GEOMETRY

JOUNI PARKKONEN

Contents

Last update: December 12, 2014.

1.1. **Metric spaces.** A function $d\colon X \times X \to [0, +\infty[$ is a *metric* in the nonempty set X if it satisfies the following properties

(1) $d(x, x) = 0$ for all $x \in X$ and $d(x, y) > 0$ if $x \neq y$,
(2) $d(x, y) = d(y, x)$ for all $x, y \in X$, and
(3) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$ (the *triangle inequality*).

The pair (X, d) is a *metric space.* Open and closed balls in a metric space, continuity of maps between metric spaces and other "metric properties" are defined in the same way as in Euclidean space, using the metrics of X and Y instead of the Euclidean metric.

Example 1.1. (a) The space $\mathbb{R}^n$ with the Euclidean distance is a metric space, see Section 1.2.
(b) The circle $\mathbb{S}^1$ with the distance between two points defined as their angle as vectors in $\mathbb{E}^2$ is a metric space, see Section 2 for details and generalisations.
(c) Let $X \neq \emptyset$. The discrete metric $d\colon X \times X \to [0, \infty[$ is defined by setting $d(x, x) = 0$ for all $x \in X$ and $d(x, y) = 1$ for all $x, y \in X$ if $x \neq y$.
(d) It is easy to check that for any $\alpha > 1$, the expression $h_\alpha(st,) = |s - t|^\alpha$ does not define a metric on $\mathbb{R}$ as it fails to satisfy the triangle inequality:

$$h_\alpha(0, 2) > 2 = 1 + 1 = h_\alpha(0, 1) + h_\alpha(1, 2)\,.$$

On the other hand, it can be shown that h_α is a metric if $0 < \alpha \leq 1$.

If (X_1, d_1) and (X_2, d_2) are metric spaces, then a map $i\colon X \to Y$ is an *isometric embedding*, if

$$d_2(i(x), i(y)) = d_1(x, y)$$

for all $x, y \in X_1$. If the isometric embedding i is a bijection, then it is called an *isometry* between X and Y. An isometry $i\colon X \to X$ is called an isometry of X.

The isometries of a metric space X form a group $\mathrm{Isom}(X)$, the *isometry group* of X, with the composition of mappings as the group law.

A map $i\colon X \to Y$ is a *locally isometric embedding* if each point $x \in X$ has a neighbourhood U such that the restriction of i to U is an isometric embedding. A (locally) isometric embedding $i\colon I \to X$ is

(1) a *(locally) geodesic segment*, if $I \subset \mathbb{R}$ is a (closed) bounded interval,
(2) a *(locally) geodesic ray*, if $I = [0, +\infty[$, and
(3) a *(locally) geodesic line*, if $I = \mathbb{R}$.

1.2. **Euclidean space.** Let us denote the *Euclidean inner product* of $\mathbb{R}^n$ by

$$(x|y) = \sum_{i=1}^{n} x_i y_i\,.$$

The *Euclidean norm* $\|x\| = \sqrt{(x|x)}$ defines the *Euclidean distance* $d(x, y) = \|x - y\|$. The triple $\mathbb{E}^n = (\mathbb{R}^n, (\cdot|\cdot), \|\cdot\|)$ is n-dimensional *Euclidean space.*

Proposition 1.2. *The Euclidean distance is a metric.*

Proof. The first two properties of a metric are clear from the expression of the metric. It suffices to show that the triangle inequality holds. Let $x, y, z \in \mathbb{E}^n$. Using the linearity and symmetry properties of the inner product and Cauchy's inequality, we get

$$\begin{aligned} d_{\mathbb{E}^n}(x,y)^2 &= (x-y|x-y) = (x-z+z-y|x-z+z-y) \\ &= (x-z|x-z) + 2(x-z|z-y) + (z-y|z-y) \\ &\le d_{\mathbb{E}^n}(x,z)^2 + 2d_{\mathbb{E}^n}(x,z)d_{\mathbb{E}^n}(z,y) + d_{\mathbb{E}^n}(z,y)^2 \\ &= \left(d_{\mathbb{E}^n}(x,z) + d_{\mathbb{E}^n}(z,y)\right)^2 , \end{aligned}$$

which implies the triangle inequality because all the terms in the inequality are positive. □

Euclidean space is a *geodesic metric space*: For any two distinct points $x, y \in \mathbb{E}^n$, the map

$$t \mapsto x + t\,\frac{y-x}{\|y-x\|},$$

is a geodesic line that passes through the points x and y. Indeed, for any $x_0 \in \mathbb{E}^n$ and any $u \in \mathbb{S}^{n-1}$, let $j_{x_0,u}\colon \mathbb{R} \to \mathbb{E}^n$ be the map

$$j_{x_0,u}(t) = x_0 + tu\,.$$

For any $s, t \in \mathbb{R}$, we have

$$d_{\mathbb{E}^n}(j_{x_0,u}(t), j_{x_0,v}(s)) = \|x_0 + tu - (x_0 + su)\| = \|(t-s)u\| = |t-s|\|u\| = d_{\mathbb{E}^1}(t,s)\,.$$

The restriction $j_{x,y}|_{[0,\|x-y\|]}$ is a geodesic segment that *connects* x to y: $j(0) = x$ and $j(\|x-y\|) = y$. In fact, this is the only geodesic segment that connects x to y up to replacing the interval of definition $[0, \|x-y\|]$ of the geodesic by $[a, a+\|x-y\|]$ for some $a \in \mathbb{R}$. More precisely: A metric space (X, d) is *uniquely geodesic*, if for any $x, y \in X$ there is exactly one geodesic segment $j\colon [0, d(x,y)] \to X$ such that $j(0) = x$ and $j(d(x,y)) = y$.

Proposition 1.3. *Euclidean space is uniquely geodesic.*

Proof. If g is a geodesic segment that connects x to y and z is a point in the image of g, then, by definition, $\|x-z\|+\|z-y\| = \|x-y\|$. But, using Cauchy's inequality, it is easy to see that the Euclidean triangle inequality becomes an equality if and only if z is in the image of the linear segment $j|_{[0,\|x-y\|]}$. □

Even if the proof of the above proposition appears obvious, it uses the connection of the Euclidean metric with the inner product in an essential way. There are plenty of examples of metric spaces arising from vector spaces endowed with a norm that are not uniquely geodesic. For example, the expression

$$d_\infty(x,y) = \max\{|x_1 - y_1|, |x_2 - y_2|\}$$

defines a metric on $\mathbb{R}^2$. It is easy to check that, among many others, the mappings $j_1, j_2\colon [0,1] \to (\mathbb{R}^2, d_\infty)$ defined by $j_1(t) = t(1,0)$ and

$$j_2(t) = \begin{cases} t(1,1), & \text{if } 0 \le t \le \frac{1}{2}, \\ (t, 1-t), & \text{if } \frac{1}{2} \le t \le 1 \end{cases}$$

are both geodesic segments in $(\mathbb{R}^2, d_\infty)$ connecting 0 to $(1,0)$.

If a metric space X is uniquely geodesic and $x, y \in X$, $x \ne y$, we denote the (image of the) unique geodesic segment connecting x to y by $[x,y]$.

1.3. **Isometries.** We will now study the isometries of Euclidean space more closely. The *(Euclidean) orthogonal group* of dimension n is

$$\begin{aligned} \mathrm{O}(n) &= \{A \in \mathrm{GL}_n(\mathbb{R}) : (Ax|Ay) = (x|x) \text{ for all } x, y \in \mathbb{E}^n\} \\ &= \{A \in \mathrm{GL}_n(\mathbb{R}) : A^T A = I_n\} . \end{aligned}$$

Recall the following basic result from linear algebra:

Lemma 1.4. *An $n \times n$-matrix $A = (a_1, \ldots, a_n)$ is in $\mathrm{O}(n)$ if and only if the vectors $a_1, \ldots, a_n$ form an orthonormal basis of $\mathbb{E}^n$.* □

It is easy to check that elements of $\mathrm{O}(n)$ give isometries on $\mathbb{E}^n$ for any $n \in \mathbb{N}$: Let $A \in \mathrm{O}(n)$ and let $x, y \in \mathbb{E}^n$. Now

$$\begin{aligned} d(Ax, Ay)^2 &= (Ax - Ay|Ax - Ay) = (A(x-y)|A(-y)) \\ &= (A^T A(x-y)|x-y) = (x-y|x-y) \\ &= d(x-y)^2 . \end{aligned}$$

For any $b \in \mathbb{R}^n$, let $t_b(x) = x + b$ be the translation by b. Again, it is easy to see that translations are isometries of $\mathbb{E}^n$. The *translation group* is

$$\mathrm{T}(n) = \{t_b : b \in \mathbb{R}^n\}$$

Orthogonal maps and translations generate the *Euclidean group*

$$\mathrm{E}(n) = \{x \mapsto Ax + b : A \in \mathrm{O}(n), b \in \mathbb{R}^n\}$$

which consists of isometries of $\mathbb{E}^n$.

If a group G acts on a space X, and x is a point in X, the set

$$G(x) = \{g(x) : g \in g\}$$

is the G-*orbit* of x. The action of a group is said to be *transitive* if $G(x) = X$ for some (and therefore for any) $x \in X$. A more elementary way to express this is that a group G acts transitively on X if for all $x, y \in X$ there is some $g \in G$ such that $g(x) = y$.

Proposition 1.5. $\mathrm{E}(n)$ *acts transitively by isometries on $\mathbb{E}^n$. In particular,* $\mathrm{Isom}(\mathbb{E}^n)$ *acts transitively on $\mathbb{E}^n$.*

Proof. The Euclidean group of $\mathbb{E}^n$ contains the group of translations $\mathrm{T}(n)$ as a subgroup. This subgroup acts transitively because for any $x, y \in \mathbb{R}^n$, we have $T_{y-x}(x) = y$. □

An *affine hyperplane* of $\mathbb{E}^n$ is a subset of the form

$$H = H(P, u) = P + u^\perp ,$$

where $P, u \in \mathbb{E}^n$ and $\|u\| = 1$. The *reflection* in H is the map

$$r_H(x) = x - 2(x - P|u)u .$$

Reflections are very useful isometries, the following results give some of their basic properties. For any mapping $f \colon X \to X$, the *fixed point set* of f is

$$\operatorname{fix} f = \{x \in X : f(x) = x\} .$$

Proposition 1.6. *Let H be an hyperplane in $\mathbb{E}^n$. Then*

(1) $r_H \circ r_H$ is the identity.

(2) $r_H \in \mathrm{E}(n)$. In particular, r_H is an isometry, and if $0 \in H$, then $r_H \in \mathrm{O}(n)$.

(3) $d(r_H(x), y) = d(x, y)$ for all $x \in \mathbb{E}^n$ and all $y \in H$.

(4) The fixed point set of r_H is H.

Proof. We will prove (3) and leave the rest as exercises. Let $x \in \mathbb{E}^n$ and $y \in H$. We have $r_H(x) = x - 2(x-y|u)u$, which implies

$$\begin{aligned} d(r_H(x), y)^2 &= (r_H(x) - y | r_H(x) - y) = (x - y - 2(x-y|u)u | x - y - 2(x-y|u)u) \\ &= (x-y|x-y) - 4(x-y|(x-y|u)u) + 4((x-y|u)u|(x-y|u)u) \\ &= (x-y|x-y) = d(x,y)^2 . \end{aligned}$$

□

The *bisector* of two distinct points p and q in $\mathbb{E}^n$ is the affine hyperplane

$$\operatorname{bis}(p,q) = \{x \in \mathbb{E}^n : d(x,p) = d(x,q)\} = \frac{p+q}{2} + (p-q)^\perp .$$

Proposition 1.7. *(1) If $r_H(x) = y$ and $x \notin H$, then $H = \operatorname{bis}(x,y)$.*
(2) If $p, q \in \mathbb{E}^n$, $p \neq q$, then $r_{\operatorname{bis}(p,q)}(p) = q$.
(3) Let $\phi \in \operatorname{Isom}(\mathbb{E}^n)$, $\phi \neq \operatorname{id}$. If $a \in \mathbb{E}^n$, $\phi(a) \neq a$, then the fixed points of ϕ are contained in $\operatorname{bis}(a, \phi(a))$.
(4) Let $\phi \in \operatorname{Isom}(\mathbb{E}^n)$, $\phi \neq \operatorname{id}$. If H is a hyperplane such that $\phi|_H$ is the identity, then $\phi = r_H$.

Proof. (1) follows from Proposition 6.10(3).
(2) From the definitions we get

$$r_{\operatorname{bis}(p,q)}(p) = p - 2(p - \frac{p+q}{2} | p - q) \frac{p-q}{\|p-q\|^2} = q .$$

(3) If $\phi(b) = b$, then $d(a,b) = d(\phi(a), \phi(b)) = d(\phi(a), b)$, so that $b \in \operatorname{bis}(a, \phi(a))$.
(4) Let $a \notin H$ be a point that is not fixed by ϕ. Claim (3) implies that H is contained in $\operatorname{bis}(a, \phi(a))$ and as the dimensions agree, we have $H = \operatorname{bis}(a, \phi(a))$. Thus, by Claim (2), $r_H(a) = \phi(a)$. But this holds for all $a \notin H$. As $r_H|_H = \phi_H = \operatorname{id}_H$, we have $\phi = r_H$. □

Next, we want to prove that all isometries of Euclidean space $\mathbb{E}^n$ are affine transformations with an orthogonal linear part.

Theorem 1.8. $\operatorname{Isom}(\mathbb{E}^n) = \operatorname{E}(n)$.

The idea of the proof is to show that each isometry of $\mathbb{E}^n$ is the composition of reflections in affine hyperplanes. In order to do this, we show that the isometry group has a stronger transitivity property than what was noted above.

Proposition 1.9. *Let $p_1, p_2, \dots, p_k, q_1, q_2, \dots, q_k \in \mathbb{E}^n$ be points that satisfy*

$$d(p_i, p_j) = d(q_i, q_j)$$

for all $i, j \in \{1, 2, \dots, k\}$. Then, there is an isometry $\phi \in \operatorname{Isom}(\mathbb{E}^n)$ such that $\phi(p_i) = q_i$ for all $i \in \{1, 2, \dots, k\}$. Furthermore, the isometry ϕ is the composition of at most k reflections in affine hyperplanes.

Proof. We construct the isometry by induction. If $p_1 = q_1$, let ϕ_1 be the identity, otherwise, let ϕ_1 be the reflection in the bisector of p_1 and q_1. Let $m > 1$ and assume that there is an isometry ϕ_m such that $\phi(p_i) = q_i$ for all $i \in \{1, 2, \dots, m\}$, which is the composition of at most m reflections.

