isomorphism of $\mathrm{GL}(n+1)$ onto the group of nonsingular $(n+1)\times(n+1)$ real matrices.

In general the algebra of $m \times n$ real matrices is topologized by regarding it as a copy of $\mathbb{R}^{mn}$ and giving it the standard metric topology of this Euclidean space. Thus, a sequence $\langle A_\nu\rangle$ of $m \times n$ matrices converges to an $m \times n$ matrix A if and only if for each i and j the $(i,j)^{th}$-entry of A_ν converges to the $(i,j)^{th}$-entry of A as $\nu \to \infty$. Because of the way in which we have associated matrices with linear transformations, the inverse isomorphism is given by $A \mapsto T_A$, in which $T_A(x) = xA^*$. In this context A^* means the transpose of A; the multiplication is ordinary matrix multiplication in which a vector x in $\mathbb{R}^{n+1}$ is regarded as a $1 \times (n+1)$ matrix.

It is traditional to let notation such as $\mathrm{GL}(n)$, $\mathrm{O}(n)$, $\mathrm{O}(1,n)$, and so forth, serve for both a group of linear transformations and for its associated matrix group. For instance, $\mathrm{O}(1,n)$ can be thought of as the group of $(n+1) \times (n+1)$ real matrices A that satisfy the identity $A^*JA = J$ (or $A^{-1} = JA^*J$), where

$$J = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & -1 & \cdots & 0 \\ & \vdots & & & \vdots \\ 0 & 0 & 0 & \cdots & -1 \end{bmatrix}.$$

The group $\mathrm{O}^+(1,n)$ consists of the matrices A in $\mathrm{O}(1,n)$ with $a_{00} > 0$.

Let T be a linear transformation from the group $\mathrm{O}^+(1,n)$. Then $f = T|\mathbb{H}^n$ maps $\mathbb{H}^n$ bijectively to itself. Moreover, if $\gamma : [a,b] \to \mathbb{R}^{n+1}$ is an arbitrary smooth path whose trajectory lies on $\mathbb{H}^n$ and if $\beta = f \circ \gamma$, then $\dot\beta(t) = T[\dot\gamma(t)]$ for each t in $[a,b]$. Because T belongs to $\mathrm{O}^+(1,n)$, we discover that

$$\ell_{\mathbb{H}}(\beta) = \int_a^b \sqrt{-Q_n[\dot\beta(t)]}\, dt = \int_a^b \sqrt{-Q_n[\dot\gamma(t)]}\, dt = \ell_{\mathbb{H}}(\gamma)\,.$$

It follows that any mapping f of this kind is an isometry of $\mathbb{H}^n$ relative to its hyperbolic metric. There are no others, as we now learn.

THEOREM 3.7.2. *Each member of the group* $\mathrm{Isom}(\mathbb{H}^n)$ *is the restriction to* $\mathbb{H}^n$ *of a uniquely determined transformation from* $\mathrm{O}^+(1,n)$.

PROOF. Since $\mathbb{H}^n$ contains a basis for $\mathbb{R}^{n+1}$, for example, $\mathbb{H}^n$ includes the vectors $e_0, \sqrt{2}\,e_0 + e_1, \sqrt{2}\,e_0 + e_2, \dots, \sqrt{2}\,e_0 + e_n$, two linear transformations of $\mathbb{R}^{n+1}$ that share the same restriction to $\mathbb{H}^n$ must coincide everywhere in $\mathbb{R}^{n+1}$. This fact shows the uniqueness assertion to be trivial. If $f = T|\mathbb{H}^n$ and $g = S|\mathbb{H}^n$, where T and S belong to $\mathrm{O}^+(1,n)$, then $g \circ f = ST|\mathbb{H}^n$. To complete the proof, therefore, it suffices to check that $\mathrm{Isom}(\mathbb{H}^n)$ has a set of generators of the stated type. It has already been observed that the correspondence $g \mapsto F^{-1} \circ g \circ F$, where $F : \mathbb{H}^n \to B^n$ is the mapping in Theorem 3.7.1, gives us an isomorphism of $\mathrm{M\ddot{o}b}(B^n)$ onto $\mathrm{Isom}(\mathbb{H}^n)$.

We shall identify two particular types of matrices A in the matrix group $\mathrm{O}^+(1,n)$ and verify that the associated isometries $f_A = T_A|\mathbb{H}^n$ generate $\mathrm{Isom}(\mathbb{H}^n)$. We do so by making certain that $\mathrm{M\ddot{o}b}(B^n)$ is generated by the mappings $g_A = F \circ f_A \circ F^{-1}$ that correspond to the aforementioned f_A under the above isomorphism. To this end, it will be useful to have a concrete representation of the function g_A for arbitrary A in $\mathrm{O}^+(1,n)$.

Given A in $\mathrm{O}^+(1,n)$ and z in B^n, write $x = F^{-1}(z)$, $y = f_A(x)$, and $w = F(y) = g_A(z)$. Then $y = xA^*$, so

$$y_i = a_{i0}x_0 + a_{i1}x_1 + \cdots + a_{in}x_n$$

for $i = 0, 1, \dots, n$. From the formula for F^{-1} we obtain

$$(1 - |z|^2)y_i = (1 + |z|^2)a_{i0} + 2(a_{i1}z_1 + a_{i2}z_2 + \cdots + a_{in}z_n)\,.$$

It is also true that

$$w_i = \frac{y_i}{1+y_0} = \frac{(1-|z|^2)y_i}{(1-|z|^2)+(1-|z|^2)y_0}$$

for $i = 1, 2, \ldots, n$. As a result, we arrive at the expression

$$w_i = \frac{(1+|z|^2)a_{i0} + 2(a_{i1}z_1 + a_{i2}z_2 + \cdots + a_{in}z_n)}{|z|^2(a_{00}-1) + 2(a_{01}z_1 + a_{02}z_2 + \cdots + a_{0n}z_n) + a_{00} + 1}$$

for z belonging to B^n.

Consider the matrix

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & U & \\ 0 & & & \end{bmatrix},$$

where U comes from $\mathrm{O}(n)$. Then A is in $\mathrm{O}^+(1,n)$ and has $\det(A) = \det(U)$. The computation above shows us that

$$w = g_A(z) = zU^* = T_U(z)$$

for every z in B^n, which implies that $g_A = T_U$ is an orthogonal linear transformation. Here we recall Corollary 3.3.11 and note that g_A belongs to $\text{Möb}^+(B^n)$ if and only if U is in $\mathrm{SO}(n)$.

Next, let R_s be the reflection in the sphere $\Sigma = \mathbb{S}^{n-1}(se_1, r)$, where $s > 1$ and $r = \sqrt{s^2-1}$. Since Σ is orthogonal to $\mathbb{S}^{n-1}$, R_s is in $\text{Möb}(B^n)$. Indeed, a moment's thought reveals that $\text{Möb}(B^n)$ is generated by $\mathrm{O}(n)$ and $\{R_s : s > 1\}$, since any reflection in a Euclidean sphere orthogonal to $\mathbb{S}^{n-1}$ is conjugate via a rotation to R_s for some $s > 1$. We claim that $R_s = g_A$ for

$$A = \begin{bmatrix} \cosh(2t) & -\sinh(2t) & 0 & \cdots & 0 \\ \sinh(2t) & -\cosh(2t) & 0 & \cdots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & I_{n-1} & \\ 0 & 0 & & & \end{bmatrix},$$

where t satisfies $\coth t = s$. (Then $\sinh t = r^{-1}$.) Certainly A is in $\mathrm{O}^+(1,n)$, with $\det(A) = -1$. Let $c = se_1$. To determine $w = g_A(z)$ for z in B^n, we first remark that

$$\begin{aligned} &|z|^2(a_{00}-1) + 2\left(\sum_{i=1}^n a_{0i}z_i\right) + a_{00} + 1 \\ &= |z|^2\big[\cosh(2t) - 1\big] - 2z_1\sinh(2t) + \cosh(2t) + 1 \\ &= 2|z|^2\sinh^2(t) - 2z_1\sinh(2t) + 2\cosh^2 t = 2|z-c|^2\sinh^2 t = \frac{2|z-c|^2}{r^2}, \end{aligned}$$

whence

$$w_i = \frac{r^2 z_i}{|z-c|^2}$$

for $i = 2, 3, \ldots, n$. When $i = 1$ we have

$$\begin{aligned} w_1 &= \frac{(1+|z|^2)\sinh(2t) - 2z_1\cosh(2t)}{2|z-c|^2\sinh^2(t)} \\ &= \frac{(|z-c|^2 + 1 - |c|^2 + 2\langle z, c\rangle)(2\sinh^2 t + 1) - 2z_1(2\cosh^2 t - 1)}{2|z-c|^2\sinh^2 t} \\ &= \coth t + \frac{r^2(z_1 - \coth t)}{|z-c|^2} = s + \frac{r^2(z_1 - s)}{|z-c|^2}. \end{aligned}$$

We conclude that

$$w = g_A(z) = c + \frac{r^2(z-c)}{|z-c|^2} = R_s(z)$$

for every z in B^n, which means that $g_A = R_s$ (Corollary 3.3.11). It follows that the group $\text{Isom}(\mathbb{H}^n)$ is generated by the mappings f_A, where A ranges over all matrices of the two types we have just considered. The proof of the theorem is thus complete. □

We can define an algebraic isomorphism φ of the matrix group $\mathrm{O}^+(1,n)$ onto $\text{Möb}(B^n)$ by setting $\varphi(A) = g_A$, where g_A is the unique member of $\text{Möb}(B^n)$ satisfying $F^{-1} \circ g_A \circ F = f_A = T_A|\mathbb{H}^n$. The upshot of the preceding deliberations is encapsulated in the following theorem.

THEOREM 3.7.3. *The function φ is a topological isomorphism of the matrix group $\mathrm{O}^+(1,n)$ onto $\text{Möb}(B^n)$. Under this isomorphism, $\mathrm{SO}^+(1,n)$ corresponds to $\text{Möb}^+(B^n)$.*

PROOF. We must verify that φ is a homeomorphism. For an arbitrary sequence $\langle A_\nu \rangle$ from $\mathrm{O}^+(1,n)$, write $f_\nu = f_{A_\nu}$ and $g_\nu = g_{A_\nu}$. If $A_\nu \to A$, then $xA_\nu^* \to xA^*$ for every x in $\mathbb{R}^{n+1}$, so $f_\nu \to f = f_A$ pointwise in $\mathbb{H}^n$. It follows that $g_\nu \to g = g_A$ pointwise in B^n. The analogue of Lemma 3.6.9 for $\text{Möb}(B^n)$ shows that $g_\nu \to g$ uniformly on $\hat{\mathbb{R}}^n$. We infer that φ is continuous. Conversely, if $g_\nu \to g$ uniformly on $\hat{\mathbb{R}}^n$, then $f_\nu \to f = F^{-1} \circ g \circ F$ pointwise on $\mathbb{H}^n$. The limit f must belong to $\text{Isom}(\mathbb{H}^n)$, whence $f = f_A$ for a unique A in $\mathrm{O}^+(1,n)$. Naturally, $\varphi(A) = g$. Since $T_{A_\nu} \to T_A$ pointwise on $\mathbb{H}^n$ and since $\mathbb{H}^n$ contains a basis for $\mathbb{R}^{n+1}$, $T_{A_\nu} \to T_A$ in the operator norm. Therefore $A_\nu \to A$ in $\mathrm{O}^+(1,n)$. This shows that φ^{-1} is also continuous, making φ a homeomorphism.

The function $\psi(g) = \det[\varphi^{-1}(g)]$ is a homomorphism of $\text{Möb}(B^n)$ onto $\{+1, -1\}$ with kernel $K = \varphi[\mathrm{SO}^+(1,n)]$, an index-two subgroup of $\text{Möb}(B^n)$. An arbitrary member g of $\text{Möb}^+(B^n)$ either belongs to $\mathrm{SO}(n)$ or has the structure $g = h \circ R_\Sigma$, where h is a sense-reversing member of $\mathrm{O}(n)$ and Σ is a Euclidean sphere that is orthogonal to $\mathbb{S}^{n-1}$. The proof of Theorem 3.7.2 demonstrates that $\varphi(g)$ is in $\mathrm{SO}^+(1,n)$ for such g; that is, $\text{Möb}^+(B^n)$ lies in K. But $\text{Möb}^+(B^n)$ is also of index two in $\text{Möb}(B^n)$, so $\text{Möb}^+(B^n) = K = \varphi[\mathrm{SO}^+(1,n)]$. □

Recalling that $\text{Möb}(n)$ is topologically isomorphic to $\text{Möb}(B^{n+1})$ via an isomorphism that maps $\text{Möb}^+(n)$ to $\text{Möb}^+(B^{n+1})$, we record a corollary of Theorem 3.7.3.

COROLLARY 3.7.4. *The Möbius groups $\text{Möb}(n)$ and $\text{Möb}^+(n)$ are isomorphic as topological groups to the matrix groups $\mathrm{O}^+(1,n+1)$ and $\mathrm{SO}^+(1,n+1)$, respectively.*

One implication of Corollary 3.7.4 is that Möb(n) carries the structure of a Lie group. For more about this side of the Möbius group we refer the reader to the recent book of D. Bump on this subject, [**26**].

3.8. Liouville's theorem

Just as conformal mappings in the plane are homeomorphic solutions to a system of partial differential equations (the Cauchy-Riemann equations in fact), so too are conformal mappings in space—the Cauchy-Riemann systems which we will come to in a moment. The Looman-Menchoff theorem describes the precise hypotheses on the derivatives of a solution to the Cauchy-Riemann equations under which one can assert the mapping is holomorphic. This is proved in R. Narasimhan's book [**132**]. In higher dimensions, $n \geq 3$, questions still remain as to whether all sufficiently regular solutions to the natural equations defining a conformal mapping are Möbius transformations. Of course this is demonstrably false in two dimensions—there are plenty of conformal mappings which are not Möbius, although entire (defined in $\mathbb{C}$ or $\hat{\mathbb{C}}$) holomorphic homeomorphisms are Möbius transformations. However in dimensions $n \geq 3$ this is basically true and reflects a remarkable higher–dimensional rigidity, with important applications which in many ways underpins the utility of the theory of quasiconformal mappings. We will explore these things not only in this section, but elsewhere in the book.

3.8.1. A little history. In 1850, the celebrated French mathematician Joseph Liouville added a short note [**99**] to a new edition of Gaspard Monge's classic work [**117**] *Application de l'Anallyse à la Géometrie*, whose publication Liouville was overseeing. The note was prompted by a series of three letters that Liouville had received in the years 1845 and 1846 from the renowned British physicist William Thomson. Thomson, better known today as Lord Kelvin, had studied in Paris under Liouville's tutelage for a period of time in the mid-1840s, so these two giants of nineteenth century science were well acquainted. In his letters, Thomson posed to Liouville a number of questions concerning inversions in spheres, questions that had arisen in conjunction with Thomson's research in electrostatics, in particular, with the so-called principle of electrical images. (It might be pointed out that the reflection in the unit sphere $\mathbb{R}^3$ is often referred to in physics circles as the "Kelvin transform".) To learn more about the interesting relationship between Thomson and Liouville, the reader is invited to consult Jesper Lützen's magnificient biography of Liouville [**101**].

Paraphrased in the language of the present chapter, the substance of Liouville's note is conveyed by the following remarkable assertion:

if D is a domain in $\mathbb{R}^n$ with $n \geq 3$, then any conformal mapping $f : D \to \mathbb{R}^n$ is the restriction to D of a Möbius transformation.

Befitting his motivation for writing the article, Liouville couched his discussion in the language of differential forms rather than mappings. As a consequence, his original formulation of the result bears scant resemblance to the preceding statement, although the relationship between the two formulations is quite transparent to anyone moderately well versed in differential geometry. Indeed, there is nothing in Liouville's work [**99**] that even has the label "theorem"! And the title of the note, "Extension au cas de trois dimensions de la question du tracé géographique"

[Extension to the case of three dimensions of the question of drawing geographic maps], gives no hint whatsoever as to its stunning contents. (The reference is to the "map projection problem" in two dimensions, whose solution Gauss had found but never published. Liouville had rediscovered the solution in 1847.) It was only later that Liouville published his theorem in a form approximating the statement of it that we have given [**100**].

Of greater importance here is the fact that the proof which Liouville outlined for his theorem makes use of certain implicit smoothness hypotheses. His argument thus translates to a proof of the stated conformal mapping interpretation of his result only under the added assumption that f be a mapping of class C^3 or better.

We initially draw attention to one significant corollary of Liouville's theorem:

the only domains in $\mathbb{R}^n$ with $n \geq 3$ that are conformally equivalent to the unit ball B^n are round Euclidean balls and half-spaces.

This stands in stark contrast to the marvelous discovery announced in 1851, a year after Liouville's note was published, by Riemann:

A simply connected proper subdomain D of the complex plane $\mathbb{C}$ can be transformed via a (complex analytic) conformal mapping to the unit disk $\mathbb{D}$.

Indeed it is the contrast between these two italicized statements immediately above which motivates the mapping problem for quasiconformal mappings. Although an arbitrary domain homeomorphic to B^n in $\mathbb{R}^n$, $n \geq 3$, cannot be mapped conformally to the unit ball, is there a "nearly conformal" mapping which effects this? The restriction to domains homeomorphic to the ball removes elementary topological obstructions, however it is fair to say that progress on this problem has been slow. We will discuss what is known in Chapter 7.

It is our intention to present here an adaptation of Rolf Nevanlinna's elementary proof, vintage 1960, for the C^4-version of Liouville's theorem [**133, 134**].

The theorem will resurface in a much more general form later in the book. In the subsection at hand we prepare the way for Nevanlinna's proof by introducing some convenient notation and by attending to a few preliminary technical details. The proof (as with even the more general approaches) is basically to find differential identities satisfied by the Jacobian determinant of any Möbius transformation. Similarities and Euclidean isometries have constant Jacobian, while we have already identified a formula for the Jacobian of an arbitrary reflection in Lemma 3.2.1. Thus we can reasonably expect to find these equations, but it is certainly not a triviality. Next one shows that a conjectured conformal mapping, say f, has a determinant which also satisfies these equations and consequently has the same form as the Jacobian of a Möbius transformation, say φ. The chain rule will show $J_{f\circ\varphi^{-1}}$ is constant while $f \circ \varphi^{-1}$ is certainly conformal. It is not too difficult to see that a conformal mapping with constant determinant is an isometry or similarity, and this implies f is Möbius. In fact we have made quite similar arguments in Lemma 3.4.2. What this discussion omits are the deep technical questions of regularity, as the equations found are of higher order than one might like, requiring higher degrees of differentiability for f. This considerable technical difficulty needs to be overcome to get the best result.

