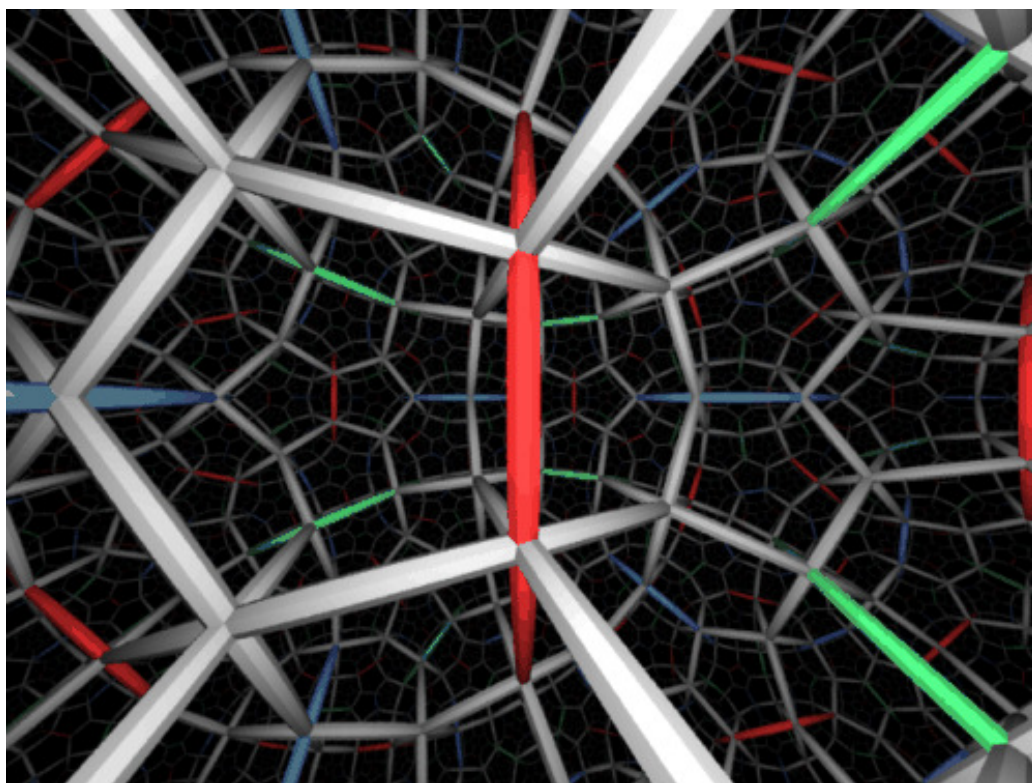# Introduction to Hyperbolic Geometry

Julien Paupert

Spring 2016

Contents

Chapter 1

Geometry of real and complex hyperbolic space

1.1 The hyperboloid model

Let $n \geqslant 1$ and consider a symmetric bilinear form of signature $(n, 1)$ on the vector space $\mathbb{R}^{n+1}$, e. g. the standard Lorentzian form:

$$\langle X, Y \rangle = x_1 y_1 + ... + x_n y_n - x_{n+1} y_{n+1}, \tag{1.1.1}$$

where: $X = (x_1, ..., x_{n+1})^T$ and $Y = (y_1, ..., y_{n+1})^T$.

Consider the following hypersurface $H \subset \mathbb{R}^{n+1}$, which we call *the hyperboloid* (more precisely, H is the upper sheet of the standard 2-sheeted hyperboloid in $\mathbb{R}^{n+1}$):

$$H = \left\{ X = (x_1, ..., x_{n+1})^T \in \mathbb{R}^{n+1} \,|\, \langle X, X \rangle = -1 \text{ and } x_{n+1} > 0 \right\}.$$

Lemma 1.1.1 *(1) H is a smooth (oriented) hypersurface.*
(2) for any $X \in H$, $T_X H = X^\perp = \{Y \in \mathbb{R}^{n+1} \,|\, \langle X, Y \rangle = 0\}$.
(3) H is (connected and) simply-connected.

Proof. (1) Let $f : \mathbb{R}^{n+1} \longrightarrow \mathbb{R}$ be the function defined by $f(X) = \langle X, X \rangle$. Then f is everywhere differentiable and -1 is a regular value of f. Indeed, noting that $f(X + Y) = f(X) + 2\langle X, Y \rangle + \langle Y, Y \rangle$, we see that the differential of f at X is $d_X f : Y \mapsto 2\langle X, Y \rangle$ (as the term $2\langle X, Y \rangle$ is linear in Y whereas $\langle Y, Y \rangle$ is quadratic - we identify here the vector space $\mathbb{R}^{n+1}$ with its tangent space at X). This has rank 1 everywhere except at $X = 0$.
(2) This also tells us that, for $X \in H$, $T_X H = \text{Ker}\, d_X f = X^\perp$.
(3) The projection $\mathbb{R}^{n+1} \longrightarrow \mathbb{R}^n$ onto the n first coordinates induces a homeomorphism from H to $\mathbb{R}^n$, as the last coordinate of a point $X = (x_1, ..., x_{n+1})^T \in H$ is given by the first n coordinates via $x_{n+1} = (1 + x_1^2 + ... + x_n^2)^{1/2}$. $\qquad\square$

Note that, given $X \in H$, the restriction of the ambient bilinear form $\langle ., . \rangle$ to $X^\perp$ is positive definite (because $\langle X, X \rangle < 0$). Obviously this product varies smoothly with the point X, as $\langle ., . \rangle$ is a smooth function of 2 variables. In fancier terms, we say that $\langle ., . \rangle$ induces a Riemannian

metric on H. We will denote H^n the resulting Riemannian manifold. We will sometimes (seldom) use the quadratic notation for this Riemannian metric:

$$ds^2 = dx_1^2 + ... + dx_n^2 - dx_{n+1}^2.$$

Recall that a Riemannian metric (an infinitesimal object) on a manifold M induces a (global) distance function on M as follows, by integrating the metric along paths (like in Calculus). The *length* of a path $\gamma : [0, 1] \longrightarrow M$ (relative to ds) is:

$$l(\gamma) := \int_0^1 ds(\gamma'(t)).dt$$

(Strictly speaking, this only makes sense for paths for which this integral is well-defined; we call such paths *rectifiable*). This defines a distance d on M by letting, for $X, Y \in M$: $d(X, Y) = \mathrm{Inf}(l(\gamma))$, where the infemum is taken over all (rectifiable) paths γ with $\gamma(0) = X$ and $\gamma(1) = Y$.

We will see that there is in fact a unique *geodesic segment* connecting any 2 given points $X, Y \in H$, i. e. a unique path realizing the infemum above. In fact, there is a very simple and convenient formula to compute distances in H, which we state without proof for now (we will prove this formula once we know more about the isometries and subspaces of H, see Proposition 1.4.1).

Proposition 1.1.1 *[Distance formula] Let $X, Y \in H$. Then $d(X, Y) = \cosh^{-1}(-\langle X, Y \rangle)$.*

1.2 Isometries of H^n

We start by recalling the definitions of isometries in the setting of metric spaces and Riemannian manifolds.

A *(metric) isometry* from a metric space (X, d_X) to another metric space (Y, d_Y) is a bijection $f : X \longrightarrow Y$ preserving distances, meaning that:

$$d_Y(f(x), f(y)) = d_X(x, y) \text{ for all } x, y \in X.$$

Given two Riemannian manifolds (M, g) and (N, h) (where g, h denote the Riemannian metrics, i.e. inner products on the tangent spaces), a *(Riemannian) isometry* from (M, g) to (N, h) is a diffeomorphism $f : M \longrightarrow N$ (i.e. a smooth bijection with smooth inverse) preserving the Riemannian metrics, meaning that:

$$h(d_x f(X), d_x f(Y)) = g(X, Y) \text{ for all } x \in M \text{ and } X, Y \in T_x M.$$

In our setting, we will use the word isometry to mean a Riemannian isometry; note that these are in particular isometries for the associated distance functions. Given an (oriented) Riemannian manifold M, we will denote $\mathrm{Isom}(M)$ the group of isometries from M to itself, and $\mathrm{Isom}^+(M)$ the subgroup of orientation-preserving isometries of M. (These groups also come with a topology, in general terms the so-called compact-open topology, but in the case of hyperbolic spaces we will identify their isometry groups with matrix groups with their usual topology).

We now return to the hyperboloid. Denote $O(n,1)$ the group of linear transformations of $\mathbb{R}^{n+1}$ preserving the ambient bilinear form. In symbols:

$$O(n,1) = \left\{ A \in GL(n+1,\mathbb{R}) \mid (\forall X, Y \in \mathbb{R}^{n+1}) \, \langle AX, AY \rangle = \langle X, Y \rangle \right\}.$$

Note that H and $-H$ are the 2 connected components of $\{X \in \mathbb{R}^{n+1} \mid \langle X, X \rangle = -1\}$, so that elements of $O(n,1)$ either preserve each of H and $-H$ or exchange them. We denote by $O^+(n,1)$ the index 2 subgroup of $O(n,1)$ consisting of those transformations which preserve H:

$$O^+(n,1) = \{ A \in O(n,1) \mid A(H) = H \}.$$

Since the metric on H is defined entirely in terms of the form $\langle ., . \rangle$, we have:

Lemma 1.2.1 $O^+(n,1) \subset \mathrm{Isom}(H^n)$ *and* $SO^+(n,1) \subset \mathrm{Isom}^+(H^n)$.

We will show that these two inclusions are in fact equalities in Theorem 1.3.1, after discussing reflections; for now this already gives us enough isometries of H^n to show the following properties.

Proposition 1.2.1 *(1)* H^n *is a homogeneous and isotropic space. In other words,* $\mathrm{Isom}(H^n)$ *acts transitively on the unit tangent bundle* UTH^n*: given any 2 points* $x, y \in H^n$ *and unit tangent vectors* $u \in T_x H^n$*,* $v \in T_y H^n$*, there exists an isometry* f *of* H^n *such that* $f(x) = y$ *and* $d_x f(u) = v$.
(2) The stabilizer in $O(n,1)$ *of any point of* H *is isomorphic to* $O(n)$.
(3) H^n *is a (Riemannian) symmetric space: given any point* $x \in H^n$ *there exists an isometry* i_x *of* H^n *such that* $i_x(x) = x$ *and* $d_x i_x = -\mathrm{Id}$.

Proof. (1) Let $x \in H$. Complete x to a (Lorentzian) orthonormal basis $(v_1, ..., v_n, x)$ and the define $A_x \in GL(n+1,\mathbb{R})$ to be (the matrix in the standard basis of) the linear transformation sending the standard basis $(e_1, .., e_{n+1})$ to $(v_1, ..., v_n, x)$. Then $A_x \in O(n,1)$ as both bases are $\langle ., . \rangle$-orthonormal. Moreover, $A_x(e_{n+1}) = x$ so A preserves H and $A_x \in O^+(n,1)$. Then, given any 2 points $x, y \in H$, $A_y A_x^{-1} \in O^+(n,1)$ sends x to y. The claim that H^n is isotropic will follow from (2).
(2) If $A \in O^+(n,1)$ fixes a point $x \in H$, it also preserves $x^\perp$ (because it preserves the form $\langle ., . \rangle$). In particular, the stabilizer of e_{n+1} in $O^+(n,1)$ consists of all block matrices A of the form:

$$A = \begin{pmatrix} \widetilde{A} & 0 \\ 0 & 1 \end{pmatrix} \text{ with } \widetilde{A} \in O(n).$$

Note that a such matrix A acts on $T_{e_{n+1}} H \simeq e_{n+1}^\perp = \mathrm{Span}(e_1, ..., e_n)$ by the usual action of $\widetilde{A}$ on $\mathbb{R}^n$, so $d_{e_{n+1}} A \simeq \widetilde{A}$. Since $O(n)$ acts transitively on unit vectors in $\mathbb{R}^n$, $\mathrm{Stab}_{O^+(n,1)}(e_{n+1})$ acts transitively on unit tangent vectors in $T_{e_{n+1}} H$. The same is then true at any other point of H: since $O^+(n,1)$ acts transitively on H, the stabilizer of any point is conjugate to the stabilizer of e_{n+1}.
(3) Define $i_{e_{n+1}} \in O^+(n,1)$ in block form as above with $\widetilde{A} = -\mathrm{Id}$. Then: $i_{e_{n+1}}(e_{n+1}) = e_{n+1}$ and $d_{e_{n+1}} i_{e_{n+1}} = -\mathrm{Id}$. Now, given any point $x \in H$ and $A_x \in O^+(n,1)$ defined as above, $i_x = A_x i_{e_{n+1}} A_x^{-1}$ satisfies: $i_x(x) = x$ and $d_x i_x = -\mathrm{Id}$. $\qquad \square$

1.3 Reflections

We start with a digression on general bilinear algebra. Let V be a (finite-dimensional) real vector space and $\langle .,. \rangle$ a *non-degenerate* symmetric bilinear form (this means that $V^\perp = \{0\}$: the only vector perpendicular to all vectors is 0). Let $W \subset V$ be a subspace such that $V = W \oplus W^\perp$ (i.e. $W \cap W^\perp = \{0\}$). Attached to such a subspace W are 2 linear maps denoted p_W (the orthogonal projection onto W) and r_W (the "reflection" across W), defined as follows. (We write "reflection" with quotes because we will reserve that term for the case where W has codimension 1). For any $v \in V$, $p_W(v)$ is defined as the unique $w \in W$ such that $v - w \in W^\perp$, and:

$$r_W(v) = 2p_W(v) - v. \tag{1.3.1}$$

Lemma 1.3.1 *(a) $p_W(v) = v \iff v \in W$, and: $p_W(v) = 0 \iff v \in W^\perp$.*
(b) $r_W(v) = v \iff v \in W$, and: $r_W(v) = 0 \iff v = 0$ (i.e. $r_W \in \mathrm{GL}(V)$).
(c) r_W has order 2, i.e. $r_W(r_W(v)) = v$ for all $v \in V$.
(d) $r_W \in \mathrm{O}(\langle .,. \rangle)$, i.e. $\langle r_W(u), r_W(v) \rangle = \langle u, v \rangle$ for all $u, v \in V$.

Proof. (a) follows from the definition of p_W.
(b) From (a), if $v \in W$ then $r_W(v) = 2v - v = v$. Conversely, if $v \in V$ satisfies $r_W(v) = v$ then $2p_W(v) = 2v$ so $v = p_W(v) \in W$.
(c) For any $v \in V$:

$$
\begin{aligned}
r_W(r_W(v)) &= 2p_W(r_W(v)) - r_W(v) \\
&= 2p_W(p_W(v) - v) - (2p_W(v) - v) \\
&= 4p_W(v) - 2p_W(v) - 2p_W(v) + v \\
&= v.
\end{aligned}
$$

(d) Recall that $p_W(v) \perp (v - p_W(v))$, i.e. $\langle p_W(v), v \rangle = \langle p_W(v), p_W(v) \rangle$. Then:

$$
\begin{aligned}
\langle r_W(v), r_W(v) \rangle &= \langle 2p_W(v) - v, 2p_W(v) - v \rangle \\
&= 4\langle p_W(v), p_W(v) \rangle - 4\langle p_W(v), v \rangle + \langle v, v \rangle \\
&= \langle v, v \rangle.
\end{aligned}
$$

The result follows by polarization (a symmetric bilinear form $B(u, v)$ is determined by its associated quadratic form $q(v) = B(v, v)$). $\qquad\square$

The case that will be of particular interest to us is when $W = e^\perp$ for some $e \in V$ with $\langle e, e \rangle \neq 0$. Then $V = \mathbb{R}e \oplus e^\perp$, and $r_{e^\perp}$ is called the *reflection across the hyperplane $e^\perp$* (sometimes called "parallel to e"). In that case there is a simple explicit formula for $r_{e^\perp}$ which we will use later:

Lemma 1.3.2 *Let $e \in V$ such that $\langle e, e \rangle \neq 0$. Then, for any $v \in V$:*

$$r_{e^\perp}(v) = v - 2\frac{\langle v, e \rangle}{\langle e, e \rangle}e. \tag{1.3.2}$$

Proof. Denote $W = e^\perp$, and let $w = v - \frac{\langle v, e \rangle}{\langle e, e \rangle} e$. Then $w \in W$, because:

$$\langle w, e \rangle = \langle v - \frac{\langle v, e \rangle}{\langle e, e \rangle} e, e \rangle = \langle v, e \rangle - \frac{\langle v, e \rangle}{\langle e, e \rangle} \langle e, e \rangle = 0.$$

Moreover, $v - w \in \mathbb{R}e = W^\perp$. Therefore $w = p_W(v)$, and $r_W(v) = 2w - v$. $\qquad\square$

Proposition 1.3.1 *Let V be a finite-dimensional real vector space and $\langle ., . \rangle$ a non-degenerate symmetric bilinear form on V. Then $O(V, \langle ., . \rangle)$ is generated by reflections.*

Proof. By induction on the dimension of V.
Let V be a real vector space of dimension $n+1$, $\langle ., . \rangle$ a non-degenerate symmetric bilinear form on V, $A \in O(V, \langle ., . \rangle)$ and $v \in V$ satisfying $\langle v, v \rangle \neq 0$.

- First assume that $\langle Av - v, Av - v \rangle \neq 0$, and consider the reflection ρ across the hyperplane $(Av - v)^\perp$. Note that $v = \frac{1}{2}(Av + v) - \frac{1}{2}(Av - v)$ with $\langle (Av + v), (Av - v) \rangle = 0$, therefore $\frac{1}{2}(Av + v) = p_{(Av-v)^\perp}(v)$ and $\rho(v) = 2p_{(Av-v)^\perp}(v) - v = Av$.

  Then $\rho \circ A(v) = v$, so $\rho \circ A$ preserves $v^\perp$, and by the induction hypothesis $(\rho \circ A)_{|v^\perp}$ is a product of reflections in $v^\perp$, say $(\rho \circ A)_{|v^\perp} = r'_1 \circ ... \circ r'_k$. Extending each r'_i to a reflection r_i in $O(V, \langle ., . \rangle)$, we get that $\rho \circ A$ and $r_1 \circ ... \circ r_k$ both preserve the decomposition $V = \mathbb{R}v \oplus v^\perp$ (as they are both in $O(V, \langle ., . \rangle)$) and agree on $v^\perp$, therefore they are equal (again, because they preserve $\langle ., . \rangle$). Thus $A = \rho \circ r_1 \circ ... \circ r_k$.

- Now, if by bad luck $\langle Av - v, Av - v \rangle = 0$, we replace A by $-A$ in the preceding argument. Indeed, noting that $\langle Av, v \rangle = \langle v, v \rangle$, we have:

$$\langle -Av - v, -Av - v \rangle = \langle Av, Av \rangle + 2\langle Av, v \rangle + \langle v, v \rangle = 4\langle v, v \rangle \neq 0.$$

  Therefore, by the preceding argument, $-A$ is a product of reflections, and so is A because $-\text{Id}$ is always a product a reflections by the following:

Lemma 1.3.3 *If $v_1, ..., v_{n+1}$ is an orthonormal basis of V then $r_{v_1^\perp} \circ ... \circ r_{v_{n+1}^\perp} = -\text{Id}$.*

Recall that *orthonormal* means that $\langle v_i, v_j \rangle = 0$ if $i \neq j$ and ± 1 if $i = j$ (with the number of $+1$'s and -1's determined by the signature of $\langle ., . \rangle$). The lemma follows by noting that:

$$r_{v_j}^\perp(v_i) = v_i - 2\frac{\langle v_i, v_j \rangle}{\langle v_j, v_j \rangle} v_j = v_i \text{ if } i \neq j \text{ and } -v_i \text{ if } i = j.$$

This concludes the proof of the proposition. $\qquad\square$

We now return to the case of the hyperboloid, with $V = \mathbb{R}^{n+1}$ and $\langle ., . \rangle$ the standard Lorentzian form (1.1.1).

Lemma 1.3.4 *Given $e \in \mathbb{R}^{n+1}$ with $\langle e, e \rangle \neq 0$, the reflection $r_{e^\perp}$ preserves H if $\langle e, e \rangle > 0$, and exchanges H and $-H$ if $\langle e, e \rangle < 0$.*

Proof. Recall that, by lemma 1.3.1 (d), $r_{e^\perp}$ is in $\mathrm{O}(n,1)$ so it either preserves each of H and $-H$ or exchanges them. The lemma follows by noting that $e^\perp$ intersects $H \cup -H$ if and only if $\langle e, e \rangle > 0$ (as the restriction of $\langle .,. \rangle$ to $e^\perp$ has signature $(n-1, 1)$ in that case, as opposed to $(n, 0)$). $\qquad\square$

Proposition 1.3.2 $\mathrm{O}^+(n,1)$ *is generated by reflections.*

Note that this means that $\mathrm{O}^+(n,1)$ is generated by the reflections it contains, i.e. by the reflections of the form $r_{e^\perp}$ with $\langle e, e \rangle > 0$.

Proof. Let $A \in \mathrm{O}^+(n,1)$. Since $\mathrm{O}(n,1)$ is generated by reflections, $A = \rho_1 \circ ... \circ \rho_k$ where each ρ_i is a reflection in $\mathrm{O}(n,1)$, i.e. $\rho_i = r_{e_i^\perp}$ for some $e_i \in \mathbb{R}^{n+1}$ with $\langle e_i, e_i \rangle \neq 0$ (but not necessarily $\langle e_i, e_i \rangle > 0$). For each $\rho_i = r_{e_i^\perp}$ which is in $\mathrm{O}(n,1)$ but not in $\mathrm{O}^+(n,1)$, we replace ρ_i with a product of reflections which are each in $\mathrm{O}^+(n,1)$ as follows. Since $\langle e_i, e_i \rangle < 0$, we can complete e_i to an orthonormal basis $(e_i, v_1, ..., v_n)$ with $\langle v_j, v_j \rangle > 0$. Then each $\sigma_j = r_{v_j}^\perp$ is a reflection in $\mathrm{O}^+(n,1)$, and from lemma 1.3.3 we have that: $\rho_i \circ \sigma_1 \circ ... \circ \sigma_n = -\mathrm{Id}$, so $\rho_i = -\sigma_1 \circ ... \circ \sigma_n$. After replacing each such ρ_i in this way, we get $A = \pm(\rho_1' \circ ... \circ \rho_l')$ where each ρ_j' is a reflection in $\mathrm{O}^+(n,1)$. Since A and each ρ_j' preserve H, the sign must be a $+$. $\quad\square$

Theorem 1.3.1 $\mathrm{Isom}(H^n) = \mathrm{O}^+(n,1)$ *and* $\mathrm{Isom}^+(H^n) = \mathrm{SO}^+(n,1)$. *In particular,* $\mathrm{Isom}(H^n)$ *is generated by reflections.*

We will prove this using the following fact that a Riemannian (local) isometry is entirely determined by the image of a single point and its differential at that point:

Proposition 1.3.3 *Let M, N be two Riemannian manifolds of the same dimension, with M connected. If $\phi_1 : M \longrightarrow N$ and $\phi_2 : M \longrightarrow N$ are two local isometries such that $\phi_1(x) = \phi_2(x)$ and $d_x\phi_1 = d_x\phi_2$ for some $x \in M$, then $\phi_1 = \phi_2$.*

Proof. Let $S = \{y \in M \,|\, \phi_1(y) = \phi_2(y)$ and $d_y\phi_1 = d_y\phi_2\}$. Then S is closed in M, and non-empty by assumption. We now prove that S is also open in M, which implies that $S = M$ because M is connected.