Assume that $\phi_m(p_{m+1}) \neq q_{m+1}$. Now, $q_1, \dots q_m \in \operatorname{bis}(\phi_m(p_{m+1}), q_{m+1})$ because for each $1 \leq i \leq m$, we have

$$d(q_i, \phi_m(p_{m+1})) = d(\phi_m(p_i), \phi_m(p_{m+1})) = d(p_i, p_{m+1}) = d(q_i, q_{m+1}) .$$

Thus, the map

$$\phi_{m+1} = r_{\operatorname{bis}(\phi_m(p_{m+1}), q_{m+1})} \circ \phi_m$$

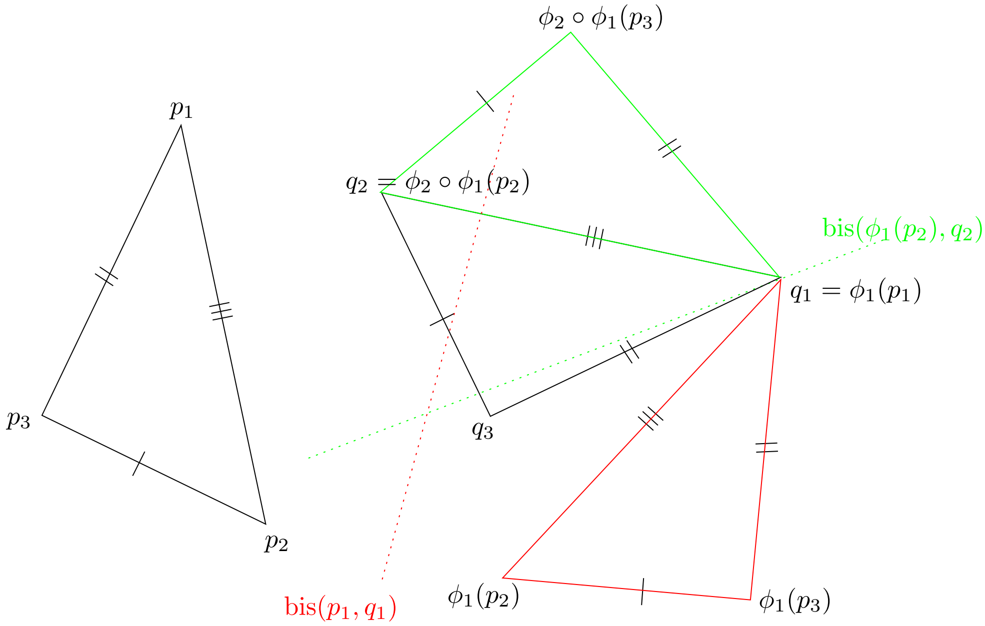

FIGURE 1.

satisfies $\phi_{m+1}(p_i) = q_i$ for all $1 \leq i \leq m+1$. $\square$

Note that Proposition 1.9 implies that if T and T' are two triangles in $\mathbb{E}^n$ with equal sides, then there is an isometry ϕ of $\mathbb{E}^n$ such that $\phi(T) = T'$.

Proof of Theorem 1.8. We already observed that elements of $\mathrm{E}(n)$ are isometries. It remains to show the opposite inclusion.

Consider the set $\{0, e_1, \ldots, e_n\}$ in $\mathbb{E}^n$. Note that this set is not contained in any affine hyperplane. $\mathbb{R}^n$.

Let $\phi \in \mathrm{Isom}(\mathbb{E}^n)$. Proposition 1.9 implies that there is an isometry $\phi_0 \in \mathrm{O}(n)$ such that $\phi_0(\phi(e_i)) = e_i$ for all $1 \leq i \leq m$ and $\phi_0(\phi(0)) = 0$. Since the set of fixed points of $\phi_0 \circ \phi$ contains the points $0, e_1, \ldots, e_n$, the fixed point set of ϕ_0 is not contained in any affine hyperplane. Proposition 6.11 implies that $\phi_0 \circ \phi$ is the identity map. Thus, $\phi = \phi_0^{-1}$. In particular, $\phi \in \mathrm{O}(n)$, which is all we needed to show. $\square$

Let X be a metric space. The *stabiliser* of a point $x \in X$ is

$$\mathrm{Stab}\, x = \{F \in \mathrm{Isom}\, X : F(x) = x\}\,.$$

Proposition 1.10. *The stabiliser in* $\mathrm{Isom}(\mathbb{E}^n)$ *of any point* $x \in \mathbb{E}^n$ *is isomorphic to* $\mathrm{O}(n)$. *An isometry* F *of* $\mathbb{E}^n$ *fixes* $b \in \mathbb{E}^n$ *if and only if there is an orthogonal linear map* F_0 *such that* $F = T_b \circ F_0 \circ T_b^{-1}$.

Proof. An element of $\mathrm{E}(n)$ fixes the origin if and only if it is an orthogonal linear transformation. Thus the claim holds for 0. If $b \in \mathbb{E}^n - \{0\}$ and $F \in \mathrm{Stab}\, b$, then $T_b^{-1} \circ F \circ T_b \in \mathrm{O}(n)$ and for any $A \in \mathrm{O}(n)$, $\quad T_b \circ A \circ T_b^{-1} \in \mathrm{fix}\, b$ $\square$

Proposition 1.11. *For each affine k-plane* P, *there is an isometry* $\phi \in \mathrm{Isom}(\mathbb{E}^n)$ *such that*

$$\phi(P) = \{x \in \mathbb{E}^n : x^{k+1} = x^{k+2} = \cdots = x^n = 0\}\,.$$

Each affine k-plane of $\mathbb{E}^n$ *is isometric with* $\mathbb{E}^k$.

Proof. This is a direct generalisation of Proposition 1.5. The details are left as an exercise. $\square$

2. The sphere

The unit sphere in $(n-1)$-dimensional Euclidean space is

$$\mathbb{S}^n = \{x \in \mathbb{E}^{n+1} : \|x\| = 1\} .$$

Let us show that the *angle distance*

$$d_{\mathbb{S}^n}(x, y) = \arccos(x|y) \in [0, \pi] \tag{1}$$

is a metric. In order to do this, we will use the analog of the Euclidean law of cosines, but first we have to define the objects that are studied in spherical geometry.

Each 2-dimensional linear subspace $T \subset \mathbb{R}^{n+1}$ intersects $\mathbb{S}^n$ in a *great circle.* If $A \in \mathbb{S}^n$ and $u \in \mathbb{S}^n$ is orthogonal to A ($u \in A^\perp$), then the path $j_{A,u} \colon \mathbb{R} \to \mathbb{S}^n$,

$$j_{A,u}(t) = A \cos t + u \sin t ,$$

parametrises the great circle $\langle A, u\rangle \cap \mathbb{S}^n$, where (A, u) is the linear span of A and u. The vectors A and u are linearly independent, so $\langle A, u\rangle$ is a 2- plane.

Lemma 2.1. *If $d_{\mathbb{S}^n}$ is a metric, then $j_{A,u}$ is a locally geodesic line.*

Proof. Observe that as A and u are unit vectors such that $(A|u) = 0$, we have

$$\begin{aligned}
(j_{A,u}(s)|j_{A,u}(t)) &= (A \cos s + u \sin s|A \cos t + u \sin t) \\
&= \|A\|^2 \cos s \cos t + (\cos s \sin t + \sin s \cos t)(A|u) + \sin s \sin t\|u\|^2 \\
&= \cos s \cos t + \sin s \sin t = \cos(s - t) . \qquad (2)
\end{aligned}$$

Thus,

$$d(j_{A,u}(s), j_{A,u}(t)) = \arccos(j_{A,u}(s)|j_{A,u}(t)) = \arccos\cos(s - t) = |s - t| ,$$

which implies that the restriction of $j_{A,u}$ to any segment of length less than π is an isometric embedding. $\square$

Note that the computation (2) applied with $s = t$ implies that the image of the mapping $j_{A,u}$ is contained in $\mathbb{S}^1$.

If $A, B \in \mathbb{S}^n$ such that $B \neq \pm A$, then there is a unique plane that contains both points. Thus, there is unique great circle that contains A and B, in the remaining cases, there are infinitely many such planes. The great circle is parametrised by the map $j_{A,u}$, with

$$u = \frac{B - (B|A)A}{\|B - (B|A)A\|} = \frac{B - (A|B)A}{\sqrt{1 - (A|B)^2}} . \tag{3}$$

Now $j(0) = A$ and $j(d(A, B)) = B$.

If $B = -A$, then there are infinitely many great circles through A and B: the map $j_{A,u}$ parametrises a great circle through A and B for any $u \in A^\perp$.

We call the restriction of any $j_{A,u}$ as above to any compact interval $[0, s]$ a *spherical segment*, and u is called the *direction* of $j_{A,u}$. Once we have proved that d is a metric, it is immediate that a spherical segment is a geodesic segment.

Our proof showing that the expression (1) defines a metric is based on the spherical law of cosines. In order to prepare for this approach, we return briefly to Euclidean geometry. A *triangle* in Euclidean space consists of three points $A, B, C \in \mathbb{E}^n$ (the *vertices*) and of the three *sides* $[A, B]$, $[B, C]$ and $[C, A]$. Let the lengths of the sides be, in the corresponding order, c, a and b, and let the angles between the sides at the vertices A, B and C be α, β and γ.

These quantities are connected via

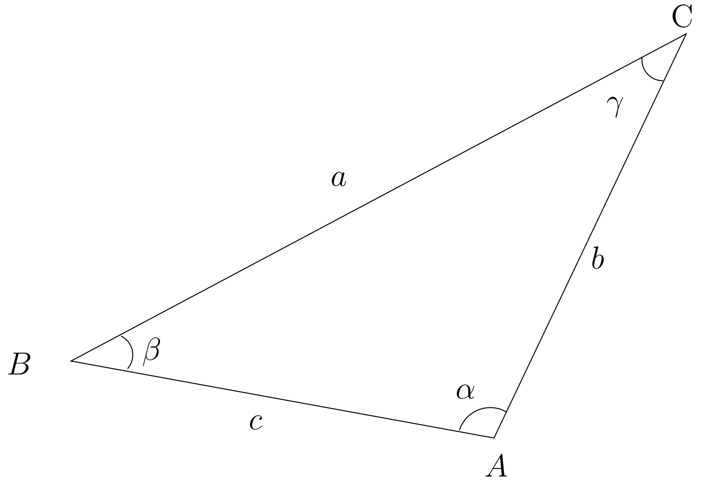

FIGURE 2.

Proposition 2.2 (The Euclidean law of cosines)**.** *In Euclidean geometry, the relation*

$$c^2 = a^2 + b^2 - 2ab\cos\gamma$$

holds for any triangle.

Proof. The proof is linear algebra:

$$\begin{aligned} c^2 &= \|B-A\|^2 = \|B-C+C-A\|^2 = b^2 + 2(B-C|C-A) + a^2 \\ &= b^2 + 2(B-C|C-A) + a^2 = b^2 - 2ab\cos\gamma + a^2\,. \end{aligned}$$ □

Proposition 2.3 (The Euclidean law of sines)**.** *In Euclidean geometry, the relation*

$$\frac{a}{\sin\alpha} = \frac{b}{\sin\beta} = \frac{c}{\sin\gamma}$$

holds for any triangle.

The law of cosines can be proved without knowing that $\mathbb{E}^n$ is uniquely geodesic. In fact, using the law of cosines, it is easy to prove that Euclidean space is uniquely geodesic, compare with the case of the sphere treated below .

A triangle in $\mathbb{S}^n$ is defined as in the Euclidean case but now the sides of the triangle are the spherical segments connecting the vertices.

Let $j_{A,u}([0, d(C,A)])$ be the side between C and A, and let $j_{A,v}([0, d(C,B)])v$ be the side between C and B. The angle between $j_{A,u}([0, d(C,A)])$ and $j_{A,v}([0, d(C,B)])$ is $\arccos(u|v)$, which is the angle at A between the sides $j_{A,u}([0, d(A,B)])$ and $j_{A,v}([0, d(A,B)])$ in the ambient space $\mathbb{E}^{n+1}$.

Now we can state and prove

Proposition 2.4 (The spherical law of cosines)**.** *In spherical geometry, the relation*

$$\cos c = \cos a\cos b + \sin a\sin b\cos\gamma$$

holds for any triangle.

Proof. Let u and v be the initial tangent vectors of the spherical segments $j_{C,u}$ from C to A and $j_{C,v}$ from C to B. As u and v are orthogonal to C, we have

$$\begin{aligned} \cos c &= (A|B) = (\cos(b)C + \sin(b)u|\cos(a)C + \sin(a)v) \\ &= \cos(a)\cos(b) + \sin(b)\sin(a)(u|v)\,. \end{aligned}$$ □

Proposition 2.5. *The angle distance is a metric on $\mathbb{S}^n$.*

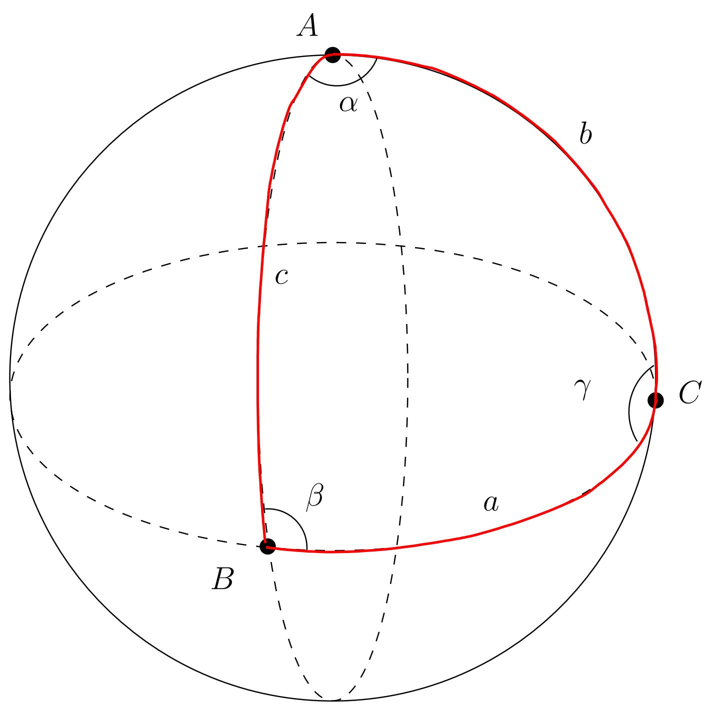

FIGURE 3.

Proof. Clearly, the triangle inequality is the only property that needs to be checked to show that the angle metric is a metric. Let $A, B, C \in \mathbb{S}^n$ be three distinct points and use the notation introduced above for triangles. The function

$$\gamma \mapsto f(\gamma) = \cos(a)\cos(b) + \sin(a)\sin(b)\cos(\gamma)$$

is strictly decreasing on the interval $[0, \pi]$, and

$$f(\pi) = \cos(a)\cos(b) - \sin(b)\sin(a) = \cos(a + b)\,.$$

Thus, the law of cosines implies that for all $\gamma \in [0, \pi]$, we have

$$\cos(c) = \cos(a)\cos(b) + \sin(a)\sin(b)\cos(\gamma) \geq \cos(a + b), \tag{4}$$

which implies $c \leq a + b$. Thus, the angle distance is a metric. □

Note that the inequality (4) is strict unless $\gamma = \pi$. This also implies that for triangles that are not completely contained in a great circle,

$$c < a + b < 2\pi - c\,. \tag{5}$$

We return to this observation in Section 4.

Proposition 2.6. *$(\mathbb{S}^n, d_{\mathbb{S}^n})$ is a geodesic metric space. If $d_{\mathbb{S}^n}(A, B) < \pi$, then there is a unique geodesic segment from A to B.*

Proof. If $x, y \in \mathbb{S}$ with $y \neq \pm x$, then, by Lemma 2.1, the spherical segment with direction given by the equation (3) is a geodesic segment that connects x to y. If the points x and y are antipodal, then it is immediate from the expression of the spherical segment that $j_{x,u}(\pi) = -x$. Thus, in this case there are infinitely many geodesic segments connecting x to y.