3.8.2. Technical considerations. Let U be an open set in $\mathbb{R}^n$, and let g belong to the class $C^2(U, \mathbb{R}^m)$. Then g is differentiable throughout U, and for each vector u in $\mathbb{R}^n$ we obtain a function $\partial_u g$ in $C^1(U, \mathbb{R}^m)$ by setting

$$\partial_u g(x) = g'(x)u = \sum_{i=1}^{n} u_i \partial_i g(x)\,. \tag{3.57}$$

In particular, we see that $\partial_{e_i} g = \partial_i g$, the partial derivative of g with respect to x_i. Now the function $\partial_u g$ is itself differentiable in U, so we can form $\partial_v(\partial_u g)$ for any vector v in $\mathbb{R}^n$. This produces a function that lies in $C(U, \mathbb{R}^m)$. We denote it by $\partial_{v,u} g$. From (3.57) we deduce that

$$\partial_{v,u} g = \sum_{i,j=1}^{n} u_i v_j \partial_{i,j} g\ , \tag{3.58}$$

which implies among other things that $\partial_{v,u} g = \partial_{u,v} g$. Should g be a C^k-function with $k > 2$, we could continue this process to obtain higher order "directional derivatives". (Since we do not insist that vectors involved in these differential operators be unit vectors, they are not directional derivatives in the strict sense of the term.)

Suppose now that $f : D \to \mathbb{R}^n$ is a conformal mapping, where D is again a domain in $\mathbb{R}^n$. For the remainder of this section we shall assume that f is a diffeomorphism of the differentiability class $C^4(D, \mathbb{R}^n)$. We write

$$\lambda = \|f'\| = |J_f|^{1/n}, \quad \text{and} \quad \rho = 1/\lambda.$$

Here J_f is the Jacobian determinant which is positive as f is assumed to be a diffeomorphism. Note that the function λ belongs to $C^3(D)$. Since $\lambda > 0$, ρ is a C^3-function in D as well. Moreover, Theorem 3.1.1 implies that

$$\langle f'(x)u, f'(x)v\rangle = \langle f'(x)^* f'(x)u, v\rangle = \langle \lambda^2(x)u, v\rangle = \lambda^2(x)\langle u, v\rangle \tag{3.59}$$

for each point x of D and for all vectors u and v in $\mathbb{R}^n$. Notice that if λ is constant, then (3.59) has $h(x) = \frac{1}{\lambda} f(x)$ and $\langle h'(x)u, h'(x)v\rangle = \langle u, v\rangle$, and as a special case of what follows we shall show that h is an isometry. The reader should compare this with what we have already established in the closely related Lemma 3.4.3.

The first step in Nevanlinna's proof is supplied by the following lemma.

LEMMA 3.8.1. *If D is a domain in $\mathbb{R}^n$ with $n \geq 3$ and $f : D \to \mathbb{R}^n$ is a conformal mapping of class C^4, then $\partial_{u,v}\rho$ vanishes identically in D whenever u and v are orthogonal vectors in $\mathbb{R}^n$.*

PROOF. Since plainly $\partial_{u,v}\rho = 0$ when $u = 0$ or $v = 0$ and since the correspondence $(u, v) \mapsto \partial_{u,v}\rho$ is bilinear, it suffices to treat the case in which both u and v are unit vectors. For any pair of orthogonal (unit) vectors u and v, we have

$$\big\langle \partial_u f(x), \partial_v f(x)\big\rangle = \big\langle f'(x)u, f'(x)v\big\rangle = \lambda^2(x)\langle u, v\rangle = 0$$

for every x in D. Thus $\langle \partial_u f, \partial_v f\rangle = 0$ in D.

If w is an arbitrary vector in $\mathbb{R}^n$, we can apply the operator ∂_w to this relation in order to obtain

$$0 = \partial_w\langle \partial_u f, \partial_v f\rangle = \langle \partial_{w,u} f, \partial_v f\rangle + \langle \partial_u f, \partial_{w,v} f\rangle$$

and conclude that

$$\langle \partial_{w,u} f, \partial_v f \rangle = -\langle \partial_{w,v} f, \partial_u f \rangle$$

in D. In particular, given that three unit vectors u, v, and w in $\mathbb{R}^n$ are mutually orthogonal (a possibility since $n \geq 3$), we find that

$$\langle \partial_{u,v} f, \partial_w f \rangle = -\langle \partial_{u,w} f, \partial_v f \rangle = \langle \partial_{v,w} f, \partial_u f \rangle = -\langle \partial_{u,v} f, \partial_w f \rangle \ ,$$

with the consequence that

$$\langle \partial_{u,v} f, \partial_w f \rangle = 0 \tag{3.60}$$

in D for every such choice of u, v, and w.

Consider orthogonal unit vectors u and v in $\mathbb{R}^n$, and fix a point x of D. If W denotes the subspace of $\mathbb{R}^n$ spanned by u and v and if $T = f'(x)$, then (3.60) and the conformality of T tell us that

$$\partial_{u,v} f(x) \in T(W^\perp)^\perp = T(W)^{\perp\perp} = T(W) \, .$$

Of course, $T(W)$ is spanned by the nonzero orthogonal vectors $\partial_u f(x) = T(u)$ and $\partial_v f(x) = T(v)$. Accordingly, the relation

$$\partial_{u,v} f = \frac{\langle \partial_{u,v} f, \partial_u f \rangle}{|\partial_u f|^2} \, \partial_u f + \frac{\langle \partial_{u,v} f, \partial_v f \rangle}{|\partial_v f|^2} \, \partial_v f$$

holds throughout D. However,

$$|\partial_u f|^2 = \langle \partial_u f, \partial_u f \rangle = \lambda^2 |u|^2 = \lambda^2$$

and

$$\langle \partial_{u,v} f, \partial_u f \rangle = \frac{\partial_v \langle \partial_u f, \partial_u f \rangle}{2} = \frac{\partial_v (\lambda^2)}{2} = \lambda \partial_v \lambda \, .$$

Similarly,

$$|\partial_v f|^2 = \lambda^2, \qquad \langle \partial_{u,v} f, \partial_v f \rangle = \lambda \partial_u \lambda \, .$$

It follows that

$$\partial_{u,v} f = \left(\frac{\partial_v \lambda}{\lambda} \right) \partial_u f + \left(\frac{\partial_u \lambda}{\lambda} \right) \partial_v f \, . \tag{3.61}$$

Remembering that $\rho = 1/\lambda$, we observe that

$$\frac{\partial_u \rho}{\rho} = -\frac{\partial_u \lambda}{\lambda} \quad , \quad \frac{\partial_v \rho}{\rho} = -\frac{\partial_v \lambda}{\lambda} \, .$$

Thus (3.61) can be rephrased as

$$\partial_v \rho \cdot \partial_u f + \partial_u \rho \cdot \partial_v f + \rho \cdot \partial_{u,v} f = 0 \tag{3.62}$$

in D, whenever u and v are orthogonal unit vectors in $\mathbb{R}^n$. (The multiplication used here is just ordinary multiplication of a vector by a scalar.)

As a next step, we look at three unit vectors u, v, and w in $\mathbb{R}^n$ that are mutually orthogonal. We apply the operator ∂_w to (3.62) and learn that

$$\partial_{w,v} \rho \cdot \partial_u f + \partial_{w,u} \rho \cdot \partial_v f + \partial_v \rho \cdot \partial_{w,u} f + \partial_u \rho \cdot \partial_{w,v} f + \partial_w \rho \cdot \partial_{u,v} f + \rho \cdot \partial_{u,v,w} f = 0$$

or, equivalently, that

$$\partial_{w,v}\rho\cdot\partial_u f+\partial_{w,u}\rho\cdot\partial_v f=-\partial_v\rho\cdot\partial_{w,u}f-\partial_u\rho\cdot\partial_{w,v}f-\partial_w\rho\cdot\partial_{u,v}f-\rho\cdot\partial_{u,v,w}f \tag{3.63}$$

in D. Because the dependence of the right-hand side of (3.63) on u, v, and w is symmetric in these quantities, the same must be true of the other side. Therefore, for instance,

$$\partial_{w,v}\rho\cdot\partial_u f+\partial_{w,u}\rho\cdot\partial_v f=\partial_{u,v}\rho\cdot\partial_w f+\partial_{w,u}\rho\cdot\partial_v f\ ,$$

demonstrating that

$$\partial_{u,v}\rho\cdot\partial_w f=\partial_{w,v}\rho\cdot\partial_u f$$

in D. Finally, for u, v, and w as specified, we can once more exploit the conformality of f to arrive at

$$\begin{aligned}\lambda^2\partial_{u,v}\rho=\partial_{u,v}\rho\,\langle\partial_w f,\partial_w f\rangle=\langle\partial_{u,v}\rho\cdot\partial_w f,\partial_w f\rangle=\langle\partial_{w,v}\rho\cdot\partial_u f,\partial_w f\rangle\\ =\partial_{w,v}\rho\cdot\langle\partial_u f,\partial_w f\rangle=0\ ,\end{aligned}$$

which demands that $\partial_{u,v}\rho=0$ throughout D. □

The function ρ satisfies another significant differential equation.

LEMMA 3.8.2. *Under the hypotheses of Lemma* 3.8.1*,*

$$n|\nabla\rho|^2=2\rho\Delta\rho \tag{3.64}$$

in D, where $\Delta\rho$ denotes the Laplacian of ρ; $\Delta\rho=\sum_{i=1}^n\partial_{i,i}\rho$.

It is possible to verify Lemma 3.8.2 by means of brute force calculation, an exercise that obscures the geometric significance of (3.64) and is bound to leave readers groping in the dark for motivation. A more illuminating approach to (3.64) takes advantage of a geometric fact that is of interest in its own right—namely, that the function $k=2n^{-1}\rho\Delta\rho-|\nabla\rho|^2$ represents the scalar curvature of a certain metric to which the conformal mapping f naturally gives rise. Rather than plunge into a lengthy series of computations, we prefer to take the differential geometric route to (3.64), even though this means that the proof we ultimately give for Lemma 3.8.2 draws on information of an elementary character from outside sources.

3.8.2.1. *Elements of Riemannian geometry.* Let D be a domain in $\mathbb{R}^n$ with $n\geq 2$, and let k be a positive integer. A *Riemannian C^k-structure* in D is a symmetric and positive definite $(0,2)$-tensor field g of class C^k in D. To express this in more mundane language, g is a rule that assigns to each point x of D an inner product $g(x)=\langle\cdot,\cdot\rangle_x$ on $\mathbb{R}^n$ and does it in such a way that $x\mapsto\langle u(x),v(x)\rangle_x$ describes a C^k-function in D whenever u and v are of the class $C^k(D,\mathbb{R}^n)$. (When $k=\infty$, the accepted terminology is that g is a *Riemannian metric* in D.)

Given a structure g of this kind, one can perform yet another variation on (3.33) and (3.34) by defining the g-length $\ell_g(\gamma)$ of any piecewise smooth path $\gamma:[a,b]\to D$ through the formula

$$\ell_g(\gamma)=\int_a^b|\dot\gamma(t)|_{\gamma(t)}\,dt\ ,$$

in which $|\cdot|_x$ denotes the norm associated with $\langle\cdot,\cdot\rangle_x$, and using this to determine the g-distance $d_g(x,y)$ between points x and y of D:

$$d_g(x,y)=\inf_\gamma\ell_g(\gamma)\ ,$$

the infimum being taken over all piecewise smooth paths γ in D that start at x and terminate at y. The pair $\langle D, d_g \rangle$ is then a metric space. Moreover, the metric topology in D corresponding to the distance d_g is the same as the relative topology induced in D by the standard topology of $\mathbb{R}^n$. A Riemannian C^k-structure g carries with it a lot of other geometric structure. These include notions of g-volume, g-surface area, g-measurement of angles, and so forth, which we only allude to in passing.

Suppose that D and D' are domains in $\mathbb{R}^n$ and that g and g_0 are Riemannian C^k-structures in D and D', respectively (the possibility that $D = D'$ and $g = g_0$ is not excluded here). By an *isometry* of the pair $\langle D, g \rangle$ onto the pair $\langle D', g_0 \rangle$ is meant a C^{k+1}-diffeomorphism f of D onto D' with the property that, giving the notation its obvious interpretation,

$$\langle u, v \rangle_x = \langle f'(x)u, f'(x)v \rangle_{f(x)} \tag{3.65}$$

for each point x of D and for all vectors u and v in $\mathbb{R}^n$. When a mapping of this kind exists, we say that $\langle D, g \rangle$ and $\langle D', g_0 \rangle$ are *isometric*. Indeed, if f is such an isometry, then it is easy to see that

$$d_g(x, y) = d_{g_0}\big[f(x), f(y)\big]$$

for all x and y in D.

Of course the basic Riemannian metric of a subdomain D of $\mathbb{R}^n$ is the restricted Euclidean metric g_e: it has $g_e(x) = \langle \cdot, \cdot \rangle$, the standard Euclidean inner product on $\mathbb{R}^n$, for every x in D.

If we set

$$g = e^{2\varphi} g_e, \tag{3.66}$$

where φ comes from the class $C^k(D)$, we obtain a Riemannian C^k-structure in D. A structure g given by (3.66) is said to be *conformally Euclidean with metric density* e^{φ} (cf. Section 3.1, where $\rho = e^{\varphi}$). A second method of constructing Riemannian C^k-structure in D is via "pull-back".

Let f be a C^{k+1}-diffeomorphism of D onto a domain D' in which a Riemannian C^k-structure g_0 is assumed to be defined. For each x in D we can use (3.65) to determine an inner product $\langle \cdot, \cdot \rangle_x$ on $\mathbb{R}^n$. Then $g(x) = \langle \cdot, \cdot \rangle_x$ describes a Riemannian C^k-structure in D. We refer to g as the *pull-back of* g_0 *under* f, a relationship that is traditionally indicated by writing $g = f^*(g_0)$. Quite clearly the diffeomorphism f is an isometry of $\langle D, g \rangle$ onto $\langle D', g_0 \rangle$ precisely when $g = f^*(g_0)$.

For technical reasons we shall henceforth assume that $k \geq 3$. With a Riemannian C^k-structure g in D are associated several different types of "curvature", each of which conveys information about the geometry introduced on D by g. Among these is the scalar curvature of g. This provides a local measure of the deviation of g-surface area from the Euclidean surface area. To be slightly more explicit, given a point x of D and $r > 0$ we let $S_r = \mathbb{S}^{n-1}(x, r)$ and $S'_r = \{y \in D : d_g(y, x) = r\}$. If σ_g and σ denote g-surface area and Euclidean surface area, respectively, it can be shown that

$$\frac{\sigma_g(S'_r)}{\sigma(S_r)} = 1 - cr^2 + o(r^2) \tag{3.67}$$

as $r \to 0$ for some constant c. The *scalar curvature* $k_g(x)$ of g at x is defined by $k_g(x) = c/c_n$, where $c_n > 0$ is a certain normalization factor that depends only on the dimension n.

C. Taubes' book [**151**] is a good place to start learning more about the differential geometry here and in particular for more precise details on the subject of scalar curvature.

Two facts are readily apparent from (3.67): first, the restricted Euclidean metric g_e in D has $k_{g_e} = 0$ throughout D; second, scalar curvature is invariant under isometry—that is, if f is an isometry of $\langle D, g\rangle$ onto $\langle D', g_0\rangle$, then $k_g(x) = k_{g_0}[f(x)]$ for every x in D.

In the event that g is a conformally Euclidean C^k-structure in D, say $g = e^{2\varphi} g_e$ with φ a function from $C^k(D)$, there is a standard formula giving its scalar curvature:

$$k_g = -n^{-1} e^{-2\varphi}\left[2\Delta\varphi + (n-2)|\nabla\varphi|^2\right]. \tag{3.68}$$

Accepting the foregoing discussion, we proceed to prove Lemma 3.8.2.

Proof of Lemma 3.8.2. It has already been noted that, under the assumptions in force in Lemma 3.8.2, the statement $\langle f'(x)u, f'(x)v\rangle = \lambda^2(x)\langle u, v\rangle$ holds for each point x of D and for all vectors u and v in $\mathbb{R}^n$. We can rephrase this by saying that $g = f^*(g_e)$, the pull-back under f of the Riemannian metric g_e in $D' = f(D)$, is just the conformally Euclidean C^3-structure in D whose metric density is e^{φ}, with $\varphi = \log\lambda = -\log\rho$. Since

$$\nabla\varphi = -\rho^{-1}\nabla\rho \quad , \quad \Delta\varphi = -\rho^{-1}\Delta\rho + \rho^{-2}|\nabla\rho|^2 \ ,$$

and $n \geq 3$, we infer from (3.68) that

$$k_g = 2n^{-1}\rho\Delta\rho - |\nabla\rho|^2 \, .$$

On the other hand, f is an isometry of $\langle D, g\rangle$ onto $\langle D', g_e\rangle$, whence $k_g = k_{g_e} \circ f = 0$. Identity (3.64) follows. ☐

With Lemma 3.8.2 in hand, we return to Liouville's theorem, but first remind the reader of an elementary fact from algebra.

Lemma 3.8.3. *If $B : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a symmetric bilinear mapping with the property that $B^n(u, v) = 0$ whenever $\langle u, v\rangle = 0$, then there is a real number σ such that $B(u, v) = \sigma\langle u, v\rangle$ for all u and v in $\mathbb{R}^n$.*

Proof. Let u and v be vectors in $\mathbb{R}^n$. We have

$$B^n(u, v) = \sum_{i,j=1}^{n} u_i v_j B^n(e_i, e_j) = \sum_{i=1}^{n} u_i v_i B^n(e_i, e_i) \ ,$$

for $\langle e_i, e_j\rangle = 0$ if $i \neq j$. It is also true that $\langle e_i - e_j, e_i + e_j\rangle = 0$, whence

$$0 = B^n(e_i - e_j, e_i + e_j) = B^n(e_i, e_i) - 2B^n(e_i, e_j) - B^n(e_j, e_j) = B^n(e_i, e_i) - B^n(e_j, e_j) \, .$$

The inference is that $B^n(e_i, e_i) = B^n(e_j, e_j)$ when $1 \leq i, j \leq n$. Letting $\sigma = B^n(e_1, e_1)$, we see that $B^n(u, v) = \sigma\sum_{i=1}^{n} u_i v_i = \sigma\langle u, v\rangle$ for all u and v in $\mathbb{R}^n$. ☐

Lemmas 3.8.1, 3.8.2, and 3.8.3 now combine to supply an important ingredient in the proof of Liouville's theorem for C^4-mappings. It establishes that the Jacobian of our conformal mapping f has the same form as the Jacobian of a Möbius transformation.

LEMMA 3.8.4. *Under the hypotheses of Lemma* 3.8.1 *there exist a point b in $\mathbb{R}^n$ and real numbers α and β with $\alpha^2+\beta^2>0$ such that $\rho(x)=\beta|x-b|^2+\alpha$ for every x in D.*

PROOF. For fixed x in D the function $B:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ defined by $B^n(u,v)=\partial_{u,v}\rho(x)$ is a symmetric bilinear mapping with the property that $B^n(u,v)=0$ whenever $\langle u,v\rangle=0$ (Lemma 3.8.1). We apply Lemma 3.8.3 at each point of D to obtain a function $\sigma:D\to\mathbb{R}$ such that $\partial_{u,v}\rho=\sigma\langle u,v\rangle$ in D for all u and v in $\mathbb{R}^n$. In particular, taking $u=v=e_i$ discloses that $\partial_{i,i}\rho=\sigma$ for $1\le i\le n$, while the choice $u=e_i$, $v=e_j$ reveals that $\partial_{i,j}\rho=0$ for $i\ne j$. The identity $\partial_{i,i}\rho=\sigma$ shows that σ belongs to $C^1(D)$. Furthermore, given $1\le i\le n$, we can choose j different from i and compute that

$$\partial_i\sigma=\partial_i\partial_{j,j}\rho=\partial_{i,j,j}\rho=\partial_{j,i,j}\rho=\partial_j\partial_{i,j}\rho=0\,.$$

It is precisely at this point where the assumption that f is of class C^4 comes into play, for the calculation requires that ρ be a C^3-function.