Let $x \in S$. $\mathrm{Im}\phi_1$ and $\mathrm{Im}\phi_2$ are open in N because M, N have the same dimension and ϕ_1, ϕ_2 are local diffeomorphisms. Therefore there exist neighborhoods U_1, U_2 of x in M and V of $\phi_1(x) = \phi_2(x)$ in N such that $\phi_i : U_i \longrightarrow V$ is bijective (and an isometry) for $i = 1, 2$. We use the existence of of a *normal neighborhood* U of x in M, which is a neighborhood such that the restriction of the exponential map to some neighborhood W of 0 in T_xM is a diffeomorphism $W \longrightarrow U$. (Recall that the exponential map $\exp_x$ at a point $x \in M$ is defined as the map:

$$\begin{aligned} \exp_x : \quad T_xM &\longrightarrow M \\ v &\longmapsto \gamma_v(1) \end{aligned}$$

where γ_v is the (unique) geodesic arc with $\gamma_v(0) = x$ and $\gamma_v'(0) = v$.) By shrinking U if necessary we may assume that $U \subset U_1$. Consider then $f = (\phi_{2|U_2})^{-1} \circ \phi_{1|U_1}$. Then f is an isometry from U_1 onto U_2, $f(x) = x$ and $d_xf = \mathrm{Id}$. Therefore, for any geodesic arc γ starting at x, $f \circ \gamma$ is

also a geodesic arc starting at x, with tangent vector $(f \circ \gamma)'(0) = d_x f(\gamma'(0)) = \gamma'(0)$. Thus $f \circ \gamma = \gamma$; since this holds for any geodesic arc γ starting at x we have: $f \circ (\exp_{x|W}) = \exp_{x|W}$, hence $f_{|U} = \mathrm{Id}_{|U}$ (as $\exp_{x|W} : W \longrightarrow U$ is a diffeomorphism). Therefore: $\phi_{1|U} = \phi_{2|U}$ (in particular, $d_p \phi_1 = d_p \phi_2$ for any $p \in U$), so that $U \subset S$. Therefore S is open in M. $\qquad \square$

Proof of Theorem 1.3.1. Let $\varphi \in \mathrm{Isom}(H^n)$ and let $x \in H^n$. Consider the linear map $A \in \mathrm{GL}(n+1, \mathbb{R})$ given in block form by:

$$A = \begin{pmatrix} d_x\varphi & 0 \\ 0 & 1 \end{pmatrix}$$

in the decompositions $x^\perp \oplus \mathbb{R}x$ (at the source) and $\varphi(x)^\perp \oplus \mathbb{R}\varphi(x)$ (at the range) of $\mathbb{R}^{n+1}$. Since $d_x\varphi$ sends $\langle .,. \rangle_{|x^\perp \times x^\perp}$ to $\langle .,. \rangle_{|\varphi(x)^\perp \times \varphi(x)^\perp}$ (as φ is an isometry) and $\langle \varphi(x), \varphi(x) \rangle = -1 = \langle x, x \rangle$, A preserves the form $\langle .,. \rangle$ so $A \in \mathrm{O}(n, 1)$. In fact $A \in \mathrm{O}^+(n, 1)$ as $A(x) = \varphi(x)$. Therefore A is an isometry of H^n, with $A(x) = \varphi(x)$ and $d_x A = d_x\varphi$, so by the previous result $A = \varphi$. $\qquad \square$

1.4 Geodesics and totally geodesic subspaces

We now determine the geodesics and totally geodesic subspaces of H^n in the hyperboloid model. Recall that a smooth parametrized curve $\gamma(t)$ in a Riemannian manifold (M, g) is *geodesic* if it is locally length-minimizing, i.e. if for all close enough $x = \gamma(t_1)$ and $y = \gamma(t_2)$, $d(x, y) = \int_{t_1}^{t_2} g(\gamma'(t)).dt$.

Proposition 1.4.1 *Let x, y be two distinct points in H.*
(a) There exists a unique geodesic segment $[xy]$ connecting x and y; it is contained in the (bi-infinite) geodesic line $H \cap \mathrm{Span}(x, y)$.
(b) Given a unit tangent vector $v \in T_x H$, the unit-speed parametrization of the unique geodesic γ satisfying $\gamma(0) = x$ and $\gamma'(0) = v$ is given by:

$$\gamma(t) = \cosh(t).x + \sinh(t).v \quad (t \in \mathbb{R}). \tag{1.4.1}$$

(c) **Distance formula:** $\cosh(d(x, y)) = -\langle x, y \rangle$.

Proof. (a) Denote $(xy) = H \cap \mathrm{Span}(x, y)$. As isometries of H^n are restrictions to H of linear maps of $\mathbb{R}^{n+1}$, any isometry fixing x and y fixes (xy) pointwise. As (xy) is a 1-dimensional submanifold of H, it is the unique geodesic line containing x and y. (Here we use the fact that geodesics always exist locally).
(b) The same argument as above tells us that, as a set, $\gamma(\mathbb{R}) = H \cap \mathrm{Span}(x, v)$. Since H^n is homogeneous and isotropic, we may normalize by an isometry so that: $x = e_{n+1}$ and $v = e_n$. Then a parametrization of $H \cap \mathrm{Span}(x, v)$ is: $\gamma(t) = \cosh(t).e_{n+1} + \sinh(t).e_n$; this satisfies $\gamma(0) = x$ and $\gamma'(0) = v$. Moreover $\gamma'(t) = \sinh(t).e_{n+1} + \cosh(t).e_n$, so $\langle \gamma'(t), \gamma'(t) \rangle = \cosh^2(t) - \sinh^2(t) = 1$.
(c) Normalize again so that $x = e_{n+1}$ and $y \in \gamma$ as above, so that $y = \cosh(T)e_{n+1} + \sinh(T)e_n$ for some $T > 0$. Then: $d(x, y) = \int_0^T \langle \gamma'(t), \gamma'(t) \rangle.dt = T$ and $\langle x, y \rangle = -\cosh(T)$. $\qquad \square$

As a corollary of (a) and the Hopf-Rinow theorem we get:

Corollary 1.4.1 H^n *is complete.*

Recall that a submanifold N of a Riemannian manifold M is *totally geodesic* if any geodesic ray starting in N and tangent to N remains in N, in other words if for any geodesic arc $\gamma : \gamma(0) = x \in N$ and $\gamma'(0) \in T_x N \Rightarrow (\forall t)\ \gamma(t) \in N$. If M is uniquely geodesic (i.e. any 2 points in M are connected by a unique geodesic segment), then this is equivalent to: for any $x, y \in N$, the geodesic segment $[xy]$ is contained in N.

A $(k+1)$-dimensional linear subspace W of $\mathbb{R}^{n+1}$ is called *hyperbolic* (resp. *parabolic*, resp. *elliptic*) if the restriction to W of the form $\langle ., . \rangle$ has signature $(k, 1)$ (resp. $(k, 0)$, resp $(k+1, 0)$).

Proposition 1.4.2 *Let N be a totally geodesic subspace of H^n.*
(a) N is of the form $H \cap W$ for some hyperbolic linear subspace W of $\mathbb{R}^{n+1}$.
(b) In particular, N is an isometrically embedded copy of H^k in H^n (with $k = \dim N$).
(c) N is the fixed-point locus of an isometry of order 2.

Proof. (a) follows from Proposition 1.4.1 (a).
(b) follows from the construction of the hyperboloid model.
(c) Take r_W as defined in Section 1.3; the fact that W is hyperbolic ensures that r_W preserves H. $\qquad\square$

1.5 The ball model and upper half-space model

1.5.1 The ball model

Denote $\mathbb{R}^{n+1,+} = \{(x_1, ..., x_{n+1}) \in \mathbb{R}^{n+1} \mid x_{n+1} > 0\}$, and let $\pi : \mathbb{R}^{n+1,+} \longrightarrow \mathbb{R}^n \times \{0\}$ denote radial projection from the point $-e_{n+1}$. Explicitly, in coordinates:

$$\pi(x_1, ..., x_{n+1}) = \frac{(x_1, ..., x_n, 0)}{1 + x_{n+1}}.$$

Lemma 1.5.1 π *induces a diffeomorphism $H \longrightarrow B$, where $B \simeq B \times \{0\}$ is the open unit ball in $\mathbb{R}^n$.*

Proof. We can write the inverse map g explicitly as $g(y_1, ..., y_n) = (\lambda y_1, ..., \lambda y_n, \lambda - 1)$ with λ chosen so that the latter point lies in H, namely $\lambda = 2/(1 - y_1^2 - ... - y_n^2)$. $\qquad\square$

We then carry over the metric on H to a metric on B by this diffeomorphism (more formally, we consider the pullback of the metric on H by π^{-1}), and denote B^n the corresponding Riemannian manifold, which is by construction isometric to H^n.

1.5.2 The upper half-space model

Let $S = S(-e_n, \sqrt{2}) \in \mathbb{R}^n$ be the sphere centered at $-e_n$ with radius $\sqrt{2}$, and consider the inversion $i = i_S$ in S (see below if you don't remember what an inversion is in $\mathbb{R}^n$). Explicitly, in coordinates:

$$i : \quad \mathbb{R}^n \setminus \{-e_n\} \quad \longrightarrow \quad \mathbb{R}^n$$
$$x \quad \longmapsto \quad 2\frac{x+e_n}{||x+e_n||^2} - e_n,$$

where $||.||$ denotes the Euclidean norm on $\mathbb{R}^n$.

Lemma 1.5.2 *i induces a diffeomorphism from the (open) unit ball B to the upper half-space $U = \{(x_1, ..., x_n) \in \mathbb{R}^n \,|\, x_n > 0\}$.*

Proof. This follows from general properties of inversions (see below), by noting that the unit sphere ∂B is a sphere containing the center $-e_n$ of S, so $i(\partial B)$ is a hyperplane, which must be $\mathbb{R}^{n-1} \times \{0\}$ (as the latter contains the $(n-2)$-sphere $\partial B \cap S$). The image of B is then one of the 2 half-spaces bounded by $\mathbb{R}^{n-1} \times \{0\}$, which must be U as $i(0) = e_n$. $\qquad\square$

Again, we carry over the metric on B to a metric on U by this diffeomorphism (more formally, we consider the pullback of the metric on B by $i^{-1} = i$), and denote U^n the corresponding Riemannian manifold, which is by construction also isometric to H^n.

1.5.3 Inversions and Möbius transformations

The isomeries in these 2 models are best described using inversions, so we now recall their definition and main properties.

Given a center $x_0 \in \mathbb{R}^n$ and a radius $R > 0$, the *inversion i_S in the sphere $S = S(x_0, R)$* - also denoted $i_{(x_0, R)}$ - is the map:

$$i_S : \quad \mathbb{R}^n \setminus \{x_0\} \quad \longrightarrow \quad \mathbb{R}^n \tag{1.5.1}$$
$$x \quad \longmapsto \quad x_0 + \frac{R^2}{||x-x_0||^2}(x - x_0),$$

where $||.||$ denotes the Euclidean norm on $\mathbb{R}^n$. We extend i_S to the one-point compactification $\widehat{\mathbb{R}^n} = \mathbb{R}^n \cup \{\infty\}$ by setting $i_S(\infty) = x_0$ and $i_S(x_0) = \infty$. For the convenience of statements relating to inversions, we define a *generalized sphere* in $\widehat{\mathbb{R}^n}$ to be either a (top-dimensional) Euclidean sphere in $\mathbb{R}^n$ or a hyperplane (affine, i.e. not necessarily containing the origin). Note that a generalized sphere is a hyperplane if and only if it contains ∞. The following proposition summarizes the main geometric properties of inversions:

Proposition 1.5.1 *Let $S = S(x_0, R)$ be a sphere and $i = i_S$ its inversion.*
(a) i_S is an involution, fixing S pointwise and exchanging interior and exterior of S.
(b) i_S induces a conformal diffeomorphism $\mathbb{R}^n \setminus \{x_0\} \longrightarrow \mathbb{R}^n \setminus \{x_0\}$.
(c) If S' is a generalized sphere, then $i(S')$ is also a generalized sphere.
(d) If S' is a generalized sphere, then: $i(S') = S' \iff S = S'$ or $S \perp S'$.

Recall that a smooth map $f : U \longrightarrow \mathbb{R}^n$ ($U \subset \mathbb{R}^n$ open) is called *conformal* if it preserves angles; more precisely this means that its differential at each point preserves dot-products up to a muliplicative factor:

$$(\forall x \in U)(\exists K(x) \in \mathbb{R})(\forall u, v \in T_x U) \; d_x f(u) \cdot d_x f(v) = K(x)\, u \cdot v.$$

We will use the following lemma, describing the composite of 2 inversions with common center; $D(x_0, R)$ denotes the dilation with center x_0 and ratio R:

Lemma 1.5.3 *If $i_1 = i_{(x_0, R_1)}$ and $i_2 = i_{(x_0, R_2)}$ then $i_2 \circ i_1 = D(x_0, \frac{R_2^2}{R_1^2})$.*

Proof of lemma. $i_2 \circ i_1(x) = i_2\left(x_0 + \frac{R_1^2}{||x - x_0||^2}(x - x_0)\right) = \frac{R_2^2}{R_1^2}(x - x_0) + x_0$. $\qquad \square$

Proof of proposition. (a) By the lemma, i_S is an involution. The second statement follows by noting that: $||x - x_0||^2 < R^2 \iff ||i(x) - x_0||^2 > R^2$ (likewise with the 2 inequalities reversed, respectively replaced by equalities).

(b) Note that $i_S = T_{x_0} \circ D(0, R^2) \circ i_{0,1} \circ T_{x_0}^{-1}$, where T_{x_0} denotes translation by x_0. Translations and dilations are conformal, therefore it suffices to show that the inversion $i_{(0,1)}$ in the unit sphere is conformal. Let $x \in \mathbb{R}^n \setminus \{0\}$ and $u \in \mathbb{R}^n \setminus \{x\}$. We claim that:

$$d_x i_{(0,1)}(u) = \frac{1}{||x||^2} r_{x^\perp}(u),$$

which gives the result as reflections are conformal. In order to see this, note that:

$$\frac{1}{||x + u||^2} = \frac{1}{||x||^2} - 2\frac{x \cdot u}{||x||^4} + o(||u||),$$

because the derivative of the function $x \longmapsto \frac{1}{||x||^2}$ is $u \longmapsto -2\frac{x \cdot u}{||x||^4}$. Therefore:

$$
\begin{aligned}
i_{(0,1)}(x + u) = \frac{x + u}{||x + u||^2} &= \frac{x + u}{||x||^2} - 2\frac{x \cdot u(x + u)}{||x||^4} + o(||u||) \\
&= \frac{x}{||x||^2} + \frac{u}{||x||^2} - 2\frac{x \cdot u x}{||x||^4} + o(||u||) \\
&= i_{(0,1)}(x) + \frac{1}{||x||^2} r_{x^\perp}(u) + o(||u||).
\end{aligned}
$$

This proves the claim. Note that in general, it follows by composing with the appropriate translations and dilation that:

$$d_x i_{(x_0, R)} = \frac{R^2}{||x - x_0||^2} r$$

for some (Euclidean) reflection r.

(c) Note that this statement contains 4 statements about Euclidean subspaces, by our conventions regarding generalized spheres and the point ∞. Namely, given a hyperplane H and a sphere S':

(1) if $x_0 \in H$ then $i(H) = H$.

(2) if $x_0 \notin H$ then $i(H) = S''$ for some sphere S'' with $x_0 \in S''$.

(3) if $x_0 \in S'$ then $i(S') = H'$ for some hyperplane H' with $x_0 \notin H'$.

(4) if $x_0 \notin S'$ then $i(S') = S''$ for some sphere S'' with $x_0 \notin S''$.

We now prove these four statements separately.

(1) follows from the definition, as x_0, x and $i(x)$ are collinear for all x.

For the remaining statements we assume that $i = i_{(0,1)}$ (as the results are invariant under translations and dilations).

(2) Let H be a hyperplane not through 0, which we write in the form $H = h + h^\perp$ (where

$h \in \mathbb{R}^n \setminus \{0\}$ is the orthogonal projection of 0 onto H). Let $C = \frac{h}{2||h||^2}$ and $R = \frac{1}{2||h||}$. Then, for $x \neq 0$:

$$
\begin{aligned}
i(x) \in S(C, R) \quad &\Longleftrightarrow \quad ||i(x) - C||^2 = R^2 \\
&\Longleftrightarrow \quad \left\| \frac{x}{||x||^2} - \frac{h}{2||h||^2} \right\|^2 = \frac{1}{4||h||^2} \\
&\Longleftrightarrow \quad \frac{1}{||x||^2} - \frac{x \cdot h}{||x||^2 ||h||^2} = 0 \\
&\Longleftrightarrow \quad h \cdot h - x \cdot h = 0 \\
&\Longleftrightarrow \quad x \in H = h + h^{\perp}.
\end{aligned}
$$

Therefore $i(H) = S(C, R)$, a sphere containing 0.

(3) follows from (2) as i is an involution.

(4) Let $S = S(C, R)$ be a sphere not containing 0 (i.e. $||C||^2 \neq R^2$). Then, for $x \neq 0$:

$$
\begin{aligned}
i(x) \in S(C, R) \quad &\Longleftrightarrow \quad \left\| \frac{x}{||x||^2} - C \right\|^2 = R^2 \\
&\Longleftrightarrow \quad \frac{1}{||x||^2} - 2\frac{x \cdot C}{||x||^2} + ||C||^2 = R^2 \\
&\Longleftrightarrow \quad \frac{1}{||C||^2 - R^2} - 2\frac{x \cdot C}{||C||^2 - R^2} + ||x||^2 = 0 \\
&\Longleftrightarrow \quad \left\| x - \frac{C}{||C||^2 - R^2} \right\|^2 = \frac{||C||^2}{(||C||^2 - R^2)^2} - \frac{1}{||C||^2 - R^2} = \frac{R^2}{(||C||^2 - R^2)^2} \\
&\Longleftrightarrow \quad x \in S\left(\frac{C}{||C||^2 - R^2}, \frac{R}{||C||^2 - R^2} \right).
\end{aligned}
$$

(in the third step we muliplied both sides by $\frac{||x||^2}{||C||^2 - R^2}$). The latter is a sphere not through 0, proving the statement. Note that the center of the image sphere is not the image of the center.

(d) First assume that $i_S(S') = S'$. Then $S \cap S' \neq \emptyset$ (or else S' would be entirely inside/outside of S, and $i(S')$ on the opposite side). Let $x \in S \cap S'$, and let $\vec{n}_S, \vec{n}_{S'}$ denote normal vectors at x to S, S' respectively. Then $d_x i(\vec{n}_S) = -\vec{n}_S$, and $d_x i(\vec{n}_{S'}) = \lambda \vec{n}_{S'}$ for some $\lambda \neq 0$, because i fixes x and preserves S'. In fact, $\lambda = \pm 1$ because $d_x i$ is a reflection, with eigenspaces $\mathbb{R}\vec{n}_S$ corresponding to -1 and $\vec{n}_S^{\perp}$ corresponding to 1. If $\lambda = -1$ then $\vec{n}_{S'}$ is collinear to $\vec{n}_S$ so S, S' are tangent and $S = S'$ (again, or else S' would be entirely inside/outside S). If $\lambda = 1$ then $\vec{n}_S \perp \vec{n}_{S'}$ and $S \perp S'$.

Conversely, assume that $S \perp S'$. If S' is a hyperplane then S' contains the center of S so $i(S') = S'$ by (a). If S' is a sphere, denote C its center and let $x \in S \cap S'$. Then $i(S')$ is a sphere, centered on the line (Cx_0) (by symmetry), containing x (which is not on (Cx_0)) and with normal vector $-\vec{n}_{S'}$, so $i(S') = S'$. $\qquad \square$

We are now ready to characterize isometries of the ball model and upper half-space model in terms of inversions. A *Möbius transformation* of $\widehat{\mathbb{R}^n}$ is a product of inversions and reflections. (From now on we will use the word "inversion" to mean generalized inversion, i.e. inversion or reflection). We denote $\text{Möb}(\widehat{\mathbb{R}^n})$ the group of all Möbius transformations of $\widehat{\mathbb{R}^n}$, and for any subset $V \subset \widehat{\mathbb{R}^n}$, $\text{Möb}(V)$ the group of all Möbius transformations preserving V. For example, by part (d) of the previous proposition, $\text{Möb}(B)$ is generated by inversions in generalized spheres S' such that $S' \perp \partial B$, and likewise $\text{Möb}(U)$ is generated by inversions in generalized spheres S' such that $S' \perp \partial U$.

Theorem 1.5.1 $\text{Isom}(B^n) = \text{Möb}(B)$ *and* $\text{Isom}(U^n) = \text{Möb}(U)$.

Recall that we passed from the hyperboloid model H to the ball model B via the radial projection:

$$\pi : \quad \begin{array}{ccc} H & \longrightarrow & B \\ (x_1, ..., x_{n+1}) & \longmapsto & \frac{(x_1, ..., x_n)}{1 + x_{n+1}}. \end{array}$$

and recall that $\mathrm{Isom}(H^n) = \mathrm{O}^+(n, 1)$ is generated by reflections. The first part of the theorem then follows from the following:

Proposition 1.5.2 *If $r \in \mathrm{O}^+(n, 1)$ is a reflection, then $\rho = \pi \circ r \circ \pi^{-1}$ is (the restriction to B of) an inversion.*

The second part of the theorem then follows by noting that we passed from the ball model B to the upper half-space model U by a reflection, which implies that $\mathrm{Isom}(B^n)$ and $\mathrm{Isom}(U^n)$ are conjugate inside $\mathrm{M\ddot{o}b}(\widehat{\mathbb{R}^n})$.

Proof of proposition 1.5.2. We will use coordinates on $\mathbb{R}^{n+1}$ of the form (X, t), with $X \in \mathbb{R}^n$ and $t \in \mathbb{R}$, so that: $\pi : (X, t) \longmapsto \frac{X}{1+t}$ and $\pi^{-1} : Y \longmapsto \left(\frac{2Y}{1 - ||Y||^2}, \frac{1 + ||Y||^2}{1 - ||Y||^2} \right) = (Y', u)$.

Recall that a reflection $r \in \mathrm{O}^+(n, 1)$ is of the form: $r_{v^\perp} : x \longmapsto x - 2\frac{\langle x, v \rangle}{\langle v, v \rangle} v$ for some $v \in \mathbb{R}^{n+1}$ with $\langle v, v \rangle > 0$. Normalizing v so that $\langle v, v \rangle = 1$, this becomes simply: $r_{v^\perp}(x) = x - 2\langle x, v \rangle v$. With coordinates $x = (X, t)$ as above, and $v = (V, s)$ (normalized with $\langle v, v \rangle = 1 = V \cdot V - s^2$), this becomes: $r(X, t) = (X, t) - 2(X \cdot V - st)(V, s)$.