If j is a geodesic segment connecting A to B, then any C in $j([0, d(A, B)])$ satisfies

$$d_{\mathbb{S}^n}(A, C) + d_{\mathbb{S}^n}(C, B) = d_{\mathbb{S}^n}(A, B)$$

by definition of a geodesic segment. Equality holds in the triangle inequality if and only if $\gamma = \pi$. In this case, all the points A, B and C lie on the same great circle and C is contained in the side connecting A to B. Thus, the spherical segments are the only geodesic segments connecting A and B. If $A \neq \pm B$, then there is exactly one 2-plane containing both points. This proves the second claim. $\square$

Note that the sphere has no geodesic lines or rays because the diameter of the sphere is π.

2.1. **More on cosine and sine laws.** The law of cosines implies that a triangle in $\mathbb{E}^n$ or $\mathbb{S}^n$ is uniquely determined up to an isometry of the space, if the lengths of the three sides are known. In Euclidean space the angles are given by

$$\cos \gamma = \frac{a^2 + b^2 - c^2}{2ab}$$

and the corresponding equations for α and β obtained by permuting the sides and angles, and in the sphere we have

$$\cos \gamma = \frac{\cos c - \cos a \cos b}{\sin a \sin b} .$$

In Euclidean space, the three angles of a triangle do not determine the triangle uniquely but in $\mathbb{S}^n$ the angles determine a triangle uniquely. This is the content of

Proposition 2.7 (The second spherical law of cosines)**.** *In spherical geometry, the relation*

$$\cos c = \frac{\cos \alpha \cos \beta + \cos \gamma}{\sin \alpha \sin \beta}$$

holds for any triangle.

Proof. This formula follows from the first law of cosines by manipulation. The first law of cosines implies

$$\sin^2 \gamma = 1 - \cos^2 \gamma = \frac{1 + 2 \cos a \cos b \cos c - (\cos^2 + \cos^2 b + \cos^2 c)}{\sin^2 a \sin^2 b} = \frac{D}{\sin^2 a \sin^2 b} ,$$

and D is symmetric in a, b and c. Thus, using the law of cosines, we get

$$\begin{aligned} (6) \qquad & \frac{\cos \alpha \cos \beta + \cos \gamma}{\sin \alpha \sin \beta} \\ &= \frac{\frac{\cos a - \cos b \cos c}{\sin b \sin c} \frac{\cos b - \cos a \cos c}{\sin a \sin c} + \frac{\cos c - \cos a \cos b}{\sin a \sin b}}{\frac{D}{\sin a \sin b \sin^2 c}} = \cos c . \end{aligned}$$

$\square$

Spherical geometry even has its own sine law

Proposition 2.8 (The spherical law of sines)**.** *In spherical geometry, the relation*

$$\frac{\sin a}{\sin \alpha} = \frac{\sin b}{\sin \beta} = \frac{\sin c}{\sin \gamma}$$

holds for any triangle.

Proof. In the proof of the second law of cosines we saw that he first law of cosines implies that

$$\left(\frac{\sin c}{\sin \gamma} \right)^2 = \frac{\sin^2 a \sin^2 b \sin^2 c}{D} .$$

The claim follows because this expression is symmetric in a, b and c. $\square$

2.2. **Isometries.**

Proposition 2.9. *The orthogonal group* $\mathrm{O}(n+1)$ *acts transitively by isometries on* $\mathbb{S}^n$*. In particular,* $\mathrm{Isom}(\mathbb{S}^n)$ *acts transitively on* $\mathbb{S}^n$*.*

Proof. Let $A \in \mathrm{O}(n+1)$ and let $x \in \mathbb{E}^{n+1}$. By definition of orthogonal matrices, we have $\|Ax\|^2 = (Ax|Ax) = \|x\|^2$. Thus, A defines a bijection of the sphere $\mathbb{S}^n$ to itself. Furthermore, for any $x, y \in \mathbb{S}^{n+1}$, again by the definition of orthogonal matrices,

$$\cosh d_{\mathbb{S}^n}(Ax, Ay) = (Ax|Ay) = (x|y) = \cosh d_{\mathbb{S}^n}(x, y)\,,$$

which implies that the above mapping is an isometry.

Transitivity follows from the fact that any element of $\mathbb{S}^n$ can be taken as the first element of an orthogonal basis of $\mathbb{E}^n$ or, equivalently, as the first column of an orthogonal matrix. □

We will prove the analog of Theorem 1.8 for the sphere.

Theorem 2.10. $\mathrm{Isom}(\mathbb{S}^n) = \mathrm{O}(n+1)$ □

The proof works as in the Euclidean case once we have defined hyperplanes and bisectors in the appropriate, natural manner.

Let H_0 be a linear hyperplane in $\mathbb{E}^n$. The intersection $H = H_0 \cap \mathbb{S}^n$ is a *hyperplane* of $\mathbb{S}^n$. Note that each hyperplane of $\mathbb{S}^n$ is isometric with $\mathbb{S}^{n-1}$. The *reflection* r_H in H is the restriction of the reflection in H_0 to the sphere: $r_H = r_{H_0}|_{\mathbb{S}^n}$. Note that by Proposition 6.10(2) and Proposition 2.9, the image of $r_{H_0}|_{\mathbb{S}^n}$ is contained in $\mathbb{S}^n$.

Proposition 2.11. *Let H be an hyperplane in $\mathbb{S}^n$. Then*

(1) $r_H \circ r_H$ is the identity.

(2) $r_H \in \mathrm{O}(n)$. In particular, r_H is an isometry.

(3) $d(r_H(x), y) = d(x, y)$ for all $x \in \mathbb{S}^n$ and all $y \in H$.

(4) The fixed point set of r_H is H.

Proof. (1), (2) and (4) are direct consequences of Proposition 6.10. We leave (3) as an exercise. □

The bisector of two distinct points $p, q \in \mathbb{S}^n$ is

$$\mathrm{bis}(p, q) = \{x \in \mathbb{S}^n : d_{\mathbb{S}^n}(x, p) = d_{\mathbb{S}^n}(x, q)\}\,.$$

Lemma 2.12. *Let $p, q \in \mathbb{S}^n$, $p \neq q$. Then* $\mathrm{bis}(p, q) = (p - q)^\perp \cap \mathbb{S}^n$*. In particular, the bisector is a hyperplane, it is the intersection of the Euclidean bisector of p and p with the $\mathbb{S}^n$.*

Proof. The points $p, q, x \in \mathbb{S}^n$ satisfy $d_{\mathbb{S}^n}(x, p) = d_{\mathbb{S}^n}(x, q)$ if and only if $(p|x) = (q|x)$, which is equivalent with $(p - q|x) = 0$. □

Proposition 2.13. *Let $x, y \in \mathbb{S}^n$ and let H be a hyperplane of $\mathbb{S}^n$.*

(1) If $r_H(x) = y$ and $x \notin H$, then $H = \mathrm{bis}(x, y)$*.*

(2) If $p, q \in \mathbb{S}^n$, $p \neq q$, then $r_{\mathrm{bis}(p,q)}(p) = q$.

(3) Let $\phi \in \mathrm{Isom}(\mathbb{S}^n)$, $\phi \neq \mathrm{id}$. If $a \in \mathbb{S}^n$, $\phi(a) \neq a$, then the fixed points of ϕ are contained in $\mathrm{bis}(a, \phi(a))$*.*

(4) Let $\phi \in \mathrm{Isom}(\mathbb{S}^n)$, $\phi \neq \mathrm{id}$. If H is a hyperplane such that $\phi|_H$ is the identity, then $\phi = r_H$.

Proof. (1) follows from Proposition 2.11(3).
(2) Using the definitions and the fact that $\frac{p+q}{2}$ is in the Euclidean bisector of p and q, we get

$$r_{\mathrm{bis}(p,q)}(p) = p - 2(p - \frac{p+q}{2}|p-q)\frac{p-q}{\|p-q\|^2} = q\,.$$

The proofs of (3) and (4) are formally the same as in the Euclidean case. □

We leave it as an exercise to check that the following result is proved in the same way as their Euclidean counterparts.

Proposition 2.14. *Let $p_1, p_2, \dots, p_k, q_1, q_2, \dots, q_k \in \mathbb{S}^n$ be points that satisfy*

$$d(p_i, p_j) = d(q_i, q_j)$$

for all $i, j \in \{1, 2, \dots, k\}$. Then, there is an isometry $\phi \in \mathrm{Isom}(\mathbb{S}^n)$ such that $\phi(p_i) = q_i$ for all $i \in \{1, 2, \dots, k\}$.

Corollary 2.15. *Any isometry of $\mathbb{S}^n$ can be represented as the composition of at most $n+1$ reflections.* □

If a group G acts on a space X A is a nonempty subset of X, the *stabiliser* of A in G is

$$\mathrm{Stab}_G\, A = \{g \in G : gA = A\}\,.$$

Clearly, stabilisers are subgroups of G.

Proposition 2.16. *The stabiliser in $\mathrm{Isom}(\mathbb{S}^n)$ of any point $x \in \mathbb{S}^n$ is isomorphic to $\mathrm{O}(n)$.*

Proof. The north pole e_{n+1} is stabilized by the subgroup of $\mathrm{O}(n)$ that consists of block diagonal matrices $\mathrm{diag}(1, A)$, where $A \in \mathrm{O}(n)$. Proposition 2.9 implies the claim as in the Euclidean case, see Proposition 1.10. □

The proof of the following result is similar to that of its Euclidean analog, Proposition 1.11.

Proposition 2.17. *Each k-plane of $\mathbb{S}^n$ is isometric with $\mathbb{S}^k$. For each k-plane P, there is an isometry $\phi \in \mathrm{Isom}(\mathbb{S}^n)$ such that*

$$\phi(P) = \{x \in \mathbb{S}^n : x^{k+2} = x^{k+3} = \dots = x^{n+1} = 0\}\,. \qquad \square$$

2.3. **Classification of isometries.** The *special orthogonal group* of dimension n is

$$\mathrm{SO}(n) = \{A \in \mathrm{O}(n) : \det A = 1\}\,.$$

Let $A \in \mathrm{O}(2)$. The columns of A form an orthonormal basis of $\mathbb{E}^2$. If we write the first column as $\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}$, then orthogonality implies that the second column is either $\begin{pmatrix}-\sin\theta\\ \cos\theta\end{pmatrix}$ or $\begin{pmatrix}\sin\theta\\ -\cos\theta\end{pmatrix}$. Therefore, there are exactly two kinds of orthogonal maps of the plane: the rotation by θ,

$$R_\theta = \begin{pmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{pmatrix} \in \mathrm{SO}(2)$$

and the reflection in the line $L = (-\cos\frac{\theta}{2}, \sin\frac{\theta}{2})^\perp$,

$$S_\theta = \begin{pmatrix}\cos\theta & \sin\theta\\ \sin\theta & -\cos\theta\end{pmatrix} \in \mathrm{O}(2) - \mathrm{SO}(2)\,.$$

The rotations form the normal subgroup $\mathrm{SO}(2)$ of $\mathrm{O}(2)$.

Remark 2.18. (1) It is easy to check that $S_\theta = R_\theta S_0$.
(2) Using complex numbers, R_θ is multiplication by $e^{i\theta}$ and the reflection S_θ is the mapping $S_\theta(z) = e^{i\theta}\bar{z}$.

As in the 2-dimensional case, $\mathrm{SO}(n)$ is a normal subgroup of $\mathrm{O}(n)$ of index 2. The following property simplifies the treatment of $\mathrm{SO}(3)$:

Proposition 2.19. *The nonidentity elements of* $\mathrm{SO}(3)$ *are rotations* $R_{v,\theta}$ *by* $\theta \in]0, 2\pi[$ *about an axis given by a unit vector* $v \in \mathbb{S}^2$.

Proof. Let $A \in \mathrm{SO}(3)$. Let us show that 1 is an eigenvalue of A by considering the characteristic polynomial χ_A. Now, using the facts that A is orthogonal, the determinant is multiplicative and $\det A = 1$, the determinant of a matrix and its transpose agree, and that we work with 3×3-matrices, we have

$$\begin{aligned}\chi_A(1) = \det(A - I_3) &= -\det(A(I_3 - A^T)) = \det(I_3 - A^T) \\ &= \det(I_3 - A) = -\det(A - I_3) = -\chi_A(1)\,.\end{aligned}$$

Thus, $\chi_A(1) = 0$. This implies that A is conjugate in $\mathrm{SO}(3)$ with a block diagonal matrix $\mathrm{diag}(A_0, 1)$ with $A \in \mathrm{SO}(2)$, and the claim follows from the above classification of the elements of $\mathrm{SO}(2)$ $\square$

3.1. **The latitude-longitude map.** Let $x \in \mathbb{S}^2$. The *latitude* of x is

$$\theta(x) = \frac{\pi}{2} - d_{\mathbb{S}^2}(x, e_3) = \frac{\pi}{2} - \arccos(x|e_3) = \frac{\pi}{2} - \arccos(x_3) \in [-\frac{\pi}{2}, \frac{\pi}{2}],$$

which is the oriented angle of x from the *equator* $\{x \in \mathbb{S}^2 : x_3 = 0\}$. The *longitude* of $x \in \mathbb{S}^2 - \{\pm e_3\}$ is

$$\phi(x) = \operatorname{sign}(x_2) \arccos(\frac{(x_1, x_2, 0)|e_1)}{\|(x_1, x_2, 0)\|}) = \operatorname{sign}(x_2) \arccos(\frac{x_1}{\sqrt{x_1^2 + x_2^2}}) \in]-\pi, \pi]$$

where $\operatorname{sign}(t) = \frac{t}{|t|}$ for nonzero t and we set $\operatorname{sign}(0) = 1$. The longitude is the oriented angle between x and the geodesic segment from the *north pole* e_3 to the *south pole* $-e_3$, called the 0-*meridian* (or the *Greenwich meridian* if we consider the Earth). Here we have chosen the value π for the longitude on the *international date line* which is the geodesic segment between the poles that passes through $-e_1$. More generally, the geodesic line between the poles determined by an equation $\phi = c$ is a *meridian* and the circle determined by an equation $\theta = c$ is a *parallel.*

The longitude and latitude of a point determine a bijection $L \colon \mathbb{S}^2 - \{\pm e_3\} \to]-\pi, \pi] \times]\frac{\pi}{2}, \frac{\pi}{2}[$,

$$L(x) = (\phi(x), \theta(x)) .$$

The inverse of this map is given by

$$L^{-1}(\phi, \theta) = (\cos\phi\cos\theta, \sin\phi\cos\theta, \sin\theta) .$$

This map is good close to the equator but distances, areas and angles are badly distorted close to the poles.

3.2. **Stereographic projection.** Let $a \in \mathbb{R} - \{0\}$ and consider the *projection plane* $P_a = \{x \in \mathbb{E}^3 : x_3 = a\}$. For any $x \in \mathbb{S}^2$, let $\mathbb{S}_0^a \colon \mathbb{S}^2 \to P_a$ be the map

$$S_0^a(x) = (1-a)\frac{x - e_3}{1 - x_3} + e_3$$

that associates to x the unique point on P_a that lies on the affine line through e_3 and x. The *stereographic projection* $S^a \colon \mathbb{S}^2 - \{e_3\} \to \mathbb{E}^2$ is $\mathrm{pr}_3 \circ S_0^a$, where $\mathrm{pr}_3(y) = (y_1, y_2)$ is the orthogonal projection of $\mathbb{E}^3$ to $\mathbb{E}^2$ identified with the hyperplane $\mathbb{E}^2 \times \{0\}$. More explicitly,

$$S^a(x) = (1-a)(\frac{x_1}{1-x_3}, \frac{x_2}{1-x_3}) .$$

Most often, one uses $a = 0$, which is the case where the projection plane passes through the origin, or $a = -1$, which is the case where the projection plane is tangent to the sphere at the south pole.