It follows that σ is a constant function, say $\sigma(x)=2\beta$ for every x in D. Thus $\partial_{i,i}\rho=2\beta$ for $i=1,2,\dots,n$. Assume first that $\beta\ne0$. In conjunction with the knowledge that $\partial_{i,j}\rho=0$ when $i\ne j$, the relation $\partial_{i,i}\rho=2\beta$ forces $\nabla\rho$ to have the form

$$\nabla\rho(x)=(2\beta x_1+c_1,2\beta x_2+c_2,\dots,2\beta x_n+c_n)=2\beta x+c$$

for some $c=(c_1,c_2,\dots,c_n)$ in $\mathbb{R}^n$. This can be rewritten as $\nabla\rho(x)=2\beta(x-b)$ by setting $b=-(2\beta)^{-1}c$. In this case it becomes clear that ρ takes the form $\rho(x)=\beta|x-b|^2+\alpha$ for some constant α.

If, on the other hand, $\beta=0$, then we find that $\Delta\rho=0$ in D. Lemma 3.8.2 mandates that $\nabla\rho=0$ in D, which tells us that ρ is constant there. Suppose that $\rho(x)=\alpha$ for all x in D. Then $\alpha>0$ and, once again, $\rho(x)=\beta|x-b|^2+\alpha$ for any b in $\mathbb{R}^n$. □

The following simple, but interesting, geometric observation plays a key role in Nevanlinna's argument.

LEMMA 3.8.5. *Let c be a point of $\mathbb{R}^n$, and let $\gamma:[t_0,t_1]\to\mathbb{R}^n\setminus\{c\}$ be a C^1-path with the property that for each t in the interval (t_0,t_1) the vector $\dot\gamma(t)$ is a normal vector to the sphere $\mathbb{S}^{n-1}(c,|\gamma(t)-c|)$. Then the trajectory of γ lies on a ray issuing from c.*

PROOF. Since the hypotheses and conclusion of the lemma are translation invariant, there is no loss of generality in assuming that $c=0$. The radial projection $P:\mathbb{R}^n\setminus\{0\}\to\mathbb{S}^{n-1}$ given by $P(x)=|x|^{-1}x$ is a C^∞-function. Moreover, for any x in $\mathbb{R}^n\setminus\{0\}$ and any scalar multiple h of x it is true that $P(x+sh)=P(x)$ whenever $|s|$ is sufficiently small, so

$$P'(x)h=\lim_{s\to0}\frac{P(x+sh)-P(x)}{s}=0\,.$$

Consider the path $\beta=P\circ\gamma$. By assumption $\dot\gamma(t)$ is a scalar multiple of $\gamma(t)$ for each t in (t_0,t_1). As a result, $\dot\beta(t)=P'[\gamma(t)]\dot\gamma(t)=0$ on (t_0,t_1). This makes β constant on $[t_0,t_1]$. This simply means that every point on the trajectory of γ projects to the same point of $\mathbb{S}^{n-1}$. The stated conclusion follows. □

The final lemma in this section represents a minor extension of Lemmas 3.4.2 and 3.4.3. It is simply a formulation of analytic continuation for Möbius transformations, while the infinitesimal version relies on Lemma 3.4.2.

LEMMA 3.8.6. *Let D be a domain in $\mathbb{R}^n$, and let $f : D \to \mathbb{R}^n$ be a function with the following property: for each point x of D there exist an open ball B_x centered at x and a transformation g_x in $\boldsymbol{E}(n)$ such that f and g_x coincide in B_x. Then f is the restriction to D of a unique member g of $\boldsymbol{E}(n)$.*

Consequently f is the restriction to D of a unique $g \in \boldsymbol{E}(n)$ if f is an injective mapping such that

$$\lim_{y \to x} |f(y) - f(x)|/|y - x| = 1$$

for each x in D.

PROOF. We initially prove the first part of the lemma. Fix x_0 in D and write g for g_{x_0}. We claim that $f = g$ in D. To see this consider $U = \{x \in D : g_x = g\}$. The set U is nonempty, for x_0 belongs to U. If x is in U and y in B_x, then $g_y = f = g_x = g$ in $B_y \cap B_x$, a nonempty open set. By Corollary 3.3.11, $g_y = g$. This implies that y is an element of U, i.e., B_x is contained in U. We have just demonstrated that U is an open set in which f and g coincide. If x is a point of $\bar{U} \cap D$, then $B_x \cap U$ is a nonempty open set in which $g_x = f = g$. Calling on Corollary 3.3.11 a second time, we conclude that $g_x = g$, so x lies in U. In other words, U is relatively closed in D. The connectedness of D dictates that $U = D$, whence $f = g|D$.

To see the second part of the lemma we simply appeal to Lemma 3.4.2. □

3.8.3. Liouville's theorem for C^4-mappings. All necessary pieces are now in place for the proof of the C^4-smooth version of Liouville's theorem. We recall that a conformal mapping $f : D \to \mathbb{R}^n$ is simply an injective function whose derivative is pointwise a scalar multiple of an orthogonal transformation, $f'(x) = \lambda(x)O(x)$, $\lambda(x) \geq 0$, and $O : D \to \mathbf{SO}(n)$.

THEOREM 3.8.7. *If D is a domain in $\mathbb{R}^n$ with $n \geq 3$ and $f : D \to \mathbb{R}^n$ is a conformal diffeomorphism that belongs to the class $C^4(D, \mathbb{R}^n)$, then f is the restriction to D of a Möbius transformation of $\hat{\mathbb{R}}^n$.*

PROOF. According to Lemma 3.8.4 there exist a point b in $\mathbb{R}^n$ and real constants α and β, not both zero, such that

$$\|f'(x)\| = \frac{1}{\beta|x - b|^2 + \alpha}$$

for every x in D. Lemma 3.8.4 applies equally well to $h = f^{-1}$, which is a conformal C^4-mapping of the domain $D' = f(D)$ onto D. Thus we can assert the existence of c in $\mathbb{R}^n$ and real numbers γ and δ with $\gamma^2 + \delta^2 > 0$ such that

$$\|h'(y)\| = \frac{1}{\gamma|y - c|^2 + \delta}$$

for every y in D'. Because $\|h'[f(x)]\| = \|f'(x)\|^{-1}$, we deduce that

$$\big(\beta|x - b|^2 + \alpha\big)\big(\gamma|f(x) - c|^2 + \delta\big) = 1 \tag{3.69}$$

whenever x belongs to D. We now distinguish two cases.

Case 1: $\beta = 0$. Here $\|f'(x)\| = \alpha^{-1}$ is constant in D. From (2.22) and the conformality of f we deduce that $L_f(x) = \ell_f(x) = \alpha^{-1}$ for each x in D, which implies that

$$\lim_{y \to x} \frac{|f(y) - f(x)|}{|y - x|} = \frac{1}{\alpha}$$

for such x. Lemma 3.8.6 tells us that αf must be the restriction to D of a Euclidean isometry from $\boldsymbol{E}(n)$. As a result, we discover that f is the restriction to D of a similarity transformation of $\mathbb{R}^n$, and hence the restriction to D of a Möbius transformation.

Case 2: $\beta \neq 0$. Identity (3.69) requires that $\gamma \neq 0$. This identity has another crucial implication: if $\mathbb{S}^{n-1}(b, t) \cap D \neq \emptyset$, then the image of this intersection under f lies on the sphere $\mathbb{S}^{n-1}(c, r)$, where r is an algebraic function of t (to be precise, $r = r(t) = \{[(\beta t^2 + \alpha)^{-1} - \delta]/\gamma\}^{1/2}$). We now fix a point x_1 of D such that $x_1 \neq b$ and $f(x_1) \neq c$. We then fix a point x_0 on the open line segment between x_1 and b, doing it in such a way that the closed line segment I with endpoints x_0 and x_1 is contained in D and that $f(I)$ does not contain c. Setting $t_0 = |x_0 - b|$, $t_1 = |x_1 - b|$, and $u = (x_0 - b)/|x_0 - b|$, we parametrize I as follows: $x = x(t) = b + tu$ for $t_0 \le t \le t_1$. From this we retrieve a smooth, injective parametrization of $f(I)$: $y = y(t) = f(b + tu)$ for $t_0 \le t \le t_1$. Now $\dot{x}(t) = u$ is a normal vector to $\mathbb{S}^{n-1}(b, t)$ for each t in (t_0, t_1), so the conformality of f ensures that the vector $\dot{y}(t) = f'(b + tu)u$ is normal to the sphere $\mathbb{S}^{n-1}(c, r(t)) = \mathbb{S}^{n-1}(c, |y(t) - c|)$ for such t. By Lemma 3.8.5, $f(I)$ lies on some ray emanating from c. We infer that $f(I)$ is the line segment whose endpoints are $y_0 = f(x_0)$ and $y_1 = f(x_1)$. Write $r_0 = |y_0 - c|$ and $r_1 = |y_1 - c|$. We proceed under the assumption that $r_0 < r_1$ (the case $r_1 < r_0$ is handled similarly). For $t_0 \le t \le t_1$ we obtain

$$r(t) - r_0 = |y(t) - y_0| = \int_{t_0}^{t} |\dot{y}(\tau)|\, d\tau = \int_{t_0}^{t} \|f'(b + \tau u)\|\, d\tau = \int_{t_0}^{t} \frac{d\tau}{\beta\tau^2 + \alpha}\,.$$

If $\alpha \neq 0$, the last integral does not define an algebraic function of t on $[t_1, t_2]$, whereas the dependence of $r(t)$ (hence, of $r(t) - r_0$) on t is known to be algebraic. We are thus led to conclude: in Case 2 we must have $\alpha = 0$, which in conjunction with (3.69) implies that b cannot be a point of D. By the same argument applied to $h = f^{-1}$, we see that $\delta = 0$ and that c is not a point of D'. Finally, we consider $H = R \circ f$, where R is an inversion in $\mathbb{S}^{n-1}(c, 1)$. Then $H : D \to \mathbb{R}^n$ is a conformal mapping, H belongs to $C^4(D, \mathbb{R}^n)$, and

$$\|H'(x)\| = \|R'\big[f(x)\big]f'(x)\| = \|R'\big[f(x)\big]\| \cdot \|f'(x)\| = \frac{1}{|f(x) - c|^2}\ \frac{1}{\beta|x - b|^2} = \gamma$$

for all x in D. To see this, use (3.1), (3.10), and (3.69), not forgetting that $\alpha = \delta = 0$. In other words, H is a mapping to which Case 1 applies. Accordingly, H is the restriction to D of a Möbius transformation, so $f = R^{-1} \circ H = R \circ H$ has the same property. □

For the sake of completeness, we should at least sketch an argument showing why the function defined on $[t_0, t_1]$ by $t \mapsto \int_{t_0}^{t} (\beta\tau^2 + \alpha)^{-1}\, d\tau$ is not an algebraic function when both α and β are nonzero and $\beta\tau^2 + \alpha$ has no roots in $[t_0, t_1]$. (Of course, if one were prepared to accept the fact that logarithm and inverse trigonometric functions are transcendental functions, such an argument would be unnecessary.)

Perhaps the easiest way to demonstrate this is to consider the analytic function ζ defined in an open disk B centered at t_0 and containing neither root of $\alpha z^2 + \beta$ by $\zeta(z) = \int_{t_0}^{z} (\beta\zeta^2 + \alpha)^{-1}\, d\zeta$, where the integration is carried out along any smooth path in B that has initial point t_0 and terminal point z. The function element (ζ, B) can be continued analytically along all paths in the complex plane that avoid the singularities of $(\beta z^2 + \gamma)^{-1}$. These singularities are exactly two simple poles at points z_1 and z_2. The continuation of (ζ, B) along a smooth path γ that starts and ends at t_0 and has winding numbers $n(\gamma, z_1) = k$ and $n(\gamma, z_2) = 0$, with k an arbitrary integer, results in a function element (ζ_k, B_k) for which $\zeta_k(t_0) = ka$, where a is the residue of $(\alpha z^2 + \beta)^{-1}$ at z_1. This implies that the Riemann surface S of the complete analytic function $\underline{\zeta}$ determined by (ζ, B) is infinitely sheeted over the complex plane.

If there were a polynomial $P(z, w)$ of positive degree in w such that $P(t, \zeta(t)) = 0$ for all t in some interval $[t_0, t_0 + \epsilon]$ with $\epsilon > 0$, then $\underline{\zeta}$ would be a complete algebraic function, and S would be finitely sheeted over the complex plane. Therefore, no such polynomial P can exist under the conditions stated, so $t \mapsto \int_0^t \frac{d\tau}{\alpha\tau^2 + \beta}$ is not an algebraic function.

Over the years Liouville's theorem has yielded only grudgingly to refinement in the form of relaxed smoothness hypotheses. In 1958, more than a century after Liouville's paper, Hartman established the result for conformal mappings of class C^1 [**71**]. The theorem for conformal mappings as we have defined them was not proven until 1962, at which time F.W. Gehring showed that, when formulated in terms of conformal invariants such as moduli, the result is true without any a priori differentiability assumptions whatsoever [**46**].

Gehring's work was further generalized by Yu.G. Reshetnyak in 1967 [**141**], who removed the injectivity assumption in the result we have presented.

The best possible version of Liouville's theorem, in a well-defined sense that we shall not go into here, was proved in 1993 by T. Iwaniec and G.J. Martin, but only for even dimensions [**85**]. Whether or not their result remains valid in odd dimensions is an intriguing open question to which there are only partial answers; see [**82, 84**].

To round out the present discussion we shall state one variant of Gehring's theorem and an extremely surprising recent refinement of it due to J. Heinonen and P. Koskela [**75**], although we shall not have the means to prove it for some chapters to come.

Let D be a domain in $\mathbb{R}^n$, and let $f : D \to \mathbb{R}^n$ be a continuous injection. The *linear dilatation* of f at the point x of D is the quantity $H_f(x)$ defined by

$$H_f(x) = \limsup_{r \to 0} \frac{L_f(x, r)}{\ell_f(x, r)}, \tag{3.70}$$

where for $0 < r < \operatorname{dist}(x, \partial D)$ we set

$$L_f(x, r) = \max_{|h| = r} |f(x + h) - f(x)|, \qquad \ell_f(x, r) = \min_{|h| = r} |f(x + h) - f(x)|.$$

Plainly $1 \le H_f(x) \le \infty$. If f happens to be differentiable at x and if $f'(x)$ is nonsingular, then from (2.16) we derive the inequalities

$$L\big[f'(x)\big] - \max_{|h|=r} |\varepsilon(x+h)| \le \frac{L_f(x,r)}{r} \le L\big[f'(x)\big] + \max_{|h|=r} |\varepsilon(x+h)|$$

and

$$\ell\big[f'(x)\big] - \max_{|h|=r} |\varepsilon(x+h)| \le \frac{\ell_f(x,r)}{r} \le \ell\big[f'(x)\big] + \max_{|h|=r} |\varepsilon(x+h)| \ ,$$

which are valid whenever $0 < r < \operatorname{dist}(x, \partial D)$. From this it is relatively straightforward to establish that

$$H_f(x) = \frac{L[f'(x)]}{\ell[f'(x)]} = H\big[f'(x)\big]. \tag{3.71}$$

In particular, $H_f(x) = 1$ would hold for every x in D, should f be a conformal diffeomorphism of D. The simplest rendering of Gehring's theorem reads:

THEOREM 3.8.8. *If D is a domain in $\mathbb{R}^n$ with $n \ge 3$ and $f : D \to \mathbb{R}^n$ is a continuous injection with the property that $H_f(x) = 1$ for every x in D, then f is the restriction to D of a Möbius transformation of $\hat{\mathbb{R}}^n$.*

The addition that the aforementioned work of Heinonen and Koskela would make would be to replace the linear distortion function $H_f(x)$ in the above theorem by the possibly smaller function

$$\underline{H}_f(x) = \liminf_{r \to 0} \ \frac{L_f(x,r)}{\ell_f(x,r)} \ . \tag{3.72}$$

The paucity of conformal mappings in dimensions three and above dictates that a reasonable surrogate for conformality be found, lest the incredible richness, flexibility, and utility of plane conformal mapping theory be totally lost in the passage to higher dimensions. The trick is to produce a theory with an abundance of mappings, yet not to retreat so far from conformality that many of the desirable geometric and analytic features of the plane theory are sacrificed in the process. The theory of quasiconformal mappings, of which we shall give an account in the remaining chapters of this book, would seem to strike a very happy medium in this regard.

CHAPTER 8

The Tukia-Väisälä Extension Theorem

In this chapter we return to a question that raised itself earlier in this book.

Given a quasiconformal self-mapping f of $\mathbb{R}^n$—when $n = 1$ and for convenience we use "quasiconformal" as a synonym for "quasisymmetric"—we asked: "Does there exist a quasiconformal mapping F of $\mathbb{H}^{n+1}$ onto itself whose homeomorphic extension F^* to $\bar{\mathbb{H}}^{n+1}$ coincides with f in $\mathbb{R}^n$?"

The case $n = 1$ of this problem was settled by Lars Ahlfors and Arne Beurling in a classic paper from 1956, [**11**]. Surprisingly the solution involved an explicit integral formula for a mapping F with the desired properties.

In the situation where f is increasing on $\mathbb{R}$, the *Beurling-Ahlfors extension* of f takes the form

$$F(x, y) = (u(x, y), v(x, y)), \tag{8.1}$$

with

$$u(x, y) = \frac{1}{2}\int_0^1 \big[f(x + ty) + f(x - ty)\big]\, dt$$

and

$$v(x, y) = \frac{1}{2}\int_0^1 \big[f(x + ty) - f(x - ty)\big]\, dt\,.$$

As a matter of fact, the Beurling-Ahlfors extension F of a quasisymmetric map f is a quasiconformal diffeomorphism of $\mathbb{H}^2$ and can be modified so as to show that any quasisymmetric homeomorphism of the real line actually admits a real-analytic quasiconformal extension.

In 1964 Ahlfors provided an affirmative response to the extension question in dimension two, [**6**].

His answer consisted in first giving an integral representation for a quasiconformal extension to $\mathbb{H}^3$ of any K-quasiconformal self-mapping of $\mathbb{R}^2$ with K suitably close to 1. He then appealed to the fact that an arbitrary quasiconformal mapping of $\mathbb{R}^2$ to itself can be realized as the finite composition of $(1 + \varepsilon)$-quasiconformal mappings of $\mathbb{R}^2$ for any preassigned $\varepsilon > 0$. It is still unknown whether a factorization of this sort is possible for quasiconformal mappings of $\mathbb{R}^n$ for dimensions $n \geq 3$.