Let $\rho = \pi \circ r \circ \pi^{-1}$. Then:

$$\begin{aligned} \rho(Y) = \pi \circ r(Y', u) &= \pi \left((Y', u) - 2(Y' \cdot V - su)(V, s) \right) \\ &= \pi \left((Y' - 2(Y' \cdot V - su)V, u - 2(Y' \cdot V - su)s) \right) \\ &= \frac{Y' - 2(Y' \cdot V - su)V}{1 + u - 2(Y' \cdot V - su)s} \\ &= \frac{Y - (2Y \cdot V - s(1 + Y \cdot Y))V}{1 - (2Y \cdot V - s(1 + Y \cdot Y))s} \end{aligned}$$

where we factored out $u = \frac{2}{1 - Y \cdot Y}$ from the numerator and denominator in the last step. We now simplify the numerator and denominator separately. First the numerator:

$$\begin{aligned} Y - (2Y \cdot V - s(1 + Y \cdot Y))V &= Y - V/s + (V/s - 2(Y \cdot V)V + sV + s(Y \cdot Y)V) \\ &= Y - V/s + s\,||Y - V/s||^2 V \end{aligned}$$

because: $\frac{V \cdot V}{s} = \frac{1 + s^2}{s} = \frac{1}{s} + s$. Now the denominator:

$$1 - (2Y \cdot V - s(1 + Y \cdot Y))s = 1 - 2sY \cdot V + s^2(1 + Y \cdot Y) = s^2\,||Y - V/s||^2.$$

Putting the pieces back together gives: $\rho(Y) = V/s + \frac{Y - V/s}{s^2||Y - V/s||^2} = i_S(Y)$ where $S = S(V/s, 1/s)$. Note that when $s = 0$, ρ is in fact a reflection (across a hyperplane containing $0 \in B$). $\qquad \square$

1.5.4 Dimensions 2 and 3: $SL(2, \mathbb{R})$ and $SL(2, \mathbb{C})$

We've seen that, in all dimensions, the group $\mathrm{Isom}(H^n)$ can be identified with the matrix group $O^+(n, 1) < GL(n+1, \mathbb{R})$. In dimensions 2 and 3 we also have the following "coincidences":

Proposition 1.5.3 $\mathrm{Isom}^+(U^2) \simeq PSL(2, \mathbb{R})$ *and* $\mathrm{Isom}^+(U^3) \simeq PSL(2, \mathbb{C})$.

Recall that $PSL(2, \mathbb{R}) = SL(2, \mathbb{R})/\{\pm \mathrm{Id}\}$ and $PSL(2, \mathbb{C}) = SL(2, \mathbb{C})/\{\pm \mathrm{Id}\}$. We first show the following:

Proposition 1.5.4 $\mathrm{M\ddot{o}b}^+(\widehat{\mathbb{R}^2}) \simeq PSL(2, \mathbb{C})$.

We identify $\mathbb{R}^2$ with $\mathbb{C}$ and consider the action of $SL(2, \mathbb{C})$ on $\mathbb{C}$ by fractional linear transformations:

$$g = \left(\begin{array}{cc} a & b \\ c & d \end{array} \right) \in SL(2, \mathbb{C})$$

acts on $\mathbb{C}$ by $z \mapsto g(z) = \frac{az+b}{cz+d}$. This action is extended to $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ in the obvious way (i.e. letting $g(\infty) = \frac{a}{c}$ and $g(-\frac{d}{c} = \infty)$). Note that $\pm \mathrm{Id}$ are the only elements acting trivially on $\mathbb{C}$, therefore this gives a faithful action of $PSL(2, \mathbb{C})$ (allowing us to identify $PSL(2, \mathbb{C})$ with a subgroup of the bijections of $\widehat{\mathbb{C}}$).

Proposition 1.5.4 follows from the following lemma, which shows that $PSL(2, \mathbb{C})$ corresponds under this action to the group generated by products of an even number of inversions of $\widehat{\mathbb{R}^2}$, namely $\mathrm{M\ddot{o}b}^+(\widehat{\mathbb{R}^2})$.

Lemma 1.5.4 *(a)* $PSL(2, \mathbb{C})$ *acts transitively on generalized circles in* $\widehat{\mathbb{C}}$.
(b) If $g \in PSL(2, \mathbb{C})$ *and* i *is an inversion,* gig^{-1} *is also an inversion.*
(c) Any element of $PSL(2, \mathbb{C})$ *is a product of 2 or 4 inversions.*
(d) Any product of 2 inversions is in $PSL(2, \mathbb{C})$.

Proof. (a) First note that $PSL(2, \mathbb{C})$ contains:

- all dilations $D_r : z \longmapsto rz$ (with $r > 0$), taking as matrix representative:

$$D_r = \left(\begin{array}{cc} \sqrt{r} & 0 \\ 0 & 1/\sqrt{r} \end{array} \right)$$

- all translations $T_v : z \longmapsto z + v$, taking as matrix representative:

$$T_v = \left(\begin{array}{cc} 1 & v \\ 0 & 1 \end{array} \right)$$

- all rotations $R_\theta : z \longmapsto e^{i\theta} z$, taking as matrix representative:

$$R_\theta = \left(\begin{array}{cc} e^{i\theta/2} & 0 \\ 0 & e^{-i\theta/2} \end{array} \right)$$

- and the involution $I : z \longmapsto -1/z$, taking as matrix representative:

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

.

Now the group generated by all dilations, translations and rotations acts transitively on the set of circles (resp. lines) in $\mathbb{C}$. Moreover, I exchanges the line $\{\operatorname{Im} z = 1\}$ and the circle $C(i/2, 1/2)$, because for $x \in \mathbb{R}$:

$$I(x + i) = -\frac{1}{x + i} = -\frac{x - i}{x^2 + 1} \quad \text{and} \quad \left\| I(x + i) - \frac{i}{2} \right\|^2 = \frac{1}{4(x^2 + 1)^2} \|i - x\|^2 = \frac{1}{2}.$$

(b) First consider $i_0 = i_{(0,1)} : z \longmapsto \frac{z}{|z|^2} = \frac{1}{\bar{z}}$, and let:

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \operatorname{SL}(2, \mathbb{C})$$

be arbitrary. We claim that $g i_0 g^{-1}$ is an inversion. To see this, we compute:

$$
\begin{aligned}
g i_0 g^{-1}(z) &= g i_0 \left(\frac{dz - b}{-cz + a} \right) = g \left(\frac{-\bar{c}\bar{z} + \bar{a}}{\bar{d}\bar{z} - \bar{b}} \right) \\
&= \frac{a\left(\frac{-\bar{c}\bar{z}+\bar{a}}{\bar{d}\bar{z}-\bar{b}} \right) + b}{c\left(\frac{-\bar{c}\bar{z}+\bar{a}}{\bar{d}\bar{z}-\bar{b}} \right) + d} = \frac{-a\bar{c}\bar{z} + |a|^2 + b\bar{d}\bar{z} - |b|^2}{-|c|^2\bar{z} + \bar{a}c + |d|^2\bar{z} - \bar{b}d} = \frac{e\bar{z} + f}{h\bar{z} - \bar{e}}
\end{aligned}
$$

with $e = b\bar{d} - a\bar{c}$, $f = |a|^2 - |b|^2$ and $h = |d|^2 - |c|^2$. If $h = 0$ then $g i_0 g^{-1}$ is a reflection across a line; if $h \neq 0$ let $z_0 = \frac{e}{h}$ and $R = \left(\frac{f}{h} + \frac{|e|^2}{h^2} \right)^{1/2}$. Then:

$$
\begin{aligned}
g i_0 g^{-1}(z) &= \frac{e\bar{z} + f}{h\bar{z} - \bar{e}} = \frac{e\bar{z}/h + f/h}{\bar{z} - \bar{e}/h} \\
&= \frac{z_0 \bar{z} - |z_0|^2 + R^2}{\bar{z} - \bar{z}_0} = z_0 + \frac{R^2}{|z - z_0|^2}(z - z_0) = i_{(z_0, R)}(z).
\end{aligned}
$$

Now, if i is any inversion then $i = h i_0 h^{-1}$ with $h = TD \in \operatorname{SL}(2, \mathbb{C})$ for some translation T and dilation D (because translations and dilations act transitively on circles). Therefore, for any $g \in \operatorname{SL}(2, \mathbb{C})$, $g i g^{-1} = g h i_0 h^{-1} g^{-1}$ is an inversion by the previous computation.

(c) From linear algebra we know that any $g \in \operatorname{SL}(2, \mathbb{C})$ is conjugate to a matrix h of the form:

$$D_{\lambda^2} = \begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix} \quad \text{for some } \lambda \in \mathbb{C}^*,$$

if it is diagonalizable, or to:

$$P = T_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

if not. Note that, writing $\lambda = re^{i\theta}$ with $r > 0$, we have: $D_{\lambda^2} = D_{r^2} R_{2\theta}$. We know that $R_{2\theta}$ and T_1 are products of 2 reflections across lines, and that D_{r^2} is a product of 2 (concentric) inversions. Therefore h is a product of 2 or 4 inversions, and by conjugation (using part (b)), so is g.

(d) Let i_1, i_2 be 2 (generalized) inversions. If they are both reflections across lines then their product is a translation or a rotation, so lies in $\mathrm{PSL}(2, \mathbb{C})$. If not, by (a) we may assume that $i_1 = i_{(0,1)}$, i.e. $i_1 : z \longmapsto 1/\bar{z}$. If i_2 is a reflection across a line we may write it as: $i_2 : z \longmapsto a\bar{z} + b$ (with $a \neq 0$). Then:

$$i_2 \circ i_1(z) = a/z + b = \frac{bz + a}{z} = \begin{pmatrix} b & a \\ 1 & 0 \end{pmatrix} z$$

and rescaling this matrix (by $1/a^{1/2}$) identifies $i_2 \circ i_1$ with an element of $\mathrm{PSL}(2, \mathbb{C})$. If i_2 is (strictly speaking) an inversion, we may write it as: $z \mapsto z_0 + \frac{R^2}{\bar{z} - \bar{z_0}}$ for some $z_0 \in \mathbb{C}$ and $R > 0$. Then:

$$i_2 \circ i_1(z) = z_0 + \frac{R^2}{1/z - \bar{z_0}} = z_0 + \frac{R^2 z}{-\bar{z_0} z + 1} = \frac{(R^2 - |z_0|^2)z + z_0}{-\bar{z_0} z + 1} = \begin{pmatrix} R^2 - |z_0|^2 & z_0 \\ -\bar{z_0} & 1 \end{pmatrix} z$$

and rescaling this matrix (by $1/R$) identifies $i_2 \circ i_1$ with an element of $\mathrm{PSL}(2, \mathbb{C})$. $\qquad \square$

Poincaré extension: This concludes the proof of Proposition 1.5.4, namely the fact that: $\mathrm{M\ddot{o}b}^+(\widehat{\mathbb{R}^2}) \simeq \mathrm{PSL}(2, \mathbb{C})$. The second statement of Proposition 1.5.3, which stated that $\mathrm{Isom}^+(U^3) \simeq \mathrm{PSL}(2, \mathbb{C})$, follows from the following general observation:

Proposition 1.5.5 *For all $n \geqslant 1$, $\mathrm{Isom}(U^{n+1}) \simeq \mathrm{M\ddot{o}b}(\widehat{\mathbb{R}^n})$.*

Proof. As the intersection of any generalized sphere in $\mathbb{R}^{n+1}$ perpendicular to $\mathbb{R}^n \times \{0\}$ is a generalized sphere in $\mathbb{R}^n \times \{0\}$, by taking restrictions we get a map from the set of all inversions of $\mathbb{R}^{n+1}$ preserving U^{n+1} to the set of inversions of $\mathbb{R}^n \times \{0\}$. This map is in fact a bijection, because each generalized sphere in $\mathbb{R}^n \times \{0\}$ is the trace of a unique generalized (half-) sphere in U^{n+1}, perpendicular to $\mathbb{R}^n \times \{0\}$. Taking the group generated by the inversions on both sides gives the result. $\qquad \square$

The process of extending an element of $\mathrm{M\ddot{o}b}(\widehat{\mathbb{R}^n})$ to an isometry of U^{n+1} is called *Poincaré extension*.

Remark: We have seen that inversions are conformal; therefore, denoting $\mathrm{Conf}(\widehat{\mathbb{R}^n})$ the group of all conformal transformations of $\widehat{\mathbb{R}^n}$, we have: $\mathrm{M\ddot{o}b}(\widehat{\mathbb{R}^n}) \subset \mathrm{Conf}(\widehat{\mathbb{R}^n})$. It can be shown in fact that this is an equality (we won't prove this; see for example chapter 1 of [BP]). Then the above correspondence can be stated in the following way:

- every isometry of U^n acts as a conformal map on ∂U^n, and

- every conformal map of ∂U^n extends to a (unique) isometry of U^n.

(The same statements hold for the other models of H^n, and in fact they can be made without refernce to a specific model once we have an intrinsic description of the boundary ∂H^n.)

Finally, the first half of Proposition 1.5.3, stating that $\mathrm{Isom}^+(U^2) \simeq \mathrm{PSL}(2, \mathbb{R})$, follows by noting that $\mathrm{SL}(2, \mathbb{R})$ is the stabilizer of the real axis in $\mathrm{SL}(2, \mathbb{C})$ (acting again on $\mathbb{C}$ by fractional linear transformations).

Classification of isometries: In the course of the above proofs we have come across different types of conjugacy classes in $\mathrm{SL}(2, \mathbb{C})$, which have different geometric and dynamical properties and are accordingly separated into 3 types. An element $g \in \mathrm{PSL}(2, \mathbb{C}) \setminus \{\mathrm{Id}\}$ is called:

- *ellipitic* if it is conjugate to $\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$ for some θ,

- *parabolic* if it is conjugate to $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, and

- *loxodromic* if it is conjugate to $\begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}$ for some $\lambda \in \mathbb{C}$ with $|\lambda| \neq 1$.

(after possibly muliplying g by $-\mathrm{Id}$). As a special case, if g is loxodromic with $\lambda \in \mathbb{R}$, it is called *hyperbolic*. Note that these 3 cases can be differentiated by traces, namely:

- g is elliptic $\iff$ $\mathrm{Tr}(g) \in (-2,2)$,

- g is parabolic $\iff$ $\mathrm{Tr}(g) = \pm 2$, and

- g is loxodromic $\iff$ $\mathrm{Tr}(g) \notin [-2,2]$ (and hyperbolic $\iff$ $\mathrm{Tr}(g) \in \mathbb{R} \setminus [-2,2]$).

The more general definition (which we will see later, in all dimensions and for $\mathbb{K} = \mathbb{R}, \mathbb{C}$ or $\mathbb{H}$) is according to the number and location of fixed points, namely $g \in \mathrm{Isom}(\mathrm{H}^n_{\mathbb{K}})$ is called:

- *elliptic* if it has a fixed point in $\mathrm{H}^n_{\mathbb{K}}$,

- *parabolic* if it has no fixed point in $\mathrm{H}^n_{\mathbb{K}}$ and a single fixed point on $\partial \mathrm{H}^n_{\mathbb{K}}$, and

- *loxodromic* if it has no fixed point in $\mathrm{H}^n_{\mathbb{K}}$ and exactly 2 fixed points on $\partial \mathrm{H}^n_{\mathbb{K}}$.

1.5.5 Geodesics, totally geodesic subspaces and metric

From the decription of reflections in the ball model B^n and upper half-space model U^n as (restrictions of) inversions, we get the following description of totally geodesic subspaces:

Proposition 1.5.6 *The totally geodesic subspaces of U^n (resp. B^n) are the (intersections with U^n, resp. B^n of) generalized spheres orthogonal to ∂U^n (resp. B^n).*

In particular:

- geodesics in U^n are either vertical half-lines or half-circles orthogonal to ∂U^n,

- geodesics in B^n are either line segments through the origin or arcs of circles orthogonal to ∂B^n.

We now give an explicit formula for the Riemannian metric on B^n and U^n (in the infinitesimal quadratic form ds^2) which will be useful later. We denote as usual $||.||$ the Euclidean norm, and as before use coodinates (X,t) for U^n with $X \in \mathbb{R}^n$ and $t > 0$. This metric on the the unit ball B is called the *Poincaré metric*.

Proposition 1.5.7 (1) Ball model: *For $Y \in B^n$ and $v \in T_Y B^n$: $ds^2_Y(v) = \left(\frac{2}{1-||Y||^2} \right)^2 ||v||^2$.*

(2) Upper half-space model: *For $(X,t) \in U^n$ and $v \in T_{(X,t)} U^n$: $ds^2_{(X,t)}(v) = \frac{||v||^2}{t^2}$.*

Proof. The metrics on U^n and B^n are defined as pullbacks of the metric on H. The brute force computation can be a little unpleasant, but we will only compute the pullback at a single point (namely $O \in B^n$ and $(O, 1) \in U^n$) and move it around by using isometries.

(1) The pullback metric on B^n is defined by (in bilinear form notation):

$$(u, v)_Y = \langle d_Y \pi^{-1}(u), d_Y \pi^{-1}(v) \rangle_{\pi^{-1}(Y)},$$

where π was defined as follows:

$$\pi : \quad \begin{aligned} H &\longrightarrow B \\ (X, t) &\longmapsto \tfrac{X}{1+t}. \end{aligned}$$

Then: $d_{(O,1)}\pi = \frac{1}{2}\mathrm{Id}_{\mathbb{R}^n}$ (identifying $T_{(O,1)}H \simeq \mathbb{R}^n$ and $T_O B \simeq \mathbb{R}^n$), so $d_O \pi^{-1} = 2\mathrm{Id}_{\mathbb{R}^n}$ and $ds_O^2(v) = 4\|v\|^2$ for $v \in T_O B^n$.

Now, for $Y \neq O$ in B^n, consider the sphere $S = S\left(\frac{Y}{\|Y\|^2}, \left(\frac{1}{\|Y\|^2} - 1\right)^{1/2}\right)$. Then $S \perp \partial B$ (by the Pythagorean theorem, as the radius R of S satisifies $R^2 + 1 = \frac{1}{\|Y\|^2}$, the square of the distance between the 2 centers). Therefore i_S induces an isometry of B^n, and:

$$i_S(O) = \frac{Y}{\|Y\|^2} + \frac{(1/\|Y\|^2) - 1}{\|Y/\|Y\|^2\|^2} \cdot \frac{(-Y)}{\|Y\|^2} = Y.$$

Moreover, as we saw when we studied inversions, $d_O i_S = \frac{(1/\|Y\|^2) - 1}{\|Y\|^2/\|Y\|^4} r = (1 - \|Y\|^2)r$ for some (Euclidean) reflection r, giving as claimed:

$$ds_Y^2(v) = \left(\frac{2}{1 - \|Y\|^2}\right)^2 \|v\|^2.$$

(2) Recall that we passed from the ball model B^n to the upper half-space model U^n by the inversion i_S, where $S = S(-e_n, \sqrt{2})$. As above, $d_O i_S$ is of the form $2r$, with r a Euclidean reflection, so the metric at the point $i_S(O) = (O, 1) \in U^n$ is given by $ds_{(O,1)}^2(v) = \|v\|^2$ (i.e. it coincides with the Euclidean metric). Now dilations $D_\lambda : (X, t) \longmapsto (\lambda X, \lambda t)$ are isometries of U^n (because they are products of 2 inversions in concentric half-spheres orthogonal to ∂U^n), and $d_{(X,t)} D_\lambda = \lambda \mathrm{Id}$, so $ds_{(O,\lambda)}^2(v) = \frac{\|v\|^2}{\lambda^2}$. The result follows for all points (X, λ) by noting that horizontal translations are also isometries of U^n (as they are products of 2 reflections across hyperplanes orthogonal to ∂U^n). $\qquad \square$

1.6 The Klein model and Hilbert metric

We now briefly decribe another model for H^n with underlying set the unit ball in $\mathbb{R}^n$ but with a different metric from the Poincaré metric described earlier. This model is usually called the Klein (or Klein-Beltrami) model. It is again obtained from the hyperboloid by radial projection onto a hyperplane, but this time the radial projection is from the origin O, and the hyperplane is $\{x_{n+1} = 1\}$. Explicitly, consider the radial projection:

$$\pi' : \quad \begin{aligned} H &\longrightarrow B' \\ (x_1, .., x_{n+1}) &\longmapsto (x_1/x_{n+1}, ..., x_n/x_{n+1}, 1) \end{aligned}$$

Lemma 1.6.1 π' *induces a diffeomorphism from H to the unit ball $B' \subset \mathbb{R}^n \times \{1\}$.*

Proof. Note that:

$$
\begin{aligned}
(x_1, ..., x_{n+1}) \in H \quad &\Longleftrightarrow \quad x_1^2 + ... + x_n^2 - x_{n+1}^2 = -1 \\
&\Longleftrightarrow \quad x_1^2/x_{n+1}^2 + ... + x_n^2/x_{n+1}^2 = 1 - 1/x_{n+1}^2,
\end{aligned}
$$

therefore if $X \in H$ then $\pi'(X) \in B'$, and the inverse map is given by $g(y_1, ..., y_n) = (\lambda y_1, ..., \lambda y_n, \lambda)$ with λ chosen so that this lies in H, namely $\lambda = \left(\frac{1}{1-||Y||^2}\right)^{1/2}$. $\qquad\square$

We denote B''^n the Riemannian manifold B' with the pullback of the metric on H by π'^{-1}.

Proposition 1.6.1 *(1) The totally geodesic subspaces of B''^n are (intersections with B' of) affine subspaces of $\mathbb{R}^n$. In particular, geodesics are Euclidean line segments.*
(2) For $X, Y \in B'$, $d(X,Y) = \frac{1}{2}\ln[X,Y;P,Q]$ where P, Q are the endpoints on $\partial B'$ of the line segment connecting X and Y (so that P, X, Y, Q appear in that order), and $[X,Y;P,Q]$ denotes the cross-ratio of (affine coordinates of) X, Y, P, Q.

Recall that the cross-ratio of 4 points $z_1, ..., z_4 \in \widehat{\mathbb{C}}$ is defined by:

$$
[z_1, z_2; z_3, z_4] = \frac{z_4 - z_1}{z_4 - z_2} \cdot \frac{z_3 - z_2}{z_3 - z_1}.
$$

In fact the formula defining the distance in (2) makes sense for pairs of points in any convex subset of $\mathbb{R}^n$, and defines a distance function on any such set, called the *Hilbert metric*. Thus hyperbolic space (in the Klein model) is a special case of a convex subset of $\mathbb{R}^n$ with the Hilbert metric.

Proof. (1) From the geometric decription of π' it is clear that it sends linear subspaces of $\mathbb{R}^{n+1}$ to affine subspaces of the hyperplane $\{x_{n+1} = 1\}$; in more formal terms this follows because π' is a *collineation*, i.e. an element of $\mathrm{PGL}(n+1, \mathbb{R})$ acting by fractional linear transformations - in fact it is the image in $\mathrm{PGL}(n+1, \mathbb{R})$ of $\mathrm{Id} \in \mathrm{GL}(n+1, \mathbb{R})$.
(2) Normalize X and Y so that $X = (0, ..., 0)$ and $Y = (t, 0, ..., 0)$ for some $t > 0$. Then $P = (-1, 0, ..., 0)$ and $Q = (1, 0, ..., 0)$ so that $[X, Y; P, Q] = \frac{1+t}{1-t}$.