3.3. **Inversion.** Let $c \in \mathbb{E}^n$ and let $\alpha \in \mathbb{R} - \{0\}$. The mapping $\iota_{c,\alpha} \colon \mathbb{E}^n - \{c\} \to \mathbb{E}^n - \{c\}$ defined by setting

$$\iota_{c,\alpha}(x) = \alpha\frac{x - c}{\|x - c\|^2} + c$$

is the *(α-)inversion with pole c.* The number α is the *power* of the inversion $\iota_{c,\alpha}$.

Lemma 3.1. $S_0^a = \iota_{e_3, 2(1-a)}|_{\mathbb{S}^2 - \{e_3\}}$.

Proof. It suffices to note that for $x \in \mathbb{S}^2$, we have

$$\|x - e_3\|^2 = \|x\|^2 - 2(x|e_3) + \|e_3\|^2 = 2(1 - x_3) . \quad \square$$

The above Lemma means that we can deduce many basic properties of the stereographic projection from those of inversions. Accordingly, we will now look at inversions more closely. The first technical properties are easy to check using directly the definitions.

Lemma 3.2. *(1)* $(x - c|\iota_{c,\alpha}(x) - c) = \alpha$ *for all* $x \in \mathbb{E}^n - \{c\}$.

(2) $\iota_{c,\alpha} \circ \iota_{c,\beta} = \frac{\alpha}{\beta}(x - c) + c$.

(3) $\iota_{c,\alpha} \circ \iota_{c,\alpha} = \mathrm{id}$*, in particular,* $\iota_{c,\alpha}$ *is a diffeomorphism.*

(4) If $\alpha > 0$*, then* $\operatorname{fix} \iota_{c,\alpha} = \mathbb{S}(c, \sqrt{\alpha})$.

Proof. Exercise. □

If $r > 0$, then ι_{c,r^2} is also called *inversion in the sphere* $\mathbb{S}(c, r)$, which is reasonable considering Lemma 3.2(4). This is the case we will mainly be interested in.

The following proposition collects a number of important mapping properties of inversions. A *generalized hyperplane* or a *generalized sphere* in $\mathbb{E}^n$ is either a hyperplane or a sphere. If U and V are open subsets of $\mathbb{E}^n$ or $\mathbb{S}^n$, a mapping $F : U \to V$ is *(locally) conformal* if it preserves angles. This happens exactly when at every point in U, the differential $DF(x)$ is a multiple of an orthogonal matrix by a nonzero real number.

Proposition 3.3. *Let* $c \in \mathbb{E}^n$ *and let* $\alpha \in \mathbb{R} - \{0\}$*. The inversion* $\iota_{c,\alpha}$

(1) stabilizes the hyperplanes containing c,

(2) maps the spheres containing c to hyperplanes not containing c and conversely.

(3) maps spheres not containing c to spheres not containing c.

(4) is conformal.

Proof. (1) is clear from the expression of the inversion.

(2) It suffices to consider the case $c = 0$. The sphere consists of the points that satisfy the equation $\|x\|^2 = 2(x|a)$ for some $a \in \mathbb{E}^n - \{0\}$. Thus, $\iota_{0,\alpha}(x) = \alpha \frac{x}{2(x|a)}$, and we have $(\iota_{0,\alpha}|a) = \frac{\alpha}{2}$, which is the equation of a hyperplane.

(3) Again assume $c = 0$. Consider $x_1, x_2 \in \mathbb{S}(a, r)$, points that are not necessarily distinct, that lie on the same line L . The orthogonal projection of a to L is $\frac{x_1+x_2}{2}$. The Pythagorean theorem gives the two equations

$$\|x_1 + x_2\|^2 + \|x_1 + x_2 - 2a\|^2 = 4\|a\|^2$$

and

$$\|x_1 - x_2\|^2 + \|x_1 + x_2 - 2a\|^2 = 4r^2 ,$$

that imply $(x_1|x_2) = \|a\|^2 - r^2$. Thus, $x_2 = \iota_{0,\|a\|^2-r^2}(x_1)$. As this holds for all such pairs of points, we have $\iota_{0,\|a\|^2-r^2}(\mathbb{S}(a, r)) = \mathbb{S}(a, r)$. Using Lemma 3.2, we have

$$\iota_{0,\alpha} = \frac{\alpha}{\|a\|^2 - r^2} \iota_{0,\|a\|^2-r^2} ,$$

which implies

$$\iota_{0,\alpha}(\mathbb{S}(a, r)) = \frac{\alpha}{\|a\|^2 - r^2} \mathbb{S}(a, r) .$$

(4) Observe that $\iota_{c,\alpha} = T_c \circ \iota_{0,\alpha} \circ T_{-c}$. Translations and dilation by α are clearly conformal mappings so it suffices to prove the claim for the standard inversion $\iota_{0,1}$.

Note that

$$
\begin{aligned}
D\iota_{0,1}(x) &= \frac{1}{\|x\|^4}\begin{pmatrix} -x_1^2+x_2^2+x_3^2 & -2x_1x_2 & -2x_1x_3 \\ -2x_1x_2 & x_1^2-x_2^2+x_3^2 & -2x_2x_3 \\ -2x_1x_3 & -2x_2x_3 & x_1^2+x_2^2-x_3^2 \end{pmatrix} \\
&= \frac{1}{\|x\|^2}I_3 - \frac{2}{\|x\|^4}\begin{pmatrix} x_1^2 & x_1x_2 & x_1x_3 \\ x_1x_2 & x_2^2 & x_2x_3 \\ x_1x_3 & x_2x_3 & x_3^2 \end{pmatrix} \\
&= \frac{1}{\|x\|^2}I_3 - \frac{2}{\|x\|^4}xx^T ,
\end{aligned}
$$

where x^T is the transpose of x as a matrix. Observe that $D\iota_{0,1}(x)^T = D\iota_{0,1}(x)$ and that

$$
D\iota_{0,1}(x)^2 = \frac{1}{\|x\|^2}I_3 - \frac{4}{\|x\|^6}xx^T + \frac{4}{\|x\|^8}xx^Txx^T = \frac{1}{\|x\|^2}I_3 .
$$

Thus, $D\iota_{0,1}(x)$ is a multiple of an orthogonal matrix. □

Corollary 3.4. *The stereographic projection is a conformal map. It maps circles on the sphere not passing through the north pole to circles. It maps the (complements of the north pole in) circles passing through the north pole to lines.* □

3.4. **Mercator's projection.** The latitude-longitude map of the sphere to a square in the plane is a standard method to determine the location of a point on the surface of the Earth. However, it is not well suited for navigation because it is not conformal. This can be seen as follows: Let $P = (0, \sin\theta, \cos\theta)$ be a point in the Eastern hemisphere. The tangent directions at P, that is, the unit sphere of $P^\perp$ consists of the vectors

$$
v(t) = (\sin t, -\cos t\sin\theta, \cos t\cos\theta) ,
$$

where t is the angle with the direction from the North, measured in the clockwise direction. Let L be the longitude-latitude map. The derivative vector of the path $L \circ j_{P,v(t)}$ at P is

$$
DL \circ j_{P,v(t)}(0) = (-\sec\theta\sin t, \cos t)
$$

Thus, the angle that $DL \circ j_{P,v(t)}(0)$ makes with the direction $(0,1)$ that points to the North on the map can be computed to be

$$
\arccos\left(\frac{\cos t}{\sqrt{\sec^2\theta\sin^2 t + \sin^2\theta}}\right) .
$$

The distortion of the angle increases rapidly as we approach the north pole.

One way to solve this problem is to stretch the north-south direction by a factor that increases close to the poles in order to define a conformal map. We will follow a different route to achieve the same goal. Let us identify the plane $\mathbb{E}^2$ with the complex plane in the usual way, identifying the point (x, y) with the complex number $z = x + iy$. The *complex logarithm function* is defined as the local inverse of the complex exponential map. More precisely, let $\log\colon \mathbb{C} - \{0\} \to \mathbb{C}$,

$$
\log(z) = \log|z| + i\arg z ,
$$

where

$$
|z| = \|(x, y)\|
$$

is the *module* of z and

$$
\arg z = \arccos\left(\frac{x}{|z|}\right) \in]-\pi, \pi]
$$

is the *argument* of z. It is a standard fact of complex analysis, that we take as given, that log is a conformal map, see [**?**].

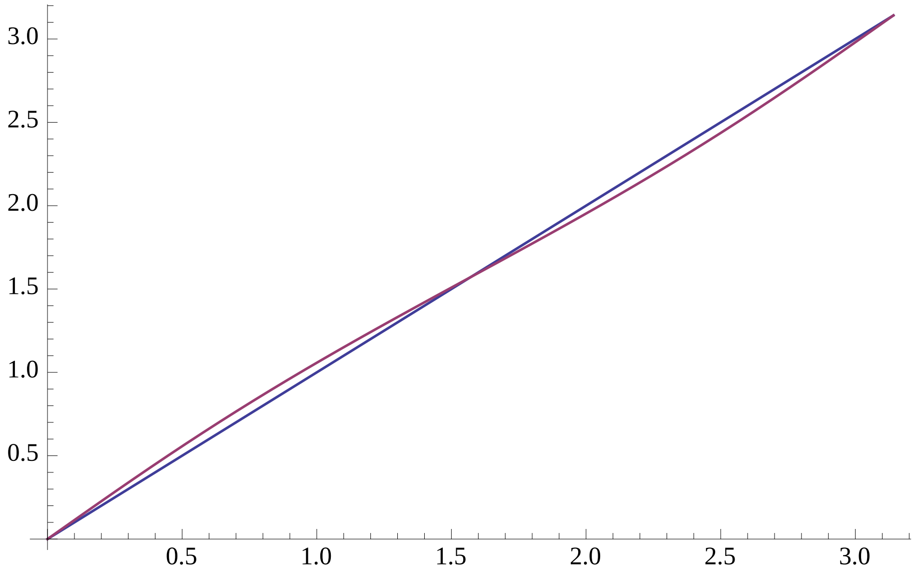

FIGURE 4. The blue curve shows the sine of the angle on the sphere and red curve shows the sine of the corresponding angle in the longitude-latitude map at longitude 0.5.

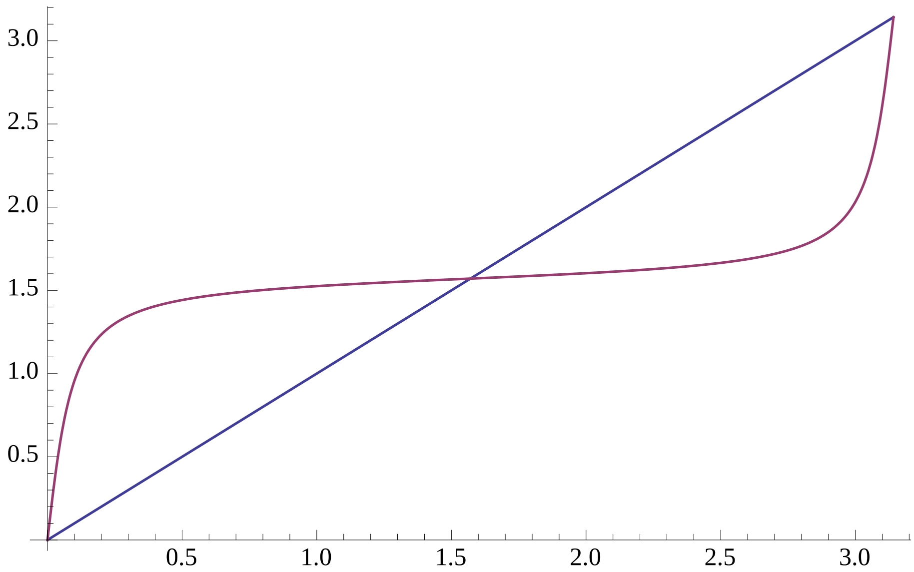

FIGURE 5. The blue curve shows the sine of the angle on the sphere and red curve shows the sine of the corresponding angle in the longitude-latitude map at longitude 1.5.

Lemma 3.5. *The complex logarithm is a bijection that is continuous outside the negative ray* $]-\infty, 0[$*. It maps circles centered at the origin to vertical segments and rays of constant argument to horizontal affine lines.* □

The *Mercator projection* is the map $M = \overline{-i \log \circ S^0} \colon \mathbb{S}^2 - \{\pm e_3\} \to \mathbb{E}^2$,

$$\begin{aligned} M(x) &= (\phi(x), \frac{1}{2}\log(x_1^2 + x_2^2) - \log(1 - x_3)) \\ &= (\phi(x), \log\cot\theta(x))\,. \end{aligned}$$

Proposition 3.6. *The Mercator projection is a bijection that is continuous away from the international date line. It maps parallels to horizontal segments and meridians to vertical affine lines.* □

We saw above that the Mercator projection is obtained from the longitude-latitude map by stretching in the North-South direction by a factor that grows towards the poles. The great advantage of using the Mercator projection is that a straight line on the map intersects the meridians at a constant angle. This is convenient for navigation purposes because it corresponds to keeping a constant compass direction. The straight lines in a Mercator map are called *loxodromes*. A loxodrome segment is almost never the shortest path between its endpoints, it coincides with a segment of a great circle only if it is a segment of the Equator.

3.5. **Some Riemannian geometry.** The *(differential geometric) length* of a piecewise continuously differentiable path $\tau\colon I \to \mathbb{S}^2$ is

$$\ell(\tau) = \int_I \|\dot{\tau}\|\,,$$

where $\dot{\tau}(t)$ is the tangent (derivative) vector of the path for each $t \in I$.

Proposition 3.7. *Let $A, B \in \mathbb{S}^2$, $A \neq B$. Let j be a spherical segment that connects A and B. Then $\ell(j) \leq \ell(\tau)$ for all piecewise continuously differentiable paths τ.*

Proof. Using an isometry of $\mathbb{S}^2$, we can assume that A and B are contained in the 0-meridian. Using longitude-latitude coordinates, consider the continuous map proj defined by $\operatorname{proj}(\phi, \theta) = (0, \theta)$ whose image is contained in the 0-meridian. Clearly, $\ell(j) \leq \ell(\operatorname{proj} \circ \tau) \leq \ell(\tau)$. □

In the computation of the length of a path τ, the norm of the tangent vector $\dot{\tau}(t)$ is computed in the tangent plane $\tau(t)^\perp$ at $\tau(t)$. Using the coordinate maps, we get

The inner product of the tangent spaces can be used to define the area of a subset of the sphere. This gives the expressions

$$\text{Area}\, A = \int_{L(A)} \cos\theta d\theta d\phi$$

in the longitude-latitude coordinates and

$$\text{Area}\, A = \int_{S^0(A)} \frac{4\, dx_1 dx_2}{(1 + \|x\|^2)^2}$$

in the coordinates given by the stereographic projection.