By combining renormalization arguments with some piecewise linear topology, L. Carleson dealt with the case $n = 3$ in 1974, at the same time contributing new proofs for the lower-dimensional cases [**28**]. Carleson's treatment of the three-dimensional situation hinged on a piecewise linear approximation result of E. Moise [**116**] that does not carry over to dimension four.

It took another 10 years before the general case of the problem was solved by P. Tukia and J. Väisälä [**158**]. They adopted and refined Carleson's renormalization technique, but they replaced the elements of piecewise linear topology in his approach with the more flexible Lipschitz methods that D. Sullivan had been developing over the previous years in order to solve other outstanding problems in the area. These methods are carefully developed and explained in [**157**].

The presentation of the Tukia-Väisälä extension theorem that follows is a blend of material from their 1982 paper [**158**] with simplifications suggested in their 1984 sequel [**159**]. The proof we give is complete apart from Sullivan's approximation theorem. This theorem follows the lines of the "furling" technique from geometric topology developed by R. Edwards and R. Kirby [**35**] and Kirby and L. Siebenmann [**92**]. In that case a key idea involves using that fact that the n-torus minus a point $\mathbb{S} \times \mathbb{S} \times \cdots \times \mathbb{S} \setminus \{x\}$ immerses in $\mathbb{R}^n$. Sullivan mimics that construction but uses hyperbolic geometry. For this he needs the existence of a compact hyperbolic n-manifold which immerses in $\mathbb{R}^n$ once a point is deleted. The technical term is parallelizable in the complement of a point. The existence of these manifolds is quite nontrivial—joint work of Sullivan and J.P. Serre—and obviously well outside the scope of this book, and so we leave the issue well alone (as did Tukia and Väisälä).

8.1. Lipschitz embeddings

As we have just suggested, the keystone of the Tukia-Väisälä scheme for extending a quasiconformal mapping from one dimension to the next is Dennis Sullivan's work on Lipschitz embeddings [**150, 157**].

Unfortunately, and as noted above, the limited scope and confines of this book do not afford us the luxury of including an in-depth account of Sullivan's ideas. We must be content to summarize and accept those of his results that are pertinent to the matter at hand, and to proceed from there.

Let X and X' be metric spaces whose distance functions we denote by d and d', respectively. A function $f : X \to X'$ is termed *bilipschitz* provided there is a constant $\lambda \geq 1$ such that

$$\lambda^{-1} d(p,q) \leq d'\big[f(p), f(q)\big] \leq \lambda d(p,q)$$

for all points p and q of X. If we wish to be explicit about λ, we shall refer to f as *λ-bilipschitz*.

We say that f is *locally λ-bilipschitz* when each point p of X has an open neighbourhood $U = U_p$ in X with the property that the restriction of f to U is λ-bilipschitz.

Assume now that $f : X \to X'$ is an embedding, meaning that f is a homeomorphism of X onto $f(X)$. We call f a *Lipschitz embedding*—or, to use the language favoured by those who regularly work with such mappings, a *LIP-embedding*—if each point p of X has an open neighbourhood $U = U_p$ in X such that the restriction of f to U is bilipschitz but not necessarily λ-bilipschitz with fixed λ.

We stress that a LIP-embedding is not required in this usage of the term to satisfy a global Lipschitz condition on X—the concept is purely a local one. A *LIP-homeomorphism* from X onto X' is a surjective LIP-embedding from X to X'.

Suppose that $f : X \to X'$ and $g : X' \to X''$ are LIP-embeddings. Then $g \circ f$ and f^{-1}, which is defined in $f(X)$, are also LIP-embeddings. If A is a relatively compact subset of X, then the restriction of f to A is locally λ-bilipschitz for some λ that is determined by f and A.

Almost exclusively we shall be concerned with Euclidean LIP-embeddings $f : U \to \mathbb{R}^n$, where U is an open set in $\mathbb{R}^n$, and hyperbolic LIP-embeddings $f : U \to \mathbb{H}^n$, where U is an open set in $\mathbb{H}^n$. We register in the form of lemmas some preparatory remarks about such mappings. The first is an immediate consequence of the following fact from (3.41): if E is a compact set in $\mathbb{H}^n$, then there are constants $a > 0$ and $b > 0$, which depend only on E, such that

$$a|x-y| \le d_{\mathbb{H}}(x,y) \le b|x-y| \tag{8.2}$$

whenever x and y are points of E.

We remind the reader that, by the notational conventions we have used, the symbol $\bar{A}$ for a subset A of $\mathbb{R}^n$ means the closure of A in $\hat{\mathbb{R}}^n$. Thus $\bar{A}$ is always a compact set.

LEMMA 8.1.1. *Let U be a nonempty open set in $\mathbb{H}^n$ such that $\bar{U}$ lies in $\mathbb{H}^n$ and let $f : U \to E$ be an embedding in which E is a compact subset of $\mathbb{H}^n$. Then f is (locally) λ-bilipschitz for some λ with respect to the Euclidean metric if and only if f is (locally) λ'-bilipschitz for some λ' with respect to the hyperbolic metric. In fact, if λ is known, then an appropriate λ' is determined by λ, U, and E, and vice versa.*

In particular, an embedding $f : V \to \mathbb{H}^n$, where V is an open set in $\mathbb{H}^n$, is a Euclidean LIP*-embedding if and only if it is a hyperbolic LIP-embedding.*

The second lemma brings to the hyperbolic context a fact we were already aware of in the Euclidean setting.

LEMMA 8.1.2. *Suppose that D is a subdomain of $\mathbb{H}^n$ and that f is an embedding of D into $\mathbb{H}^n$ which is locally λ-bilipschitz with respect to the hyperbolic metric. Then f is a λ^{2n-2}-quasiconformal mapping of D onto $D' = f(D)$.*

PROOF. It suffices to check that $f|B$ is λ^{2n-2}-quasiconformal whenever B is an open ball such that $\bar{B}$ is contained in D and $f|B$ is λ-bilipschitz in the hyperbolic metric. Fix a ball B of this type. Lemma 8.1.1 implies that $f|B$ is bilipschitz with respect to the Euclidean metric and hence quasiconformal by Lemma 6.2.3.

For any point x of B at which f is differentiable with $J_f(x) \neq 0$, we compute

$$\begin{aligned}
\|f'(x)\| &= \limsup_{h\to 0} \frac{|f(x+h)-f(x)|}{|h|} \\
&= \limsup_{h\to 0} \left(\frac{|f(x+h)-f(x)|}{d_{\mathbb{H}}[f(x+h),f(x)]} \cdot \frac{d_{\mathbb{H}}[f(x+h),f(x)]}{d_{\mathbb{H}}(x+h,x)} \cdot \frac{d_{\mathbb{H}}(x+h,x)}{|h|} \right) \\
&\le \frac{\lambda f_n(x)}{x_n}\,.
\end{aligned}$$

Similarly, for any such x we obtain

$$\ell\big[f'(x)\big] \ge \frac{f_n(x)}{\lambda x_n}\ ,$$

whence

$$H_f(x) = H\big[f'(x)\big] = \frac{\|f'(x)\|}{\ell[f'(x)]} \le \lambda^2 .$$

Since the last inequality is true for almost every point x of B, Corollary 6.4.19 yields $K(f|B) \le \lambda^{2n-2}$, as desired. □

Finally, we point out one situation in which a locally λ-bilipschitz mapping is automatically λ-bilipschitz on a global scale.

LEMMA 8.1.3. *Suppose that f is a homeomorphism of a hyperbolically convex subdomain D of $\mathbb{H}^n$ onto a domain D' of the same type. If f is locally λ-bilipschitz with respect to the hyperbolic metric, then f is λ-bilipschitz in that metric.*

PROOF. Let x and y be distinct points of D, and let A be the hyperbolic geodesic segment joining x and y. We choose points $x = z_0, z_1, \ldots, z_{r-1}, z_r = y$ in sequence along A in such a way that f is hyperbolically λ-lipschitz on each of the hyperbolic segments into which these points partition A. Then

$$d_{\mathbb{H}}\big[f(x), f(y)\big] \le \sum_{j=1}^{r} d_{\mathbb{H}}\big[f(x_{j-1}), f(x_j)\big] \le \lambda \sum_{j=1}^{r} d_{\mathbb{H}}(x_{j-1}, x_j) = \lambda d_{\mathbb{H}}(x, y) .$$

We conclude that f is hyperbolically λ-lipschitz in D. As the same is true of f^{-1} in D', f is a λ-bilipschitz mapping with respect to the hyperbolic metric. □

Lemma 8.1.3 has an obvious Euclidean analogue.

8.1.1. Sullivan's theorem. Sullivan's theory of LIP-embeddings contains much more information than is actually required to deal with the extension problem for quasiconformal mappings.

Following Tukia and Väisälä's presentation of this material, we state—and subsequently treat as a given—a quantitative version of the particular Sullivan result that gives precisely the information we need. A few words concerning notation: if S is a set and $\varphi : S \to \mathbb{R}^n$, then for any nonempty subset A of S we set $\|\varphi\|_A = \sup\{|\varphi(x)| : x \in A\}$; we denote the identity mapping of $\mathbb{R}^n$ by *id*.

THEOREM 8.1.4. *Suppose that U is an open set in $\mathbb{R}^n$, C is a compact subset of U, V is a neighbourhood of C in U, and $\varepsilon > 0$. Then there is a $\delta > 0$, which depends only on U, C, V, and ε, such that for each* LIP-*embedding $f : V \to \mathbb{R}^n$ satisfying $\|f - \text{identity}\|_V \le \delta$, there exists a* LIP-*homeomorphism $f^* : \mathbb{R}^n \to \mathbb{R}^n$ endowed with the following properties:*

(i) $\|f^* - \text{identity}\|_{\mathbb{R}^n} \le \varepsilon$;
(ii) $f^* = f$ *on* C;
(iii) $f^* = \text{identity}$ *in* $\mathbb{R}^n \setminus V$.

Furthermore, if f is locally λ-bilipschitz, then f^* can be chosen so that it is λ^*-bilipschitz for some λ^* which is solely a function of λ and n.

Theorem 8.1.4 and its proof are found in [**157**], which contains a wealth of material related to the quasiconformal extension problem and also provides a reasonably self-contained exposition of the requisite Sullivan theory.

We shall say that an embedding $g : U \to \mathbb{R}^n$, U being an open set in $\mathbb{R}^n$, is *uniformly approximable by* LIP-*embeddings* if for each $\varepsilon > 0$ there exists a LIP-embedding $h : U \to \mathbb{R}^n$ for which $\|h - g\|_U < \varepsilon$.

We apply Theorem 8.1.4 to derive a rather technical—but, as things turn out, crucial—approximation and smoothing result.

Lemma 8.1.5. *Let U, U_0, V and W be open sets in $\mathbb{H}^n$ that exhibit the following relationships:*

$$\bar{U} \subset \mathbb{H}^n \ , \quad W \subset V \subset U \ , \quad U \cap \bar{W} \subset V \ , \quad \emptyset \neq \bar{U}_0 \subset U \, .$$

Assume that $\mathcal{G}$ is a family of embeddings $g : U \to \mathbb{H}^n$ which are uniformly approximable by LIP*-embeddings. Assume, further, that $\mathcal{G}$ is compact in the topology of Euclidean uniform convergence on U and that*

$$E = \overline{\{g(x) : g \in \mathcal{G}, \ x \in U\}} \subset \mathbb{H}^n.$$

To each $\varepsilon > 0$ there corresponds a $\delta > 0$, where δ depends on n, ε, $\mathcal{G}$, and the four sets listed above, for which the following statement is true.

If $g \in \mathcal{G}$ and $h : V \to \mathbb{H}^n$ is a LIP*-embedding satisfying $\|h - g\|_V \leq \delta$, then there exists a* LIP*-embedding $\hat{h} : U \to \mathbb{H}^n$ such that $\|\hat{h} - g\|_{U_0} \leq \varepsilon$ and such that $\hat{h} = h$ in $U_0 \cap W$.*

Moreover, if h is locally λ-bilipschitz, then $\hat{h}$ can be chosen so that its restriction to U_0 is locally $\hat{\lambda}$-bilipschitz for some $\hat{\lambda}$ determined by λ, n, ε, $\mathcal{G}$ and the sets U, U_0, V, and W.

Proof. Fix open sets A and B in U for which

$$\bar{U}_0 \subset A \subset \bar{A} \subset B \subset \bar{B} \subset U \, .$$

This can be done in a manner which makes A and B completely dependent on U and U_0. For instance, $A = \{x : \operatorname{dist}(x, U_0) < r\}$ and $B = \{x : \operatorname{dist}(x, U_0) < 2r\}$, where $r = \operatorname{dist}(U_0, \partial U)/4)$.

Since the family $\mathcal{G}$ is compact and since the function $g \mapsto \operatorname{dist}[g(\bar{A}), \partial g(B)]$ is positive and continuous on $\mathcal{G}$, we can be certain that

$$m = \min\Big\{ d\big[g(\bar{A}), \partial g(B)\big] : g \in \mathcal{G} \Big\} > 0 \, . \tag{8.3}$$

Let $\varepsilon > 0$ be given. The Arzelá-Ascoli theorem ensures that $\mathcal{G}$ is equicontinuous at each point of U—hence, uniformly equicontinuous on the compact set $\bar{B}$. In particular, we can choose η in $(0, d(\bar{A}, \partial B))$ such that

$$|g(x) - g(y)| \leq \varepsilon/4 \tag{8.4}$$

whenever g belongs to $\mathcal{G}$, x is a point of A, and y satisfies $|y - x| \leq \eta$.

The choice of η is controlled completely by ε, $\mathcal{G}$, U, and U_0. Next we apply Theorem 8.1.4 to the open set $A \cap V$, its compact subset $C = \bar{U}_0 \cap \bar{W}$, the neighbourhood $A \cap V$ of C, and the number $\eta > 0$ to obtain $\zeta > 0$ for which the following assertion is valid: if $f : A \cap V \to \mathbb{R}^n$ is a LIP-embedding with $\|f - \text{identity}\|_{A \cap V} \leq \zeta$, then there exists a LIP-homeomorphism f^* of $\mathbb{R}^n$ onto itself such that

$$\|f^* - \text{identity}\|_{\mathbb{R}^n} \leq \eta \ , \tag{8.5}$$

$$f^* = f \quad \text{in} \quad \bar{U}_0 \cap \bar{W} \tag{8.6}$$

and

$$f^* = \text{identity} \quad \text{in} \quad \mathbb{R}^n \setminus (A \cap V) \, . \tag{8.7}$$

In fact (8.7) carries with it the implication that any such f^* has $f^*(A) = A$ and $f^*(U) = U$.

The ζ arising here is determined by η, U, U_0, V, and W. Thus, ζ is dependent on ε, $\mathcal{G}$, and the sets appearing in the statement of the lemma. If, in addition, f is known to be locally λ-bilipschitz, then we may assume that f^* is λ^*-bilipschitz for some λ^* which is a function of λ and n.

Finally, we choose δ in $(0, m/4)$ so that the following condition is met:

$$|g^{-1}(w) - g^{-1}(z)| \leq \zeta \tag{8.8}$$

whenever g is in $\mathcal{G}$, z is in $g(A)$, and $|w - z| < 4\delta$. Note that by (8.3) any such w lies in $g(B)$, and if this were not possible, we could find sequences $\langle g_\nu \rangle$ in $\mathcal{G}$, $\langle x_\nu \rangle$ in A, and $\langle y_\nu \rangle$ in B for which $|y_\nu - x_\nu| > \zeta$, yet for which $|g_\nu(y_\nu) - g_\nu(x_\nu)| < d/(4\nu + 1)$. Passing to subsequences, we could assume that $g_\nu \to g$ uniformly on U, where g is a member of $\mathcal{G}$, while $x_\nu \to x$ and $y_\nu \to y$, with x and y being points of $\bar{B}$ with $|x - y| \geq \zeta$. Thus $x \neq y$, whereas

$$g(x) = \lim_{\nu\to\infty} g_\nu(x_\nu) = \lim_{\nu\to\infty} g_\nu(y_\nu) = g(y) ,$$

contradicting the fact that g is an embedding. We can therefore choose δ with the stated property.

We claim that this δ, which depends only on the parameters specified, satisfies the requirements of the lemma.

The compactness of $\mathcal{G}$ entitles us to select a finite set of mappings $g_1, g_2, \ldots, g_p$ from $\mathcal{G}$ with the feature that

$$\min_{1\leq j\leq p} \|g - g_j\|_U < \min\{\varepsilon/8,\ \delta/2\} \tag{8.9}$$

whenever g belongs to $\mathcal{G}$. If we assume, as we may, that the number p of mappings involved here is the minimal one for which relation (8.9) can hold, then p and the mappings $g_1, g_2, \ldots, g_p$ are controlled by $\mathcal{G}$, ε, and δ—hence, by the proper data. By hypothesis we are free to fix for $j = 1, 2, \ldots, p$ a LIP-embedding $\tilde{g}_j : U \to \mathbb{R}^n$ such that

$$\|\tilde{g}_j - g_j\|_U < \min\bigl\{\varepsilon/8,\ \delta/2,\ \mathrm{dist}(E, \partial\mathbb{H}^n)\bigr\} . \tag{8.10}$$

Thus $\tilde{g}_j(U)$ lies in $\mathbb{H}^n$ and, as B is a relatively compact subset of U, we can fix a $\tilde{\lambda}$ for which all of the mappings $\tilde{g}_j$ become locally $\tilde{\lambda}$-bilipschitz when restricted to B.

Now consider a member g of $\mathcal{G}$ and a LIP-embedding $h : V \to \mathbb{H}^n$ that satisfies the condition $\|h - g\|_V \leq \delta$. In view of (8.9) and (8.10) we can pick a j such that $\tilde{g} = \tilde{g}_j$ has

$$\|\tilde{g} - g\|_U < \min\{\varepsilon/4,\ \delta\} . \tag{8.11}$$

Moreover,

$$|\tilde{g}^{-1}(w) - \tilde{g}^{-1}(z)| \leq \zeta \tag{8.12}$$

whenever z is a point of $\tilde{g}(A)$ and $|w - z| \leq 2\delta$. To see this, observe that by (8.11)

$$d\bigl[\tilde{g}(\bar{A}), \partial\tilde{g}(B)\bigr] > d\bigl[g(\bar{A}), \partial g(B)\bigr] - 2\delta \geq m - 2\delta > 4\delta - 2\delta = 2\delta .$$

In particular, $\tilde{g}(B)$ includes any point w within a distance 2δ of $\tilde{g}(A)$. Let $z = \tilde{g}(x)$ with x from A, and let w satisfy $|w - z| \leq 2\delta$. Then $w = \tilde{g}(y)$ for a point y of B, and

$$|g(y) - g(x)| \leq |g(y) - \tilde{g}(y)| + |\tilde{g}(y) - \tilde{g}(x)| + |\tilde{g}(x) - g(x)| < 4\delta .$$

By the choice of δ,

$$|\tilde{g}^{-1}(w) - \tilde{g}^{-1}(z)| = |y - x| = |g^{-1}[g(y)] - g^{-1}[g(x)]| \leq \zeta\,.$$

On the basis of the information that

$$\|\tilde{g} - h\|_{V\cap A} \leq \|\tilde{g} - g\|_U + \|g - h\|_V < 2\delta\ ,$$

we conclude that $h(A\cap V)$ is contained in $\tilde{g}(B)$. It follows that $f = \tilde{g}^{-1}\circ h \mid A\cap V$ is a well-defined LIP-embedding and that, because of (8.12),

$$|f(x) - x| = |\tilde{g}^{-1}[h(x)] - \tilde{g}^{-1}[\tilde{g}(x)]| \leq \zeta$$

for every x in $A\cap V$; i.e., $\|f - \text{identity}\|_{A\cap V} \leq \zeta$.