On the other hand, we lift X and Y to points $\tilde{X}$ and $\tilde{Y}$ in H and compute their distance there by the distance formula, Proposition (1.1.1). This gives: $\tilde{X} = (0, ..., 0, 1)$ and $\tilde{Y} = (\lambda t, 0, ..., 0, \lambda)$ with $\lambda = (1 - t^2)^{-1/2}$. Then:

$$
\langle \tilde{X}, \tilde{Y} \rangle = -(1 - t^2)^{-1/2} = -\frac{e^d + e^{-d}}{2}
$$

by the distance formula, with $d = d(\tilde{X}, \tilde{Y}) = d(X, Y)$. Write $y = e^d + e^{-d}$ and $x = e^d$; then by the quadratic formula: $x = y/2 + \sqrt{y^2/4 - 1} = (1 - t^2)^{-1/2} + t(1 - t^2)^{-1/2} = \left(\frac{1+t}{1-t}\right)^{1/2}$. $\qquad\square$

1.7 Curvature

We first give a brief qualitative overview of the negative curvature properties of hyperbolic space. We will then define these properties more precisely/quantitatively and prove them. Hyperbolic space H^n (with $n \geqslant 2$)) has *negative curvature* in any of the following senses:

- **(a) Local Riemannian**: H^n has negative sectional curvature; this is the most classical meaning of (negative) curvature. In fact, H^n has constant negative sectional curvature κ (equal to -1); here constant means both constant at all points and in all (pairs of) directions. Moreover, H^n is the *model space* of constant negative curvature in the sense that it is the unique simply-connected (complete) Riemannian manifold of constant curvature -1. (Likewise, Euclidean space E^n and the sphere S^n with its standard round metric are the model spaces of constant null, respectively positive sectional curvature).

- **(b) Global Riemannian**: The angle sum in geodesic triangles is $< \pi$. In fact, we will see that any geodesic triangle T with internal angles α, β, γ (is contained in a copy of H^2 and) satisfies:
$$\text{Area}(T) = \pi - (\alpha + \beta + \gamma). \tag{1.7.1}$$
We will see that one can recover from this property the fact that $\kappa \equiv -1$.

- **(c) Metric**: Distance functions are strictly convex. More precisely, this means that, given a point P and a geodesic $\gamma(t)$ not containing P (resp. 2 geodesics $\gamma_1(t)$, $\gamma_2(t)$), parametrized with unit speed, (1) the function $t \mapsto d(P, \gamma(t))$ is strictly convex and (2) the function $t \mapsto d(\gamma_1(t), \gamma_2(t))$ is strictly convex. In fact these properties are shared by Euclidean space, except (2) in the case of intersecting or parallel lines (in which case the distance function is linear).

In fact there are more general notions of negative (or non-positive) curvature for metric spaces which are modelled on the properties of H^n. These properties have played an increasingly important role in geometric group theory since Gromov introduced them in the 1980's (following a long line of ideas from people like E. Cartan, A.D. Alexandrov and V.A. Toponogov - hence the letters CAT).

- **CAT(κ) spaces:** A metric space is said to be CAT(κ) if its triangles are at least as thin as in the correponding model space of constant curvature κ (the most common cases are $\kappa = 0$ and -1). More precisely, given a geodesic triangle T with vertices x, y, z in a (geodesic) metric space X, we define a *comparison triangle* to be any geodesic triangle with the same side-lengths as T in X', the model space of constant curvature κ (this requires additional conditions on the 3 side-lengths if $\kappa > 0$). One then compares the thinness of the 2 triangles by comparing the distances between pairs of points along their various sides. More precisely, given a vertex and the pair of sides of T emanating from that vertex, parametrized with unit speed from the vertex as $\gamma_1(t_1)$ and $\gamma_2(t_2)$, with $\gamma_1'(t_1)$, $\gamma_2'(t_2)$ the unit-speed parametrizations of the 2 corresponding sides of T' in X', we say that T is at least as thin as T' along that pair of sides if $d(\gamma_1(t_1), \gamma_2(t_2)) \leqslant d'(\gamma_1'(t_1), \gamma_2'(t_2))$ for all t_1, t_2 in the appopriate intervals. We say that T is at least as thin as T' if this holds for all 3 pairs of sides, or that T satisfies the CAT(κ) inequality. Finally, we say that X is a CAT(κ) space if all of its geodesic triangles satisfy the CAT(κ) inequality.

- **δ-hyperbolic spaces:** A metric space is said to be δ-hyperbolic (for some fixed $\delta > 0$) if its geodesic triangles are uniformly δ- thin. This means that, in any geodesic triangle, each side is entirely contained in a δ-neighborhood of the union of the 2 other sides.

Examples: Hyperbolic space H^n is both CAT(-1) (by definition) and δ-hyperbolic for some $\delta > 0$ (Exercise: prove this and in particular compute a suitable δ). Trees are CAT(κ) for any $\kappa \leqslant 0$ and are 0-hyperbolic (triangles are tripods). Further examples include Cayley graphs of certain finitely generated groups - e.g. free groups corresponding to trees - and higher-dimensional cell complexes such as CAT(0) cube complexes.

We now prove properties (a), (b) and (c) announced above, starting with (c):

(c) Convexity of distance functions:

Recall that a function $f : \mathbb{R} \longmapsto \mathbb{R}$ is convex $\iff f(\frac{t_1+t_2}{2}) \leqslant \frac{f(t_1)+f(t_2)}{2}$ for all $t_1, t_2 \in \mathbb{R}$. We claimed above that, given a point $O \in H^n$ and a geodesic $\gamma(t)$ not containing O (parametrized with unit speed), the function $t \mapsto d(O, \gamma(t))$ is convex. This can be restated as follows, denoting $P_1 = \gamma(t_1)$, $P_2 = \gamma(t_2)$ and $M = \gamma(\frac{t_1+t_2}{2})$ (the condition about unit speed guarantees that M is the midpoint of $[P_1 P_2]$ - constant speed would suffice).

Lemma 1.7.1 *Let O, P_1, P_2 be 3 points in H^n, and M the midpoint of $[P_1 P_2]$. Then: $2d(O, M) < d(O, P_1) + d(O, P_2)$.*

Proof. Let s denote the central involution at M. Then $s(P_1) = P_2$; denote $O' = s(O)$. By the triangle inequality: $2d(O, M) = d(O, O') < d(O, P_1) + d(P_1, O') = d(O, P_1) + d(O, P_2)$. $\qquad\square$

The second convexity property of distance functions is the following:

Proposition 1.7.1 *Let $\gamma_1(t)$, $\gamma_2(t)$ be 2 distinct geodesics in H^n, parametrized with unit speed. Then the function $t \mapsto d(\gamma_1(t), \gamma_2(t))$ is strictly convex.*

Proof. Let $t_1 \neq t_2 \in \mathbb{R}$, $P_1 = \gamma_1(t_1)$, $Q_1 = \gamma_1(t_2)$, $M_1 = \gamma_1(\frac{t_1+t_2}{2})$, and likewise $P_2 = \gamma_2(t_1)$, $Q_2 = \gamma_2(t_2)$, $M_2 = \gamma_2(\frac{t_1+t_2}{2})$. The goal is to show that:

$$2d(M_1, M_2) < d(P_1, P_2) + d(Q_1, Q_2).$$

Consider $g = s_{M_2} \circ s_{M_1}$ (where s_x denotes the central involution at x), and let $M = g(M_1) = s_{M_2}(M_1)$. Then M_2 is the midpoint of $[M_1 M]$, so $d(M_1, M) = 2d(M_1, M_2)$. Let $P = g(P_1) = s_{M_2}(Q_1)$; then:

$$d(P_1, P) \leqslant d(P_1, P_2) + d(P_2, P) = d(P_1, P_2) + d(Q_1, Q_2),$$

so it suffices to show that $d(M_1, M) < d(P_1, P)$, in other words: $d(M_1, g(M_1)) < d(P_1, g(P_1))$ (geometrically, this says that points on the axis of g are moved a smaller distance than points off the axis). We pass to the upper half-space model U^n to show the latter inequality, normalizing so that M_1, M_2 lie on the vertical axis $\{O\} \times \mathbb{R}^+$. Then $g = s_{M_2} \circ s_{M_1}$ is a dilations D_λ for some $\lambda > 0$ (as can be seen by writing each of s_{M_1}, s_{M_2} as a product of 3 inversions/reflections, 2 of them across orthogonal vertical hyperplanes being common to both products, and the third in each product being an inversions in a sphere centered at $O \in \partial U^n$). By assumption, P_1 does not lie on the vertical axis, and the distance $d(P_1, g(P_1))$ is unchanged by dilations so we may

asume that P_1 is at the same height (the second factor in $U^n = \mathbb{R}^n \times \mathbb{R}^+$) as M_1, from which it follows that $g(P_1)$ is also at the same height as $g(M_1)$. It then follows from the explicit form of the metric in U^n that $d(P_1, g(P_1)) > d(M_1, g(M_1))$ because $\int(dx_1^2 + ... + dx_n^2) > \int dx_n^2$ along any non-vertical curve. $\qquad \square$

(a) Riemannian sectional curvature I:

We will not define the Riemannian sectional curvature $\kappa_x(X, Y)$ at a point $x \in H^n$ in the direction of the 2-plane in $T_x H^n$ spanned by X, Y, but we will use 2 of its characteristic properties (see below).

Proposition 1.7.2 *The sectional curvature at any point $x \in H^n$ in the direction of any 2-plane in $T_x H^n$ is -1.*

Proof. First observe that $\kappa_x(X, Y)$ is independent of the point x and the direction $\mathrm{Span}(X, Y)$, as $\mathrm{Isom}(H^n)$ acts transitively on H^n and doubly transitively on orthonormal pairs $(X, Y) \in (T_x H^n)^2$, because $O(n)$ ($\simeq \mathrm{Stab}(x)$) acts transitively on orthogonal pairs of vectors in the unit sphere in $\mathbb{R}^n$.

We now compute the value of κ at the point O in the ball model of H^2, using the classical fact (see e.g. [GHL]) that if Ω is a domain in $\mathbb{R}^2$ with metric of the form $ds_x^2 = \alpha(x)^2 ds_{E^2}^2$, then the sectional curvature of Ω is given by: $\kappa(x) = -\frac{1}{\alpha(x)^2}\left(\Delta \log \alpha\right)(x)$.

In the ball model B^2, we've seen that: $\alpha(x) = \frac{2}{1 - ||x||^2}$ so $\alpha(O) = 4$, $\Delta \log \alpha(O) = 4$ and $\kappa(O) = -1$. $\qquad \square$

(b) Angle-sum in geodesic triangles: First observe the following:

Lemma 1.7.2 *Any m points in H^n (with $m \leqslant n$) belong to an isometrically embedded copy of H^m.*

Proof. This is clear in the description of totally geodesic subspaces in the hyperboloid model, as any m points in $\mathbb{R}^{n+1}$ belong to a linear subspace of dimension m. $\qquad \square$

In particular, any 3 points (and the geodesics connecting them) lie in a copy of H^2. Therefore we only need to analyze gedesic triangles in H^2:

Proposition 1.7.3 *Any geodesic triangle T in H^2 with internal angles α, β, γ satisfies:*

$$\mathrm{Area}(T) = \pi - (\alpha + \beta + \gamma).$$

In particular, $\alpha + \beta + \gamma < \pi$.

Proof. Consider the upper half-plane model U^2. Note that this is a conformal model (the metric is a multiple of the Euclidean one), so the angles we "see" are the intrinsic (Riemannian) angles betwen geodesic rays.

23

First assume that T has a vertex at ∞; we may assume (after maybe applying an isometry of U^2) that the other 2 vertices lie on the unit cicle, say $P = (\cos(\pi - \alpha), \sin(\pi - \alpha))$ and $Q = (\cos(\beta), \sin(\beta)$ with $\beta < \pi - \alpha$. Then:

$$T = \{(\cos(\theta), y) \mid \beta \leqslant \theta \leqslant \pi - \alpha, \ y \geqslant \sin(\theta)\},$$

so that:

$$\text{Area}(T) = \iint_T \frac{dxdy}{y^2} = \int_\beta^{\pi-\alpha} \sin\theta d\theta \cdot \int_{\sin\theta}^\infty \frac{dy}{y^2} = \int_\beta^{\pi-\alpha} d\theta = \pi - \alpha - \beta.$$

Now, if T has all 3 vertices P, Q, R in H^2, we "subdivide" to express T as an algebraic sum of triangles with a vertex at ∞ as follows:

$$\begin{aligned}
\text{Area}(T) &= \text{Area}(PQ\infty) + \text{Area}(QR\infty) - \text{Area}(PR\infty) \\
&= (\pi - \alpha - \alpha' - \beta_1) + (\pi - \beta_2 - \gamma - \gamma') - (\pi - \alpha' - \gamma') \\
&= \pi - \alpha - \beta - \gamma.
\end{aligned}$$

$\square$

(a') Riemannian sectional curvature II: (by parallel transport around boundaries of triangles)

It can be shown (see e.g. [GHL]) that on an oriented Riemannian surface S, for any embedded closed disk with piecewise linear boundary:

$$\int_D \kappa(x) dA = (v, \widehat{T_{\partial D}(v)})$$

for any $p \in \partial D$ and $v \in T_p S$, where $T_{\partial D}$ denotes parallel transport along ∂D, counterclockwise. Denote by $\phi(\partial D)$ the angle on the right-hand side (which is, by this fact, independent of p and v). By this formula, given a point $x \in S$, if (D_n) is a decreasing sequence of disks as above, with $x \in \overset{\circ}{D}_n$ for all n and $\text{Diam}(D_n) \longrightarrow 0$, then: $\kappa(x) = \lim \frac{\phi(\partial D_n)}{Area(D_n)}$. Using a decreasing sequence of triangles T_n, the formula for the area of a triangle gives: $\kappa(x) = -1$. Indeed, parallel transport along the geodesic sides does not contribute to the angle $\phi(\partial T_n)$; only the external angles at the vertices do, giving: $\phi(\partial T_n) = \alpha_n + \beta_n + \gamma_n - \pi$.

1.8   Boundary at infinity

Recall that a *geodesic ray* is a half-infinite geodesic line; a geodesic ray γ is entirely determined by a point $p \in H^n$ and a unit tangent vetor $v \in T_p H^n$, namely γ is then parametrized with unit speed by $\gamma(t) = \exp_p(tv)$ for $t \geqslant 0$.

Let $\mathcal{R} = \{$geodesic rays in H^n parametrized with unit speed$\}$, and define an equivalence relation $\sim$ on $\mathcal{R}$ by:

$$\gamma_1 \sim \gamma_2 \text{ if } \underset{t \in \mathbb{R}^+}{\text{Sup}}\, d(\gamma_1(t), \gamma_2(t)) < \infty.$$

Two rays equivalent under this relation are called *asymptotic*. The (Gromov) boundary ∂H^n of H^n is then defined as $\mathcal{R}/\sim$. We denote by $[\gamma]_\sim$ the equivalence class of $\gamma \in \mathcal{R}$ under $\sim$.

We then define a topology on $\overline{H^n} = H^n \cup \partial H^n$ by requiring that the inclusion $H^n \hookrightarrow \overline{H^n}$ is a homeomorphism onto its image, and that a basis of neighborhoods for any point $p \in \partial H^n$ is of the form $(U(x, V, r))_{(x,V,r)}$ where each $U(x, V, r)$ is defined, given a choice of representative γ (such that $p = [\gamma]_\sim$), $x = \gamma(0)$, V a neighborhood of $\gamma'(0)$ in $UT_x H^n$, and $r > 0$ by:

$$U(x, V, r) = \bigcup \gamma((r, \infty)) \cup \bigcup [\gamma]_\sim,$$

where each union is taken over rays γ such that $\gamma(0) = x$ and $\gamma'(0) \in V$. (The verification that this family forms a neighborhood basis is straightforward and is left to the reader).

This definition of the boundary is intrinsic, i.e. it only depends on the Riemannian manifold H^n and not on a particular choice of model; we now shows that it agrees with the obvious boundary in the ball model (and hence also in the upper half-space model). We denote as before the unit ball in $\mathbb{R}^n$ by B.

Proposition 1.8.1 ∂H^n *is homeomorphic to* $\partial B = S^{n-1}$, *and* $\overline{H^n}$ *is homeomorphic to* $\overline{B}$.

This will easily follow from the following description of asymptotic rays in the upper half-space model. Given a geodesic ray γ in U^n, we denote by $\gamma(\infty)$ the endpoint on $\partial U^n = (\mathbb{R}^{n-1} \times \{0\}) \cup \{\infty\}$ of the line segment/arc of circle γ.

Lemma 1.8.1 *Let* γ_1, γ_2 *be 2 geodesic rays in* U^n. *Then:* $\gamma_1 \sim \gamma_2 \iff \gamma_1(\infty) = \gamma_2(\infty)$.

Proof. First assume that $\gamma_1(\infty) = \gamma_2(\infty)$ and normalize this common endpoint to be ∞. Then γ_1, γ_2 are 2 vertical half-lines, with unit speed parametrizations of the form $\gamma_1(t) = (X_1, e^{t-t_1})$ (with $t \geqslant 0$) and $\gamma_2(t) = (X_2, e^{t-t_2})$ (with $t \geqslant 0$), for some $X_1, X_2 \in \mathbb{R}^{n-1}$ and $t_1, t_2 \in \mathbb{R}$. Then, for all $t \geqslant 0$:

$$\begin{aligned}
d\big(\gamma_1(t), \gamma_2(t)\big) &\leqslant d\big((X_1, e^{t-t_1}), (X_1, e^{t-t_2})\big) + d\big((X_1, e^{t-t_2}), (X_2, e^{t-t_2})\big) \\
&\leqslant |t_1 - t_2| + d\big((X_1, e^{-t_2}), (X_2, e^{-t_2})\big)
\end{aligned}$$

which is a uniform bound, independent of t.

Conversely, assume that $\gamma_1(\infty) \neq \gamma_2(\infty)$, and normalize so that $\gamma_1(\infty) = \infty$. Then γ_1 is a vertical half-line as above, and γ_2 is an arc of circle, terminating on $\mathbb{R}^{n-1} \times \{0\}$, whose unit-speed parametrizations are of the form: $\gamma_1(t) = (X_1, f_1(t))$ and $\gamma_2(t) = (X_2(t), f_2(t))$ where f_1 increases to ∞ and f_2 eventually decreases to 0. Therefore, for large enough t:

$$d\big(\gamma_1(t), \gamma_2(t)\big) \geqslant \int_{f_2(t)}^{f_1(t)} \frac{dx_{n+1}}{x_{n+1}} \xrightarrow[t \to \infty]{} \infty.$$

$\square$

Proof of Proposition 1.8.1. By the lemma, the map:

$$\begin{aligned}
\phi : \partial H^n &\longrightarrow \partial B \\
[\gamma]_\sim &\longmapsto \gamma(\infty)
\end{aligned}$$

(where we've identified $\partial B \simeq \partial U$) is well-defined and a bijection. It induces a homeomorphism from $\overline{H^n}$ to $\overline{B}$, because the basis of neighborhoods of any point $p \in \partial H^n$ is sent by ϕ to a basis of neighborhoods of $\phi(p)$ in ∂B. $\square$

1.9 Complex hyperbolic space

We now describe complex hyperbolic space $H^n_{\mathbb{C}}$ in the projective model. Start with the Hermitian vector space $\mathbb{C}^{n,1}$ which is $\mathbb{C}^{n+1}$ endowed with a Hermitian form of signature $(n,1)$, for example the standard form:

$$\langle X, Y \rangle = x_1\overline{y_1} + ... + x_n\overline{y_n} - x_{n+1}\overline{y_{n+1}}, \tag{1.9.1}$$

where: $X = (x_1, ..., x_{n+1})^T$ and $Y = (y_1, ..., y_{n+1})^T$.

Let $V_- = \{X \in \mathbb{C}^{n,1} \,|\, \langle X, X \rangle < 0\}$ and $V_0 = \{X \in \mathbb{C}^{n,1} \,|\, \langle X, X \rangle = 0\}$. As a set, we define: $H^n_{\mathbb{C}} = \pi(V_-)$ where $\pi : \mathbb{C}^{n+1} \setminus \{0\} \longrightarrow \mathbb{C}P^n$ denotes projectivization (then we will have as before $\partial H^n_{\mathbb{C}} = \pi(V_0)$). For example, taking the standard Hermitian as above, this gives in the affine chart $\{[z_1 : ... : z_{n+1}] \,|\, z_{n+1} \neq 0\}$ of $\mathbb{C}P^n$:

$$H^n_{\mathbb{C}} = \big\{(z_1, ..., z_n, 1) \,\big|\, |z_1|^2 + ... + |z_n|^2 < 1\big\},$$

the unit ball $B^n_{\mathbb{C}}$ in $\mathbb{C}^n$. Note that we cannot directly use the hyperboloid:

$$H = \big\{X \in \mathbb{C}^{n+1} \,\big|\, \langle X, X \rangle = -1\big\}$$

as a model for $H^n_{\mathbb{C}}$ as it is not a complex manifold (it has real dimension $2n + 1$). However, we may still use H to define a (Riemannian or Hermitian) metric on $H^n_{\mathbb{C}}$ as follows.

Note that $\pi_{|H} : H \longrightarrow H^n_{\mathbb{C}}$ is a submersion (in fact it is a fiber bundle, induced from the so-called tautological bundle $\mathbb{C}^* \longrightarrow \mathbb{C}^{n+1} \setminus \{0\} \xrightarrow{\pi} \mathbb{C}P^n$), and that for $X \in H$:

$$T_X H = \big\{Y \in \mathbb{C}^{n,1} \,\big|\, \mathrm{Re}(\langle X, Y \rangle) = 0\big\} \supset X^\perp$$

(this was an equality in the real case).

Lemma 1.9.1 *The subspace $X^\perp$ of $T_X H$ is transverse to the fiber of π at X, therefore $d_X \pi_{|H}$ induces an isomorphism $X^\perp \xrightarrow{\sim} T_{\pi(X)} H^n_{\mathbb{C}}$.*

Proof. Denote $F = \pi^{-1}(\pi(X))$ the fiber of π at X. Then: $F = \mathbb{C}^* X$ so $T_X F \simeq \mathbb{C}X$. Writing, for $\epsilon \in \mathbb{C}$: $\langle (1 + \epsilon)X, (1 + \epsilon)X \rangle = (1 + \epsilon)(1 + \bar{\epsilon})\langle X, X \rangle = -1 - 2\mathrm{Re}(\epsilon) - |\epsilon|^2$, we see that $T_X F \cap T_X H = i\mathbb{R}X$ and $T_X F \cap X^\perp = \{0\}$. $\square$

Now, as before $X^\perp$ is equipped with a positive definite Hermitian form, namely the restriction of the ambient form $\langle ., . \rangle_{X^\perp \times X^\perp}$ (which induces a positive definite quadratice form, $\mathrm{Re}\langle ., . \rangle_{X^\perp \times X^\perp}$). By the lemma these forms carry over to $T_{\pi(X)} H^n_{\mathbb{C}}$, inducing a Hermitian metric (and a Riemannian metric) on $H^n_{\mathbb{C}}$. (Concretely, given $x \in H^n_{\mathbb{C}}$, lift x to $X \in \mathbb{C}^{n,1}$ with $\langle X, X \rangle = -1$; then $T_x H^n_{\mathbb{C}} \simeq X^\perp$ and the Hermitian, resp. Riemannian, metric at x is given by $\langle ., . \rangle_{X^\perp \times X^\perp}$, resp. $\mathrm{Re}\langle ., . \rangle_{X^\perp \times X^\perp}$.)