Proposition 3.8. *The area of $\mathbb{S}^2$ is 4π.*

3.6. **Cylindrical projection.** The mapping $C\colon \mathbb{S}^2 - \{\pm e_3\} \to \mathbb{S}^1 \times]-1, 1[\subset \mathbb{E}^3$,

$$C(x) = \left(\frac{x_1}{\sqrt{x_1^2 + x_2^2}}, \frac{x_2}{\sqrt{x_1^2 + x_2^2}}, x_3 \right)$$

maps the complement of the poles to a cylinder bijectively. As $x_3 = \cos\theta$, this projection preserves area. On the other hand, the distortion of angles is even worse than in the longitude-latitude map.

Let $0 < \alpha < \pi$. The area of the *(spherical) sector* $S_\alpha = \{x \in \mathbb{S}^2 : 0 \leq \phi(x) \leq \alpha\}$ and any of its isometric images is easily seen to be $\frac{\alpha}{2\pi} 4\pi = 2\alpha$.

Proposition 4.1 (Girard)**.** *The area of a triangle with angles α, β and γ is $\alpha + \beta + \gamma - \pi$.*

Proof. Let A, B and C be the vertices of the triangle. The antipodal points $-A$, $-B$ and $-C$ determine a triangle $(-A)(-B)(-C)$ that is isomorphic with ABC. The three great circles $\langle A, B\rangle \cap \mathbb{S}^2$, $\langle B, C\rangle \cap \mathbb{S}^2$ and $\langle C, A\rangle \cap \mathbb{S}^2$ determine six sectors with angles $\alpha, \alpha, \beta, \beta, \gamma, \gamma$ that cover the sphere. In the complement of the great circles, the triangles ABC and $(-A)(-B)(-C)$ are both covered by three sectors, other points are contained in one sector. Thus,

$$4\pi = \operatorname{Area} \mathbb{S}^2 = 2(\operatorname{Area} S_\alpha + \operatorname{Area} S_\beta + \operatorname{Area} S_\gamma) - 4 \operatorname{Area} ABC = 4\alpha - 4 \operatorname{Area} ABC \,,$$

which gives the claim. □

We know that the sum of the angles of a triangle in $\mathbb{E}^2$ is π. (Prove this!) On the sphere, the situation is different:

Proposition 4.2. *The sum of the angles of a triangle in $\mathbb{S}^2$ is greater than π.*

Girard's result implies the claim but we give a proof that avoids the use of arguments involving area or integration. In order to do this, we introduce a useful construction.

Let ABC be a triangle in $\mathbb{S}^2$ such that the vertices do not all lie on the same great circle. Let $A^*, B^*, C* \in \mathbb{S}^2$ be the unique points that satisfy the conditions

$$\begin{aligned} (A^*|B) &= 0 = (A^*|C), \quad (A^*|A) > 0 \\ (B^*|C) &= 0 = (B^*|A), \quad (B^*|B) > 0 \\ (C^*|A) &= 0 = (C^*|B), \quad (C^*|C) > 0 \,, \end{aligned} \tag{7}$$

that is, for each vertex of the triangle, we pick the intersection point of the line orhogonal to the plane that contains the other two vertices, on the same side of the plane as the original vertex. We call A^*, B^* and C^* the *polar vertices* and $(ABC)^* = A^*B^*C^*$ the *polar triangle* of ABC. Let a^*, b^* and c^* be the side lengths and α^*, β^* and γ^* be the angles of $(ABC)^*$.

Lemma 4.3. *The polar vertices are linearly independent and $((ABC)^*)^* = ABC$.*

Proposition 4.4. *Let ABC be a triangle in $\mathbb{S}^2$ such that the vertices do not all lie on the same great circle. Then*

$$a + \alpha^* = b + \beta^* = c + \gamma^* = a^* + \alpha = b^* + \beta = c^* + \gamma = \pi \,.$$

Proof. The situation is completely symmetric so it suffices to prove $a + \alpha^* = \pi$. Let $u, v \in A^\perp = \langle B^*, C^*\rangle$ be the directions of the edges AB and AC, respectively. Recall that $(u|v) = \cos \alpha$ and $(B^*|C^*) = \cos a^*$.

Now, $u \in \langle A, B\rangle$ implies that $(u|C^*) = 0$ and similarly we have $(v|B^*) = 0$. Furthermore,

$$(u|B^*) = \Big(\frac{B - (B|A)A}{\|B - (B|A)A\|}\Big|B^*\Big) = \frac{(B|B^*)}{\|B - (B|A)A\|} > 0$$

and similarly $(v|C^*) > 0$. Thus, we have either the points u, B^*, C^* and v on the circle $\langle B^*, C^*\rangle$ in this order or in the order B^*, u, v and C^* with the right angles between u and B^* and v and C^* overlapping in both cases. The claim follows easily. □

Lemma 4.5. *The perimeter of a spherical triangle is at most 2π. If the perimeter is 2π, then the vertices are all contained in the same great circle.*

Proof. This follows from the inequality (5) and the fact that this inequality is an equality if and only if $\gamma = \pi$. $\square$

We can now give a second proof of Proposition 4.2: Proposition 4.4 implies that $\alpha + \beta + \gamma + a^* + b^* + c^* = 3\pi$. As $a^* + b^* + c^* < 2\pi$ by Lemma 4.5, we get the claim of Proposition 4.2. $\square$

The following converse of Lemma 4.5 holds

Proposition 4.6. *Let $0 < a, b, c < \pi$. If $a + b > c$, $b + c > a$, $c + a > b$ and $a + b + c < 2\pi$, then there is a triangle in $\mathbb{S}^2$ with side lengths a, b and c. All such triangles are isometric.*

Proof. We use the law of cosines in the construction: Note that if such a triangle exists, then the angle at C satisfies the cosine law. Therefore, we can compute it if we know that

$$|\frac{\cos c - \cos a \cos b}{\sin a \sin b}| < 1\,, \tag{8}$$

because then $\frac{\cos c - \cos a \cos b}{\sin a \sin b}$ is in the range of cos, and we can proceed with the construction. The pair of inequalities $c < a + b < 2\pi - c$ implies

$$\cos c > \cos(a + b) = \cos a \cos b - \sin a \sin b\,.$$

The inequalities $b + c > a$ and $c + a > b$ give $|a - b| < c$, which implies

$$\cos c < \cos(a - b) = \cos a \cos b + \sin a \sin b\,.$$

These two inequalities give

$$-\sin a \sin b < \cos c - \cos a \cos b < \sin a \sin b\,,$$

which implies the inequality (8). Now we can place the sides of length a and b starting at C in the correct angle γ. The cosine law implies that the lengths of the side opposite to C is indeed c.

The triangles are isometric by Proposition 2.14 $\square$

5.1. **Bilinear forms and Minkowski space.** Let V and W be real vector spaces. A map $\Phi\colon V \times W \to \mathbb{R}$ is a *bilinear form*, if the maps $v \mapsto \Phi(v, w_0)$ and $v \mapsto \Phi(v_0, w)$ are linear for all $w_0 \in W$ and all $v_0 \in V$. A bilinear form Φ is *nondegenerate* if

- $\Phi(x, y) = 0$ for all $y \in W$ only if $x = 0$, and
- $\Phi(x, y) = 0$ for all $x \in V$ only if $y = 0$.

If $W = V$, then Φ is *symmetric* if $\Phi(x, y) = \Phi(y, x)$ for all $x, y \in V$. It is

- *positive semidefinite* if $\Phi(x, x) \geq 0$ for all $x \in V$,
- *positive definite* if $\Phi(x, x) > 0$ for all $x \in V - \{0\}$,
- *negative (semi)definite* if $-\Pi$ is positive (semi)definite, and
- *indefinite* otherwise.

The quadratic form corresponding to a bilinear form $\Phi\colon V \times V \to \mathbb{R}$ is the function $q\colon V \to \mathbb{R}$, $q(x) = \Phi(x, x)$. A positive definite symmetric bilinear form is often called an inner product or a scalar product.

If V is a vector space with a symmetric bilinear form Φ, we say that two vectors $u, v \in V$ are *orthogonal* if $\Phi(u, v) = 0$, and this is denoted as usual by $u \perp v$. The *orthogonal complement* of $u \in V$ is

$$u^{\perp} = \{v \in V : u \perp v\}.$$

Let us consider the indefinite nondegenerate symmetric bilinear form $\langle\cdot|\cdot\rangle$ on $\mathbb{R}^{n+1}$ given by

$$\langle x|y\rangle = -x_0 y_0 + \sum_{i=1}^{n} x_i y_i = -x_0 y_0 + (\bar{x}|\bar{y}) = x^T J y,$$

where $J_{1,n} = \operatorname{diag}(-1, 1, \ldots, 1)$ and $x = (x_0, x_1, \ldots, x_n) = (x_0, \bar{x})$. We call $\langle\cdot|\cdot\rangle$ the *Minkowski bilinear form*, and the pair

$$\mathbb{M}^{1,n} = \left(\mathbb{R}^{n+1}, \langle\cdot|\cdot\rangle\right)$$

is the $n+1$-dimensional *Minkowski space*.

We say that a vector is

- *lightlike* if $\langle x|x\rangle = 0$,
- *timelike* if $\langle x|x\rangle < 0$, and
- *spacelike* if $\langle x|x\rangle > 0$.

The names come from Einstein's special theory of relativity, which lives in $\mathbb{M}^{1,3}$. Minkowski space has a number of geometrically significant subsets: The subset of *null-vectors* is the *light cone*

$$\mathscr{L}^n = \left\{x \in \mathbb{M}^{1,n} : \langle x|x\rangle = 0\right\}.$$

The variety

$$\mathscr{H}_-^n = \{x \in \mathbb{M}^{1,n} : \langle x|x\rangle = -1\}$$

is a two-sheeted hyperboloid, and its upper sheet is

$$\mathbb{H}^n = \{x \in \mathbb{M}^{1,:} \langle x|x\rangle = -1,\ v_0 > 0\}.$$

The variety

$$\mathscr{H}_+^n = \{x \in \mathbb{M}^{1,n} : \langle x|x\rangle = 1\}$$

is a one-sheeted hyperboloid.

The following is an important observation on time-like vectors.

Lemma 5.1. *If $u, v \in \mathbb{H}^n$, then $\langle u|v\rangle \leq -1$ with equality only if $u = v$.*

Proof. Using the Cauchy inequality for the Euclidean inner product in $\mathbb{R}^n$ for the first inequality and a simple calculation for the second, we have

$$\begin{aligned}\langle u|v\rangle &= -u_0v_0 + \sum_{i=1}^n u_iv_i \le -u_0v_0 + \sqrt{\sum_{i=1}^n u_i^2}\sqrt{\sum_{i=1}^n v_i^2} \\ &= -u_0v_0 + \sqrt{u_0^2-1}\sqrt{v_0^2-1} \le -1\,.\end{aligned}$$

Cauchy's inequality is an equality if and only if u and v are parallel, and the final inequality is an equality if and only if $u_0 = v_0$. This implies the claim on equality. □

Proposition 5.2. *Let $v_1, v_2 \in \mathbb{M}^{1,n}$.*
(1) If v_1 and v_2 are timelike, then $\langle v_1, v_2\rangle \neq 0$.
(2) if v_1 lightlike and v_2 is timelike, then $\langle v_1|v_2\rangle \neq 0$.

Proof. (1) is a direct consequence of Lemma 5.1. We leave (2) as an exercise. □

Corollary 5.3. *The restriction of the Minkowski bilinear form to the orthogonal complement of a timelike vector is positive definite.* □

5.2. **The orthogonal group.** The *orthogonal group* of the Minkowski bilinear form is

$$\begin{aligned}\mathrm{O}(1,n) &= \{A \in \mathrm{GL}_n(\mathbb{R}) : \langle Ax|Ay\rangle = \langle x|y\rangle \text{ for all } x,y \in \mathbb{M}^{1,n}\} \\ &= \{A \in \mathrm{GL}_n(\mathbb{R}) : {}^T\!AJ_{1,n}A = J_{1,n}\}\,.\end{aligned}$$

Clearly, the linear action of $\mathrm{O}(1,n)$ on $\mathbb{M}^{1,n}$ preserves the light cone and the two-sheeted hyperboloid $\mathscr{H}^n$.

Let us write an $(n+1)\times(n+1)$-matrix A in terms of its column vectors $A = (a_0, a_1, \dots, a_n)$. If $A \in \mathrm{O}(1,n)$, then $a_0 = A(x_0)$ for $(x_0 = 1, 0, \dots, 0) \in \mathbb{H}^n$. Thus $A(x_0) \in \mathbb{H}^n$ if and only if $A_{00} > 0$, and therefore the stabiliser in $\mathrm{O}(1,n)$ of the upper sheet $\mathbb{H}^n$ is

$$\begin{aligned}\mathrm{O}^+(1,n) &= \{A \in O(1,n) : A\mathbb{H}^n = \mathbb{H}^n\} \\ &= \{A \in \mathrm{GL}_n(\mathbb{R}) : A_{00} > 0,\ \langle Ax|Ay\rangle = \langle x|y\rangle \text{ for all } x,y \in \mathbb{M}^{1,n}\} \\ &= \{A \in \mathrm{GL}_n(\mathbb{R}) : A_{00} > 0,\ {}^T\!AJ_{1,n}A = J_{1,n}\}\,,\end{aligned}$$

which is the identity component of $\mathrm{O}(1,n)$.

A basis $\{v_0, v_1, \dots, v_n\}$ of $\mathbb{M}^{1,n}$ is *orthonormal* if the basis elements are pairwise orthogonal and if $\langle v_0|v_0\rangle = -1$ and $\langle v_i|v_i\rangle = 1$ for all $i \in \{1, 2, \dots, n\}$. The following observation is proved in the same way as its Euclidean analog:

Lemma 5.4. *An $(n+1)\times(n+1)$-matrix $A = (a_0, a_1, \dots, a_n)$ is in $\mathrm{O}(1,n)$ if and only if the vectors $a_0, a_1, \dots, a_n$ form an orthonormal basis of $\mathbb{M}^{1,n}$. Furthermore, $A \in \mathrm{O}^+(1,n)$ if and only if $A \in \mathrm{O}(1,n)$ and $a_0 \in \mathbb{H}^n$.* □

Example 5.5. (1) Let $t \in \mathbb{R}$. The matrix

$$L_t = \begin{pmatrix}\cosh t & \sinh t & 0\\ \sinh t & \cosh t & 0\\ 0 & 0 & 1\end{pmatrix} \in \mathrm{SO}(1,2)$$

stabilises any affine hyperplane

$$H_c = \{x \in \mathbb{M}^{1,2} : x_2 = c\}\,.$$

(2) For any $\theta \in \mathbb{R}$, let $\widehat{R}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in \mathrm{O}(2)$, and let

$$R_\theta = \mathrm{diag}(1, \widehat{R}_\theta) = \begin{pmatrix} 1 & 0 \\ 0 & \widehat{R}(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \in \mathrm{O}^+(1,2)\,.$$

This mapping is a Euclidean rotation around the vertical axis by the angle θ. The rotation R_θ stabilizes each affine hyperplane

$$E_r = \{x \in \mathbb{M}^{1,2} : x_0 = r\}\,.$$

Another important mapping that comes by extension from $\mathrm{O}(2)$ is given by the matrix $\mathrm{diag}(1,1,-1)$, which is a reflection in the hyperplane H_0 defined above.