We can thus select a LIP-homeomorphism $f^* : \mathbb{R}^n \to \mathbb{R}^n$ for which (8.5), (8.6), and (8.7) hold. If h happens to be locally λ-bilipschitz, then f is locally $(\tilde{\lambda}\lambda)$-bilipschitz, so we may take f^* to be λ^*-bilipschitz for some λ^* dependent only on λ, $\tilde{\lambda}$, and n. As noted earlier, $f^*(U) = U$ and $f^*(A) = A$. We infer that $\hat{h} = \tilde{g}\circ f^*$ is a well-defined LIP-embedding of U into $\mathbb{H}^n$ which, in the case of a locally λ-bilipschitz h, is locally $\hat{\lambda}$-bilipschitz in A for $\hat{\lambda} = \tilde{\lambda}\lambda^*$, a number depending only on the allowable parameters. For x in $U_0\cap W$ we conclude on the grounds of (8.6) and the definition of f that

$$\hat{h}(x) = \tilde{g}[f^*(x)] = \tilde{g}[f(x)] = h(x)\,.$$

In other words, $\hat{h} = h$ in $U_0\cap W$.

To complete the proof we must only verify that $\|\hat{h} - g\|_{U_0} \leq \varepsilon$. The inequalities (8.5) and (8.4) imply that

$$|g[f^*(x)] - g(x)| \leq \varepsilon/4$$

for every x in A, meaning that $\|g\circ f^* - g\|_A \leq \varepsilon/4$. From this fact, from the knowledge that $f^*(A) = A$, and from (8.11) we learn that

$$\begin{aligned}\|\hat{h} - g\|_A &\leq \|\hat{h} - \tilde{g}\|_A + \|\tilde{g} - g\|_A = \|\tilde{g}\circ f^* - \tilde{g}\|_A + \|\tilde{g} - g\|_A\\ &\leq \|\tilde{g}\circ f^* - g\circ f^*\|_A + \|g\circ f^* - g\|_A + 2\|\tilde{g} - g\|_A\\ &= \|\tilde{g} - g\|_A + \|g\circ f^* - g\|_A + 2\|\tilde{g} - g\|_A \leq \varepsilon\,.\end{aligned}$$

Since U_0 is contained in A, $\|\hat{h} - g\|_{U_0} \leq \varepsilon$. □

As we intend to apply Lemma 8.1.5 in a hyperbolic setting, a reformulation appropriate to that context will prove convenient. We use the following notation: if S is a set and if φ and ψ are functions from S into $\mathbb{H}^n$, then for any nonempty subset A of S we define

$$d_{\mathbb{H}}(\varphi,\psi;A) = \sup_{x\in A} d_{\mathbb{H}}\big[\varphi(x),\psi(x)\big]\,.$$

In light of (8.2) it is clear that if $\varphi(A)$ and $\psi(A)$ are subsets of a compact set E in $\mathbb{H}^n$, then

$$a\|\varphi - \psi\|_A \leq d_{\mathbb{H}}(\varphi,\psi;A) \leq b\|\varphi - \psi\|_A \tag{8.13}$$

for constants $a > 0$ and $b > 0$ that depend only on E. The result we seek is an immediate consequence of Lemma 8.1.5, (8.13), and Lemma 8.1.1.

LEMMA 8.1.6. *Let U, U_0, V, W and $\mathcal{G}$ be as in Lemma* 8.1.5. *To each $\varepsilon > 0$ there corresponds a $\delta > 0$, depending on ε and the items just listed, for which the following statement holds:*

If $g \in \mathcal{G}$ and $h : V \to \mathbb{H}^n$ is a LIP*-embedding satisfying $d_{\mathbb{H}}(h, g; V) \leq \delta$, then there exists a* LIP*-embedding $\hat{h} : U \to \mathbb{H}^n$ such that $d_{\mathbb{H}}(\hat{h}, g; U_0) \leq \varepsilon$ and such that $\hat{h} = h$ in $U_0 \cap W$. Moreover, if h is locally λ-bilipschitz with respect to the hyperbolic metric, then $\hat{h}$ can be chosen so that its restriction to U_0 is locally $\hat{\lambda}$-bilipschitz relative to the hyperbolic metric, where $\hat{\lambda}$ is determined by λ, n, ε, $\mathcal{G}$, and the sets U, U_0, V, and W.*

8.2. Preliminaries

We now fix the dimension $n \geq 1$ and keep it fixed for the rest of this chapter. In this section we put together a number of preliminary results.

8.2.1. The extension F_f. Given a domain D in $\mathbb{R}^n$ and a number $r \geq 1$ we write $D^*(r)$ for the subset of $\mathbb{R}^{n+1}_+ = \mathbb{R}^n \times [0, \infty)$ defined as follows:

$$D^*(r) = \left\{(x,t) : x \in D\ ,\ 0 \leq t < r^{-1}\mathrm{dist}(x, \partial D)\right\}.$$

When $D = \mathbb{R}^n$ we have $\mathrm{dist}(x, \partial D) = \infty$ for every x in D, so $D^*(r) = \mathbb{R}^{n+1}_+$.

The set $D^*(r)$ is connected and is relatively open in $\mathbb{R}^{n+1}_+$. Clearly $D^*(s)$ is contained in $D^*(r)$ when $r \leq s$. For the rest of this chapter the symbol D^* will serve as an abbreviation for $D^*(1)$. Notice that if (x, t) is a point of $D^*(r)$, then the ball $\bar{B}^n(x, rt)$ is contained in D.

In particular the following observation helps clarify the geometric situation:

$$\bar{B}^n(x,t) \subset D \iff (x,t) \in D^*.$$

Observe, too, that if φ is a member of $\mathrm{Sim}(\mathbb{H}^{n+1})$, the group of similarity transformations of $\mathbb{R}^{n+1}$ which preserve $\mathbb{H}^{n+1}$ (hence, leave $\mathbb{R}^n$ and $\mathbb{R}^{n+1}_+$ invariant as well), then $\varphi(D)^* = \varphi(D^*)$.

Let $f : D \to \mathbb{R}^n$ be an embedding, where once again D is a domain in $\mathbb{R}^n$. We define a continuous function $\tau_f : D^* \to [0, \infty)$ through the rule of correspondence

$$\tau_f(x,t) = \max_{|h|=t} |f(x+h) - f(x)|\,.$$

Thus $\tau_f(x,t) = L_f(x,t) > 0$ when $t > 0$, whereas $\tau_f(x, 0) = 0$. With the aid of τ_f we can create an embedding F_f of D^* into $\mathbb{R}^{n+1}_+$:

$$F_f(x,t) = \big(f(x), \tau_f(x,t)\big)\,.$$

Certainly $F_f(x,t) \neq F_f(y,s)$ when $x \neq y$. On the other hand, if (x, t) and (x, s) lie in D^* and if $0 \leq s < t$, then $f[\bar{B}^n(x, s)]$ is a compact subset of $f[B^n(x, t)]$, which implies that $\tau_f(x,s) < \tau_f(x,t)$ and, as a result, that $F_f(x,t) \neq F_f(x,s)$. This demonstrates that the function F_f is injective. The continuity of F_f is evident, while the fact that F_f is an open mapping of D^* into $\mathbb{R}^{n+1}_+$ is readily verified. Hence F_f is an embedding.

Treating $\mathbb{R}^n$ in the standard way as a subset of $\mathbb{R}^{n+1}$ and remarking that $F_f(x, 0) = (f(x), 0)$ for x in D, we regard F_f as an extension of f from D to D^*.

The correspondence $f \mapsto F_f$ is continuous in a sense that the next lemma makes precise.

LEMMA 8.2.1. *Suppose that D is a domain in $\mathbb{R}^n$, that f and g are embeddings of D into $\mathbb{R}^n$, and that A is a domain whose closure lies in D. Then*

$$\|F_g - F_f\|_{A^*} \leq 3\|g - f\|_A \,. \tag{8.14}$$

In particular, if $\langle f_\nu \rangle$ is a sequence of embeddings of D into $\mathbb{R}^n$ that converges in the Euclidean locally uniform sense to an embedding f of D into $\mathbb{R}^n$, then $\langle F_{f_\nu} \rangle$ is locally uniformly convergent in D^* to F_f.

PROOF. Let (x, t) be a point of A^*. If $|y - x| = t$, then y belongs to A and

$$\begin{aligned} |g(y) - g(x)| &\leq |g(y) - f(y)| + |f(y) - f(x)| + |f(x) - g(x)| \\ &\leq \|g - f\|_A + \tau_f(x, t) + \|g - f\|_A \ , \end{aligned}$$

from which we infer that

$$\tau_g(x, t) \leq \tau_f(x, t) + 2\|g - f\|_A \,.$$

By similar reasoning

$$\tau_f(x, t) \leq \tau_g(x, t) + 2\|g - f\|_A \ ,$$

so

$$|\tau_g(x, t) - \tau_f(x, t)| \leq 2\|g - f\|_A \,.$$

It follows that

$$\begin{aligned} |F_g(x, t) - F_f(x, t)| &\leq |g(x) - f(x)| + |\tau_g(x, t) - \tau_f(x, t)| \\ &\leq \|g - f\|_A + 2\|g - f\|_A = 3\|g - f\|_A \end{aligned}$$

for each element (x, t) of A^*, whence

$$\|F_g - F_f\|_{A^*} \leq 3\|g - f\|_A \,.$$

The convergence statement is a direct consequence of the estimate (8.14) and the observation that each compact subset of D^* lies in A^* for some domain A whose closure is contained in D. □

The following lemma points out the naturality of the correspondence $f \mapsto F_f$ with respect to the action of the similarity group, hardly unexpected when one reflects for an instant on the definition of F_f.

LEMMA 8.2.2. *Let D be a domain in $\mathbb{R}^n$, and let f be an embedding of D into $\mathbb{R}^n$. Then*

$$F_{\psi \circ f \circ \varphi} = \psi \circ F_f \circ \varphi \tag{8.15}$$

whenever φ and ψ are similarity transformations of $\mathbb{R}^{n+1}$ that preserve $\mathbb{R}^{n+1}_+$.

PROOF. The domains of the functions appearing on the opposite sides of (8.15) are the same—namely, the set $A = \varphi^{-1}(D)^* = \varphi^{-1}(D^*)$.

Let λ and μ be the dilation factors of φ and ψ, respectively. Given a point (x, t) of A, we set $y = \varphi(x)$ and $z = f(y)$. Recalling Theorem 3.3.13, we write

$$\varphi(x, t) = (\varphi(x), \lambda t) \quad \text{and} \quad \psi(z, s) = (\psi(z), \mu s).$$

For w in $\varphi^{-1}(D)$ we see that $|w - x| = t$ if and only if $|\varphi(w) - y| = \lambda t$. Accordingly,

$$\tau_{\psi \circ f \circ \varphi}(x, t) = \mu \tau_f(y, \lambda t) \ ,$$

which leads to

$$F_{\psi \circ f \circ \varphi}(x, t) = \big(\psi(z), \mu \tau_f(y, \lambda t)\big) \,.$$

As for the other side of (8.15), we have

$$\psi \circ F_f \circ \varphi(x,t) = \psi \circ F_f(y, \lambda t) = \psi\big(z, \tau_f(y, \lambda t)\big) = \big(\psi(z), \mu\tau_f(y,\lambda t)\big) \ ,$$

which is what we wanted to show. □

8.2.2. The families $\mathcal{G}_K$ and $\mathcal{G}_K^*$. For any point z of $\mathbb{H}^{n+1}$ we let α_z denote the similarity transformation of $\mathbb{R}^{n+1}$ given by the formula

$$\alpha_z(x) = z + z_{n+1}(x - e_{n+1}) \ .$$

Put differently, α_z is the unique similarity that maps e_{n+1} to z and has the structure $x \mapsto \lambda x + b$ with $\lambda > 0$ and b in $\mathbb{R}^n$.

Suppose that D is a subdomain of $\mathbb{R}^n$. With each interior point z of D^* and each embedding $f : D \to \mathbb{R}^n$, we associate a similarity transformation $\beta_{z,f}$ belonging to $\sigma_{\mathbb{S}}(H^{n+1})$—namely,

$$\beta_{z,f} = \alpha_w^{-1}, \qquad w = F_f(z) \, .$$

To be more explicit,

$$\beta_{z,f}(x) = e_{n+1} + w_{n+1}^{-1}(x - w) \, . \tag{8.16}$$

We next establish some notation that will be in use for the remainder of this chapter. First, we set

$$r_0 = 3 + \sqrt{n}, \qquad B_0 = B^n(r_0), \qquad D_0 = \operatorname{int}(B_0^*) = B_0^* \cap \mathbb{H}^{n+1} \, .$$

The choice of r_0 ensures that the cube

$$Q = [-1,1] \times \cdots \times [-1,1] \times [0,2] \subset B_0^* \subset \mathbb{R}^{n+1}.$$

A small observation will prove its worth in upcoming constructions.

Lemma 8.2.3. *If D is a subdomain of $\mathbb{R}^n$ and z is an interior point of $D^*(r_0)$, then $\alpha_z(D_0)$ is contained in D^*.*

Proof. Write $z = (c,t)$, where c is in D and $0 < tr_0 < d(c, \partial D)$. Then

$$\alpha_z(x) = (c_1 + tx_1, \ldots, c_n + tx_n, tx_{n+1}) \, .$$

Consider a point x of D_0, say $x = (b,s)$ in which b lies in B_0 and $0 \le s < r_0 - |b|$. Notice that

$$|tb| < \frac{d(c,\partial D)}{r_0} r_0 = d(c, \partial D) \ ,$$

whence $c + tb$ belongs to D. Moreover,

$$d(c + tb, \partial D) \ge d(c, \partial D) - t|b| > tr_0 + t(s - r_0) = ts \, .$$

Therefore

$$\alpha_z(x) = (c + tb, ts) \in D^* \ ,$$

as maintained. □

When D is a domain in $\mathbb{R}^n$ and $1 \le K < \infty$ the notation $\mathcal{Q}_K(D)$ denotes the family of all K-quasiconformal embeddings of D into $\mathbb{R}^n$. Note that if $n = 1$ we take "K-quasiconformal" to be synonymous with "K-quasisymmetric".

Corresponding to each domain D in $\mathbb{R}^n$, each interior point z of $D^*(r_0)$, and each embedding f of D into $\mathbb{R}^n$, we introduce embeddings f_z of B_0 into $\mathbb{R}^n$ and $F_{z,f}$ of D_0 into $\mathbb{H}^{n+1}$ by setting

$$f_z = \beta_{z,f} \circ f \circ \alpha_z|_{B_0}, \qquad F_{z,f} = \beta_{z,f} \circ F_f \circ \alpha_z|_{D_0} \, .$$

Recalling Lemma 8.2.2, we note that $F_{z,f}$ is just the restriction to D_0 of F_g, where $g = f_z$. We then define for K in $[1,\infty)$ the families of mappings

$$\mathcal{G}_K = \left\{ f_z : D \text{ is a domain in } \mathbb{R}^n \ , \ z \in \operatorname{int}\left[D^*(r_0)\right] \ , \ f \in \mathcal{Q}_K(D) \right\}$$

and

$$\begin{aligned} \mathcal{G}_K^* &= \left\{ F_{z,f} : D \text{ is a domain in } \mathbb{R}^n \ , \ z \in \operatorname{int}\left[D^*(r_0)\right] \ , \ f \in \mathcal{Q}_K(D) \right\} \\ &= \{ F_g|_{D_0} : g \in \mathcal{G}_K \} \, . \end{aligned}$$

One can view $\mathcal{G}_K$ as the family of "germs" of K-quasiconformal mappings in $\mathbb{R}^n$, since each member f of $\mathcal{Q}_K(D)$ sees its essential geometric structure near any point of D captured, modulo renormalization, in some member of $\mathcal{G}_K$.

Our next order of business is to demonstrate that the families $\mathcal{G}_K$ and $\mathcal{G}_K^*$ are relatively compact in the topology of (Euclidean) locally uniform convergence. As we proceed with the verification of this fact, two technical lemmas will come in handy.

Lemma 8.2.4. *There exists a constant $H = H(n, K)$ for which the following statement is true: if D is a domain in $\mathbb{R}^n$, if f is a member of the family $\mathcal{Q}_K(D)$, if x belongs to D, and if $t > 0$ is such that the ball $B^n(x, r_0 t)$ lies in D, then*

$$\frac{L_f(x,t)}{\ell_f(x,t)} \le H \, . \tag{8.17}$$

Proof. We assume that $n \ge 2$, the assertion being trivial when $n = 1$. Because of the way in which the ratio $L_f(x,t)/\ell_f(x,t)$ behaves when f is composed with similarity transformations, it suffices to consider the situation for $D = B_0$, $x = 0$, $t = 1$ and to produce a constant H such that (8.17) holds for every member of the family $\mathcal{F} = \{ f \in \mathcal{Q}_K(B_0) : f(0) = 0, \ \ell_f(0,1) = |f(e_1)| = 1 \}$. Since $q[f(0), \infty] = 2$ and $q[f(e_1), \infty] = \sqrt{2}$ for each f in $\mathcal{F}$, Theorem 6.6.12 (ii) informs us that $\mathcal{F}$ is a normal family in the chordal metric. The condition $f(0) = 0$, in conjunction with Theorem 6.6.26, shows that $\mathcal{F}$ is also normal with respect to the Euclidean metric.

We define

$$H = \sup_{f \in \mathcal{F}} L_f(0,1) \, .$$

To complete the proof we must verify that $H < \infty$. Let $\langle f_\nu \rangle$ be a sequence from $\mathcal{F}$ such that $L_{f_\nu}(0,1) \to H$. By passing to a subsequence we may assume that $f_\nu \to f$ in a Euclidean locally uniform fashion in B_0, where $f : B_0 \to \mathbb{R}^n$ is continuous. In particular, $f_\nu \to f$ uniformly on $\mathbb{S}^{n-1}$, so

$$H = \lim_{\nu \to \infty} L_{f_\nu}(0,1) = L_f(0,1) < \infty \, .$$

As defined, the constant H depends on n, K, and the number $r_0 = 3 + \sqrt{n}$, hence, entirely on n and K. □

The second technical result we require reads as follows.