As in the real hyperbolic case, since the metric is entirely defined in terms of the form $\langle ., . \rangle$, the group:

$$U(n, 1) = \big\{A \in \mathrm{GL}(n + 1, \mathbb{C}) \,\big|\, \langle AX, AY \rangle = \langle X, Y \rangle \ (\forall X, Y \in \mathbb{C}^{n,1})\big\}$$

acts by (holomorphic) isometries on $H^n_{\mathbb{C}}$, i.e. $\mathrm{PU}(n, 1) \subset \mathrm{Isom}^+(H^n_{\mathbb{C}})$, where $\mathrm{PU}(n, 1) = U(n, 1)/\{\lambda \mathrm{Id} \,|\, \lambda \in U(1)\}$ and $\mathrm{Isom}^+(H^n_{\mathbb{C}})$ now denotes the subgroup of holomorphic isometries

(not to be confused with orientation-preserving – for even n all isometries of $H^n_{\mathbb{C}}$ preserve orientation, whereas for odd n orientation-preserving isometries coincide with holomorphic isometries).

In fact it can be shown that $\mathrm{Isom}^+(H^n_{\mathbb{C}}) = \mathrm{PU}(n,1)$, and the full isometry group $\mathrm{Isom}(H^n_{\mathbb{C}})$ is generated by $\mathrm{PU}(n,1)$ and a single antiholomorphic isometry, such as the standard *real reflection* $\sigma_0 : (z_1, ..., z_n) \longmapsto (\overline{z_1}, ..., \overline{z_n})$ in ball coordinates.

Analogously to the real case (Proposition 1.2.1), the isometries of $H^n_{\mathbb{C}}$ given by $\mathrm{PU}(n,1)$ are enough to show that $H^n_{\mathbb{C}}$ enjoys the following properties:

Proposition 1.9.1 *(1)* $\mathrm{U}(n,1)$ *acts transitively on* V_-, *in particular* $H^n_{\mathbb{C}}$ *is homogeneous.*
(2) The stabilizer in $\mathrm{PU}(n,1)$ *of any point of* $H^n_{\mathbb{C}}$ *is isomorphic to* $U(n)$.
(3) $\mathrm{PU}(n,1)$ *acts transitively on* $\mathrm{UTH}^n_{\mathbb{C}}$.
(4) $H^n_{\mathbb{C}}$ *is a symmetric space.*

The proof is analogous to the real case and is left to the reader. Note that a more precise statement of (2) is that the stabilizer of $\mathbb{C}e_{n+1}$ in $\mathrm{U}(n,1)$ is $\mathrm{U}(n) \times \mathrm{U}(1)$, so that the stabilizer of $O = \pi(e_{n+1})$ (the origin in the ball model) in $\mathrm{PU}(n,1)$ is $\mathrm{P}\big(\mathrm{U}(n) \times \mathrm{U}(1)\big) \simeq \mathrm{U}(n)$.

The main difference from the real case is that now $\mathrm{U}(n)$ acts transitively on the unit sphere in $\mathbb{C}^n \simeq \mathbb{R}^{2n}$, but not transitively on pairs of orthogonal unit vectors (contrary to $\mathrm{O}(n)$). Geometrically, this has several manifestations in $H^n_{\mathbb{C}}$, for example in the fact that sectional curvature is *not* constant in $H^n_{\mathbb{C}}$ (all unit tangent vectors are equivalent, but not all directions spanned by 2 unit tangent vectors); one can show that it is in fact pinched between the constants -1 (in the directions of complex projective lines) and $-1/4$ (in the directions of totally real planes – see below). A similar phenomenon appears more globally, in the fact that triangles in $H^n_{\mathbb{C}}$ (triples of points and the geodesic segments connecting them) are not in general contained in a 2-dimensional totally geodesic subspace (as they are in H^n, where every triangle is contained in a copy of H^2). What remains true is that any triangle in $H^n_{\mathbb{C}}$ is conatined in a copy of $H^2_{\mathbb{C}}$, but this has (real) dimension 4.

Note that, as a consequence of part (2) of the proposition, if an element $g \in \mathrm{PU}(n,1)$ is *elliptic* (i.e. has a fixed point in $H^n_{\mathbb{C}}$) then any of its lifts to $\mathrm{U}(n,1)$ is diagonalizable, with unit modulus eigenvalues. The following are special kinds of elliptic isometries:

- *Central involutions* are elements of $\mathrm{PU}(n,1)$ with a lift conjugate to the diagonal matrix $\mathrm{Diag}(-1, ..., -1, 1) \in \mathrm{U}(n,1)$.

- More generally, *complex reflections in a point* are elements of $\mathrm{PU}(n,1)$ with a lift conjugate to a diagonal matrix of the form $\mathrm{Diag}(e^{i\theta}, ..., e^{i\theta}, 1) \in \mathrm{U}(n,1)$ for some $\theta \in \mathbb{R}$.

- *Complex reflections* are elements of $\mathrm{PU}(n,1)$ with a lift conjugate to a diagonal matrix of the form $\mathrm{Diag}(e^{i\theta}, 1, ..., 1) \in \mathrm{U}(n,1)$.

Note that these special isometries all act on $\mathbb{C}^{n,1}$ according to the following formula:

$$X \longmapsto X + (e^{i\theta-1})\frac{\langle X, e\rangle}{\langle e, e\rangle}e, \tag{1.9.2}$$

for some $e \in \mathbb{C}^{n,1}$ with $\langle e, e\rangle \neq 0$, reminiscent of the formula in Equation (1.3.2). Here, when $\langle e, e\rangle < 0$ this formula gives the complex reflection in the point $\pi(e)$ through angle $-\theta$; when $\langle e, e\rangle > 0$ the formula gives the complex reflection across the (complex) hyperplane $e^{\perp}$, through angle θ.

1.10 Totally geodesic subspaces of $\mathrm{H}_{\mathbb{C}}^n$

In this section we determine the totally geodesic submanifolds of complex hyperbolic space $\mathrm{H}_{\mathbb{C}}^n$. As in the real hyperbolic case these will be projective images of linear subspaces of the Hermitian vector space $\mathbb{C}^{n,1}$, but the situation is more subtle due to the interaction between the real and complex structures on this space. In order to state the main result we need a few definitions.

Definitions: An $\mathbb{R}$-linear subspace W of $\mathbb{C}^{n,1}$ is called *totally real* if $\langle X, Y \rangle \in \mathbb{R}$ for all $X, Y \in W$ (in other words, if the restriction of $\langle ., . \rangle$ to $W \times W$ is a quadratic form). A $\mathbb{C}$-linear (resp. totally real) subspace W of $\mathbb{C}^{n,1}$ is called *hyperbolic* if the restriction of $\langle ., . \rangle$ to $W \times W$ has signature $(k, 1)$ for some $k \geqslant 1$; *elliptic* if it is positive definite; *parabolic* if it is positive semidefinite (but not definite). Note that if W is a hyperbolic $\mathbb{C}$-linear subspace of $\mathbb{C}^{n,1}$ with $\dim_{\mathbb{C}} W = k + 1$ (resp. a hyperbolic totally real subspace with $\dim_{\mathbb{R}} W = k + 1$), then $\pi(W \cap V_-) \subset \mathrm{H}_{\mathbb{C}}^n$ is a copy of $\mathrm{H}_{\mathbb{C}}^k$ (resp. $\mathrm{H}_{\mathbb{R}}^k$) – in fact we will see that these copies are isometrically embedded. Such subspaces $\pi(W \cap V_-)$ are called $\mathbb{C}^k$-*planes*, resp. $\mathbb{R}^k$-*planes*. The main result is the following:

Theorem 1.10.1 *The totally geodesic submanifolds of $\mathrm{H}_{\mathbb{C}}^n$ are exactly the $\mathbb{C}^k$-planes and $\mathbb{R}^k$-planes with $0 \leqslant k \leqslant n$.*

This has the following consequence, which in the context of discrete groups of isometries makes the construction of fundamental polyhedra more interesting, but alo more challenging than in constant curvature:

Corollary 1.10.2 *There do not exist totally geodesic real hypersurfaces in $\mathrm{H}_{\mathbb{C}}^n$ when $n \geqslant 2$.*

We will prove Theorem 1.10.1 in several steps, the main tool being the following:

Lemma 1.10.1 *Let X be a symmetric space, and $M \subset X$ a submanifold. Then, denoting s_x the central symmetry at a $x \in X$: M is totally geodesic $\iff s_x(M) = M$ for all $x \in M$.*

Recall that M is *totally geodesic* if it contains every geodesic line which is tangent to it (in particular, totally geodesic submanifolds are geodesically complete).

Proof. The direction $\Rightarrow$ follows from the definition of the central symmetry s_x. More explicitly, assume that M is totally geodesic in X and let $x \in M$. Then, for any geodesic line γ through x and tangent to M at x, $s_x(\gamma) = \gamma$ is contained in M by hypothesis. Therefore $s_x(M)$ is contained in the union of such geodesic lines γ, which is contained in M. But $s_x(M)$ is a geodesically complete submanifold of X of the same dimension as M, therefore $s_x(M) = M$.

Conversely, assume that $s_x(M) = M$ for all $x \in M$. We use the equivalent formulation that M is totally geodesic if and only if, locally, geodesic segments connecting pairs of points in M are in M. Let $x_0 \in M$ and let N be a neighborhood of x_0 in X such that any 2 points of N are connected by a unique geodesic segment in X; likewise, let N_1 be a neighborhood of x_0 in M such that any 2 points of N_1 are connected by a unique geodesic segment in M. We may assume that $N_1 \subset N$ (after shrinking it if necessary).

Let $p, q \in N_1$, and denote by $\alpha(t)$ with $t \in [0, a]$ (resp. $\beta(t)$ with $t \in [0, b]$) the unique geodesic segment in X (resp. M) connecting p to q, parametrized with unit speed. Let $m =$

$\beta(b/2) \in M$ be the midpoint of the geodesic segment β; then s_m preserves β (as it is an isometry preserving M) and exchanges p and q. Since s_m exchanges p and q, it also preserves α and therefore also fixes the midpoint $m' = \alpha(a/2)$ of α. But central symmetries have an isolated fixed point, so (after shrinking N if necessary) $m = m'$, i.e. $\alpha(a/2) = \beta(b/2)$. By iterating this process, we get that $\alpha(ta) = \beta(tb)$ for any t of the form $k/2^n$ (for some $n \in \mathbb{N}$ and $1 \leqslant k \leqslant 2^n - 1$). But such points are dense in $[0,1]$, therefore by continuity $\alpha(ta) = \beta(tb)$ for all $t \in [0,1]$, giving ($a = b$ and) $\alpha = \beta$. $\qquad \square$

We now state and prove the easier half of Theorem 1.10.1:

Proposition 1.10.1 *If W is a hyperbolic $\mathbb{C}$-linear or totally real subspace of $\mathbb{C}^{n,1}$, then $\pi(W \cap V_-)$ is a totally geodesic subspace of $\mathrm{H}^n_{\mathbb{C}}$.*

Proof. Let $M = \pi(W \cap V_-)$ and $p \in M$; lift p to $P \in W$ with $\langle P, P \rangle = -1$. Then the central symmetry s_p lifts to $\mathrm{U}(n,1)$ as $S_P : X \longmapsto X - 2\langle X, P \rangle P$, which preserves W (because W is $\mathbb{C}$-linear or totally real). Therefore $s_p(M) = M$, and by Lemma 1.10.1 M is totally geodesic. $\square$

Before proving the second half of Theorem 1.10.1 we use this to give a concrete description of geodesics in $\mathrm{H}^n_{\mathbb{C}}$:

Proposition 1.10.2 *(1) Geodesics through the origin O in the ball model are exactly straight line segments.*
(2) $\mathrm{PU}(n,1)$ acts transitively on geodesics in $\mathrm{H}^n_{\mathbb{C}}$.
(3) There exists a unique geodesic connecting any 2 distinct points in $\mathrm{H}^n_{\mathbb{C}}$.
(4) Given 2 distinct points $p, q \in \mathrm{H}^n_{\mathbb{C}}$ with lifts P, Q to $\mathbb{C}^{n,1}$ normalized so that $\langle P, Q \rangle \in \mathbb{R}$, the geodesic line (pq) containing p and q is $\pi\big(\mathrm{Span}_{\mathbb{R}}(P,Q) \cap V_-\big)$.

Note: In practice, part (4) allows us to parametrize $(pq) \setminus \{q\}$ by $\pi(P + tQ)$ (with t in the interval giving $\langle P + tQ, P + tQ \rangle < 0$), and likewise $(pq) \setminus \{p\}$ by $\pi(tP + Q)$. These are however not unit speed parametrizations. The most convenient way to parametrize a unit speed geodesic in $\mathrm{H}^n_{\mathbb{C}}$ is to use its endpoints $p_\infty, q_\infty \in \partial \mathrm{H}^n_{\mathbb{C}}$ (or, as in the real hyperbolic case, paramatrize a standard geodesic then carry it over by an isometry). Namely, it can easily be checked that given lifts $P_\infty, Q_\infty \in V_0$, normalized so that $\langle P_\infty, Q_\infty \rangle = -1/2$, the following is a unit speed parametrization of the geodesic line connecting p_∞ and q_∞:

$$\gamma(t) = e^{-t}P_\infty + e^t Q_\infty.$$

Proof. (1) Take $W = \mathrm{Span}_{\mathbb{R}}\big((1,0,...,0,1),(-1,0,...,0,1)\big) \subset \mathbb{C}^{n,1}$. This is a hyperbolic totally real subspace of $\mathbb{C}^{n,1}$, of real dimension 2, so by Proposition 1.10.1 $\gamma_0 = \pi(W \cap V_-)$ is a totally geodesic subspace of dimension 1, i.e. a geodesic line. Now any straight line through O is in the orbit of γ_0 under $\mathrm{Stab}_{\mathrm{PU}(n,1)}(O)$ because, as we've seen, $\mathrm{Stab}_{\mathrm{PU}(n,1)}(O)$ acts transitively on unit tangent vectors at O. Conversely any geodesic through O is a straight line, as there can only be one geodesic through a given point with a given tangent vector.
(2) and (3) follow from transitivity of the action of $\mathrm{PU}(n,1)$ on $\mathrm{H}^n_{\mathbb{C}}$.
(4) Any geodesic line γ in $\mathrm{H}^n_{\mathbb{C}}$ is the intersection of a complex line and a totally real plane,

because γ_0 has this property and this property is invariant under the action of $\mathrm{PU}(n,1)$. This means exactly that: $\gamma = \pi\big(\mathrm{Span}_{\mathbb{R}}(P,Q) \cap V_-\big)$ for some $P, Q \in \mathbb{C}^{n,1}$ with $\langle P, Q \rangle \in \mathbb{R}$. $\qquad\square$

We now state and prove the more difficult half of Theorem 1.10.1:

Proposition 1.10.3 *Any totally geodesic submanifold of* $\mathrm{H}^n_{\mathbb{C}}$ *is a* $\mathbb{C}^k$*-plane or an* $\mathbb{R}^k$*-plane for some* k *with* $0 \leqslant k \leqslant n$.

Proof. We break down the proof into a sequence of lemmas. Let M be a totally geodesic submanifold of $\mathrm{H}^n_{\mathbb{C}}$ in the ball model $B \subset \mathbb{C}^n$; by applying an isometry we may assume that M contains the origin O.

Lemma 1.10.2 *M is of the form* $W \cap B$ *for some* $\mathbb{R}$*-linear subspace* W *of* $\mathbb{C}^n$.

Proof of Lemma 1.10.2: We know that the geodesic lines through O are exactly the straight line segments through O contained in B. Therefore, $M = W \cap B$ where $W = \mathbb{R}M$ is a subset of $\mathbb{C}^n$ closed under multiplication by real numbers. Now let $u, v \in W$ and assume in fact (after maybe rescaling) that $u, v \in W \cap B$. Then, since W is totally geodesic, by the same dyadic subdivision argument as in the proof of Lemma 1.10.1, the whole line segment $[u, v]$ is contained in M. In particular, $u + v \in W$ and W is an $\mathbb{R}$-linear subspace of $\mathbb{C}^n$. $\qquad\square$

In what follows, given a subspace W of $\mathbb{C}^n$, we identify W with $W \times \{0\} \subset \mathbb{C}^n$. Given $z \in V_-$, recall from Equation (1.9.2) that the central symmetry $s_{\pi(z)}$ at $\pi(z) \in \mathrm{H}^n_{\mathbb{C}}$ lifts to $S_z \in \mathrm{U}(n,1)$ acting on $\mathbb{C}^{n+1}$ by:

$$S_z(w) = w - 2\frac{\langle z, w \rangle}{\langle z, z \rangle}z.$$

Lemma 1.10.3 *Let W be an $\mathbb{R}$-linear subspace of $\mathbb{C}^n$. Then:* $W \cap B$ *is totally geodesic* $\iff$ $S_z(W + \mathbb{R}e_{n+1}) \subset \mathbb{C} \cdot (W + \mathbb{R}e_{n+1})$ *for all* $z \in V_- \cap (W + \mathbb{R}e_{n+1})$.

Proof of Lemma 1.10.3: This follows directly from Lemma 1.10.1 and the fact that $W = \pi(W + \mathbb{R}e_{n+1}) = \pi(\mathbb{C} \cdot (W + \mathbb{R}e_{n+1}))$ (again, identifying W with $W \times \{0\} \subset \mathbb{C}^n$). $\qquad\square$

Lemma 1.10.4 *Let W be an $\mathbb{R}$-linear subspace of $\mathbb{C}^n$ such that $W \cap B$ is totally geodesic, and identify as above W with $W \times \{0\} \subset \mathbb{C}^n$. For any $w \in W \setminus \{0\}$, define:*

$$F(w) = \big\{\lambda \in \mathbb{C} \,|\, \lambda w \in W\big\}$$

(a) For any $w \in W \setminus \{0\}$, $F(w) = \mathbb{R}$ or $\mathbb{C}$.
(b) For any $w, w' \in W \setminus \{0\}$, $\langle w, w' \rangle \in F(w) \cap F(w')$.
(c) If $\lambda \in F(w)$ then $\lambda\langle w, w' \rangle \in F(w')$.
(d) If $\langle w, w' \rangle \in \mathbb{R}$ then $F(w) = F(w')$.
(e) Either $F(w) = \mathbb{R}$ for all $w \in W \setminus \{0\}$, or $F(w) = \mathbb{C}$ for all $w \in W \setminus \{0\}$.

Proof of Lemma 1.10.4: (a) For any $w \in W \setminus \{0\}$, $F(w)$ is an $\mathbb{R}$-linear subspace of $\mathbb{C}$, containing $\mathbb{R}$ (because W is an $\mathbb{R}$-linear subspace). Therefore $F(w) = \mathbb{R}$ or $\mathbb{C}$.

(b) Let $w, w' \in W \setminus \{0\}$, $\lambda = \langle w, w' \rangle$ and assume (after possibly rescaling by a real number) that $\langle w, w \rangle < 1$. This guarantees that $x = w + e_{n+1} \in V_- \cap (W + \mathbb{R}e_{n+1})$, because:

$$\langle x, x \rangle = \langle w, w \rangle + 2\mathrm{Re}\langle w, e_{n+1} \rangle + \langle e_{n+1}, e_{n+1} \rangle = \langle w, w \rangle - 1.$$

Given any $r \in \mathbb{R}$, denote $y = rw' + e_{n+1}$. Then by Lemma 1.10.3, $s_x(y) \in \mathbb{C} \cdot (W + \mathbb{R}e_{n+1})$, i.e. $y - \nu x = \mu(w'' + te_{n+1})$ for some $\mu \in \mathbb{C}$ and $t \in \mathbb{R}$, with $\nu = 2\frac{\langle x, y \rangle}{\langle x, x \rangle}$.

In terms of w, w', r this becomes:

$$rw' + e_{n+1} - \nu(w + e_{n+1}) = \mu(w'' + te_{n+1}), \text{ or :}$$

$$rw' - \nu w + (1 - \nu)e_{n+1} = \mu w'' + t\mu e_{n+1}.$$

Since e_{n+1} and W are ($\mathbb{C}$-)linearly independent, this gives:

$$\begin{cases} 1 - \nu = t\mu \\ rw' - \nu w = \mu w'' \end{cases} \tag{1.10.1}$$

The first line gives $\mathbb{R}(\nu) = \mathbb{R}(\mu)$, and on the other hand $\mathbb{R}(\nu) = \mathbb{R}(\lambda)$ if $r \neq 0$ (because $\langle x, x \rangle \nu / 2 = \langle x, y \rangle = \langle w + e_{n+1}, rw' + e_{n+1} \rangle = r\langle w, w' \rangle - 1$), so that $\mathbb{R}(\mu) = \mathbb{R}(\lambda)$.

Combining the 2 lines of (1.10.1) gives: $w'' = \mu^{-1}rw' + (t - \mu^{-1})w \in W$. Since $tw \in W$, this gives: $\mu^{-1}(rw' - w) \in W$, so $\mu^{-1} \in F(rw' - w)$, hence $\lambda \in F(rw' - w)$ (because $\mathbb{R}(\mu) = \mathbb{R}(\lambda)$). In other words, $\lambda(rw' - w) \in W$. Taking $r = \pm 1$, we get $\lambda w, \lambda w' \in W$ as claimed.

(c) By part (b), $\langle \lambda w, w' \rangle = \lambda \langle w, w' \rangle \in F(w')$ if $\lambda \in F(w)$.

(d) First assume that $\langle w, w' \rangle = r \in \mathbb{R} \setminus \{0\}$. By part (c), for any $\lambda \in F(w)$ we have $r\lambda \in F(w')$ so $\lambda \in F(w')$, giving $F(w) \subset F(w')$. Likewise, $F(w') \subset F(w)$, reversing the roles of w and w'. Now, if $\langle w, w' \rangle = 0$ then $\langle w, w + w' \rangle = \langle w, w \rangle \neq 0$ (as $\langle ., . \rangle$ is positive definite on W). Then by the previous argument $F(w) = F(w + w')$ and likewise $F(w') = F(w + w')$ so that $F(w) = F(w')$.

(e) Let $w, w' \in W \setminus \{0\}$ and $\lambda = \langle w, w' \rangle$. If $\lambda = 0$ then by part (d) $F(w) = F(w')$. If not, then $\langle \lambda^{-1}w, w' \rangle = 1$ so by part (d) $F(\lambda^{-1}w) = F(w')$. But $\lambda \in F(w)$ by part (b), and $F(w)$ is a field by part (a). Therefore $F(w) = F(w')$. $\qquad\square$

This concludes the proof of Proposition 1.10.3, except the bound $0 \leqslant k \leqslant n$ for the dimension of an $\mathbb{R}^k$-plane in $\mathrm{H}_{\mathbb{C}}^n$. Note that $\mathbb{R}^k$-planes through O in the ball model are of the form $W \cap B$ for some totally real $\mathbb{R}$-linear subspace of $\mathbb{C}^n$. $\mathbb{R}$-linear subspaces of $\mathbb{C}^n$ may have any real dimension $\leqslant 2n$, but the bound on the dimension of totally real subspaces of $\mathbb{C}^n$ comes from the following observation:

Lemma 1.10.5 *Let W be an $\mathbb{R}$-linear subspace of $\mathbb{C}^n$. Then: W is totally real if and only if $W \perp_{\mathrm{Re}\langle ., . \rangle} iW$.*

Proof of Lemma 1.10.5: W is totally real $\iff$ $\langle X, Y \rangle \in \mathbb{R}$ for all $X, Y \in W$, which is equivalent to $\mathrm{Re}\langle X, iY \rangle = 0$ for all $X, Y \in W$, as $\langle X, iY \rangle = -i\langle X, Y \rangle$. $\qquad\square$

Now $\mathrm{Re}\langle ., . \rangle$ is positive definite on $\mathbb{C}^n$, therefore if W is totally real Lemma 1.10.5 implies $W \cap iW = \{0\}$, so that W has (real) dimension at most n. $\qquad\square$

1.11   Bisectors

Following Mostow, we call equidistant hypersurfaces in $\mathrm{H}_{\mathbb{C}}^n$ *bisectors*. Explicitly, given 2 distinct points $p_1, p_2 \in \mathrm{H}_{\mathbb{C}}^n$, the *bisector* equidistant from p_1, p_2 is:

$$B(p_1, p_2) = \big\{ p \in \mathrm{H}_{\mathbb{C}}^n \,|\, d(p, p_1) = d(p, p_2) \big\}.$$

The *complex spine* Σ of $B = B(p_1, p_2)$ is the complex line spanned by p_1, p_2; the *real spine* σ of B is the real geodesic $B \cap \Sigma$.