(3) For each $v \in \mathscr{L}^2$ and $c < 0$, the affine hyperplane

$$P_{v,t} = \{x \in \mathbb{H}^2 : \langle v|x\rangle = t\}$$

The mapping given by the matrix

$$N_s = \begin{pmatrix} 1+\frac{s^2}{2} & -\frac{s^2}{2} & s \\ \frac{s^2}{2} & 1-\frac{s^2}{2} & s \\ s & -s & 1 \end{pmatrix} \in \mathrm{O}(1,2)$$

maps each horosphere based at $(1,1,0) \in \mathscr{L}^2$ to itself.

(4) All of the above examples can be generalised to higher dimensions:

- L_t is extended as the identity on the last coordinates to $\mathrm{diag}(L_t, I_{n-2}) \in \mathrm{O}(1,n)$.
- Any Euclidean orthogonal matrix $A \in \mathrm{O}(n)$ gives an isometry $\mathrm{diag}(1,A) \in \mathrm{O}(1,n)$.
- N_s is extended as the identity on the last coordinates to $\mathrm{diag}(N_s, I_{n-2}) \in \mathrm{O}(1,n)$.

A modification of the proof of Proposition 1.5 gives the following result.

Proposition 5.6. *The identity component of the orthogonal group of Minkowski space acts transitively on $\mathbb{H}^n$.*

Proof. We use the notation of Example 5.5. If $x \in \mathbb{H}^n$, then $x = (\sqrt{\|\bar{x}\|^2 - 1}, \bar{x})$. There is some $\widehat{R}_\theta \in \mathrm{O}(n)$ such that $R_\theta\bar{x} = \|\bar{x}\|e_1$, and thus, $R_\theta(x) = (\sqrt{\|\bar{x}\|^2 - 1}, \|\bar{x}\|e_1)$. Furthermore,

$$L_{\mathrm{arsinh}\,\|\bar{x}\|}e_0 = (\sqrt{\|\bar{x}\|^2 - 1}, \|x\|e_1)\,.$$

This implies that $\mathbb{H}^n$ is contained in the $\mathrm{O}^+(1,n)$-orbit of e_0. $\square$

The metric space $(\mathbb{H}^n, d)$, where

$$d(x, y) = \operatorname{arcosh}(-\langle x|y\rangle) \in [0, \infty[\,,$$

is the *hyperboloid model* of n-dimensional *(real) hyperbolic space.* The metric d is the *hyperbolic metric.*

We still need to show that the hyperbolic metric is a metric. The proof follows the same idea that was used to treat the angle metric for the sphere $\mathbb{S}^n$.

Let $a \in \mathbb{H}^n$, and let $u \in a^\perp$ such that $\langle u|u\rangle = 1$. Recall that the restriction of the Minkowski bilinear form to $a^\perp$ is positive definite by Corollary 5.3. The mapping $j_{a,u} \colon \mathbb{R} \to \mathbb{H}^n$,

$$j_{a,u}(t) = a\cosh(t) + u\sinh(t)\,,$$

is the *hyperbolic line* through a in *direction* u. It is easy to check that, indeed, the image of $j_{a,u}$ is contained in $\mathbb{H}^n$ and that for all $s, t \in \mathbb{R}$, we have

$$d(j_{a,u}(t), j_{a,u}(s)) = |s - t|\,. \tag{9}$$

As in section 2 for the sphere, if we show that d is a metric, then $j_{a,u}$ is a geodesic line. We define hyperbolic segments and rays as the appropriate restrictions of the geodesic line.

Lemma 6.1. *For any $a \in \mathbb{H}^n$ and any $u \in a^\perp$, $j_{a,u}(\mathbb{R}) = \mathbb{H}^n \cap \langle a, u\rangle$. If a 2-plane T intersects $\mathbb{H}^n$, then $T \cap \mathbb{H}^n$ is the image of a hyperbolic line.*

Proof. Clearly, the image of $j_{a,u}$ is contained in the 2-plane $\langle a, u\rangle$.

On the other hand, if a plane $T = \langle u, v\rangle$ intersects $\mathbb{H}^n$ at two distinct points p and q, the geodesic line $j_{p,u}$ with

$$u = \frac{q + \langle p|q\rangle p}{|q + \langle p|q\rangle p|}$$

passes through p and q. If we fix $p \in \mathbb{H}^n$, there are exactly two unit tangent vectors v and $-v$ in $T_p\mathbb{H}^n \cap T$, and the hyperbolic lines $j_{p,v}$ and $j_{p,-v}$ defined by these vectors have the same image. Therefore, all points in $\mathbb{H}^n \cap T$ are contained in $j_{p,v}(\mathbb{R})$ for any $p \in \mathbb{H}^n$. $\square$

Lemma 6.2. *For any $a \in \mathbb{H}^n$, the tangent space $T_a\mathbb{H}^n$ of $\mathbb{H}^n$ at a coincides with $a^\perp$.*

Proof. The orthogonal complement $a^\perp$ has dimension n because the Minkowski bilinear form is nondegenerate. Each vector in $a^\perp$ is the tangent vector at a of a smooth curve contained in $\mathbb{H}^n$. $\square$

We define the *angle* $\measuredangle(u, v)$ of any two vectors $u, v \in T_a\mathbb{H}^n = a^\perp - \{0\}$, using the inner product induced from the Minkowski bilinear form:

$$\measuredangle(u, v) = \arccos(\langle u|v\rangle)$$

The inner product induces a norm

$$|u| = \sqrt{\langle u|u\rangle}$$

on $a^\perp$ for all $a \in \mathbb{H}^n$.

Proposition 6.3 (The first hyperbolic law of cosines)**.**

$$\cosh c = \cosh a \cosh b - \sinh a \sinh b \cos \gamma\,.$$

Proof. Let u and v be the initial tangent vectors of the hyperbolic segments from C to A and from C to B. As u and v are orthogonal to C, we have as in the spherical case,

$$\begin{aligned}\cosh c &= -\langle A|B\rangle = -\langle \cosh(b)C + \sinh(b)u|\cosh(a)C + \sinh(a)v\rangle \\ &= \cosh(a)\cosh(b) - \sinh(b)\sinh(a)\langle u|v\rangle\,. \end{aligned}$$ □

Theorem 6.4. *Hyperbolic space is a uniquely geodesic metric space. Hyperbolic lines, rays and segments are geodesic lines, rays and segments.*

Proof. The fact that the hyperbolic metric is indeed a metric is proved in the same way as Proposition 2.5 in the spherical case. Now we consider the increasing function

$$\gamma \mapsto \cosh a \cosh b - \sinh a \sinh b \cos \gamma,$$

which attains its maximum value $\cosh(a + b)$ when $\gamma = \pi$. The claim on hyperbolic lines, rays and segments follows from equation (9).

If p and q are distinct points in $\mathbb{H}^n$, there is a unique 2-plane through them. Thus, there is exactly one hyperbolic line through these points. As in the spherical case, we see that the triangle inequality in hyperbolic geometry is an equality if and only if the third point z lies in the hyperbolic segment between x and y. □

The law of cosines implies that a triangle in $\mathbb{E}^n$, $\mathbb{S}^n$ or $\mathbb{H}^n$ is uniquely determined up to an isometry of the space, if the lengths of the three sides are known. In Euclidean space, the three angles of a triangle do not determine the triangle uniquely. In $\mathbb{S}^n$ and $\mathbb{H}^n$ the angles determine a triangle uniquely.

For $\mathbb{H}^n$, this is the content of

Proposition 6.5 (The second hyperbolic law of cosines)**.**

$$\cosh c = \frac{\cos\alpha\cos\beta + \cos\gamma}{\sin\alpha\sin\beta}\,.$$

Proof. This formula follows from the first law of cosines by a lengthy manipulation analogous to the proof of Proposition 2.7. □

Proposition 6.6 (The hyperbolic law of sines)**.**

$$\frac{\sinh a}{\sin\alpha} = \frac{\sinh b}{\sin\beta} = \frac{\sinh c}{\sin\gamma}\,.$$

Proof. The first law of cosines implies that

$$\left(\frac{\sinh c}{\sin\gamma}\right)^2 = \frac{\sinh^2 a \sinh^2 b \sinh^2 c}{2\cosh a\cosh b\cosh c - \cosh^2 a - \cosh^2 b - \cosh^2 c + 1}\,.$$

The claim follows because this expression is symmetric in a, b and c. □

6.1. **Isometries.**

Proposition 6.7. $\mathrm{O}^+(1, n)$ *acts transitively by isometries on* $\mathbb{H}^n$*. In particular,* $\mathrm{Isom}(\mathbb{H}^n)$ *acts transitively on* $\mathbb{H}^n$*.*

Proof. Let $g \in \mathrm{O}^+(1, n)$, and let $x, y \in \mathbb{H}^n$. By the definition of the hyperbolic metric and of $\mathrm{O}^+(1, n)$, we have

$$d(g(x), g(y)) = \operatorname{arcosh}(-\langle g(x)|g(y)\rangle = \operatorname{arcosh}(-\langle x|y\rangle = d(x, y)\,.$$

Transitivity follows from the fact that any orthonormal basis of $\mathbb{M}^{1,n}$ whose first vector is in $\mathbb{H}^n$ can be mapped to any other similar one by a transformation in

$O^+(1,n)$: If $p \in \mathbb{H}^n$, and the vectors $v_1, v_2, \dots, v_n$ form an orthogonal basis of $p^\perp = T_p\mathbb{H}^n$, then the matrix $A = (p, v_1, \dots, v_n) \in O(1,n)$ gives an isometry which maps $(1,0,\dots,0)$ to p. □

Example 6.8. (1) Let $t \in \mathbb{R}$. The matrix

$$L_t = \begin{pmatrix} \cosh t & \sinh t & 0 \\ \sinh t & \cosh t & 0 \\ 0 & 0 & 1 \end{pmatrix} \in O^+(1,2)$$

acts on $\mathbb{H}^2$ as an isometry that preserves the intersection of $\mathbb{H}^2$ with any 2-plane $\{x \in \mathbb{M}^{1,2} : x_2 = c\}$, in particular, it stabilises the geodesic line

$$\ell = \{x \in \mathbb{H}^3 : x_2 = 0\}.$$

For any point $p = (a,b,0) \in \ell$, we have

$$d(L_t(p), p) = \operatorname{arcosh}(-\langle L_t p | p \rangle) = \operatorname{arcosh}((-a^2+b^2)\cosh(t)) = |t|.$$

(2) For any $\theta \in \mathbb{R}$, let $\widehat{R}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in SO(2)$, and let

$$R_\theta = \operatorname{diag}(1, \widehat{R}_\theta) = \begin{pmatrix} 1 & 0 \\ 0 & \widehat{R}(\theta) \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix} \in O^+(1,2).$$

This mapping rotates the hyperboloid around the vertical axis by the angle θ.
(3) For each $v \in \mathscr{L}^2$ and $c < 0$, the set

$$\{x \in \mathbb{H}^2 : \langle v | x \rangle = c\}$$

is called a *horosphere* based at v. The mapping given by the matrix

$$N_s = \begin{pmatrix} 1+\frac{s^2}{2} & -\frac{s^2}{2} & s \\ \frac{s^2}{2} & 1-\frac{s^2}{2} & s \\ s & -s & 1 \end{pmatrix} \in O(1,2)$$

maps each horosphere based at $(1,1,0) \in \mathscr{L}^2$ to itself.
(4) Composing some number of the above mappings we obtain further examples of isometries of the hyperbolic plane. For example, if $p \in \mathbb{H}^2$, then there is some $\theta \in \mathbb{R}$ such that $R_\theta(p) \in \ell$. Now, $L^{-1}_{d(o,p)}(R_\theta(p)) = L_{-d(o,p)}(R_\theta(p)) = (1,0,0)$, and for any $\phi \in \mathbb{R}$, the mapping $S = R_{-\theta} \circ L_{d(o,p)} \circ R_\phi \circ L^{-1}_{d(o,p)} \circ R_\theta$ is an isometry that fixes p and maps each sphere centered at p to itself. The mapping S is *conjugate* to R_ϕ in $\operatorname{Isom}(\mathbb{H}^n)$.

The isometries introduced above are classified according to the conic sections they correspond to. The mapping L_t and any of its conjugates in $\operatorname{Isom}(\mathbb{H}^n)$ is called *hyperbolic* because L_t maps each affine plane parallel to the (x_0, x_1)-plane in $\mathbb{M}^{1,2}$ to itself, and these planes intersect the lightcone in hyperbola, which is degenerate for the (x_0, x_1)-plane itself.

The mapping $R(\theta)$ and any of its conjugates is called *elliptic* because it preserves all horizontal hyperplanes in $\mathbb{M}^{1,2}$ and their intersections with $\mathscr{L}^2$, which are circles centered at $(1,0,0)$.

The mapping N_s and any of its conjugates is called *parabolic* because it preserves all affine hyperplanes $\{x \in \mathbb{M}^{1,2} : \langle v | x \rangle = c\}$, which intersect $\mathscr{L}^2$ in a parabola when $c < 0$.

As in the Euclidean and spherical geometries, we wil now study a fundamental class of isometries, reflections in a hyperplane. If T is an $(m+1)$-dimensional linear subspace of $\mathbb{R}^{n+1}$ that intersects $\mathbb{H}^n$, then $T \cap \mathbb{H}^n$ is an m-dimensional *hyperbolic subspace* of $\mathbb{H}^n$. If $m = n-1$, then T is a *hyperplane.* A modification of the proof of Proposition 1.5 gives

Proposition 6.9. *Any two hyperbolic subspaces of $\mathbb{H}^n$ can be mapped to each other by isometries of $\mathbb{H}^n$. In particular, a k-dimensional hyperbolic subspace of $\mathbb{H}^n$ is isometric to $\mathbb{H}^k$.* $\square$

Any hyperplane T in $\mathbb{M}^{1,n}$ is of the form $T = u^\perp$ for some $u \in \mathbb{M}^{1,n} - \{0\}$ because the Minkowski bilinear form is nondegenerate. Let $H = u^\perp \cap \mathbb{H}^n$ be a hyperbolic hyperplane. Since H intersects $\mathbb{H}^n$, it contains a vector v for which $\langle v|v\rangle = -1$. Proposition 5.2 implies that $\langle u|u\rangle > 0$, and after normalising, we may assume that u is a unit vector. The *reflection* in H is the map

$$r_H(x) = x - 2\langle x|u\rangle u\,. \tag{10}$$

The proofs of the basic properties of reflections are natural modifications of those in the spherical case. Note that the expression (10) defines a mapping in Minkowski space, fixing the hyperplane T. The reflection in hyperbolic space is, in fact, the restriction of a reflection of Minkowski space.

Proposition 6.10. *Let H be a hyperbolic hyperplane. Then*

(0) r_H imaps $\mathbb{H}^n$ into itself.

(1) $r_H \circ r_H$ is the identity.

(2) $r_H \in O^+(1,n)$.

(3) $d(r_H(x), y) = d(x, y)$ for all $x \in \mathbb{H}^n$ and all $y \in H$.

(4) The fixed point set of r_H is H

Proof. (0) Let $x \in \mathbb{H}^n$. Using bilinearity and symmetry of the Minkowski form and the fact that u is a unit vector, we get

$$\begin{aligned}\langle r_H(x)|r_H(x)\rangle &= \langle x - 2\langle x|u\rangle u | x - 2\langle x|u\rangle u\rangle \\ &= \langle x|x\rangle - 2\langle x|u\rangle\langle x|u\rangle - 2\langle x|u\rangle\langle u|x\rangle + 4\langle x|u\rangle\langle x|u\rangle\langle u|u\rangle \\ &= \langle x|x\rangle = -1\,.\end{aligned}$$

Thus, $r_H(x) \in \mathscr{H}^n_-$. Furthermore, for any $v \in H$,

$$r_H(v) = v - 2\langle v|u\rangle u = v,$$

so there are points in $\mathbb{H}^n$ which are mapped to $\mathbb{H}^n$. Since r_H is continuous and preserves the Minkowski form, $r_H(\mathbb{H}^n) \subset \mathbb{H}^n$.