Lemma 8.2.5. *If g belongs to the family $\mathcal{G}_K$, then $g(0) = 0$ and*

$$\frac{1}{H} \le |g(e_1)| \le 1 \ , \tag{8.18}$$

in which H is the constant from Lemma 8.2.4.

PROOF. Write $g = f_z|B_0$, where f comes from $\mathcal{Q}_K(D)$ for some subdomain D of $\mathbb{R}^n$ and z from the interior of $D^*(r_0)$, and let $w = F_f(z)$. If we express z in the form $z = (x,t)$ with x in D and $0 < r_0 t < \operatorname{dist}(x, \partial D)$, then $w = (f(x), \tau_f(x,t))$. By using (8.16) and in the last step abbreviating $(y,0)$ to y, we obtain

$$\begin{aligned} g(0) &= \beta_{z,f} \circ f \circ \alpha_z(0) = \beta_{z,f}\big[\big(f(x),0\big)\big] \\ &= e_{n+1} + w_{n+1}^{-1}\big[\big(f(x),0\big) - w\big] = w_{n+1}^{-1}\big[f(x) - f(x)\big] = 0\,. \end{aligned}$$

Similarly,

$$g(e_1) = w_{n+1}^{-1}\big[f(x+te_1) - f(x)\big] = \tau_f(x,t)^{-1}\big[f(x+te_1) - f(x)\big]\,.$$

Now $B^n(x, r_0 t)$ is contained in D, so we can assert on the strength of Lemma 8.2.4 that

$$\frac{1}{H} \le \frac{\ell_f(x,t)}{L_f(x,t)} \le \frac{|f(x+te_1) - f(x)|}{\tau_f(x,t)} \le 1\ ,$$

leading to

$$\frac{1}{H} \le |g(e_1)| \le 1$$

and thereby confirming (8.18). □

Armed with Lemma 8.2.5, we can establish the relative compactness of $\mathcal{G}_K$.

THEOREM 8.2.6. *The family $\bar{\mathcal{G}}_K$, the closure of $\mathcal{G}_K$ with respect to the topology of Euclidean locally uniform convergence in B_0, is a compact family of K-quasiconformal embeddings of B_0 into $\mathbb{R}^n$.*

PROOF. We treat the case $n \ge 2$ and leave the case $n = 1$ as an exercise for the reader. From Lemma 8.2.5 we extract the information that

$$q\big[g(0), \infty\big] = 2, \qquad q\big[g(e_1), \infty\big] \ge \sqrt{2}$$

whenever g is a member of $\mathcal{G}_K$. Once again invoking Corollary 6.6.12 (ii), we deduce that $\mathcal{G}_K$ is a normal family relative to the chordal metric. As $g(0) = 0$ holds for each g in $\mathcal{G}_K$, Theorem 6.6.26 confirms that the same is true for the Euclidean metric.

The upshot of the foregoing statements is that, as far as sequences from $\mathcal{G}_K$ are concerned, chordal and Euclidean locally uniform convergence in B_0 are equivalent notions.

For a mapping $\bar{g} : B_0 \to \mathbb{R}^n$ to qualify as a member of $\bar{\mathcal{G}}_K$ we require that there be a sequence $\langle g_\nu \rangle$ in $\mathcal{G}_K$ such that $g_\nu \to \bar{g}$ in the Euclidean—hence, also, chordal—locally uniform sense in B_0. In particular, $\bar{g}(0) = 0$. Lemma 8.2.5 implies that $|g(e_1)| \ge \frac{1}{H} > 0$ for any such function, so $\bar{g}$ is definitely nonconstant in B_0. By Theorem 6.6.23, each $\bar{g}$ in $\bar{\mathcal{G}}_K$ must be a K-quasiconformal mapping of B_0 into $\mathbb{R}^n$.

Let $\langle \bar{g}_\nu \rangle$ be an arbitrary sequence from $\bar{\mathcal{G}}_K$. We can choose for each ν a member g_ν of $\mathcal{G}_K$ with the property that $|\bar{g}_\nu(x) - g_\nu(x)| \le \nu^{-1}$ whenever $|x| \le r_0 - \nu^{-1}$. The normality of $\mathcal{G}_K$ guarantees the existence of a subsequence $\langle g_{\nu_k} \rangle$ of $\langle g_\nu \rangle$ such that $g_{\nu_k} \to \bar{g}$ locally uniformly in B_0. By definition $\bar{g}$ belongs to $\bar{\mathcal{G}}_K$ and by construction $\bar{g}_{\nu_k} \to \bar{g}$ locally uniformly in B_0. This shows $\bar{\mathcal{G}}_K$ to be a compact family of mappings relative to the topology of locally uniform convergence in B_0. □

It is now an easy task to verify that the family $\mathcal{G}_K^*$ is itself relatively compact.

THEOREM 8.2.7. *The family $\bar{\mathcal{G}}_K^*$, the closure of $\mathcal{G}_K^*$ with respect to the topology of Euclidean locally uniform convergence in D_0, is a compact family of embeddings of D_0 into $\mathbb{H}^{n+1}$.*

Each member of $\bar{\mathcal{G}}_K^*$ is the restriction to D_0 of F_g for some $g \in \bar{\mathcal{G}}_K$.

PROOF. Let $\langle \bar{G}_\nu \rangle$ be a sequence from $\bar{\mathcal{G}}_K^*$. We must produce a subsequence $\langle \bar{G}_{\nu_k} \rangle$ such that $\bar{G}_{\nu_k} \to \bar{G}$ locally uniformly in D_0, where $\bar{G}$ belongs to $\bar{\mathcal{G}}_K^*$.

By the definition of $\bar{\mathcal{G}}_K^*$ we can choose a sequence $\langle G_\nu \rangle$ from $\mathcal{G}_K^*$ such that $\|\bar{G}_\nu - G_\nu\|_A \to 0$ as $\nu \to \infty$ for each compact set A in D_0. We can then write $G_\nu = F_{g_\nu}|D_0$, where g_ν is a mapping from $\mathcal{G}_K$.

By Theorem 8.2.6, there is a subsequence $\langle g_{\nu_k} \rangle$ of $\langle g_\nu \rangle$ such that $g_{\nu_k} \to \bar{g}$, a member of $\bar{\mathcal{G}}_K$, locally uniformly in B_0. Appealing to Lemma 8.2.1 we conclude that $G_{\nu_k} \to \bar{G} = F_{\bar{g}}|D_0$ locally uniformly in D_0, which places $\bar{G}$ in $\bar{\mathcal{G}}_K^*$. Moreover, from the choice of G_ν it is clear that $\bar{G}_\nu \to \bar{G}$ locally uniformly in D_0. Finally, since each $\bar{G}$ in $\bar{\mathcal{G}}_K^*$ can be obtained in the above way, each has the structure suggested by the theorem. □

Consider a mapping f from the family $\mathcal{Q}_K(D)$. The mapping F_f is not the extension of f we seek: unfortunately F_f is not generally quasiconformal in $V \cap \mathbb{H}^{n+1}$ for any open neighbourhood V of D in $\mathbb{R}^{n+1}$.

Instead, F_f plays an auxiliary role of providing us with something defined $\mathbb{H}^{n+1}$ to work with. In particular we will seek to approximate F_f in the hyperbolic metric by a quasiconformal mapping. Notice that any quasiconformal mapping a bounded hyperbolic distance from F_f agrees with f on $\mathbb{R}^{n+1}$. The other nice property F_f has is that at large scales it behaves like a mapping which is bilipschitz in the hyperbolic metric.

The next lemma is typical of the way in which we use these properties of F_f: roughly speaking, it asserts that a mapping which is sufficiently close to F_f on a set A and is locally injective in A must be globally injective on A.

LEMMA 8.2.8. *Let M be a number satisfying*

$$0 < M < d_{\mathbb{H}}(e_{n+1}, \partial D_0 \cap \mathbb{H}^{n+1}).$$

There exists a constant $\rho > 0$, which depends only on n, K, and M, with the following property: if D is a domain in $\mathbb{R}^n$, if $f \in \mathcal{Q}_K(D)$, if A is a nonempty set in the interior of D^*, and if $h : A \to \mathbb{H}^{n+1}$ is a function such that $d_{\mathbb{H}}(h, F_f : A) \leq \rho$ and such that $h(z) \neq h(z')$ whenever z and z' are points of A with $0 < d_{\mathbb{H}}(z, z') \leq M$, then h is injective.

PROOF. The sphere $S = \mathbb{S}_{\mathbb{H}^{n+1}}(e_{n+1}, M)$ is a compact subset of D_0. Therefore, the set $\bar{\mathcal{G}}_K^* \times S$ is compact in the product topology, while the function $(g, x) \mapsto d_{\mathbb{H}}[g(x), g(e_{n+1})]$ is continuous and positive on $\bar{\mathcal{G}}_K^* \times S$. It follows that

$$\rho = \frac{1}{3} \min\Big\{ d_{\mathbb{H}}\big[g(x), g(e_{n+1})\big] : x \in S \ , \ g \in \bar{\mathcal{G}}_K^* \Big\} > 0 \, .$$

Now let f and h be as in the statement of the theorem, and let z and z' be distinct points of A. We claim that $h(z) \neq h(z')$. This is true by hypothesis if $d_{\mathbb{H}}(z, z') \leq M$, so we suppose that $d_{\mathbb{H}}(z, z') > M$. Because F_f is an embedding, it is evident that

$$d_{\mathbb{H}}\big[F_f(z), F_f(z')\big] \geq d_{\mathbb{H}}\big[F_f(z), F_f(S')\big] \ ,$$

where $S' = S^{n-1}_{\mathbb{H}}(z, M) = \alpha_z(S)$. Let $g = \beta_{z,f} \circ F_f \circ \alpha_z | D_0$, a member of $\mathcal{G}^*_K$. As $\beta_{z,f}$ is a hyperbolic isometry we deduce that

$$\begin{aligned} d_{\mathbb{H}}\big[F_f(z), F_f(S')\big] &= d_{\mathbb{H}}\big[\beta_{z,f} \circ F_f \circ \alpha_z(e_{n+1})\ ,\ \beta_{z,f} \circ F_f \circ \alpha_z(S)\big] \\ &= d_{\mathbb{H}}\big[g(e_{n+1}), g(S)\big] \geq 3\rho\,. \end{aligned}$$

According to our assumptions the function h satisfies the condition $d_{\mathbb{H}}(h, F_f; A) \leq \rho$.

The triangle inequality thus tells us that

$$\begin{aligned} d_{\mathbb{H}}\big[h(z), h(z')\big] &\geq d_{\mathbb{H}}\big[F_f(z), F_f(z')\big] - d_{\mathbb{H}}\big[h(z), F_f(z)\big] - d_{\mathbb{H}}\big[h(z'), F_f(z')\big] \\ &\geq 3\rho - \rho - \rho = \rho > 0\ , \end{aligned}$$

whence $h(z) \neq h(z')$. □

We close the present section by identifying the last purely topological ingredient in the proof of the extension theorem. For $n \neq 3$ the result we shall merely quote follows directly from standard, albeit nontrivial, approximation theorems in piecewise linear topology; see [**158**]. When $n = 3$ a special argument is called for: one that rests on Lemma 8.2.1 and the easy-to-verify fact, true in all dimensions, that $F_f : B^*_0 \to \mathbb{R}^{n+1}_+$ is a LIP-embedding whenever $f : B_0 \to \mathbb{R}^n$ is.

LEMMA 8.2.9. *If D is a domain in $\mathbb{H}^{n+1}$ such that $\bar{D}$ lies in D_0 and if g is a member of $\bar{\mathcal{G}}^*_K$, then $g|D$ is uniformly approximable by* LIP*-embeddings.*

Since the set $g(\bar{D})$ arising in Lemma 8.2.9 is a compact subset of $\mathbb{H}^{n+1}$, it is clear that by applying this result we can substitute uniform approximation in the hyperbolic metric for uniform convergence in the Euclidean metric. Thus, for each $\varepsilon > 0$ there is a LIP-embedding $h : D \to \mathbb{H}^{n+1}$ for which $d_{\mathbb{H}}(g, h; D) < \varepsilon$.

8.2.3. The Whitney decomposition of $\mathbb{H}^{n+1}$. Let Q be a closed cube in $\mathbb{R}^{n+1}$ whose edges are parallel to the coordinate axes. We use z_Q to designate the center of Q and $2\lambda_Q$ to indicate its edge-length.

Thus we can represent Q in the manner $Q = z_Q + \lambda_Q I^{n+1}$, where $I = [-1, 1]$. For $t > 0$ we denote

$$Q(t) = z_Q + t\lambda_Q(-1, 1)^{n+1}.$$

Thus $Q(t)$ is the open cube concentric with Q but t times it in size.

The symbol $\mathcal{K}$ will stand for the collection of all closed cubes in $\mathbb{R}^{n+1}$ with unit edge-length whose vertices lie in the set $\mathbb{Z}^n \times \{0, 1\}$. In other words, $\mathcal{K}$ consists of all translates of the cube $[0, 1]^{n+1}$ by vectors from the lattice $\mathbb{Z}^n$ in $\mathbb{R}^n$.

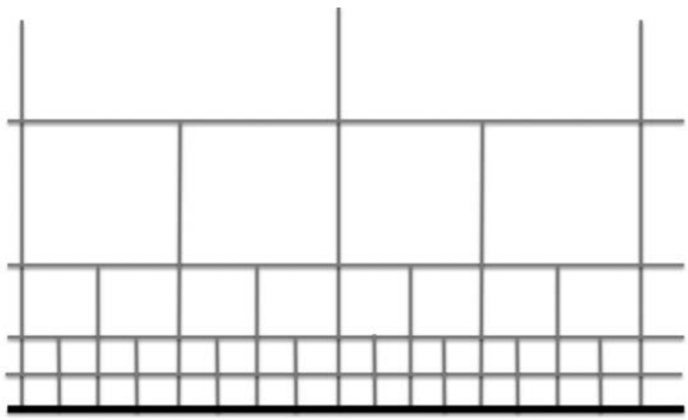
Distinct members of $\mathcal{K}$ are nonoverlapping, and the union of all the cubes in $\mathcal{K}$ is $\mathbb{R}^n \times [0, 1]$. We denote by $\mathcal{W}$ the collection of all cubes of the type $2^\nu e_{n+1} + 2^\nu Q$, where Q comes from $\mathcal{K}$ and ν is an integer:

$$\mathcal{W} = \{2^\nu e_{n+1} + 2^\nu Q : Q \in \mathcal{K}, \nu \in \mathbb{Z}\}.$$

As with $\mathcal{K}$, distinct cubes in $\mathcal{W}$ do not overlap; the union of all the cubes in $\mathcal{W}$ is $\mathbb{H}^{n+1}$. We shall refer to $\mathcal{W}$ as the *Whitney decomposition* of $\mathbb{H}^{n+1}$.

Notice that the members of $\mathcal{W}$ with fixed ν provide a decomposition $\mathcal{K}_\nu$ of $\mathbb{R}^n \times [2^\nu, 2^{\nu+1}]$ which, except for scale, mimics the decomposition of $\mathbb{R}^n \times [0, 1]$ given by $\mathcal{K}$.

Each cube in this "level ν" of $\mathcal{W}$ has below it 2^n cubes from "level $\nu - 1$". Observe, too, that any two cubes in $\mathcal{W}$ are congruent in the hyperbolic geometry of



The Whitney decomposition of $\mathbb{H}^2$.

$\mathbb{H}^{n+1}$. For ease of reference we report several elementary facts about $\mathcal{W}$ in a lemma. In the lemma—and from this point on in the chapter—we take $N = 2^{n+1}$. The first two assertions of the lemma are readily confirmed by straightforward geometric arguments.

LEMMA 8.2.10. *The Whitney decomposition $\mathcal{W}$ of $\mathbb{H}^{n+1}$ enjoys the following properties:*

(i) *if Q belongs to $\mathcal{W}$, then $(z_Q)_{n+1} = 3\lambda_Q$;*
(ii) *if Q and Q' are disjoint members of $\mathcal{W}$, then $Q(3/2)$ and $Q'(3/2)$ are also disjoint;*
(iii) *there exist nonempty, pairwise disjoint subcollections $\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_N$ of $\mathcal{W}$ such that $\mathcal{W} = \mathcal{W}_1 \cup \mathcal{W}_2 \cup \cdots \cup \mathcal{W}_N$ and such that the cubes in each $\mathcal{W}_j$ are pairwise disjoint.*

PROOF OF (iii). Let $\mathcal{L}$ denote the collection of closed unit cubes in $\mathbb{R}^n$ that comprises all translates of the cube $J^n = [0,1]^n$ by vectors from the lattice $\mathbb{Z}^n$. It is a trivial observation that we can write

$$\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2 \cup \cdots \cup \mathcal{L}_M \ ,$$

where $M = 2^n$, where $\mathcal{L}_1, \mathcal{L}_2, \ldots, \mathcal{L}_M$ are nonempty, pairwise disjoint subcollections of $\mathcal{L}$, and where the cubes in each $\mathcal{L}_j$ are pairwise disjoint.

The simplest way to arrange this is to let each $\mathcal{L}_j$ consist of Q_j, one of the 2^n cubes of $\mathcal{L}$ whose union is $2J^n$, together with all translates of Q_j by vectors from the lattice $2\mathbb{Z}^n$. This decomposition of $\mathcal{L}$ leads immediately to a corresponding decomposition for the collection $\mathcal{K}$; namely, $\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2 \cup \cdots \cup \mathcal{K}_M$, where

$$\mathcal{K}_j = \{Q \times [0,1] : Q \in \mathcal{L}_j\} \, .$$

By translating and scaling we can transport the above decomposition of $\mathcal{K}$ to each level of $\mathcal{W}$, thereby obtaining a decomposition of $\mathcal{K}_\nu$ into subcollections $\mathcal{K}_{\nu,1}$, $\mathcal{K}_{\nu,2}, \ldots, \mathcal{K}_{\nu,M}$ with features analogous to those of $\mathcal{K}_1, \mathcal{K}_2, \ldots, \mathcal{K}_M$. Let

$$\mathcal{W}_j^0 = \{Q : Q \in \mathcal{K}_{\nu,j} \, , \ \nu \text{ odd}\}, \qquad \mathcal{W}_j^e = \{Q : Q \in \mathcal{K}_{\nu,j} \, , \ \nu \text{ even}\},$$

and for $j = 1, 2, \ldots, N$ take $\mathcal{W}_j = \mathcal{W}_j^0$ if j is odd and $\mathcal{W}_j = \mathcal{W}_{j/2}^e$ if j is even. Then the subcollections $\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_N$ of $\mathcal{W}$ have the stated properties. $\square$

Some additional abbreviations and notation will later turn out to be helpful. First, we abbreviate

$$\alpha_{z_Q} = \alpha_Q, \qquad \beta_{z_Q,f} = \beta_{Q,f}\,.$$

Next, we set

$$Q_0 = e_{n+1} + (1/3)I^{n+1}$$

and remark that $Q_0(3)$, which is the interior of $Q = e_{n+1} + I^{n+1}$, lies in D_0. From Lemma 8.2.10(i) it follows easily that

$$Q_0 = \alpha_Q^{-1}(Q)$$

for every Q in $\mathcal{W}$, while part (ii) of the same lemma ensures that

$$Q_0(3/2) \cap \alpha_Q^{-1}\big[Q'(3/2)\big] = \emptyset \tag{8.19}$$

whenever Q and Q' are disjoint members of $\mathcal{W}$. One consequence of (8.19) is the following lemma.