As we've seen, bisectors cannot be totally geodesic (as they would be in constant curvature spaces) because they are real hypersurfaces (Corollary 1.10.2), but they admit 2 foliations by totally geodesic subspaces as follows:

Proposition 1.11.1 *Let p_1, p_2 be 2 distinct points in $\mathrm{H}_{\mathbb{C}}^n$, $B = B(p_1, p_2)$ the bisector equidistant from p_1 and p_2, Σ the complex spine of B and σ its real spine. Then:*
(1) $B = \pi_{\Sigma}^{-1}(\sigma)$, where π_{Σ} denotes orthogonal projection onto Σ, and
(2) B is the union of all $\mathbb{R}^n$-planes containing σ.

Part (1) tells us that B is foliated by the $\mathbb{C}^{n-1}$-planes $\pi_{\Sigma}^{-1}(\{x\})$ for $x \in \sigma$; these are called the *slices* of B; the $\mathbb{R}^n$-planes containing σ are called the *meridians* of B (the meridian decompoisition is not literally a foliation as all meridians share the real spine σ). Note also that part (1) tells us that the real and complex spines only depend on the bisector itself (and not on the pair of points used to define it). Also, given a bisector B, the pairs of points from which B is equidistant are much more constrained than in constant curvature, namely they are exactly pairs of points in $\Sigma \setminus \sigma$ which are symmetric with repsect to σ.

We will skip the proof of Proposition 1.11.1 for the sake of time. Part (1) is due to Mostow and part (2) to Goldman. The following gives a complete description of the geodesic non-convexity of bisectors; it is also due to Goldman and we will also skip its proof:

Proposition 1.11.2 *Given a bisector B in $\mathrm{H}_{\mathbb{C}}^n$ and 2 distinct points $p, q \in B$, the geodesic (pq) is contained in B iff p, q are in a common slice or meridian of B.*

Our motivation to study these equidistant hypersurfaces comes from the world of discrete groups and fundamental domains, where they arise as bounding hypersurfaces for *Dirichlet domains*, which we define and study in the next part.

Chapter 2

Discrete subgroups and lattices in $\mathrm{Isom}(\mathrm{H}^n_{\mathbb{R}})$ and $\mathrm{Isom}(\mathrm{H}^n_{\mathbb{C}})$

2.1 Discontinuous group actions and fundamental domains

We start with a few general definitions about group actions on metric spaces. Let Γ be a group acting by isometries on a metric space X. The definitions and setup are in complete generality, however most of the properties that we are interested in will require X to be a *proper* metric space, meaning that all closed metric balls in X are compact (note that such a space is in particular locally compact).

Definitions: We will say that Γ acts *properly discontinuously* on X if for any compact $K \subset X$, the set $\{\gamma \in \Gamma \,|\, \gamma K \cap K \neq \emptyset\}$ is finite.

A simply-connected subset $D \subset X$ with non-empty interior will be called a *fundamental domain* for the action of Γ on X if:

$$(a) \ \bigcup_{\gamma \in \Gamma} \gamma \overline{D} = X, \text{ and}$$

$$(b) \ \gamma_1 \neq \gamma_2 \in \Gamma \implies \gamma_1 \mathring{D} \cap \gamma_2 \mathring{D} = \emptyset.$$

When these conditions are satisfied, we say that the images of D under Γ *tessellate* X. Note that traditionally, a *domain* is a simply-connected open set, but we do not require D to be open, just to have non-empty interior. There is a related notion that is a bit more technical but useful in practice: D as above is called a *fundamental domain for the action of Γ modulo a subgroup $H < \Gamma$* if it satisfies condition (a) and the following condition in place of (b):

(b') For all $\gamma_1, \gamma_2 \in \Gamma$, either $\gamma_1 \mathring{D} \cap \gamma_2 \mathring{D} = \emptyset$, or $\gamma_1 \mathring{D} = \gamma_2 \mathring{D}$ and $\gamma_1 \gamma_2^{-1} \in H$.

The following is a general construction of fundamental domains for groups acting properly discontinuously on a (proper) metric space. Given a point $p_0 \in X$, the *Dirichlet domain* for Γ centered at p_0 is defined as:

$$D_\Gamma(p_0) = \{x \in X \,|\, d(x, p_0) \leqslant d(x, \gamma p_0) \ \forall \gamma \in \Gamma\}. \tag{2.1.1}$$

Note: We will be mostly interested in the case where the group Γ is infinite, in which case it is not clear whether or not $D_\Gamma(p_0)$ has non-empty interior (or even whether or not it is reduced to $\{p_0\}$), and if so, whether or not it has finitely many faces, i.e. whether or not a finite collection of elements $\gamma \in \Gamma$ suffice in (2.1.1).

For any two distinct points $x, y \in X$, the space X is divided into the following 3 regions:

$$\mathcal{B}(p_1, p_2) = \big\{x \in X \mid d(x, p_1) = d(x, p_2)\big\} \tag{2.1.2}$$
$$\mathcal{B}^-(p_1, p_2) = \big\{x \in X \mid d(x, p_1) < d(x, p_2)\big\} \tag{2.1.3}$$
$$\mathcal{B}^+(p_1, p_2) = \big\{x \in X \mid d(x, p_1) > d(x, p_2)\big\} \tag{2.1.4}$$

We will say that X is *balanced* if for any two distinct points $p_1, p_2 \in X$, every point of $\mathcal{B}(p_1, p_2)$ is an accumulation point of both $\mathcal{B}(p_1, p_2)^-$ and $\mathcal{B}(p_1, p_2)^+$.

Proposition 2.1.1 *Let Γ be a group acting by isometries on a proper balanced metric space X.*
(1) If Γ acts properly discontinuously on X, then for all $p_0 \in X$, $D_\Gamma(p_0)$ contains an open ball centered at p_0.
(2) Assume moreover that (a) Γ is a finitely generated linear group acting faithfully on X, and (b) for any isometry $f \neq \mathrm{Id}$ of X, $\mathrm{Fix}(f)$ is nowhere dense in X (respectively, X admits a nonzero measure m such that for any isometry $f \neq \mathrm{Id}$ of X, $m(\mathrm{Fix}(f)) = 0$). If for all $p_0 \in X$ $D_\Gamma(p_0)$ contains an open ball centered at p_0, then Γ acts properly discontinuously on X.
(3) If $D_\Gamma(p_0)$ contains an open ball centered at p_0 then $D_\Gamma(p_0)$ is a fundamental domain for the action of Γ modulo $\mathrm{Stab}_\Gamma(p_0)$.

We start by stating and proving the following auxiliary results:

Lemma 2.1.1 *(a) $D_\Gamma(p_0)$ is closed.*
(b) If $x \in \overset{\circ}{D}_\Gamma(p_0)$ then $d(x, p_0) < d(x, \gamma p_0)$ for all $\gamma \in \Gamma$ such that $\gamma p_0 \neq p_0$.
(c) For any $\gamma_0 \in \Gamma$, $\gamma_0 D_\Gamma(p_0) = D_\Gamma(\gamma_0 p_0)$.
(d) For $\gamma_1, \gamma_2 \in \Gamma$, if $\gamma_1 p_0 \neq \gamma_2 p_0$ then $\gamma_1 \overset{\circ}{D}_\Gamma(p_0) \cap \gamma_2 \overset{\circ}{D}_\Gamma(p_0) = \emptyset$.

Proof. When the dependence on p_0 is not relevant (as it is in (c)) we will write D for $D_\Gamma(p_0)$.

(a) For each $\gamma \in \Gamma$, $\big\{x \in X \mid d(x, p_0) \leqslant d(x, \gamma p_0)\big\}$ is closed, therefore D is an intersection of closed sets and is closed.

(b) Let $x \in D_\Gamma(p_0)$ and assume that $d(x, p_0) = d(x, \gamma p_0)$ for some $\gamma \in \Gamma$ such that $\gamma p_0 \neq p_0$. Then $x \in \mathcal{B}(p_0, \gamma p_0)$, and since X is assumed balanced any ball centered at x contains points of both $\mathcal{B}^-(p_0, \gamma p_0)$ and $\mathcal{B}^+(p_0, \gamma p_0)$. Therefore x is not an interior point of $D_\Gamma(p_0)$.

(c) Note that:

$$
\begin{aligned}
x \in D_\Gamma(p_0) &\iff (\forall \gamma \in \Gamma)\; d(\gamma_0 x, \gamma_0 p_0) \leqslant d(\gamma_0 x, \gamma_0 \gamma p_0) \\
&\iff \gamma_0 x \in D_\Gamma(\gamma_0 p_0),
\end{aligned}
$$

where the first equivalence follows from the fact that γ_0 is an isometry and the second from the fact that $\gamma \longmapsto \gamma_0\gamma$ is a bijection $\Gamma \longrightarrow \Gamma$.

(d) From parts (b) and (c), for $i = 1, 2$: if $x \in \gamma_i\mathring{D}$ then for all $\gamma \in \Gamma$ such that $\gamma\gamma_ip_0 \neq \gamma_ip_0$, $d(x, \gamma_ip_0) < d(x, \gamma\gamma_ip_0)$. Assume that there exists a point $x \in \gamma_1\mathring{D} \cap \gamma_2\mathring{D}$. Then, taking $\gamma = \gamma_2\gamma_1^{-1}$ in the above condition for $i = 1$ gives $d(x, \gamma_1p_0) < d(x, \gamma_2p_0)$. Likewise, taking taking $\gamma = \gamma_1\gamma_2^{-1}$ in the above condition for $i = 2$ gives $d(x, \gamma_2p_0) < d(x, \gamma_1p_0)$, a contradiction. $\square$

Proof of Proposition 2.1.1: (1) Assume Γ acts properly discontinuously on X, let $p_0 \in X$ and $r > 0$ and denote $B = B(p_0, r)$. Then $\bar{B}$ is compact by properness of X, so $\{\gamma \in \Gamma \mid \gamma\bar{B} \cap \bar{B} \neq \emptyset\}$ is finite. Denote $\{\gamma_1, ..., \gamma_n\} = \{\gamma \in \Gamma \mid \gamma\bar{B} \cap \bar{B} \neq \emptyset \text{ and } \gamma p_0 \neq p_0\}$. After possibly shrinking r, we may assume that $p_0 \notin \gamma_1\bar{B} \cup ... \cup \gamma_n\bar{B}$. Note that:

- if $\gamma\bar{B} \cap \bar{B} = \emptyset$, then $\bar{B} \subset \{x \in X \mid d(x, p_0) \leqslant d(x, \gamma p_0)\}$, and

- if $\gamma\bar{B} \cap \bar{B} \subset \partial B$, then $B \subset \{x \in X \mid d(x, p_0) \leqslant d(x, \gamma p_0)\}$.

Therefore it suffices to take $\varepsilon \leqslant \underset{1 \leqslant i \leqslant n}{\text{Min}} \left(d(p_0, \gamma_i\bar{B})\right)$ to get $B(p_0, \varepsilon) \subset D_\Gamma(p_0)$.

(2) Assume the action of Γ on X is not properly discontinuous. Then there exists a compact subset K of X such that $\{\gamma \in \Gamma \mid \gamma K \cap K \neq \emptyset\}$ is infinite. This provides a sequence (γ_n) of *distinct* elements of Γ and a sequence (x_n) of points of K such that $x_n \in \gamma_n K \cap K$ for all n. Since K is compact, we may assume after passing to a subsequence that the sequence (x_n) (resp. $(\gamma_n^{-1}x_n)$) converges, say to x_0 (resp. x_1). By continuity of the action $\Gamma \times X \longrightarrow X$, $(\gamma_n^{-1}x_0)$ converges to x_1. If this sequence is eventually constant then $\text{Stab}_\Gamma(x_0)$ is infinite; if not then the orbit Γx_0 has an accumulation point. We treat these 2 cases separately.

Lemma 2.1.2 *If the orbit Γx_0 has an accumulation point then $D_\Gamma(x_0)$ does not contain any open ball centered at x_0.*

Proof of Lemma 2.1.2: Assume that $D = D_\Gamma(x_0)$ contains an open ball $B(x_0, \varepsilon)$ centered at x_0, and let $(\gamma_n x_0)$ be a sequence of *distinct* points converging to some point $p \in X$, where $\gamma_n \in \Gamma$ for all n. Then, for m, n large enough: $d(\gamma_m x_0, \gamma_n x_0) \leqslant \varepsilon$. On the other hand, $\gamma_m D$ (resp. $\gamma_n D$) contains $B(\gamma_m x_0, \varepsilon)$ (resp. $B(\gamma_n x_0, \varepsilon)$), which would imply that $\gamma_m\mathring{D} \cap \gamma_n\mathring{D} \neq \emptyset$, contradicting Lemma 2.1.1 (d), as $\gamma_m x_0 \neq \gamma_n x_0$. $\square$

Lemma 2.1.3 *If some point of X has an infinite stabilizer, then Γ has an orbit with an accumulation point.*

Proof of Lemma 2.1.3: Assume that $|\Gamma_p| = \infty$ for some $p \in X$, denoting $\Gamma_p = \text{Stab}_\Gamma(p)$.

Claim: Γ_p has an infinite orbit.

Lemma 2.1.3 then follows from the claim, as the orbits of Γ_p are contained in spheres centered at p, which are compact by properness of X (as they are closed subsets of closed balls, which are assumed to be compact).

We now prove the claim. Assume that all orbits of Γ_p are finite. Then in particular:

$$(\forall x \in X)\,(\forall \gamma \in \Gamma_p)\,(\exists n \geqslant 1)\,\gamma^n x = x. \tag{2.1.5}$$

Fix $\gamma \in \Gamma_p$, and for $n \geqslant 1$ let $X_n = \mathrm{Fix}(\gamma_n) = \{x \in X \mid \gamma^n x = x\}$. Then by (2.1.5): $\bigcup_{n \geqslant 1} X_n = X$. But by assumption, each X_n is either all of X, or nowhere dense (resp. has measure 0). Assuming that $X_n \neq X$ for all n, this would imply by the Baire Category Theorem (as X is locally compact Hausdorff) that X is nowhere dense (resp. that X has measure 0 by subadditivity of measure), a contradiction. Therefore $X_n = X$ for some $n \geqslant 1$, hence γ has finite order as the action is faithful. This proves that all elements of Γ_p have finite order.

Problem: Is it true that a (finitely generated) group Γ all of whose elements have finite order is finite?

This is a famous question, known as the *Burnside problem*. Burnside proved that the answer is yes if Γ is linear (i.e. a subgroup of $\mathrm{GL}(n, \mathbb{C})$ for some n) and the orders of all elements are uniformly bounded; Schur then extended this result by removing the bounded orders assumption. (The answer is no in general: Golod and Shafarevich constructed counter-examples in the 1970's, with unbounded orders, and subsequently Olshanskii and Ivanov constructed counter-examples with bounded orders).

In our case, under the assumption that Γ is a finitely generated linear group, we cannot directly apply Schur's theorem to Γ_p (as it could be infinitely generated). However, if we look more closely at the proof of Schur's theorem we see that the result also applies to subgroups of finitely generated linear groups (see e.g. Curtis-Reiner). Alternatively, we could appeal to Selberg's lemma (Theorem 2.3.1), which says that Γ has a finite-index torsion-free subgroup H; then $H \cap \Gamma_p$ has finite index in Γ_p and is trivial, so Γ_p is finite.

This proves the claim and concludes the proof of Proposition 2.1.1(2).

Proposition 2.1.1(3) follows from Lemma 2.1.1(d) and the following:

Lemma 2.1.4 *If $D_\Gamma(p_0)$ contains an open ball centered at p_0, then $D_\Gamma(p_0)$ contains an element of each orbit.*

Proof of Lemma 2.1.4: Let $x \in X$ and $d = \mathrm{Inf}\{d(x, \gamma p_0) \mid \gamma \in \Gamma\}$.

- If d is attained, i.e. $d = d(x, \gamma_0 p_0)$ for some $\gamma_0 \in \Gamma$, then $x \in D_\Gamma(\gamma_0 p_0)$ which by Lemma 2.1.1(c) is equal to $\gamma_0 D_\Gamma(p_0)$.

- If not, there exists a sequence (γ_n) of distinct elements of Γ such that $d(x, \gamma_n p_0)$ decreases to d. Then, for n sufficently large $\gamma_n p_0 \in \overline{B(p_0, 2d)}$ which is compact by properness of X, hence $(\gamma_n p_0)$ has a converging subsequence. Now by assumption $D_\Gamma(p_0)$ contains an open ball $B(p_0, \varepsilon)$ centered at p_0. Then, for m, n large enough we have $d(\gamma_m p_0, \gamma_n p_0) \leqslant \varepsilon$, which implies that $\gamma_m \mathring{D}_\Gamma(p_0) \cap \gamma_n \mathring{D}_\Gamma(p_0) \neq \emptyset$, contradicting Lemma 2.1.1(d). $\square$

2.2 Hyperbolic manifolds and orbifolds

We now describe the structure that arises on a quotient of the form X/Γ where X is a proper metric space and $\Gamma < \mathrm{Isom}(X)$ is a discrete subgroup. Here discrete refers to the *compact-open topology* on $\mathrm{Isom}(X)$, which is generated by sets of the form $U_{K,V} = \big\{ g \in \mathrm{Isom}(X) \,|\, g(K) \subset V \big\}$, where K (resp. V) is a fixed compact (resp. open) subset of X.

Proposition 2.2.1 *Let X be a proper metric space and $\Gamma < \mathrm{Isom}(X)$. Then:*
(1) Γ is discrete (in the compact-open topology) $\iff$ Γ acts properly discontinuously on X.
(2) If Γ is discrete then X/Γ is Hausdorff. If moreover Γ acts freely (i.e. without fixed points) then $p : X \longmapsto X/\Gamma$ is a covering map and a local isometry.

Notes: (a) When as in part (2) Γ is discrete and acts freely on X, if X is a (topological, resp. smooth,...) manifold then so is X/Γ, and X/Γ is locally homeomorphic (resp. diffeomorphic,...) to X (as p is then also a local diffeomorphism). Loosely speaking, such an X/Γ inherits any "geometric structure" that X has. We now make this more precise in the case where $X = \mathrm{H}_{\mathbb{R}}^n$ or $\mathrm{H}_{\mathbb{C}}^n$.

(b) When $\Gamma < \mathrm{Isom}(\mathrm{H}_{\mathbb{R}}^n)$ (resp. $\mathrm{Isom}(\mathrm{H}_{\mathbb{C}}^n)$) is discrete and acts freely the manifold X/Γ is locally isometric to $\mathrm{H}_{\mathbb{R}}^n$ (resp. $\mathrm{H}_{\mathbb{C}}^n$). Such manifolds are called *hyperbolic manifolds* (resp. *complex hyperbolic manifolds*). More generally, a *hyperbolic orbifold* (resp. *complex hyperbolic orbifold*) is a quotient space $\mathrm{H}_{\mathbb{R}}^n/\Gamma$ (resp. $\mathrm{H}_{\mathbb{C}}^n/\Gamma$) where Γ is a discrete subgroup of $\mathrm{Isom}(\mathrm{H}_{\mathbb{R}}^n)$ (resp. $\mathrm{Isom}^+(\mathrm{H}_{\mathbb{C}}^n)$). Moreover, it can be shown that the compact-open topology on $\mathrm{Isom}(\mathrm{H}_{\mathbb{R}}^n)$ (resp. $\mathrm{Isom}^+(\mathrm{H}_{\mathbb{C}}^n)$) coincides with the usual matrix topology on $\mathrm{O}^+(n,1)$ (resp. the quotient topology on $\mathrm{PU}(n,1)$) – this is left to the reader as an exercise.

(c) In particular, if $M = \mathrm{H}_{\mathbb{R}}^n/\Gamma$ is a hyperbolic manifold then M is a complete Riemannian manifold with constant negative curvature -1. It can be shown that $\mathrm{H}_{\mathbb{R}}^n$ is the only complete, simply-connected Riemannian n-manifold with constant negative curvature -1, from which it follows that if M is a complete Riemannian manifold with constant negative curvature -1, then M is isometric to $\mathrm{H}_{\mathbb{R}}^n/\Gamma$ for some discrete $\Gamma < \mathrm{Isom}(\mathrm{H}_{\mathbb{R}}^n)$ acting freely on $\mathrm{H}_{\mathbb{R}}^n$, i.e. M is a hyperbolic manifold as we've defined it. This gives an equivalent definition of hyperbolic manifolds which is the standard definition in differential geometry.

We will use the following lemmas in the proof of Proposition 2.2.1:

Lemma 2.2.1 *$\Gamma < G = \mathrm{Isom}(X)$ is discrete $\iff$ any sequence of elements of Γ converging to Id is eventually constant.*

$$
\begin{aligned}
\textit{Proof} : \Gamma \text{ is discrete} \quad &\iff\quad (\forall \gamma \in \Gamma) \ \{\gamma\} \text{ is open in } \Gamma \\
&\iff\quad (\forall \gamma \in \Gamma) \ (\exists U_\gamma \text{ a neighborhood of } \gamma \text{ in } G) \ \Gamma \cap U_\gamma = \{\gamma\} \\
&\iff\quad (\forall \gamma \in \Gamma) \text{ any sequence of elements of } \Gamma \text{ converging to } \gamma \\
&\qquad\qquad \text{is eventually constant} \\
&\iff\quad \text{any sequence of elements of } \Gamma \text{ converging to } \mathrm{Id} \\
&\qquad\qquad \text{is eventually constant.}
\end{aligned}
$$

Note for the last point that for any γ, the map $g \longmapsto \gamma g$ is a self-homeomorphism of G. $\qquad\square$

Lemma 2.2.2 *The compact-open topology on $G = \mathrm{Isom}(X)$ is the topology of pointwise convergence. In other words, given a sequence (g_n) of elements of G and $g \in G$:*

$$g_n \longrightarrow g \text{ for the compact-open topology on } G \iff (\forall x \in X) \ g_n(x) \longrightarrow g(x).$$

Proof. We will use the following general fact about the compact-open topology: if Y is a metric space (and X any topological space) then the compact-open topology on $C(X,Y) = \{\text{continuous functions } X \longrightarrow Y\}$ is the topology of uniform convergence on compact subsets of X (see e.g. []). This immediately gives the forward implication, namely if $g_n \longrightarrow g$ for the compact-open topology on G then $g_n(x) \longrightarrow g(x)$ for all $x \in X$.