(1) This easy computation is left as an exercise.

(2) Clearly, r_H is a linear mapping, and it is a bijection by (1). Again, using bilinearity and symmetry of the Minkowski form and the fact that u is a unit vector, we get

$$\begin{aligned}\langle r_H(x)|r_H(y)\rangle &= \langle x - 2\langle x|u\rangle u | y - 2\langle y|u\rangle u\rangle \\ &= \langle x|y\rangle - 2\langle y|u\rangle\langle x|u\rangle - 2\langle x|u\rangle\langle u|y\rangle + 4\langle x|u\rangle\langle y|u\rangle\langle u|u\rangle \\ &= \langle x|y\rangle\,.\end{aligned}$$

Thus, $r_H \in \mathrm{O}(1,n)$. Claim (0) gives $r_H \in \mathrm{O}^+(1,n)$.

(3) For any $x \in \mathbb{H}^n$ and all $y \in H$, we have

$$\langle r_H(x)|y\rangle = \langle x - 2\langle x|u\rangle u|y\rangle = \langle x|y\rangle - 2\langle x|u\rangle\langle u|y\rangle = \langle x|y\rangle,$$

where the final equality follows from the assumption $u \in H^\perp$.
(4) This follows immediately from (3) by taking $x = y \in H$. $\square$

The *bisector* of two distinct points p and q in $\mathbb{H}^n$ is the hyperplane

$$\operatorname{bis}(p,q) = \{x \in \mathbb{H}^n : d(x,p) = d(x,q)\} = (p-q)^\perp \cap \mathbb{H}^n.$$

Proposition 6.11. *(1) For any $p, q \in \mathbb{H}^n$, the bisector $\operatorname{bis}(p,q)$ is a hyperbolic hyperplane.*
(2) If $r_H(x) = y$ and $x \notin H$, then $H = \operatorname{bis}(x,y)$.
(3) If $p, q \in \mathbb{H}^n$, $p \neq q$, then $r_{\operatorname{bis}(p,q)}(p) = q$.
(4) Let $\phi \in \operatorname{Isom}(\mathbb{H}^n)$, $\phi \neq \operatorname{id}$. If $a \in \mathbb{H}^n$, $\phi(a) \neq a$, then the fixed points of ϕ are contained in $\operatorname{bis}(a, \phi(a))$.
(5) Let $\phi \in \operatorname{Isom}(\mathbb{H}^n)$, $\phi \neq \operatorname{id}$. If H is a hyperplane such that $\phi|_H$ is the identity, then $\phi = r_H$.

Proof. (1) Lemma 5.1 implies that

$$\langle p - q | p - q \rangle = -2 - 2\langle p | q \rangle > 0 \,.$$

Thus, $(p-q)^\perp$ contains a vector v for which $\langle v|v\rangle < 0$, and therefore the hyperplane $(p-q)^\perp$ of $\mathbb{M}^{1,n}$ intersects $\mathbb{H}^n$.
(2) follows from Proposition 6.10(3).
(3) Now, $2\langle p | p-q\rangle = 2(\langle p|p\rangle - \langle p|q\rangle) = -2 - 2\langle p|q\rangle = |p-q|^2$. Thus,

$$r_{\operatorname{bis}(p,q)}(p) = p - 2\langle p|p-q\rangle \frac{p-q}{|p-q|^2} = q \,. \quad \square$$

(4) If $\phi(b) = b$, then $d(a,b) = d(\phi(a), \phi(b)) = d(\phi(a), b)$, so that $b \in \operatorname{bis}(a, \phi(a))$.
(5) is an instructive exercise.

Next, we prove that all isometries of hyperbolic space are restrictions to $\mathbb{H}^n$ of linear automorphisms of $\mathbb{M}^{1,n}$:

Theorem 6.12. $\operatorname{Isom}(\mathbb{H}^n) = \mathrm{O}^+(1,n)$.

The idea of the proof is to show that each isometry of $\mathbb{H}^n$ is the composition of reflections in hyperbolic hyperplanes. Again, the proof follows the same ideas as in the Euclidean and spherical cases.

Proposition 6.13. *Let $p_1, p_2, \dots, p_k, q_1, q_2, \dots, q_k \in \mathbb{H}^n$ be points that satisfy*

$$d(p_i, p_j) = d(q_i, q_j)$$

for all $i, j \in \{1, 2, \dots, k\}$. Then, there is an isometry $\phi \in \operatorname{Isom}(\mathbb{H}^n)$ such that $\phi(p_i) = q_i$ for all $i \in \{1, 2, \dots, k\}$. Furthermore, the isometry ϕ is the composition of at most k reflections in hyperplanes.

Proof. The proof is formally exactly the same as that of Proposition 1.9. $\square$

Note that Proposition 1.9 implies that if T and T' are two triangles in $\mathbb{H}^n$ with equal angles or equal sides, then there is an isometry ϕ of $\mathbb{H}^n$ such that $\phi(T) = T'$.

Proof of Theorem 6.12. Let $\{a_0, a_1, \dots, a_n\}$ be a set of points in $\mathbb{H}^n$ which is not contained in any proper hyperbolic subspace. This is achieved by choosing them so that they generate $\mathbb{M}^{1,n}$ as a vector space. Proposition 6.13 implies that there is an isometry $\phi_0 \in \mathrm{O}^+(1,n)$ such that $\phi_0(\phi(a_1)) = a_i$ for all $1 \le i \le m+1$. Since the set of fixed points of $\phi_0 \circ \phi$ contains the points $a_0, a_1, \dots, a_{n+1}$, the fixed point set of ϕ_0 is not contained in a proper hyperbolic subspace. Proposition 6.11 implies that

$\phi_0 \circ \phi$ is the identity map. Thus, $\phi = \phi_0^{-1}$. In particular, $\phi \in \mathrm{O}^+(1, n)$, which is all we needed to show. $\square$

Corollary 6.14. *Any isometry of $\mathbb{H}^n$ can be represented as the composition of at most $n + 1$ reflections.*

Proposition 6.15. *The stabiliser of any point $x \in \mathbb{H}^n$ is isomorphic to $\mathrm{O}(n)$.*

Proof. Again, we follow the proof of the spherical case. The details are left as an exercise. $\square$

In this section we consider a number of other models for hyperbolic space, that is, metric spaces (X, d_X) for which there is an isometry $h\colon (X, d_X) \to (\mathbb{H}^n, d)$. Hyperbolic space is the class of all metric spaces isometric with the hyperboloid model $(\mathbb{H}^n, d)$, and we can use any model that is best suited for the geometric problem at hand. After this section we will often talk about the "upper half plane model of $\mathbb{H}^n$" etc.

7.1. **Klein's model.** Each line in $\mathbb{M}^{1,n}$ through the origin which intersects $\mathbb{H}^n$, intersects it in exactly one point, and it also intersects the embedded copy $\{1\} \times \mathbb{B}^n$ in $\mathbb{M}^{1,n}$ of the Euclidean n-dimensional unit ball $\mathbb{B}^n(0,1)$ in exactly one point. This correspondence determines a bijection $K\colon \mathbb{B}(0,1) \to \mathbb{H}^n$, which has the explicit expression

$$K(x) = \frac{(1,x)}{\sqrt{1-\|x\|^2}} \,.$$

The map K becomes an isometry when we define a metric on $\mathbb{B}(0,1)$ by setting

$$d_K(x,y) = d(K(x),K(y)) = \operatorname{arcosh} \frac{1-(x|y)}{\sqrt{1-\|x\|^2}\sqrt{1-\|y\|^2}} \,.$$

The metric space $(\mathbb{B}(0,1), d_K)$ is the *Klein model* of n-dimensional hyperbolic space.

Proposition 7.1. *As sets, the geodesic lines of the Klein model are Euclidean segments connecting two points in the Euclidean unit sphere.*

Proof. A geodesic line in $\mathbb{H}^n$ is the intersection of $\mathbb{H}^n$ with a 2-plane in $\mathbb{M}^{1,n}$. The intersection of this plane with $\mathbb{B}^n(0,1) \times \{1\}$ is the preimage under K of the geodesic line. $\square$

The above observation makes it easy to show that the parallel axiom does not hold in $\mathbb{H}^n$.

7.2. **Poincaré's ball model.** Each affine line that passes through the point $(-1,0) \in \mathbb{R} \times \mathbb{R}^n = \mathbb{M}^{1,n}$ which intersects $\mathbb{H}^n$, intersects it in exactly one point, and it also intersects the n-dimensional ball $\{0\} \times B^n(0,1)$ embedded in $\mathbb{M}^{1,n}$ in exactly one point. This correspondence determines a bijection $P\colon \mathbb{B}(0,1) \to \mathbb{H}^n$,

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

This expression is found by computing for any $x \in \mathbb{B}(0,1)$ that the point $y_t = (0,x) + t(1,x)$ on the line through the points $(0,x)$ and $(-1,0)$ of $\mathbb{R} \times \mathbb{R}^n = \mathbb{M}^{1,n}$ is in $\mathbb{H}^n$ if and only if $t = \frac{1+\|x\|^2}{1-\|x\|^2}$.

The map P becomes an isometry when we define a metric on $\mathbb{B}(0,1)$ by setting

$$d_P(x,y) = d(P(x),P(y)) = \operatorname{arcosh}\Big(1 + 2\frac{\|x-y\|^2}{(1-\|x\|^2)(1-\|y\|^2)}\Big) \,.$$

The metric space $(\mathbb{B}(0,1), d_P)$ is the *Poincaré model* of n-dimensional hyperbolic space.

Proposition 7.2. *(1) The mapping P is conformal.*

(2) As sets, the geodesic lines of the Poincaré model are the intersections of the Euclidean unit ball with Euclidean circles which are orthogonal to the unit sphere.

Proof. (1) A computation shows that for all tangent vectors u, v in $T_xB(0,1)$, we have

$$\langle DP(x)u|DP(x)v\rangle = \frac{4(u|v)}{(1-\|x\|^2)^2}\,.$$

(2) The map $h = K^{-1}\circ P$ is an isometry between the Poincaré and Klein models. A computation shows that

$$h(x) = \frac{2x}{1+\|x\|^2}.$$

(This can be done by observing that h is a radial map and then solving the equation

$$\frac{(1,y)}{\sqrt{1-y^2}} = \big(\frac{1+x^2}{1-x^2}, \frac{2x}{1-x^2}\big)$$

with $0\le x, y<1$.) On the other hand, the inversion $\iota_{(-1,0),2}$ in the sphere centered at $(-1,0)$ of radius $\sqrt{2}$ has the expression

$$\iota_{(-1,0),2}(x) = \big(\frac{1-\|x\|^2}{1+\|x\|^2}, \frac{2x}{1+\|x\|^2}\big)\,,$$

so that if $\mathrm{pr}\colon \mathbb{E}^{n+1}\to\mathbb{E}^n$ is the Euclidean orthogonal projection, we have

$$h = \mathrm{pr}\circ\iota_{(-1,0),2}\,.$$

Note that $\{0\}\times\mathbb{S}(0,1)$ is contained in the fixed sphere of $\iota_{(-1,0),2}$. The inversion $\iota_{(-1,0),2}$ maps any circle orthogonal to $\{0\}\times\mathbb{S}(0,1)$ to a circle on the unit sphere in $\mathbb{E}^{n+1}$ orthogonal to $\{0\}\times\mathbb{S}(0,1)$. These circles are the intersections of the unit sphere with 2-planes parallel to the x_0-axis, and thus, pr maps them to the geodesic lines of the Klein model. As h is an isometry, the result follows. □

Recall that the restriction to $\{0\}\times B(0,1)$ of the mapping $\iota_{(-1,0),2}$ in the proof of the above result is (the inverse) of the stereographic projection.

7.3. **The upper halfspace model.** Let

$$\mathbb{U}^n = \{x\in\mathbb{R}^n : x_n>0\}$$

be the n-dimensional *upper halfspace*. Let $\iota_{(0,-1),2}$ be the inversion in the sphere of center $(0,-1)\in\mathbb{R}^{n-1}\times\mathbb{R}=\mathbb{E}^n$ of radius $\sqrt{2}$. Now, the map

$$F = \iota_{(-1,0),2}|_{\mathbb{B}(0,1)}\colon \mathbb{B}(0,1)\to\mathbb{U}^n$$

is a bijection, which becomes an isometry if we use the metric

$$d_{\mathbb{U}}(x,y) = d_P(F^{-1}(x),F^{-1}(y)) = \operatorname{arcosh}\big(1+\frac{\|x-y\|^2}{2\,x_ny_n}\big) \tag{11}$$

in $\mathbb{U}^n$. The metric space $(\mathbb{U}^n, d_{\mathbb{U}})$ is the *upper halfspace model* of n-dimensional hyperbolic space.

Proposition 7.3. *As sets, the geodesic lines of the upper halfspace model are the intersections of $\mathbb{U}^n$ with Euclidean circles and lines which are orthogonal to $\mathbb{E}^{n-1}\times\{0\}$.*

Proof. The inversion used in the definition of the upper halfspace model maps lines and circles to lines or circles and preserves angles. □

In practical applications, it is good to remember that a circle is perpendicular to $\mathbb{E}\times\{0\}\subset\mathbb{E}^2$ if and only if its center is in $\mathbb{E}\times\{0\}$.

8. Some geometry and techniques

In this section, we investigate a number of geometric properties of hyperbolic space using the various models according to the needs of the situation.

8.1. **Triangles.** The mappings used to define the Poincaré model and the upper haplspace model are conformal is very useful. In particular, the angle between two tangent vectors is in these models is the same as the Euclidean angle. This allows us to prove the following facts on triangles in hyperbolic space. We say that a triangle is degenerate if it is contained in a geodesic segment.

Proposition 8.1. *(1) The sum of the angles of a nondegenerate triangle in hyperbolic space is strictly less than π.*

(2) For any $0 < a, b, ci$ for which $a + b > c$, $b + c > a$ and $c + a > b$, there is a triangle with side lengths a, b and c. Any two such triangles are isometric.

(3) For any $0 < \alpha, \beta, \gamma < \pi$ for which $\alpha + \beta + \gamma < \pi$, there is a triangle with angles α, β and γ. Any two such triangles are isometric.

Proof. Any three points in the hyperboloid model $\mathbb{H}^n$ are contained in the intersection of $\mathbb{H}^n$ with a 3-dimensional linear subspace of $\mathbb{M}^{1,n}$, which is an isometrically embedded copy of the hyperbolic plane. Furthermore, the geodesic arc between any two of these points in is contained in the same 2-plane. Thus, any triangle is always contained in an isometrically embedded copy of $\mathbb{H}^2$ in $\mathbb{H}^n$, so in the proof below, it suffices to consider the hyperbolic plane. We may assume that one of the vertices A is the origin in the Poincaré disk model. Thus, two sides of the triangle are contained in two radii of the ball and the third one is contained in a circle which is orthogonal to the boundary of $\mathbb{B}(0, 1)$.