LEMMA 8.2.11. *For t in $(0, 3/2]$, the collection $\mathcal{C}_t$ of (nonempty) open $(n+1)$-intervals $Q_0(t)\cap\alpha_Q^{-1}[Q'(1+2^{-j})]$, generated as Q and Q' vary over $\mathcal{W}$ and j ranges over $\{1, 2, \ldots, N+2\}$, is a finite collection, one that depends entirely upon n and t.*

Indeed, it is perfectly clear that $\mathcal{C}_t$ can be arrived at by fixing Q and letting Q' vary over the finitely many members of $\mathcal{W}$ that intersect Q.

Recalling that $r_0 = 3 + \sqrt{n}$, we introduce numbers r_1 and r_2:

$$r_1 = 2r_0 + \sqrt{n} = 6 + 3\sqrt{n}, \qquad r_2 = (3r_1 + \sqrt{n}\,)/9 = 9 + 5\sqrt{n}\,.$$

Then $1 < r_0 < r_1 < r_2$, so

$$D^*(r_2) \subset D^*(r_1) \subset D^*(r_0) \subset D^*$$

whenever D is a domain in $\mathbb{R}^n$. For any such domain D, we define

$$\mathcal{W}(D) = \big\{Q \in \mathcal{W} : z_Q \in D^*(r_1)\big\}$$

and make two observations.

LEMMA 8.2.12. *Let D be a domain in $\mathbb{R}^n$. Then $\overline{Q(3)}$ is contained in $D^*(r_0)$ whenever Q belongs to $\mathcal{W}(D)$. Moreover, the collection $\mathcal{W}(D)$ covers the interior of $D^*(r_2)$.*

PROOF. First let Q be a member of $\mathcal{W}(D)$ and let z be a point of $\overline{Q(3)}$. Write $z_Q = (x, t)$, where x is in D and $0 < r_1 t < \operatorname{dist}(x, \partial D)$, and $z = (y, s)$, where y is in $\mathbb{R}^n$ and $s \geq 0$. Since $z - z_Q$ lies in the set $3\lambda_Q I^{n+1}$, we see that $y - x$ is in $3\lambda_Q I^n$. In view of Lemma 8.2.10(i),

$$|y - x| \leq 3\lambda_Q\sqrt{n} = t\sqrt{n} < r_1 t < d(x, \partial D)\ ,$$

so y belongs to D. Furthermore,

$$\begin{aligned} \operatorname{dist}(y, \partial D) &\geq \operatorname{dist}(x, \partial D) - |y - x| \geq \operatorname{dist}(x, \partial D) - t\sqrt{n} \\ &\geq \Big(1 - \frac{\sqrt{n}}{r_1}\Big)\operatorname{dist}(x, \partial D) = \frac{2r_0 \operatorname{dist}(x, \partial D)}{r_1}\ , \end{aligned}$$

which in conjunction with the inequality $|s - t| \leq 3\lambda_Q = t$ implies that

$$s \le 2t < \frac{2\mathrm{dist}(x, \partial D)}{r_1} < \frac{\mathrm{dist}(y, \partial D)}{r_0}$$

and thus places z in $D^*(r_0)$.

Next, consider an interior point z of the set $D^*(r_2)$. Again write $z = (y, s)$, where now y belongs to D and $0 < r_2 s < \mathrm{dist}(y, \partial D)$. Choose a cube Q from $\mathcal{W}$ such that z lies in Q and, as earlier, let $z_Q = (x, t)$. We show that z_Q is a point of $D^*(r_1)$, which makes Q a member of $\mathcal{W}(D)$. Since z belongs to Q it is definitely the case that $t \le 3s/2$. In the present situation, $x - y$ is a point of $\lambda_Q I^n$, implying that

$$|x - y| \le \lambda_Q \sqrt{n} = \frac{t\sqrt{n}}{3} \le \frac{s}{2}\sqrt{n} < r_2 s < \mathrm{dist}(y, \partial D)\,.$$

Accordingly, x lies in D and

$$\begin{aligned} \mathrm{dist}(x, \partial D) &\ge \mathrm{dist}(y, \partial D) - |x - y| \ge \mathrm{dist}(y, \partial D) - \frac{s\sqrt{n}}{2} \\ &\ge \Big(1 - \frac{\sqrt{n}}{2r_2}\Big)\mathrm{dist}(y, \partial D) = \frac{3r_1 \mathrm{dist}(y, D)}{2r_2} \\ &\ge \frac{3r_1 s}{2} \ge r_1 t\,. \end{aligned}$$

We conclude that z_Q is in $D^*(r_1)$, as asserted. □

8.3. The Tukia-Väisälä extension theorem

Finally we come to the proof of the extension theorem.

8.3.1. The construction. To begin we fix, once and for all, the subcollections $\mathcal{W}_1, \mathcal{W}_2, \ldots, \mathcal{W}_N$ of the Whitney decomposition $\mathcal{W}$ of $\mathbb{H}^{n+1}$ as specified in Lemma 8.2.12 (iii) and the proof thereof.

Let D be a subdomain of $\mathbb{R}^n$. We write $\mathcal{W}_j(D) = \mathcal{W}_j \cap \mathcal{W}(D)$. Our choice of $\mathcal{W}_j$ makes certain that $\mathcal{W}_j(D)$ is a nonempty collection of pairwise disjoint cubes. For $j = 1, 2, \ldots, N$ we define

$$\mathcal{W}_j^*(D) = \bigcup_{k=1}^{j} \mathcal{W}_k(D)\,.$$

Therefore we have

$$\mathcal{W}_1(D) = \mathcal{W}_1^*(D) \subset \mathcal{W}_2^*(D) \subset \cdots \subset \mathcal{W}_N^*(D) = \mathcal{W}(D)\,.$$

Next we set

$$V_0(D) = \bigcup_{Q \in \mathcal{W}_1(D)} Q(3/2), \qquad W_0(D) = \bigcup_{Q \in \mathcal{W}_1(D)} Q(5/4),$$

and for $j = 1, 2, \ldots, N$ define sets $V_j(D)$ and $W_j(D)$ by

$$V_j(D) = \bigcup_{Q \in \mathcal{W}_j^*(D)} Q(1 + 2^{-j-1}), \qquad W_j(D) = \bigcup_{Q \in \mathcal{W}_j^*(D)} Q(1 + 2^{-j-2})\,.$$

Finally, for $j = 1, 2, \ldots, N$, for Q in $\mathcal{W}_j(D)$, and for t in $(0, 3/2]$, we let

$$V_Q(D, t) = Q_0(t) \cap \alpha_Q^{-1}\big[V_{j-1}(D)\big] \quad , \quad W_Q(D, t) = Q_0(t) \cap \alpha_Q^{-1}\big[W_{j-1}(D)\big]\,.$$

Thus we make the following remark.

LEMMA 8.3.1. *For fixed t in $(0, 3/2]$ the number of different sets $V_Q(D,t)$ (respectively, $W_Q(D,t)$), generated as D varies over all subdomains of $\mathbb{R}^n$ and Q over all cubes in $\mathcal{W}(D)$, is finite and depends only on n and t.*

PROOF. Any set $V_Q(D,t)$ is the union of sets of the type $Q_0(t) \cap \alpha_Q^{-1}[Q'(1+2^{-j})]$, where Q and Q' come from $\mathcal{W}$ and where $1 \leq j \leq N+1$; i.e., $V_Q(D,t)$ is either empty or is the union of sets from the finite collection $\mathcal{C}_t$. This fact dictates that at most finitely many different sets $V_Q(D,t)$ arise. The same reasoning obviously applies to sets of the kind $W_Q(D,t)$. □

The next result is where Sullivan's theorem is used in the solution of the extension problem.

LEMMA 8.3.2. *Let $\varepsilon > 0$ and $K \geq 1$ be given. There exists a δ in $(0, \varepsilon]$, where δ depends entirely on n, K, and ε, for which the following statement holds:*

- *if D is a domain in $\mathbb{R}^n$,*
- *if Q is a cube from $\mathcal{W}(D)$ such that $V = V_Q(D, 3/2)$ is nonempty,*
- *if g is a member of the family $\bar{\mathcal{G}}_K^*$, and*
- *if $h : V \to \mathbb{H}^{n+1}$ is a* LIP*-embedding for which $d_{\mathbb{H}}(h, g; V) \leq \delta$,*

then there exists a LIP*-embedding*

$$\hat{h} : Q_0(3/2) \to \mathbb{H}^{n+1}$$

such that $d_{\mathbb{H}}[\hat{h}, g; Q_0(4/3)] \leq \varepsilon$ and such that $\hat{h} = h$ in $W = W_Q(D, 4/3)$.

Furthermore, if h is locally λ-bilipschitz with respect to the hyperbolic metric, then $\hat{h}$ can be chosen so that its restriction to $Q_0(4/3)$ is locally $\hat{\lambda}$-bilipschitz in the hyperbolic metric, where $\hat{\lambda}$ is a function of λ, n, K, and ε.

PROOF. Fix V and W as indicated. We wish to apply Lemma 8.1.6 with $U = Q_0(3/2)$, $U_0 = Q_0(4/3)$, V, W, and $\mathcal{G} = \{g|U : g \in \bar{\mathcal{G}}_K^*\}$. Theorem 8.2.7 tells us that $\mathcal{G}$ is compact in the topology of Euclidean uniform convergence on U. In the present situation $E = \{g(x) : g \in \bar{\mathcal{G}}_K^*,\ x \in \bar{U}\}$, which is a compact set because $\bar{\mathcal{G}}_K^* \times \bar{U}$ is compact in the product topology and the function $(g, x) \mapsto g(x)$ is continuous on $\bar{\mathcal{G}}_K^* \times \bar{U}$. The appeal to Lemma 8.1.6 is therefore justified.

Although its wording is not this explicit, Lemma 8.1.6 guarantees the existence of a function

$$\Delta_n^{V,W} : [1, \infty) \times (0, \infty) \to (0, \infty),$$

which we may assume satisfies $\Delta_n^{V,W}(s,t) \leq t$, so that $\delta = \Delta_n^{V,W}(K, \varepsilon)$ fulfills the requirements of the present lemma for the particular pair of sets (V, W).

At the same time it ensures the existence of a function

$$\Lambda_n^{V,W} : [1, \infty) \times [1, \infty) \times (0, \infty) \to [1, \infty).$$

In this instance we may suppose that $\Lambda_n^{V,W}(s,t,u) \geq s$, with the property that if the given embedding h is locally λ-bilipschitz relative to the hyperbolic metric, then $\hat{h}$ can be chosen so that its restriction to $Q_0(4/3)$ is locally $\hat{\lambda}$-bilipschitz with respect to the hyperbolic metric for $\hat{\lambda} = \Lambda_n^{V,W}(\lambda, K, \varepsilon)$.

Lemma 8.3.1 points out that the number of pairs (V, W) that can turn up in the context of Lemma 8.3.2 is finite. Letting Δ_n denote the minimum of the functions $\Delta_n^{V,W}$ for all pairs (V, W) that actually occur, we see that $\delta = \Delta_n(K, \varepsilon)$ is a number with property stated in the lemma.

Similarly, if Λ_n denotes the maximum of the functions $\Lambda_n^{V,W}$, then $\hat{\lambda} = \Lambda_n(\lambda, K, \varepsilon)$ meets the final stipulation of the lemma. □

In an attempt to simplify the statement of the lemma that gives the details of the extension procedure, we establish ahead of time certain parameters that enter into either the result or its proof.

Thus, given $\varepsilon > 0$ and $K \geq 1$, we do the following:

(i) We fix $\rho = \rho(n, K)$ corresponding to $M = d_{\mathbb{H}}[Q_0(5/4), \partial Q_0(4/3)]$ as in Lemma 8.2.8.

(ii) Letting $\delta_N = \min\{\varepsilon, \rho\}$, we define $\delta_{N-1}, \delta_{N-2}, \ldots, \delta_2, \delta_1$ by

$$\delta_j = \Delta_n(K, \delta_{j+1}),$$

where Δ_n is the function that appeared in the proof of Lemma 8.3.2. Then

$$\delta_N \geq \delta_{N-1} \geq \cdots \geq \delta_1 = \delta.$$

(iii) Exploiting the compactness of $\bar{\mathcal{G}}_K^*$ (Theorem 8.2.7) and invoking (8.13) for the compact set $E = \{g(x) : g \in \bar{\mathcal{G}}_K^*,\ x \in \overline{Q_0(3/2)}\,\}$, we select a finite set of mappings $g_1, g_2, \ldots, g_p$ from $\bar{\mathcal{G}}_K^*$—the choice will vary, of course, with K and ε—so that

$$\min_{1 \leq \nu \leq p} d_{\mathbb{H}}\big[g_\nu, g; Q_0(3/2)\big] < \delta/2 \tag{8.20}$$

holds for every g in $\bar{\mathcal{G}}_K^*$.

(iv) For each domain D in $\mathbb{R}^n$, each cube Q in $\mathcal{W}(D)$, and each mapping f from $\mathcal{Q}_K(D)$ we use (6.19) to fix an integer $\nu(Q, f)$ in $\{1, 2, \ldots, p\}$ for which the inequality

$$d_{\mathbb{H}}\big[g_{\nu(Q,f)}, \beta_{Q,f} \circ F_f \circ \alpha_Q; Q_0(3/2)\big] < \delta/2 \tag{8.21}$$

is valid.

(v) We apply Lemma 8.2.9 and choose for $\nu = 1, 2, \ldots, p$ a LIP-embedding $h_\nu : Q_0(3/2) \to \mathbb{H}^{n+1}$ with the property that

$$d_{\mathbb{H}}\big[g_\nu, h_\nu; Q_0(3/2)\big] < \delta/2\,. \tag{8.22}$$

(vi) Lemma 8.1.1 enables us to fix a number $\lambda_1 \geq 1$ so that the restriction of each of the embeddings $h_1, h_2, \ldots, h_p$ to $Q_0(5/4)$ is locally λ_1-bilipschitz with respect to the hyperbolic metric. We then define $\lambda_2, \lambda_3, \ldots, \lambda_N$ by $\lambda_j = \Lambda_n(\lambda_{j-1}, K, \delta_{j-1})$, where Λ_n is the other function that surfaced in the proof of Lemma 8.3.2. Here $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$.

The stage is now set for the construction.

LEMMA 8.3.3. *Let $\varepsilon > 0$ and $K \geq 1$ be given, and let the preconditions specified by* (i)–(vi) *above hold. If D is a domain in $\mathbb{R}^n$ and f is a member of the family $\mathcal{Q}_K(D)$, then there exists for $j = 1, 2, \ldots, N$ an embedding $F_f^j : V_j(D) \to \mathbb{H}^{n+1}$ such that F_f^j is locally λ_j-bilipschitz with respect to the hyperbolic metric and such that $d_{\mathbb{H}}[F_f^j, F_f : V_j(D)] \leq \delta_j$.*

PROOF. The proof proceeds by finite induction. Let Q be a cube from $\mathcal{W}_1^*(D)$. From (8.20) and (8.21) we infer that

$$d_{\mathbb{H}}\big[h_{\nu(Q,f)}, \beta_{Q,f} \circ F_f \circ \alpha_Q; Q_0(3/2)\big] \leq \delta\,.$$

Equivalently, since α_Q and $\beta_{Q,f}$ are hyperbolic isometries,

$$d_{\mathbb{H}}\big[\beta_{Q,f}^{-1} \circ h_{\nu(Q,f)} \circ \alpha_Q^{-1}, F_f; Q(3/2)\big] \le \delta\,. \tag{8.23}$$

We define $F_f^1 : V_1(D) \to \mathbb{H}^{n+1}$ by requiring that, for each cube Q belonging to $\mathcal{W}_1^*(D)$, F_f^1 coincide in $Q(5/4)$ with the mapping $\beta_{Q,f}^{-1} \circ h_{\nu(Q,f)} \circ \alpha_Q$.

Because $Q(5/4)$ and $Q'(5/4)$ are disjoint whenever Q and Q' are disjoint cubes from $\mathcal{W}$ by Lemma 8.2.10, F_f^1 is a well-defined function that satisfies the condition

$$d_{\mathbb{H}}\big[F_f^1, F_f; V_1(D)\big] \le \delta_1\,. \tag{8.24}$$

Certainly F_f^1 is one-to-one in $Q(5/4)$ for each cube Q in $\mathcal{W}_1^*(D)$. Indeed, we infer from (vi) and the definition of F_f^1 that the restriction of F_f^1 to $Q(5/4)$ is an embedding which is locally λ_1-bilipschitz with respect to the hyperbolic metric.

Let z and z' be distinct points of $V_1(D)$ for which $d_{\mathbb{H}}(z,z') \le M$. For the reason that

$$M = d_{\mathbb{H}}\big[Q(5/4), \partial Q(4/3)\big] < d_{\mathbb{H}}\big[Q(5/4), Q'(5/4)\big]$$

whenever Q and Q' are different cubes from $\mathcal{W}$, the choice of M dictates that both z and z' be in $Q(5/4)$ for some Q from $\mathcal{W}_1^*(D)$, implying that $F_f^1(z) \ne F_f^1(z')$.

In conjunction with (8.24) and the assumption that $\delta \le \rho$, this fact tells us that F_f^1 is an injection as per Lemma 8.2.8. Consequently, F_f^1 is an embedding of $V_1(D)$ into $\mathbb{H}^{n+1}$ which is locally λ_1-bilipschitz in the hyperbolic metric.

Assume now that $2 \le j \le N$ and that an embedding $F_f^{j-1} : V_{j-1}(D) \to \mathbb{H}^{n+1}$ has been constructed which is locally λ_{j-1}-bilipschitz with respect to the hyperbolic metric and which satisfies the condition

$$d_{\mathbb{H}}\big[F_f^{j-1}, F_f; V_{j-1}(D)\big] \le \delta_{j-1}\,. \tag{8.25}$$

We use F_f^{j-1} in the construction of an embedding $F_f^j : V_j(D) \to \mathbb{H}^{n+1}$ with the desired characteristics.

Noting that

$$V_j(D) = W_{j-1}(D) \cup \bigcup_{Q \in \mathcal{W}_j(D)} Q(1 + 2^{-j-1})\ ,$$

we start by setting $F_f^j(z) = F_f^{j-1}(z)$ for z in $W_{j-1}(D)$. Then F_f^j embeds $W_{j-1}(D)$ into $\mathbb{H}^{n+1}$ in a locally λ_{j-1}-bilipschitz fashion with respect to the hyperbolic metric, and the inequality

$$d_{\mathbb{H}}\big[F_f^j, F_f; W_{j-1}(D)\big] \le d\big[F_f^{j-1}, F_f; V_{j-1}(D)\big] \le \delta_{j-1} \le \delta \tag{8.26}$$

holds.