Conversely, assume that $(\forall x \in X) \ g_n(x) \longrightarrow g(x)$ and let $K \subset X$ be compact. Assuming that the convergence is not uniform on K, there would exist $\varepsilon > 0$, a subsequence (g_{n_i}) of (g_n) and a sequence of points $x_i \in K$ such that $d\big(g_{n_i}(x_i), g(x_i)\big) \geqslant \varepsilon$ for all i. But, K being compact, we may assume (after possibly passing to a subsequence) that (x_i) converges, say to x. Then, for i large enough, $d(x_i, x) < \varepsilon/4$ and $d\big(g_{n_i}(x), g(x)\big) < \varepsilon/2$, giving:

$$
\begin{aligned}
d\big(g_{n_i}(x_i), g(x_i)\big) &\leqslant d\big(g_{n_i}(x_i), g_{n_i}(x)\big) + d\big(g_{n_i}(x), g(x)\big) + d\big(g(x), g(x_i)\big) \\
&= 2d(x_i, x) + d\big(g_{n_i}(x), g(x)\big) < \varepsilon,
\end{aligned}
$$

a contradiction. (Note that we used in the last step that g and the g_{n_i} are isometries of X.) $\square$

Lemma 2.2.3 *Let $K_1, K_2 \subset X$ be compact. Then: $G_{K_1,K_2} = \big\{g \in G \,|\, g(K_1) \subset K_2\big\}$ is compact (for the compact-open topology). In particular, $\mathrm{Stab}_{\mathrm{Isom}(X)}(x)$ is compact for all $x \in X$.*

Proof. This follows immediately from the Arzela-Ascoli theorem, as G_{K_1,K_2} is:

(a) equicontinuous (as all its elements are isometries),

(b) pointwise relatively compact, i.e. $(\forall x \in K_1) \ G_{K_1,K_2}$ is relatively compact as it is contained in K_2, and

(c) closed, by continuity of the orbit map $(g, x) \longmapsto gx$. $\square$

Proof of Proposition 2.2.1: (1) First assume that Γ is not discrete in $G = \mathrm{Isom}(X)$. Then, by Lemma 2.2.1 there exists a sequence (γ_n) of distinct elements of Γ converging to Id. Let $x_0 \in X$ and $r > 0$; then for n large enough $\gamma_n x_0 \in B(x_0, r) = B$ by Lemma 2.2.2, in particular $\big\{\gamma \in \Gamma \,|\, \gamma\overline{B} \cap \overline{B} \neq \emptyset\big\}$ is infinite. But $\overline{B}$ is compact by properness of X, so the action of Γ on X is not properly discontinuous.

Conversely, assume that the action of Γ on X is not properly discontinuous. As previously, this gives either an orbit with an accumulation point, or a point with an infinite stabilizer. In the latter case, Γ is not discrete as point stabilizers in G are compact by Lemma 2.2.3. If Γ has an orbit with an accumulation point, there exists a sequence of distinct points $(\gamma_n x_0)$ converging to a point $p \in X$ (for some $\gamma_n \in \Gamma$). Fix $r > 0$; then for n large enough we have $d(\gamma_n x_0, p) \leqslant r$ hence $\gamma_n\big(\overline{B(x_0, r)}\big) \subset \overline{B(p, 2r)}$ (indeed, if $y \in X$ satisfies $d(y, \gamma_n x_0) \leqslant r$ then $d(y, p) \leqslant d(y, \gamma_n x_0) + d(\gamma_n x_0, p) \leqslant 2r$). In other words, for n large enough we have $\gamma_n \in G_{\overline{B(x_0,r)}, \overline{B(p,2r)}}$ which is compact by Lemma 2.2.3 and by properness of X. Therefore (γ_n)

has a converging subsequence, therefore Γ is not discrete as the γ_n are distinct.

(2) First assume that Γ is discrete; then by (1) Γ acts properly discontinuously on X. Let $x, y \in X$ with $p(x) \neq p(y)$, i.e. $x \notin \Gamma y$ (and equivalently, $y \notin \Gamma x$). Let U, V be neighborhoods of x, y in X such that $\overline{U}, \overline{V}$ are compact. Then $\{\gamma \in \Gamma \mid \gamma(\overline{U} \cup \overline{V}) \cap (\overline{U} \cup \overline{V}) \neq \emptyset\}$ is finite, say equal to $\{\gamma_1, ..., \gamma_r\}$. Now by assumption $x \neq \gamma_i(y)$ (resp. $y \neq \gamma_i(x)$) for $i = 1, ..., r$, so we may assume after possibly shrinking U (resp. V) that $x \notin \gamma_i(\overline{V})$ (resp. $y \notin \gamma_i(\overline{U})$) for $i = 1, ..., r$. Let $U' = U \setminus \bigcup_{i=1}^{r} \gamma_i(\overline{V})$ and $V' = V \setminus \bigcup_{i=1}^{r} \gamma_i(\overline{U})$. Then U' is a neighborhood of x in X not intersecting $\Gamma V'$, and V' is a neighborhood of y in x not intersecting $\Gamma U'$, therefore $p(U')$ and $p(V')$ are disjoint neighborhoods of $p(x)$ and $p(y)$ in X/Γ, hence X/Γ is Hausdorff.

Now assume moreover that Γ acts freely on X. The fact that the quotient map p is a covering map then follows from the following:

Lemma 2.2.4 *Given $x \in X$, there exists a neighborhood U of x in X such that:*

$$\left(\forall \gamma \in \Gamma \setminus \{\mathrm{Id}\}\right) \ \gamma U \cap U = \emptyset.$$

Proof. This follows essentially the same lines as the previous argument. Let V be a neighborhood of x in X with compact closure; by proper discontinuity of the action of Γ on X, the set $\{\gamma \in \Gamma \mid \gamma \overline{V} \cap \overline{V} \neq \emptyset\}$ is finite, say equal to $\{\mathrm{Id}, \gamma_1, ..., \gamma_r\}$. Note that for $i = 1, ..., r$, $\gamma_i^{-1}(x) \neq x$. Let $W \subset V$ be another neighborhood of x with compact closure, such that $\gamma_i^{-1}(x) \notin \overline{W}$ for $i = 1, ..., r$. Let $U = W \setminus \bigcup_{i=1}^{r} \gamma_i^{-1}(\overline{W})$. Then U is a neighborhood of x in X, and: (a) $\gamma_i U \cap U = \emptyset$ for $i = 1, ..., r$ by definition of U, and (b) $\gamma U \cap U = \emptyset$ for $\gamma \in \Gamma \setminus \{\mathrm{Id}, \gamma_1, ..., \gamma_r\}$ by definition of W. $\qquad \square$

Finally, the fact that the covering map p is also a local isometry follows from the fact that each element of Γ is an isometry. This concludes the proof of Proposition 2.2.1. $\qquad \square$

Moral: Given a discrete group $\Gamma < \mathrm{Isom}(\mathrm{H}_{\mathbb{K}}^n)$ for $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ (resp. a discrete group acting freely), we get a hyperbolic orbifold (resp. manifold) $\mathrm{H}_{\mathbb{K}}^n/\Gamma$. This raises the following question:

Question: How can one find/construct discrete subgroups of $\mathrm{Isom}(\mathrm{H}_{\mathbb{K}}^n)$ for $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$?

We will investigate this question in detail in the remaining sections, where we give an overview of two main families of constructions: arithmetic constructions, leading to so-called *arithmetic lattices*, and more geometric constructions such as Coxeter groups and more generally reflection groups.

We finish this section by pointing out that for discrete subgroups of $\mathrm{Isom}(\mathrm{H}_{\mathbb{K}}^n)$ for $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$, acting freely is nothing more than being torsion-free (i.e. having no nontrivial element of finite order):

Lemma 2.2.5 *Let $\Gamma < \mathrm{Isom}(\mathrm{H}_{\mathbb{K}}^n)$ be discrete (with $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$). Then, for $\gamma \in \Gamma \setminus \{\mathrm{Id}\}$:*
γ has a fixed point in $\mathrm{H}_{\mathbb{K}}^n \iff \gamma$ has finite order. In particular:
Γ acts freely on $\mathrm{H}_{\mathbb{K}}^n \iff \Gamma$ is torsion-free.

Note that this is not true in general, for example in $\mathrm{Isom}(S^n)$ the antipodal map $x \mapsto -x$ has order 2 but has no fixed point in S^n.

Proof. If γ has a fixed point p then $\langle \gamma \rangle = \{\gamma^n \,|\, n \in \mathbb{Z}\} \subset \mathrm{Stab}_G(p) \cap \Gamma$ is finite as $\mathrm{Stab}_G(p)$ is compact and Γ discrete, hence γ has finite order.

Conversely, if γ has finite order then γ has a fixed point in $\mathrm{H}_{\mathbb{K}}^n$. This follows from the classification of isometries (into elliptic, parabolic and loxodromic isometries) by noting that parabolic and loxodromic isometries have infinite order. $\qquad\square$

2.3 Structure theorems for finitely generated linear groups and hyperbolic lattices

In this section we give a brief overview of structure theorems for finitely generated linear groups and hyperbolic lattices, namely: the Selberg lemma and Tits alternative for f.g. linear groups, the Margulis lemma fro discrete groups of hyperbolic isometries, and Mostow rigidity and finite presentation of hyperbolic lattices.

2.3.1 Finitely generated linear groups

Theorem 2.3.1 (Selberg's lemma) *If Γ is a finitely generated subgroup of $\mathrm{GL}(n, \mathbb{C})$ for some n, then Γ has a torsion-free subgroup Γ_0 of finite index.*

Using Proposition 2.2.1 and Lemma 2.2.5, this can be rephrased as follows for discrete groups of hyperbolic isometries (which we have seen to be linear):

Corollary 2.3.2 *If Γ is a finitely generated subgroup of $\mathrm{Isom}(\mathrm{H}_{\mathbb{K}}^n)$, the hyperbolic orbifold $\mathrm{H}_{\mathbb{K}}^n / \Gamma$ is finitely covered by a hyperbolic manifold $\mathrm{H}_{\mathbb{K}}^n / \Gamma_0$.*

Note: In Selberg's lemma one can in fact require Γ_0 to be normal in Γ as well, giving a normal finite cover in the Corollary.

Idea of Proof: Reduce entries mod. maximal ideals of the ring generated (over $\mathbb{Z}$) by the entries of the generators of Γ – since Γ is finitely generated this ring is finitely generated. For example, if $\Gamma < \mathrm{GL}(n, \mathbb{Z})$, consider $\Gamma_p = \Gamma \cap \mathrm{Ker}\phi_p$, where $\phi_p : \mathrm{GL}(n, \mathbb{Z}) \longmapsto \mathrm{GL}(n, \mathbb{Z}/p\mathbb{Z})$ is reduction mod. a prime p. Then Γ_p (is normal and) has finite index in Γ, and one can show that for p large enough Γ_p is torsion-free.

For the next statement we recall the definitions of solvable and virtually solvable groups. A group G is *solvable* if its derived series terminates, i.e. if there exists $n \geqslant 1$ such that $G_n = \mathrm{Id}$, where the derived series (G_n) is defined inductively as follows:

$$\begin{cases} G_0 = G \\ G_{n+1} = [G_n, G_n], \end{cases}$$

where $[H, K]$ denotes the subgroup generated by all commutators of the form $[h, k] = hkh^{-1}k^{-1}$ with $h \in H$ and $k \in K$. The group G is called *virtually solvable* if it has a solvable subgroup of finite index.

Theorem 2.3.3 (Tits alternative) *If Γ is a subgroup of $\mathrm{GL}(n,\mathbb{C})$ for some n, then either Γ contains a non-abelian free group, or Γ is virtually solvable.*

Idea of Proof: The main ingredient is what Tits called the *Ping-Pong lemma*, which is a reformulation of Klein's combination theorem, and which in this formulation is very flexible and can be applied in many different contexts (see for example [dlH]):

Lemma 2.3.1 (Ping-Pong lemma) *Let Γ be a group acting on a set X, and let $a_1, a_2 \in \Gamma$. Assume that there exist 4 non-empty, pairwise disjoint subsets $A_1^+, A_1^-, A_2^+, A_2^- \subset X$ such that, for $i = 1, 2$:*

$$\begin{cases} a_i(X \setminus A_i^-) \subset A_i^+ \\ a_i^{-1}(X \setminus A_i^+) \subset A_i^-. \end{cases}$$

Then $\langle a_1, a_2 \rangle \simeq F_2$.

Sketch of Proof of Ping-Pong lemma: Let $g = a_1^{k_1} a_2^{k_2} ... a_1^{k_{r-1}} a_2^{k_r}$ be a (non-trivial) reduced word in $\langle a_1, a_2 \rangle$, with all $k_i \neq 0$ except possibly k_1 and k_r. Assume for instance that $k_1 \neq 0$ and $k_r \neq 0$. Note that $a_1^k(A_2^\pm) \subset A_1^\pm$ and $a_2^k(A_1^\pm) \subset A_2^\pm$, therefore $g(A_2^\pm) \subset A_1^\pm$, hence $g \neq \mathrm{Id}$. The other cases are handled similarly. $\qquad\square$

Idea of proof of Tits alternative: One can show that, unless Γ is virtually solvable, Γ contains 2 non-commuting matrices a_1, a_2, each with at least 2 eigenvalues of distinct absolute values. Then the eigenspaces for the eigenvalues with extremal absolute values, $\lambda_i^{max}/\lambda_i^{min}$ correspond to attracting/repulsing fixed points for a_i acting on $\mathbb{C}P^{n-1}$. One can then apply the Ping-Pong lemma, taking $A_i^\pm$ to be suitable neighborhoods of the attracting/repulsing fixed points of a_i.

2.3.2 Discrete hyperbolic isometry groups

We first introduce an important invariant of isometries of metric spaces. Given an isometry γ of a metric space X, the *displacement* (or *translation length*) of γ is defined as:

$$d_\gamma = \mathrm{Inf}\big\{d(x, \gamma x) \,|\, x \in X\big\}.$$

This allows to roughly classify all isometries of any metric space into 3 types. An isometry γ of X is called:

- *hyperbolic* if $d_\gamma > 0$ and is attained, i.e. $(\exists x \in X)\, d(x, \gamma x) = d_\gamma$,

- *elliptic* if $d_\gamma = 0$ and is attained, i.e. γ has a fixed point in X, and

- *parabolic* if d_γ is not attained.

If X is negatively curved then in the latter case one can show that $d_\gamma = 0$. It is left to the reader as an exercise that this coincides with the previous classification of isometries in the case of $X = \mathrm{H}_\mathbb{K}^n$, where we defined γ to be *loxodromic* (rather than hyperbolic) if it has no fixed point in X and exactly 2 fixed points on ∂X, and *parabolic* if it has no fixed point in X and exactly 1 fixed point on ∂X.

The following result describes the structure of hyperbolic isometry groups for which some point is not moved very much by the generators; informally one can think of it as giving a uniform lower bound on the size of an open ball contained in any hyperbolic orbifold of a given dimension. It is usually called the Margulis lemma, but is essentially due to Kazhdan-Margulis, based on earlier work by Zassenhaus.

For the statement we recall the definitions of nilpotent and virtually nilpotent groups. A group G is *nilpotent* if its lower central series terminates, i.e. if there exists $n \geqslant 1$ such that $G'_n = \mathrm{Id}$, where the lower central series (G'_n) is defined inductively as follows:

$$\begin{cases} G'_0 = G \\ G'_{n+1} = [G'_n, G], \end{cases}$$

where as before $[H, K]$ denotes the subgroup generated by all commutators of the form $[h, k] = hkh^{-1}k^{-1}$ with $h \in H$ and $k \in K$. The group G is called *virtually nilpotent* if it has a nilpotent subgroup of finite index.

Theorem 2.3.4 (Margulis lemma) *Given $\mathbb{K} = \mathbb{R}$ or $\mathbb{C}$ and $n \geqslant 1$, there exists a constant $\mu = \mu(\mathbb{K}, n)$ such that, for any discrete subgroup $\Gamma < \mathrm{Isom}(\mathrm{H}^n_{\mathbb{K}})$, if there exists a point $x \in \mathrm{H}^n_{\mathbb{K}}$ and a generating set S for Γ such that $d(x, \gamma x) \leqslant \mu$ for all $\gamma \in S$, then Γ is virtually nilpotent.*

2.3.3 Hyperbolic lattices

If G is a Lie group (or more generally, a locally compact topological group), a subgroup $\Gamma < G$ is called a *lattice in G* if Γ is discrete in G and G/Γ carries a finite G-invariant measure (where G acts on G/Γ by left translation). This holds in particular if G/Γ is compact.

For a Lie group G with a *Haar measure* (a regular measure invariant under left- or right-translation), this is equivalent to requiring that the Haar measure on G descends to a finite measure on G/Γ. For a connected semisimple Lie group G with associated symmetric space $X = K\backslash G$ (e.g. $G = \mathrm{Isom}^+(\mathrm{H}^n_{\mathbb{K}})$ and $X = \mathrm{H}^n_{\mathbb{K}}$), this is in turn equivalent to the condition that Γ acts on X with finite covolume, i.e. that X/Γ has finite volume.

Examples: The most basic example of a lattice is $\mathbb{Z}$ in $\mathbb{R}$, and more generally $\mathbb{Z}^n$ in $\mathbb{R}^n$. We'll see many other examples in the remaining sections, including the following: $\mathrm{SL}(2, \mathbb{Z})$ is a lattice in $\mathrm{SL}(2, \mathbb{R})$, more generally $\mathrm{SL}(n, \mathbb{Z})$ is a lattice in $\mathrm{SL}(n, \mathbb{R})$ and this is the prototype of what is called an *arithmetic lattice*. We will see other examples provided by reflection groups (which also generalize $\mathrm{SL}(2, \mathbb{Z})$, in a different direction).

The following notion gives a convenient way of showing that a group acting on a topological space is finitely generated or finitely presented. Assume Γ is a group acting properly discontinuously on a topological space X. A subset $D \subset X$ is called a *coarse fundamental domain* for the action of Γ on X if:

$$(a) \ \bigcup_{\gamma \in \Gamma} \gamma D = X, \ \text{and}$$

$$(b) \ \{\gamma \in \Gamma \,|\, \gamma D \cap D \neq \emptyset\} \ \text{is finite.}$$

The fact that Γ is finitely generated or finitely presented can be obtained via the structure of a coarse fundamental domain as follows (see [WM] for this formulation, and ideas going back to Weil):

Proposition 2.3.1 *Let Γ be a group acting properly discontinuously on a topological space X.*
(1) If X is connected and there exists an open coarse fundamental domain for Γ acting on X, then Γ is finitely generated.
(2) If X is simply-connected and there exists a connected open coarse fundamental domain for Γ acting on X, then Γ is finitely presented.

The idea for part (1) is to show that Γ is generated by the elements γ such that $\gamma D \cap D \neq \emptyset$. The second part is more complicated, and in fact for hyperbolic lattices (more precisely, lattices in rank-1 semisimple real Lie groups), Garland and Raghunathan proved the following result by producing a general construction of fundamental domains with finitely many faces for such groups:

Theorem 2.3.5 *Lattices in $\mathrm{Isom}(\mathrm{H}_{\mathbb{K}}^n)$ ($\mathbb{K} = \mathbb{R}, \mathbb{C}$, $n \geqslant 1$) are finitely presented.*

The following result is fundamental for the study of the geometry and topology of hyperbolic manifolds. It states that finite-volume hyperbolic manifolds are determined up to isometry by their fundamental group, in dimensions $\geqslant 3$. This is a strong rigidity result, implying in particular that one cannot deform the hyperbolic structure on a hyperbolic manifold with fixed topology – except in the case of surfaces, see below (the latter statement is known as *local rigidity*). In general it is not even known if (closed, aspherical) manifolds with isomorphic fundamental groups are necessarily homeomorphic; this is known as the *Borel conjecture*. The following result is originally due to Mostow, in the case where $\mathrm{H}_{\mathbb{K}}^n/\Gamma$ is compact, and was extended by Prasad to the case where $\mathrm{H}_{\mathbb{K}}^n/\Gamma$ is non-compact bu thas finite volume.

Theorem 2.3.6 (Mostow rigidity) *Let Γ_1, Γ_2 be two lattices in $\mathrm{Isom}(\mathrm{H}_{\mathbb{K}}^n)$ (with $n \geqslant 3$ if $\mathbb{K} = \mathbb{R}$ and $n \geqslant 2$ if $\mathbb{K} = \mathbb{C}$). If Γ_1, Γ_2 are isomorphic as groups then there exists $g \in \mathrm{Isom}(\mathrm{H}_{\mathbb{K}}^n)$ such that $\Gamma_2 = g\Gamma_1 g^{-1}$; in particular the orbifolds $\mathrm{H}_{\mathbb{K}}^n/\Gamma_1$ and $\mathrm{H}_{\mathbb{K}}^n/\Gamma_2$ are isometric.*

The restriction on dimension is necessary because it is well-known that, in the case of $\mathrm{H}_{\mathbb{R}}^2 \simeq \mathrm{H}_{\mathbb{C}}^1$, hyperbolic surfaces can be deformed (in fact, they have a rich deformation theory). More precisely, given a closed srface Σ_g with $g \geqslant 2$, there exists a $(6g - 6)$-dimensional space of pairwise non-isometric hyperbolic structures on Σ_g, called the *Teichmüller space* of Σ_g (and this space is path-connected, in fact homeomorphic to $\mathbb{R}^{6g-6}$, so any 2 such structres can be defomed to each other). Succinctly, this can be seen by decomposing the surface into $2g-2$ *pairs of pants* (spheres minus 3 disks, with geodesic boundary). The hyperbolic structure on eacg such pair of pants is determined by 3 real numbers (the lengths of the boundary components, or cuffs); when we glue the various pieces by isometries we require the lengths of the cuffs that we glue to be equal, but there is an extra "twist" parameter at each such gluing. The $6g - 6$ length and twist parameters are called *Fenchel-Nielsen coordinates* on Teichmüller space.

2.4 Arithmetic lattices

We start by giving examples of arithmetic lattices, in families of increasing complexity, before giving the general definition of arithmetic lattices in linear algebraic groups and semisimple real

Lie groups.

(1) First examples:

If we take a discrete subring such as $\mathbb{Z} \subset \mathbb{R}$ or $\mathcal{O}_d \subset \mathbb{C}$ (the ring of integers of $\mathbb{Q}[\sqrt{-d}]$ for some square-free integer d, e.g. $\mathbb{Z}[i]$ when $d = 1$ and $\mathbb{Z}[e^{2i\pi/3}]$ when $d = 3$), we get discrete matrix groups by taking entries in the discrete subring, e.g. $\mathrm{SL}(n, \mathbb{Z}) < \mathrm{SL}(n, \mathbb{R})$ and $\mathrm{SL}(n, \mathcal{O}_d) < \mathrm{SL}(n, \mathbb{C})$, or in the hyperbolic world $\mathrm{SO}(n, 1, \mathbb{Z}) < \mathrm{SO}(n, 1)$ and $\mathrm{SU}(n, 1, \mathcal{O}_d) < \mathrm{SU}(n, 1)$. Discreteness of such subgroups is obvious (because the topology on the matrix group is the same as the product topology on the entries, i.e. a sequence of matrices converges to Id if and only if each entry converges to the correponding entry of Id). It turns out that all of the subgroups above are in fact lattices in the corresponding Lie group; in full generality this follows from the Borel–Harish-Chandra theorem (Theorem 2.4.1 below), but some cases were previously known. The case of $\mathrm{SL}(2, \mathbb{Z}) < \mathrm{SL}(2, \mathbb{R})$ is essentially due to Gauss, and Siegel proved that $\Gamma = \mathrm{SL}(n, \mathbb{Z})$ has finite covolume in $G = \mathrm{SL}(n, \mathbb{R})$ by exhibiting an explicit region (in coordinates in the KAN decomposition, now called a *Siegel region*) with finite Haar measure in G, whose cosets under Γ cover G.