(1) Consider the Euclidean triangle with the same vertices as T. The angles β and γ are strictly smaller than the corresponding angles in the Euclidean triangle. This implies the result as the angles of an Euclidean triangle sum to π.

(2) The proof is analogous with that of Proposition 4.6 without the upper bound on the lengths. We use the hyperbolic law of cosines in the construction. If a triangle with the asserted properties exists, then the angle at C satisfies the cosine law. Therefore, we can compute what this angle needs to be if we know that

$$|\frac{\cosh a \cosh b - \cosh c}{\sinh a \sinh b}| < 1\,. \tag{12}$$

The inequality $c < a + b$ implies

$$\cosh c < \cos(a + b) = \cosh a \cosh b + \sinh a \sinh b\,,$$

which gives

$$-1 < \frac{\cosh a \cosh b - \cosh c}{\sinh a \sinh b}\,.$$

The inequalities $b + c > a$ and $c + a > b$ give $|a - b| < c$, which implies

$$\cosh c > \cosh(a - b) = \cosh a \cosh b + \sinh a \sinh b\,,$$

and we get

$$\frac{\cosh a \cosh b - \cosh c}{\sinh a \sinh b} < 1\,.$$

Now we can place the sides of length a and b starting at C in the correct angle γ. The cosine law implies that the length of the side opposite to C is indeed c.

The triangles are isometric by Proposition 2.14

(3) Sketch: Fix the side containing A and B to be contained in the positive real line. Then consider the family of circles C_s, $s \in [0,1[$ that are orthogonal to the Euclidean unit circle and form an angle β with the segment $[0,1[$ at the point of intersection s. When s is small, the side from B to C is very close (in the underlying Euclidean space) to the euclidean segment connecting B and C. When s increases, there is a unique parameter $0 < t < 1$ for which the circle C_t is tangent to the ray that forms an angle α with the positive real line. Continuity implies that all angles $0 < \gamma < \pi - \alpha - \beta$ are realised for some parameter in $]0,t[$. □

In the proof of the above result we made the following observation which is important in itself:

> If the sides of a triangle in hyperbolic space are all short, then the angle sum is almost π.

A related observation that uses the hyperbolic law of cosines, equality of angles and the second order Taylor polynomials of the hyperbolic functions is

> If the sides of a triangle in hyperbolic space are all short, then the sides satisfy the Euclidean law of cosines up to a small error.

8.2. **Geodesic lines and isometries.** We already know that the geodesic lines of the upper halfspace model are, as sets, the intersections with $\mathbb{H}^n$ with Euclidean circles and lines that are orthogonal to $\mathbb{E}^{n-1} = \mathbb{E} \times \{0\}$. The following easy lemma records the expressions of the geodesics as mappings:

Lemma 8.2. *Let $x \in \mathbb{R}^{n-1}$ and $y > 0$. The mapping $\gamma\colon \mathbb{R} \to \mathbb{H}^n$,*

$$\gamma(t) = (x, ye^t)$$

is a geodesic line in the upper halfspace model. For any isometry $g \in \operatorname{Isom} \mathbb{H}^n$, the mapping $g \circ \gamma$ is a geodesic line. □

In the upper halfspace model, it is often convenient to move a geodesic line by an isometry such that the endpoints of the geodesic in the model are 0 and ∞. The following results on isometries allow to do that and a bit more.

Lemma 8.3. *Let $a \in \mathbb{E}^{n-1} \times \{0\}$, and let $r > 0$.*

(1) The inversion in the sphere $\mathbb{S}^{n-1}(a,r)$ preserves the upper halfspace and its restriction to the upper halfspace model is an isometry.
(2) The Euclidean reflection in a hyperplane orthogonal to $\mathbb{E}^{n-1} \times \{0\}$ preserves the upper halfplane and its restriction to the upper halfspace model is an isometry.

Proof. Let us prove (1): The first claim is clear. To prove the second, it is enough to show that the expression $\frac{\|x-y\|^2}{x_n y_n}$ is invariant under the inversion. Now if ι is the inversion, we have

$$\frac{\iota(x) - \iota(y)}{r^2} = \frac{x-a}{\|x-a\|^2} - \frac{y-a}{\|y-a\|^2} = \frac{(x-a)\|y-a\|^2 - (y-a)\|x-a\|^2}{\|x-a\|^2\|y-a\|^2},$$

which gives

$$\frac{\|f(x)-f(y)\|^2}{f(x)_n f(y)_n} = \frac{\frac{\|x-a\|^2\|y-a\|^4 - 2(x-a|y-a)\|x-a\|^2\|y-a\|^2 + \|x-a\|^4\|y-a\|^2}{\|x-a\|^4\|y-a\|^4}}{\frac{x_n y_n}{\|x-a\|^2\|y-a\|^2}} = \frac{\|x-y\|^2}{x_n y_n}\,.$$

The proof of (2) is an exercise. ☐

Proposition 8.4. *The maps*

- $T_b(x) = x + b$ *for any* $b \in \mathbb{R}^{n-1} = \mathbb{R}^n \times \{0\} \subset \mathbb{R}^{n+1}$, *(horizontal translations)*
- $\iota(x) = x/\|x\|^2$, *(inversion in the Euclidean unit sphere)*
- $L_\lambda(x) = \lambda x$ *for any* $\lambda > 0$ *(dilation fixing* 0*), and*
- $Q(\bar{x}, x_n) = (Q_0(\bar{x}), x_n)$ *for any* $Q_0 \in \mathrm{O}(n-1)$

are isometries of the upper halfspace model.

Proof. Exercise, compute directly or use Proposition 8.3. ☐

Corollary 8.5. *The subgroup of* $\mathrm{Isom}(\mathbb{H}^n)$ *generated by dilations fixing* 0 *and horizontal translations acts transitively on the upper half plane model of* $\mathbb{H}^n$.

Proof. If x is in the upper half plane,

$$T_{-(x_1,x_2,\dots x_{n-1},0)}(x) = (0,\dots,x_n) = L_{x_n}(0,\dots,0,1)\,.$$

Thus,

$$x = T_{(x_1,x_2,\dots x_{n-1},0)} \circ L_{x_n}(0,\dots,0,1)\,. \quad \square$$

Proposition 8.6. *Let* x_1, x_2, x_3 *and* y_1, y_2, y_3 *be two triples of distinct points in* $\mathbb{R}^{n-1} \cup \{\infty\}$. *There is an isometry of the upper halfspace model of* $\mathbb{H}^n$ *which is the restriction of a continuous map* $g\colon \mathbb{R}^n \cup \{\infty\} \to \mathbb{R}^n \cup \{\infty\}$ *such that* $g(x_i) = y_i$ *for all* $i \in \{1,2,3\}$.

Proof. The mappings given in Proposition 8.4 are clearly continuous mappings of the one point compactification of $\mathbb{R}^n$ to itself.

It suffices to show that we can use a combination of these isometries to map x_1, x_2, x_3 to $\infty, 0, (1, 0, \dots, 0)$. The claim then follows by composing such a map with the inverse of another one. If all points x_1, x_2, x_3 are finite, map x_1 by a translation T_{-x_1} to 0 and then by the inversion ι to ∞. Rename $\iota \circ T_{-x_1}(x_2)$ and $\iota \circ T_{-x_1}(x_3)$ to x_2 and x_3. Map x_2 to 0 by a translation. This map keeps ∞ fixed. Map x_3 (again renamed) to the unit sphere by a dilation and then to $(1, 0, \dots, 0)$ by the extension of an orthogonal map of $\mathbb{E}^{n-1}$. These two maps fix ∞ and 0. ☐

Corollary 8.7. *Let* x_1, x_2, x_3 *and* y_1, y_2, y_3 *be two triples of distinct points in* $\mathbb{S}^{n-1} \cup \{\infty\}$. *There is an isometry of the ball model of* $\mathbb{H}^n$ *which is the restriction of a continuous map of the closed unit ball of* $\mathbb{E}^{n+1}$ *to itself such that* $g(x_i) = y_i$ *for all* $i \in \{1,2,3\}$. ☐

8.3. **Balls.** In this section, we observe that hyperbolic balls in the upper halfspace model and in the Poincaré ball model are balls in the Euclidean metric.

Proposition 8.8. *Balls in the upper halfspace model and in the Poincaré ball model are Euclidean balls in the model space.*

Proof. By symmetry, balls centered at the origin of the ball model are Euclidean balls. The inversion that maps the ball model to the upper halfspace model is an isometry, and on the other hand it preserves generalized spheres. Thus, the images of the balls centered at the origin are hyperbolic and Euclidean balls. The hyperbolic center of these balls can be mapped to any other point in $\mathbb{H}^n$ by one of the isometries of Proposition 8.4. These mappings preserve generalized spheres, which implies that all balls in the upper halfspace model are Euclidean balls. The rest of the claim follows by one more application of the inversion that maps the ball model to the upper halfspace model. $\square$

Note that the Euclidean radii and centers of the balls hardly ever coincide with the hyperbolic ones.

The restriction of the Minkowski bilinear form $\langle \cdot | \cdot \rangle$ to the tangent space $T_p\mathbb{H}^n = p^\perp$ of any point in the hyperboloid model is positive definite, and it defines a Riemannian metric on the hyperboloid.

The *Riemannian length* of a piecewise smooth path $\gamma \colon [a,b] \to \mathbb{H}^n$ is

$$\ell(\gamma) = \int_a^b \sqrt{\langle \dot\gamma(t) | \dot\gamma(t) \rangle}\, dt\,.$$

The length metric of the Riemannian metric of $\mathbb{H}^n$ is

$$d_{\mathrm{Riem}}(x,y) = \inf \ell(\gamma),$$

where the infimum is taken over all piecewise smooth paths that connect x to y.

By definition of the Riemannian metric as the restriction of the Minkowski bilinear form to each tangent space, the group $\mathrm{O}(1,n)$ acts by Riemannian isometries on the hyperboloid. Thus, it is not surprising that the following result holds:

Proposition 9.1. *The length metric of the Riemannian metric of hyperbolic space is the hyperbolic metric.*

Proposition 9.1 allows us to use the Riemannian structure of hyperbolic space in any of the models introduced above. The expressions in the Poincaré and upper halfspace models are particularly useful.

The Riemannian structure defines a natural volume form and a volume measure on hyperbolic space: If V is for example an open subset of n-dimensional hyperbolic space, and λ_n is the n-dimensional Lebesgue measure, the volume of V is

$$\mathrm{Vol}(V) = \int_V \frac{2^n \, d\lambda_n(x)}{(1 - \|x\|^2)^n}$$

in the Poincaré ball model and

$$\mathrm{Vol}(V) = \int_V \frac{d\lambda_n(x)}{x_n^n}$$

in the upper halfspace model.

Proposition 9.2. *The volume of a ball in hyperbolic space is*

$$vol(\mathbb{B}(x,r)) = \mathrm{Vol}(\mathbb{S}^{n-1}) \int_0^r \sinh^{n-1} t\, dt.$$

In the hyperbolic plane, we have

$$\mathrm{Vol}(B^2(0,r)) = 4\pi \sinh^2 \frac{r}{2}\,.$$

Proof. As the isometry group acts transitively, the volume of each ball of a fixed radius is the same. Thus, it suffices to consider one ball that has a convenient center. The Euclidean radius of a ball of hyperbolic radius r centered at 0 in the Poincaré model is obtained by solving for R in the equation

$$r = d(0,(R,0)) = \int_0^R \frac{2s}{1-s^2} = \log \frac{1+R}{1-R}\,.$$

This shows that the Euclidean radius of a hyperbolic ball centered at the origin of the Poincaré model is $\tanh \frac{r}{2}$. In order to compute the volume of the ball of radius r, recall that the Lebesgue measure is given in the spherical coordinates ($x \leftrightarrow (r,u)$)

by $d\lambda_n(x) = r^{n-1} d\operatorname{Vol}_{\mathbb{S}^{n-1}}(u)$, and thus, using a change of variables $s \leftrightarrow \tanh \frac{t}{2}$, we get

$$\begin{aligned}
\operatorname{Vol}(\mathbb{B}(x,r)) = \operatorname{Vol}(\mathbb{B}(0,r)) &= \operatorname{Vol}(\mathbb{S}^{n-1}) \int_0^{\tanh \frac{r}{2}} \frac{2^n s^{n-1}}{(1+s^2)^n} ds \\
&= 2^{n-1} \operatorname{Vol}(\mathbb{S}^{n-1}) \int_0^r \sinh^{n-1} \frac{t}{2} \cosh^{n-1} \frac{t}{2}\, dt \\
&= \operatorname{Vol}(\mathbb{S}^{n-1}) \int_0^r \sinh^{n-1} t\, dt. \qquad \square
\end{aligned}$$

It is clear from the expression of the volume, that for all $x \in \mathbb{H}^n$, we have

$$\operatorname{Vol}(\mathbb{B}^n(x,r)) \sim \frac{\operatorname{Vol}(\mathbb{S}^n)}{2^{n-1}} e^{(n-1)r} ,$$

as $r \to \infty$. Thus, the volume of balls in hyperbolic space grows exponentially with the radius, much faster than in Euclidean space.

In the proof of the previous result, we obtained the following useful observation

Lemma 9.3. *The Euclidean radius of a hyperbolic ball centered at the origin of the Poincaré model is* $\tanh \frac{r}{2}$. $\square$

Proposition 9.4. *The area of a triangle is* $\pi - (\alpha + \beta + \gamma)$.

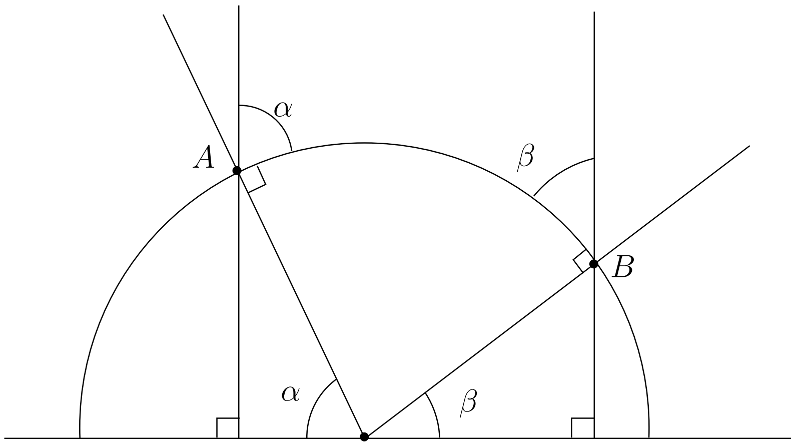

FIGURE 6.

Proof. Any triangle T can be described as the difference of two "triangles with one vertex at infinity" by which we mean subsets of the hyperbolic plane as in Figure 6 and their isometric images. By the additivity of area and angles in the hyperbolic plane, we may restrict to this special case. Using Proposition 8.6, we can assume that that A and B are on the Euclidean unit circle and that the vertex C has been moved to infinity. Now, the area of T is

$$\int_T \frac{d\lambda_2(x)}{x_2^2} = \int_{-\cos(\alpha)}^{\cos\beta} \int_{\sqrt{1-x_1^2}}^{\infty} \frac{dx_1 dx_2}{x_2^2} = \int_{\cos(\pi-\alpha)}^{\cos\beta} \frac{dx_1}{\sqrt{1-x_1^2}} = \pi - \alpha - \beta\,. \quad \square$$

Proposition 9.4 gives a new proof of Proposition 8.1(1).