Next, if Q is a cube from $\mathcal{W}_j(D)$ that is disjoint from every cube in $\mathcal{W}_{j-1}(D)$ (by Lemma 8.2.10, this means that $Q(1+2^{-j-1})$ does not meet $W_{j-1}(D)$), then we can revert to the procedure used to construct F_f^1 and demand that F_f^j coincide in $Q(1+2^{-j-1})$ with $\beta_{Q,f}^{-1} \circ h_{\nu(Q,f)} \circ \alpha_{Q,f}$. In this case the restriction of F_f^j to $Q(1+2^{-j-1})$ is locally λ_1-bilipschitz relative to the hyperbolic metric. Moreover, (8.23) applies here and informs us that

$$d_{\mathbb{H}}\big[F_f^j, F_f; Q(1+2^{-j-1})\big] \le \delta \le \delta_j \tag{8.27}$$

for any cube Q of the variety under discussion.

Finally, we consider cubes Q from $\mathcal{W}_j(D)$ that have nonempty intersection with one or more cubes from $\mathcal{W}^*_{j-1}(D)$. This guarantees that $V_Q = V_Q(D, 3/2)$ is not empty, so for each cube of the type in question we obtain an embedding $h_Q : V_Q \to \mathbb{H}^{n+1}$ by setting

$$h_Q = \beta_{Q,f} \circ F_f^{j-1} \circ \alpha_Q | V_Q\,.$$

The embedding h_Q is locally λ_{j-1}-bilipschitz with respect to the hyperbolic metric. By (8.25),

$$d_{\mathbb{H}}(h_Q, \beta_{Q,f} \circ F_f \circ \alpha_Q; V_Q) \leq d_{\mathbb{H}}\big[F_f^{j-1}, F_f; V_{j-1}(D)\big] \leq \delta_{j-1}\,.$$

According to Lemma 8.3.2 we can produce for each Q in this category a LIP-embedding $\hat{h}_Q : Q_0(3/2) \to \mathbb{H}^{n+1}$ such that $\hat{h}_Q = h_Q$ in $W_Q = W_Q(D, 4/3) = Q_0(4/3) \cap \alpha_Q^{-1}[W_{j-1}(D)]$ and such that

$$d_{\mathbb{H}}\big[\hat{h}_Q, \beta_{Q,f} \circ F_f \circ \alpha_Q; Q_0(4/3)\big] \leq \delta_j\,. \tag{8.28}$$

Here we recall that $\delta_{j-1} = \Delta_n(K, \delta_j)$.

Furthermore, we may assume that the restriction of $\hat{h}_Q$ to $Q_0(4/3)$ is locally $\hat{\lambda}$-bilipschitz in the hyperbolic metric, where $\hat{\lambda} = \Lambda_n(\lambda_{j-1}, K, \delta_{j-1}) = \lambda_j$. For cubes Q of the present class we define F_f^j in $Q(1 + 2^{-j-1})$ by insisting that it agree in this set with the mapping $\beta_{Q,f}^{-1} \circ \hat{h}_Q \circ \alpha_Q^{-1}$. Since $\hat{h}_Q = h_Q$ in W_Q, this definition forces $F_f^j(z) = F_f^{j-1}(z)$ to hold at any point z of $Q(1 + 2^{-j-1}) \cap W_{j-1}(D)$, so F_f^j is well defined in $Q(1 + 2^{-j-1})$.

Note that if Q and Q' are different cubes from $\mathcal{W}_j(D)$, then $Q(3/2)$ and $Q'(3/2)$ are disjoint. Thus no problems arise concerning the well-definedness of F_f^j in $Q(1 + 2^{-j-1}) \cap Q'(1 + 2^{-j-1})$. The restriction of F_f^j to $Q(1 + 2^{-j-1})$ is locally λ_j-bilipschitz with respect to the hyperbolic metric, and by (8.28)

$$d_{\mathbb{H}}\big[F_f^j, F_f; Q(1 + 2^{-j-1})\big] = d_{\mathbb{H}}\big[\hat{h}_Q, \beta_{Q,f} \circ F_f \circ \alpha_Q; Q_0(1 + 2^{-j-1})\big] \leq \delta_j\,. \tag{8.29}$$

We conclude that F_f^j is a well-defined mapping of $V_j(D)$ into $\mathbb{H}^{n+1}$, that F_f^j is locally λ_j-bilipschitz with respect to the hyperbolic metric as $\lambda_1 \leq \lambda_{j-1} \leq \lambda_j$, and, because of (8.26), (8.27), and (8.29), that

$$d_{\mathbb{H}}\big[F_f^j, F_f; V_j(D)\big] \leq \delta_j\,. \tag{8.30}$$

To complete the induction step—and, with it, the proof of the lemma—we need only confirm that F_f^j is injective.

In view of (8.30), the assumption that $\delta_j \leq \delta_N \leq \rho$, and Lemma 8.2.8, it suffices to check that $F_f^j(z) \neq F_f^j(z')$ whenever z and z' are points of $V_j(D)$ for which $0 < d_{\mathbb{H}}(z, z') \leq M$.

Fix z and z' fitting this description. If z and z' lie in $W_{j-1}(D)$, then $F_f^j(z) = F_f^{j-1}(z) \neq F_f^{j-1}(z') = F_f^j(z')$ by the induction hypothesis. Assume, therefore, that one of the two points z or z' is in $Q(1 + 2^{-j-1})$, where Q is a cube from $\mathcal{W}_j(D)$. For definiteness, say z has this property. Since z then belongs to $Q(5/4)$ and since

$$d_{\mathbb{H}}(z, z') \leq M = d_{\mathbb{H}}\big[Q(5/4), \partial Q(4/3)\big] < d_{\mathbb{H}}\big[Q(1 + 2^{-j-1}), Q'(1 + 2^{-j-1})\big]$$

for any Q' in $\mathcal{W}_j(D)$ other than Q, either z and z' are both points of $Q(1+2^{-j-1})$, in which case it is evident that $F_f^j(z) \neq F_f^j(z')$, or z' is a point of $W_{j-1}(D) \cap Q(4/3)$.

In the latter event $w = \alpha_Q^{-1}(z)$ and $w' = \alpha_Q^{-1}(z')$ are distinct points of W_Q, a set in which $h_Q = \hat{h}_Q$. Accordingly,

$$F_f^j(z') = F_f^{j-1}(z') = \beta_{Q,f}^{-1} \circ h_Q(w') = \beta_{Q,f}^{-1} \circ \hat{h}_Q(w') \neq \beta_{Q,f}^{-1} \circ \hat{h}_Q(w) = F_f^f(z)\,.$$

We infer that F_f^j is an injection, which finishes the proof. □

8.3.2. The extension theorem. The effort that went into Lemma 8.3.3 has an immediate payoff, as we are about to find out. For a domain D in $\mathbb{R}^n$ we shall use $\hat{D}$ as an abbreviation for the interior of $D^*(r_2)$. The extension theorem can now be formulated as follows:

THEOREM 8.3.4. *Suppose that n is a positive integer, that $1 \leq K < \infty$, and that $\varepsilon > 0$. Then there exists a constant $\lambda \geq 1$, determined entirely by n, K, and ε, for which the following assertion is true: each K-quasiconformal embedding $f : D \to \mathbb{R}^n$, where D is a domain in $\mathbb{R}^n$, admits an extension to an embedding $F^* : D^*(r_2) \to \mathbb{R}_+^{n+1}$ whose restriction F to the domain $\hat{D}$ is a locally λ-bilipschitz mapping with respect to the hyperbolic metric, and hence a λ^{2n}-quasiconformal mapping of $\hat{D}$ onto a domain in $\mathbb{H}^{n+1}$, which satisfies the condition that $d_{\mathbb{H}}(F, F_f; \hat{D}) \leq \varepsilon$.*

PROOF. Fix a domain D in $\mathbb{R}^n$ and a K-quasiconformal embedding $f : D \to \mathbb{R}^n$. Let δ_N, $\lambda = \lambda_N$ and $F_f^N : V_N(D) \to \mathbb{H}^{n+1}$ be as in Lemma 8.3.3. It follows from Lemma 8.2.12 that $V_N(D)$ contains the domain $\hat{D}$. As a result, $F = F_f^N|\hat{D}$ provides an embedding of $\hat{D}$ into $\mathbb{H}^{n+1}$ which is locally λ-bilipschitz in the hyperbolic metric—Lemma 8.1.2 certifies that F is a λ^{2n}-quasiconformal mapping—and obeys the condition $d_{\mathbb{H}}(F, F_f; \hat{D}) \leq \delta_N \leq \varepsilon$. This inequality makes it plain that

$$\lim_{z \to x} F(z) = \lim_{z \to x} F_f(z) = f(x)$$

for every x in D, so the function $F^* : D^*(r_2) \to \mathbb{R}_+^{n+1}$ given by $F^*(z) = F(z)$ if z belongs to $\hat{D}$ and $F^*(x) = f(x)$ if x is in D provides an embedding with the desired properties. □

There is a special case of Theorem 8.3.4 that deserves a statement of its own; see also Theorem 6.5.20.

THEOREM 8.3.5. *Suppose that n is a positive integer, that $1 \leq K < \infty$, and that $\varepsilon > 0$.*

Then there is a constant λ, which depends only on n, K, and ε, such that each K-quasiconformal self-mapping f of $\mathbb{R}^n$ admits an extension to a homeomorphism F^* of $\mathbb{R}_+^{n+1}$ whose restriction F to $\mathbb{H}^{n+1}$ is a λ-bilipschitz homeomorphism of $\mathbb{H}^{n+1}$ with respect to the hyperbolic metric, and hence is a λ^{2n}-quasiconformal self-mapping of $\mathbb{H}^{n+1}$, that satisfies the condition

$$d_{\mathbb{H}}(F, F_f; \mathbb{H}^{n+1}) \leq \varepsilon.$$

PROOF. We apply Theorem 8.3.4 in the case $D = \mathbb{R}^n$. Then $D^*(r_2) = \mathbb{R}_+^{n+1}$ and $\hat{D} = \mathbb{H}^{n+1}$. Let f be a K-quasiconformal mapping of $\mathbb{R}^n$ onto itself, and let F^* be an extension of f that meets the specifications of Theorem 8.3.4. The mapping F^* can be extended by reflection to a λ^{2n}-quasiconformal embedding $\hat{F}$ of $\mathbb{R}^{n+1}$ into itself. By the remark after Theorem 6.4.25, $\hat{F}$ maps $\mathbb{R}^{n+1}$ onto itself. It follows that $F^*(\mathbb{R}^{n+1}_+) = \mathbb{R}^{n+1}_+$ and $F(\mathbb{H}^{n+1}) = \mathbb{H}^{n+1}$. Lemma 8.1.3 accounts for the fact that F is globally λ-bilipschitz in $\mathbb{H}^{n+1}$ with respect to the hyperbolic metric. □

A refinement of Theorem 8.3.5 also bears mentioning.

Corollary 8.3.6. *Suppose that $n \geq 2$ and that $1 \leq K < \infty$. Then there is a constant λ, which is a function of n and K, such that each K-quasiconformal self-mapping f of $\hat{\mathbb{R}}^n$ admits an extension to a homeomorphism F^* of $\bar{\mathbb{H}}^{n+1}$ whose restriction F to $\mathbb{H}^{n+1}$ is a λ-bilipschitz homeomorphism of $\mathbb{H}^{n+1}$ with respect to the hyperbolic metric and is thus a λ^{2n}-quasiconformal self-mapping of $\mathbb{H}^{n+1}$.*

Proof. Let λ be a constant that has the feature described in Theorem 8.3.5 for $\varepsilon = 1$. Given a K-quasiconformal mapping f of $\hat{\mathbb{R}}^n$ onto itself, we select a mapping φ from Möb$(\mathbb{H}^{n+1})$ that transforms $f(\infty)$ to ∞. We apply Theorem 8.3.5 to the function $g = \varphi \circ f|\mathbb{R}^n$ and arrive at an extension G^* of g to a homeomorphism of $\mathbb{R}^{n+1}_+$ whose restriction to $\mathbb{H}^{n+1}$ is a λ-bilipschitz mapping when viewed from the perspective of the hyperbolic metric.

Moreover, by defining $G^*(\infty) = \infty$ we turn G^* into a homeomorphism of $\bar{\mathbb{H}}^{n+1}$ as the extension $\hat{G}$ of G^* obtained via reflection in $\mathbb{R}^n$ is a quasiconformal mapping of $\mathbb{R}^{n+1}$ onto itself, so $\hat{G}(x) \to \infty$ as $x \to \infty$.

Since φ is both a 1-quasiconformal mapping and a hyperbolic isometry, $F^* = \varphi^{-1} \circ G^*$ yields an extension of f with the stated properties. □

Corollary 8.3.6 has an analogue on the sphere $\mathbb{S}^n$. Indeed, when $n \geq 2$ there is a natural definition of quasiconformality for a homeomorphism f of $\mathbb{S}^n$: f is K-quasiconformal if and only if $\pi^{-1} \circ f \circ \pi$ is K-quasiconformal, where $\pi : \hat{\mathbb{R}}^n \to \mathbb{S}^n$ is the stereographic projection.

For example, if F is a K-quasiconformal mapping of B^{n+1} onto itself and F^* is the homeomorphic extension of F to $\overline{B}^{n+1}$, then $f = F^*|\mathbb{S}^n$ is K-quasiconformal. More generally, if $0 < r < 1$ and F^* is an embedding of $\overline{B}^{n+1} \setminus \overline{B}^{n+1}(r)$ into $\hat{\mathbb{R}}^{n+1}$ such that $F^*(\mathbb{S}^n) = \mathbb{S}^n$ and such that F^* is K-quasiconformal in $B^{n+1} \setminus \overline{B}^{n+1}(r)$, then $f = F^*|\mathbb{S}^n$ is K-quasiconformal.

To see this, let Φ be the Möbius transformation of $\hat{\mathbb{R}}^{n+1}$ that maps $\mathbb{H}^{n+1}$ to B^{n+1} and coincides with π on $\hat{\mathbb{R}}^n$. Theorem 6.5.20 and the remark subsequent to it, applied to $\Phi^{-1} \circ F^* \circ \Phi$, show that $\pi^{-1} \circ f \circ \pi$ is K-quasiconformal. Here we note that the proof of Theorem 6.5.20 and the remark alluded to are actually valid in the more general situation indicated above.

The counterpart of Corollary 8.3.6 for $\mathbb{S}^n$ provides a converse to this observation.

Corollary 8.3.7. *Suppose that $n \geq 2$ and that $1 \leq K < \infty$. There is a constant λ, which is a function of n and K, such that each K-quasiconformal self-mapping of $\mathbb{S}^n$ admits an extension to a homeomorphism F^* of $\overline{B}^{n+1}$ whose restriction to B^{n+1} is a λ-bilipschitz homeomorphism of B^{n+1} with respect to the hyperbolic metric and is thus a λ^{2n}-quasiconformal self-mapping of B^{n+1}.*

Proof. Let Φ be the Möbius transformation of $\hat{\mathbb{R}}^{n+1}$ that entered into the discussion immediately preceding the statement of the corollary. We have noted that Φ is an isometry of $\mathbb{H}^{n+1}$ onto B^{n+1} when these spaces are given their hyperbolic metrics. If f is a K-quasiconformal self-mapping of $\mathbb{S}^n$, then Corollary 8.3.7 informs us that $g = \pi^{-1} \circ f \circ \pi$ has an extension to a homeomorphism G^* of $\overline{\mathbb{H}}n + 1$

whose restriction to $\mathbb{H}^{n+1}$ is hyperbolically λ-bilipschitz, where λ depends only on n and K.

Clearly $F^* = \Phi \circ G^* \circ \Phi^{-1}$ is a homeomorphic extension of f to $\overline{B}^{n+1}$ that in B^{n+1} is λ-bilipschitz relative to the hyperbolic metric. □

Corollaries 8.3.6 and 8.3.7 are by no means the end of the story as far as quasiconformal extensions from $\hat{\mathbb{R}}^n$ to $\mathbb{H}^{n+1}$ or from $\mathbb{S}^n$ to B^{n+1} are concerned. In fact, these results have triggered a hunt for more direct methods of constructing quasiconformal extensions and more explicit ways of representing them. Moreover, it is sometimes required to find extensions that obey certain side conditions. As they stand, Corollaries 8.3.6 and 8.3.7 do not have such flexibility. What one ultimately seeks are extension operators—integral operators, for instance, like the one yielding the Beurling-Ahlfors extension—that produce extensions when the boundary maps are fed into them.

In the case of B^n ($n = 2$ is allowed here) one looks for an operator $\mathbf{E}$ that assigns to each homeomorphism f of $\mathbb{S}^{n-1}$ a homeomorphic extension $F^* = \mathbf{E}(f)$ to $\overline{B}^n$ with the following properties:

(i) $\mathbf{E}(\text{identity}) = \text{identity}$.

(ii) If f arises as the boundary correspondence of some quasiconformal self-mapping of B^n (when $n \geq 3$ this is equivalent to saying that f is a quasiconformal homeomorphism of $\mathbb{S}^{n-1}$), then $F = F^*|B^n$ is itself a quasiconformal mapping and $K(F)$ depends completely on data derived from f.

(iii) $\mathbf{E}$ is continuous with respect to the topology of uniform convergence on $\mathbb{S}^{n-1}$ and some reasonable topology for $C(\overline{B}^n, \mathbb{R}^n)$—for instance the topology of Euclidean uniform convergence on $\overline{B}^n$ or the topology of Euclidean locally uniform convergence in B^n.

The above are the minimal requirements for $\mathbf{E}$. There are numerous other special features which it would be highly desirable for $\mathbf{E}$ to incorporate. We continue our list with several of these.

(iv) F is always a (real analytic?) diffeomorphism of B^n.

(v) If f is a homeomorphism of $\mathbb{S}^{n-1}$ induced by a quasiconformal self-mapping of B^n, then F is λ-bilipschitz in the hyperbolic metric, where λ depends only on data associated with f.

(vi) $\mathbf{E}$ is natural in relation to the action of Möb(B^n); that is, with the obvious meaning

$$\mathbf{E}(\psi \circ f \circ \varphi) = \psi \circ \mathbf{E}(f) \circ \varphi$$

whenever φ and ψ are Möbius transformations of $\hat{\mathbb{R}}^n$ that preserve B^n—hence, also leave $\mathbb{S}^{n-1}$ invariant Thus, in conjunction with (i), (vi) implies that $\mathbf{E}(\varphi) = \varphi$ for every φ from Möb(B^n).

(vii) $\mathbf{E}(g \circ f) = \mathbf{E}(g) \circ \mathbf{E}(f)$ for all homeomorphisms f and g of $\mathbb{S}^{n-1}$ or, at least, for any such homeomorphisms that admit quasiconformal extensions to B^n.

In the case $n = 2$ an operator $\mathbf{E}$ with properties (i)–(vi) was discovered by Douady and Earle [**34**]. Although the Douady-Earle operator does have direct analogues in all dimensions, only in the plane is it certain to produce homeomorphic extensions. The existence of an operator enjoying all seven of these properties remains an open question even when $n = 2$.