(2) The Galois trick, part I:

The subgroup $\mathbb{Z}[\sqrt{2}]$ is dense in $\mathbb{R}$ (this follows from the fact that additive subgroups of $\mathbb{R}$ are either cyclic or dense). However, one can realize $\mathbb{Z}[\sqrt{2}]$ as a discrete subgroup (in fact a lattice) of $\mathbb{R}^2$, by the following embedding:

$$\varphi : \quad \begin{aligned} \mathbb{Z}[\sqrt{2}] &\longrightarrow \mathbb{R}^2 \\ (a + b\sqrt{2}) &\longmapsto (a + b\sqrt{2}, a - b\sqrt{2}). \end{aligned}$$

The fact that $\mathrm{Im}\,\varphi$ is a lattice in $\mathbb{R}^2$ can be easily seen by writing it as $\mathbb{Z}(1, 1)^T \oplus \sqrt{2}\mathbb{Z}(1, -1)^T$. Note that the projection of $\mathrm{Im}\,\varphi$ to each factor is dense.

Analogously, denoting $\sigma : a + b\sqrt{2} \longmapsto a - b\sqrt{2}$ the non-trivial Galois conjugation in $\mathbb{Z}[\sqrt{2}]$ and $^\sigma M$ the matrix obtained from a matrix M by applying σ to all entries, the embedding:

$$\varphi : \quad \begin{aligned} \mathrm{SL}(2, \mathbb{Z}[\sqrt{2}]) &\longrightarrow \mathrm{SL}(2, \mathbb{R}) \times \mathrm{SL}(2, \mathbb{R}) \\ M &\longmapsto (M, {}^\sigma M) \end{aligned}$$

realizes $\mathrm{SL}(2, \mathbb{Z}[\sqrt{2}])$ as a lattice $\Gamma = \varphi\big(\mathrm{SL}(2, \mathbb{Z}[\sqrt{2}])\big)$ in $\mathrm{SL}(2, \mathbb{R}) \times \mathrm{SL}(2, \mathbb{R})$ (this will follow from the Borel–Harish-Chandra theorem). Such quotients $(\mathrm{H}^2_{\mathbb{R}} \times \mathrm{H}^2_{\mathbb{R}})/\Gamma$ are called *Hilbert modular surfaces*.

More generally, for any totally real number field E with $\mathrm{Gal}(E/\mathbb{Q}) = \{\sigma_1, ..., \sigma_r\}$ (and $n \geqslant 1$), we have an embedding:

$$\varphi : \quad \begin{aligned} \mathrm{SL}(n, \mathcal{O}_E) &\longrightarrow \mathrm{SL}(n, \mathbb{R}) \times ... \times \mathrm{SL}(n, \mathbb{R}) \\ M &\longmapsto ({}^{\sigma_1} M, ..., {}^{\sigma_r} M). \end{aligned}$$

The Borel–Harish-Chandra theorem will tell us that $\varphi\big(\mathrm{SL}(n, \mathcal{O}_E)\big)$ is a lattice in $\mathrm{SL}(n, \mathbb{R}) \times ... \times \mathrm{SL}(n, \mathbb{R})$. This is perfectly fine, but this construction produces lattices in *higher-rank* Lie groups, whereas the hyperbolic isometry groups we are interested in have rank 1. More precisely, the *real rank* of a linear Lie group $G < \mathrm{GL}(n, \mathbb{C})$ is the maximal dimension of an $\mathbb{R}$-*split torus*, i.e. a closed, connected subgroup of G all of whose elements are simultaneously diagonalizable over $\mathbb{R}$. For example, a maximal $\mathbb{R}$-split torus in $\mathrm{SL}(n, \mathbb{R})$ is the subgroup of diagonal matrices, so $\mathrm{SL}(n, \mathbb{R})$ has real rank $n - 1$. Geometrically, if G is semisimple the real rank of G can be seen in the symmetric space $X = G/K$ associated to G (where K is a maximal compact subgroup of G), namely the real rank of G is then the maximal dimension of a *flat* in X, which is an isometrically embedded, totally geodesic Euclidean subspace. This shows that the hyperbolic spaces $\mathrm{H}_{\mathbb{K}}^n$ have real rank 1, as do their isometry groups which are (up to finite index) $\mathrm{SO}(n, 1)$ and $\mathrm{SU}(n, 1)$. Note that real rank is additive under products, so even in the case of $\mathrm{SL}(2, \mathbb{R})$ (which has rank 1), when we take products of more than one copy we get a higher-rank Lie group. However the Galois trick can be upgraded to produce lattices in rank 1 Lie groups, as we now illustrate.

(3) The Galois trick, part II:

When considering Lie groups definied by a quadratic or Hermitian form such as $\mathrm{SO}(n, 1)$ and $\mathrm{SU}(n, 1)$, the second part of the Galois trick is to use a form which changes under the Galois conjugations of the number field under consideration. We start again with the example of $E = \mathbb{Q}[\sqrt{2}]$ with its ring of integers $\mathcal{O}_E = \mathbb{Z}[\sqrt{2}]$. The quadratic form on $\mathbb{R}^{n+1}$ given by the symmetric matrix $Q = \mathrm{Diag}(1, ..., 1, -\sqrt{2})$ has signature $(n, 1)$, so the group $\mathrm{SO}(Q)$ is conjugate to $\mathrm{SO}(n, 1)$ in $\mathrm{GL}(n + 1, \mathbb{R})$, whereas the quadratic form corresponding to $^\sigma Q = \mathrm{Diag}(1, ..., 1, \sqrt{2})$ has signature $(n + 1, 0)$, so the group $\mathrm{SO}(^\sigma Q)$ is conjugate to $\mathrm{SO}(n + 1)$ and in particular is compact. As above, the Borel–Harish-Chandra theorem will tell us that $\varphi\big(\mathrm{SO}(Q, \mathbb{Z}[\sqrt{2}])\big)$ is a lattice in $\mathrm{SO}(Q) \times \mathrm{SO}(^\sigma Q)$, where as before:

$$\varphi: \quad \mathrm{SO}(Q, \mathbb{Z}[\sqrt{2}]) \quad \longrightarrow \quad \mathrm{SO}(Q) \times \mathrm{SO}(^\sigma Q)$$
$$M \quad \longmapsto \quad (M, {}^\sigma M).$$

But now the projection to the first factor $p_1 : \mathrm{SO}(Q) \times \mathrm{SO}(^\sigma Q) \longrightarrow \mathrm{SO}(Q)$ has compact kernel, hence it maps discrete sets to discrete sets, and lattices to lattices, by the following lemma, whose proof is left to the reader as an excercise.

Lemma 2.4.1 *Let G,H be 2 (locally compact) topological groups, and $\phi : G \longrightarrow H$ a continuous homomorphism with compact kernel. Then, if $\Gamma < G$ is a discrete subgroup (resp. a lattice), then $\phi(\Gamma) < H$ is a discrete subgroup (resp. a lattice).*

Therefore $\mathrm{SO}(Q, \mathbb{Z}[\sqrt{2}])$ is a lattice in $\mathrm{SO}(Q) \simeq \mathrm{SO}(n, 1)$. More generally, the same argument gives the following:

Proposition 2.4.1 *Let E be a totally real number field and Q a quadratic form of signature $(n, 1)$ with coefficents in E. If for all $\sigma \in \mathrm{Gal}(E/\mathbb{Q}) \setminus \{\mathrm{Id}\}$ the form $^\sigma Q$ is definite then $\mathrm{SO}(Q, \mathcal{O}_E)$ is a a lattice in $\mathrm{SO}(Q) \simeq \mathrm{SO}(n, 1)$.*

Sketch of proof: The Borel–Harish-Chandra theorem will tell us as before that $\varphi\big(\mathrm{SO}(Q, \mathcal{O}_E)\big)$ is a lattice in $\mathrm{SO}(Q) \times \mathrm{SO}(^{\sigma_1}Q) \times ... \times \mathrm{SO}(^{\sigma_r}Q)$, denoting $\mathrm{Gal}(E/\mathbb{Q}) = \{\mathrm{Id}, \sigma_1, ..., \sigma_r\}$, via the embedding:

$$\varphi : \begin{array}{ccc} \mathrm{SO}(Q, \mathcal{O}_E) & \longrightarrow & \mathrm{SO}(Q) \times \mathrm{SO}(^{\sigma_1}Q) \times ... \times \mathrm{SO}(^{\sigma_r}Q) \\ M & \longmapsto & (M, ^{\sigma_1}M, ..., ^{\sigma_r}M). \end{array}$$

But the projection to the first factor $p_1 : \mathrm{SO}(Q) \times \mathrm{SO}(^{\sigma_1}Q) \times ... \times \mathrm{SO}(^{\sigma_r}Q) \longrightarrow \mathrm{SO}(Q)$ has compact kernel, by the assumption that all the nontrivial Galois conjugates $^\sigma Q$ are positive definite or negative definite, hence $\mathrm{SO}(Q, \mathcal{O}_E)$ is a a lattice in $\mathrm{SO}(Q)$ by Lemma 2.4.1. □

The analogous statement in the unitary setting is the following:

Proposition 2.4.2 *Let E be a purely imaginary quadratic extension of a totally real number field F and H a Hermitian form of signature $(n, 1)$ with coefficients in E. If for all $\sigma \in \mathrm{Gal}(E/\mathbb{Q})$ not inducing the identity on F the form $^\sigma H$ is definite then $\mathrm{SU}(H, \mathcal{O}_E)$ is a a lattice in $\mathrm{SU}(H) \simeq \mathrm{SU}(n, 1)$.*

This is in fact a characterization of all arithmetic lattices in $\mathrm{GL}(n+1, \mathbb{C})$ with entries in $\mathcal{O}_E$ and preserving a Hermitian form of signature $(n, 1)$ defined over E, see Proposition 2.7.4 below.

(4) General arithmetic lattices:

Recall that a *real linear algebraic group defined over $\mathbb{Q}$* is a subgroup G of $\mathrm{GL}(n, \mathbb{R})$ for some n, such that the elements of G are precisely the solutions of a set of polynomial equations in the entries of the matrices, with the coefficients of the polynomials lying in $\mathbb{Q}$; one denotes $G(\mathbb{R}) = G$ and $G(\mathbb{Z}) = G \cap \mathrm{GL}(n, \mathbb{Z})$.

Theorem 2.4.1 (Borel–Harish-Chandra) *If G is a linear algebraic group defined over $\mathbb{Q}$ then $G(\mathbb{Z})$ is a lattice in $G(\mathbb{R})$.*

This result implies that any real semisimple Lie group has infinitely many (distinct commensurability classes of) lattices, both cocompact and non cocompact, see [WM].

One then obtains the general definition by extending this notion to all groups equivalent to such groups $G(\mathbb{Z})$ in the following sense:

Definition: Let G be a semisimple Lie group, and Γ a subgroup of G. Then Γ is an *arithmetic lattice* in G if there exist an algebraic group S defined over $\mathbb{Q}$ and a continuous homomorphism $\phi : S(\mathbb{R})^0 \longrightarrow G$ with compact kernel such that Γ is commensurable to $\phi(S(\mathbb{Z}) \cap S(\mathbb{R})^0)$.

Note that by definition, if $\Gamma_1, \Gamma_2 < G$ are commensurable then Γ_1 is an arithmetic lattice in G if and only if Γ_2 is. (We say that 2 subgroups $\Gamma_1, \Gamma_2 < G$ are *commensurable* if $\Gamma_1 \cap g\Gamma_2 g^{-1}$ has finite index in Γ_1 and in $g\Gamma_2 g^{-1}$ for some $g \in G$).

The fact that a subgroup Γ as in the definition is a lattice in G follows from the Borel–Harish-Chandra theorem. We now briefly explain why the groups constructed with the Galois tricks, parts I and II above, are instances of this general definition.

The main ingredient is *restriction of scalars*; in its simplest form, this is the fact that in a field extension E/F, E naturally has the structure of a vector space over F. In the particular

case where E is a number field (i.e. a finite extension of $\mathbb{Q}$), denoting $d = [E : \mathbb{Q}]$, selecting a basis for E over $\mathbb{Q}$ gives an embedding $i : E \hookrightarrow \mathrm{GL}(d, \mathbb{Q})$ by sending any element $x \in E$ to multiplication by x (seen as a $\mathbb{Q}$-linear endomorphism of the $\mathbb{Q}$-vector space E). Now if we choose this basis to be a $\mathbb{Z}$-basis for $\mathcal{O}_E$, then additionally: $i(\mathcal{O}_E) = i(F) \cap \mathrm{GL}(d, \mathbb{Z})$.

The groups appearing in the Galois trick part I are of this form, taking SL rather than GL which is obviously a linear algebraic group defined over $\mathbb{Q}$ (as the polynomial defining it is the determinant, which is an integer polynomial in the entries of a matrix). For the groups appearing in the Galois trick, part II, we only need to argue additionally that the group $S = \mathrm{SO}(Q) \times \mathrm{SO}(^{\sigma_1}Q) \times ... \times \mathrm{SO}(^{\sigma_r}Q)$ (respectively $S = \mathrm{SU}(H) \times \mathrm{SU}(^{\sigma_1}H) \times ... \times \mathrm{SU}(^{\sigma_r}H)$) is a linear algebraic group defined over $\mathbb{Q}$; the continuous homomorphism $\phi : S(\mathbb{R})^0 \longrightarrow G$ with compact kernel appearing in the definition of an arithmetic lattice is then simply projection onto the first factor of S. (As pointed out above this has compact kernel by the assumption that all non-trivial Galois conjugates of Q (resp. H) are definite).

The fact that such groups S are linear algebraic groups defined over $\mathbb{Q}$ follows again from restriction of scalars. More precisely, a product of the form $S = \mathrm{SO}(Q) \times \mathrm{SO}(^{\sigma_1}Q) \times ... \times \mathrm{SO}(^{\sigma_r}Q)$ can be embedded in block form into $\mathrm{SL}((n + 1)(r + 1), \mathbb{R})$ (where $(n + 1)$ is the rank of Q and $(r + 1)$ the number of factors). The condition of being a block matrix of a given shape is given by polynomial equations over $\mathbb{Q}$ in the entries (namely, that the appropriate entries are 0). Now for each block $A_i \in \mathrm{SL}(n + 1, \mathbb{R})$, the condition of being in $\mathrm{SO}(^{\sigma_i}Q)$ is given by the equation $A_i^T \, {}^{\sigma_i}Q A_i = \, {}^{\sigma_i}Q$, which corresponds to $(n + 1)^2$ polynomial equations over E in the entries of A_i. Now as above, choosing a $\mathbb{Q}$-basis for E allows us to write each of these equations as a system of d polynomial equations over $\mathbb{Q}$. Putting the pieces back together, we see that $S = \mathrm{SO}(Q) \times \mathrm{SO}(^{\sigma_1}Q) \times ... \times \mathrm{SO}(^{\sigma_r}Q)$ is a subgroup of $\mathrm{SL}((n+1)(r+1), \mathbb{R})$ given by a collection of polynomial equations over $\mathbb{Q}$, i.e. S is a linear algebraic group defined over $\mathbb{Q}$.

2.5 Arithmeticity vs. non-arithmeticity

It turns out that the arithmetic construction is the only way to produce lattices in higher-rank simple Lie groups; this is the following deep and celebrated result of Margulis, for which he was awarded a Fields medal in 1978 (see Section 2.4 (2) for the definition of real rank).

Theorem 2.5.1 (Margulis arithmeticity theorem) *If G is a simple real Lie group of real rank at least 2, then any lattice in G is arithmetic.*

This result also holds when G is a semisimple real Lie group of real rank at least 2, under the additional assumption that the lattice is *irreducilble*, i.e. its projection onto each simple factor of G is dense. This excludes cases like $\Gamma \times \Gamma < G \times G$ with Γ a non-arithmetic lattice in a rank-1 simple group G, e.g. $G = \mathrm{PSL}(2, \mathbb{R})$.

We now examine to which extent the simple Lie groups of real rank 1, where the Margulis arithmeticity theorem does not apply, are known to admit non-arithmetic lattices. The first observation is the following:

Observation: $\mathrm{SL}(2, \mathbb{R})$ has uncountably many non-arithmetic lattices.

Indeed, as pointed out in the discussion of Mostow rigidity, most lattices in $\mathrm{SL}(2,\mathbb{R})$ (including all closed surface groups) have families of continuous deformations to isomorphic but non-conjugate lattices (the Teichmüller space of the surface). But there are only countably many arithmetic lattices up to conjugacy in any given Lie group (finitely many defined over any given number field); this follows for example from Tits' classification of linear algebraic groups defined over $\mathbb{Q}$ ([Ti]), and in the case of $\mathrm{SL}(2,\mathbb{R})$ it follows from a result of Borel ([Borel]) that there are in fact only finitely many arithmetic lattices in the Teichmüller space of a closed surface with fixed genus.

From the classification of simple real Lie groups (due to Killing and E. Cartan), the complete list of simple real Lie groups of real rank 1 is the following. For simplicity, we list groups G which are subgroups of $\mathrm{GL}(n,\mathbb{C})$ for some n; they are not simple but have finite center, the correpsonding simple group is $G/Z(G)$ which is finitely covered by G. We also list the corresponding symmetric spaces G/K (with $K < G$ a maximal compact subgroup), which are exactly the hyperbolic spaces over $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$ and the exotic octave plane.

Simple real Lie groups of rank 1: (up to finite index)

Group:	$\mathrm{SO}(n,1)$	$\mathrm{SU}(n,1)$	$\mathrm{Sp}(n,1)$	$F_4^{(-20)}$
Symmetric space:	$\mathrm{H}_{\mathbb{R}}^n$	$\mathrm{H}_{\mathbb{C}}^n$	$\mathrm{H}_{\mathbb{H}}^n$	$\mathrm{H}_{\mathbb{O}}^2$

The Margulis arithmeticity theorem was extended as follows to the groups corresponding to quaternions and octaves by Corlette and Gromov–Schoen:

Theorem 2.5.2 (Corlette, Gromov–Schoen) *Any lattice in $\mathrm{Sp}(n,1)$ or $F_4^{(-20)}$ is arithmetic.*

In real hyperbolic space, beyond dimension 2, Makarov and Vinberg gave the first examples of non-arithmetic lattices in $\mathrm{SO}(n,1)$ for $n = 3,4,5$ in the 1960's. Their examples were hyperbolic Coxeter groups, whose construction will be discussed in detail in the next section. Subsequently, Gromov and Piatetskii-Shapiro gave in [GPS] a construction producing non-arithmetic lattices in $\mathrm{SO}(n,1)$ for all $n \geqslant 2$ (and in fact, infinitely many, non-commensurable, both cocompact and non-cocompact in each dimension).

Theorem 2.5.3 (Gromov–Piatetskii-Shapiro) *There exist non-arithmetic lattices in $\mathrm{SO}(n,1)$ for all $n \geqslant 2$.*

We briefly sketch the idea of the construction. The Gromov–Piatetski-Shapiro construction, which they call interbreeding of 2 arithmetic lattices (often referred to as hybridation), produces a lattice $\Gamma < \mathrm{PO}(n,1)$ from 2 lattices Γ_1 and Γ_2 in $\mathrm{PO}(n,1)$ which have a common sublattice $\Gamma_{12} < \mathrm{PO}(n-1,1)$. Geometrically, this provides 2 hyperbolic n-manifolds $V_1 = \Gamma_1 \backslash H_{\mathbb{R}}^n$ and $V_2 = \Gamma_2 \backslash H_{\mathbb{R}}^n$ with a hyperbolic $(n-1)$-manifold V_{12} which is isometrically embedded in V_1 and V_2 as a totally geodesic hypersurface. This allows one to produce the hybrid manifold V by gluing $V_1 - V_{12}$ and $V_2 - V_{12}$ along V_{12}. The resulting manifold is also hyperbolic because the gluing took place along a totally geodesic hypersurface, and its fundamental group Γ is therefore a lattice in $\mathrm{PO}(n,1)$. The point is then that if Γ_1 and Γ_2 are both arithmetic but non-commensurable, their hybrid Γ is non-arithmetic.

Much less is known in the remaining case of complex hyperbolic space, corresponding to $SU(n,1)$. The first constructions of complex hyperbolic lattices (beyond dimension 1) go back to Picard, in [Pic1] and [Pic2]. In fact it turns out that some of the groups constructed in [Pic2] are non-arithmetic lattices, but this was only discovered later. Mostow constructed the first known non-arithmetic lattices in $SU(2,1)$ in [Mos1]; we will decsribe his construction in more detail in the last section. Deligne and Mostow subsequently extended Picard's construction (which was based on monodromy groups for so-called hypergeometric functions) up to (complex) dimension 5 in [DM] (Mostow pushed the construction to dimension 9 in [Mos2]). It turns out that, beyond Picard's groups, there is only one non-arithmetic lattice in the Deligne–Mostow construction in dimension 3 (it is non-cocompact). We summarize this in the following:

Theorem 2.5.4 (1) (Mostow) *There exist non-arithmetic lattices in* $SU(2,1)$.
(2) (Deligne–Mostow) *There exists a non-arithmetic lattice in* $SU(3,1)$.

In the 2-dimensional case, the Picard and Mostow groups each provided 7 non-arithmetic lattices; considerations of trace fields show that these fell in between 7 and 9 distinct commensurability classes, and it was shown in [KM] that there are in fact 9. Recently, Deraux, Parker and the author have extended Mostow's construction to produce 5 additional commensurability classes of non-arithmetic lattices in $SU(2,1)$ (see [DPP1], [DPP2] and the last section).

The following outstanding questions remain open in the complex hyperbolic case:

Open questions:

(1) Do there exist non-arithmetic lattices in $SU(n,1)$ when $n \geqslant 4$?

(2) For fixed $n \geqslant 2$, do there exist infinitely many non-commensurable non-arithmetic lattices in $SU(n,1)$?

2.6 Hyperbolic Coxeter groups

Let X denote one of the 3 model spaces of constant curvature, $X = S^n$, E^n or H^n. A *polyhedron* in X is a subset P of X which is bounded by finitely many hyperplanes, in other words P is the intersection of finitely many closed half-spaces. A *wall* of P is a hyperplane H such that $P \cap H$ has dimension $n-1$. In that case we will say that $P \cap H$ is a (maximal) face of P; in turn this polyhedron of dimension $n-1$ has faces, and inductively this defines the set of all faces of P. (Alternatively, in such a constant curvature space one can also construct P from the bottom up, by taking convex hulls of vertices). To avoid all confusion we will call *k-face* of P any of its faces of dimension k.

A *Coxeter polyhedron* in X is a polyhedron P all of whose dihedral angles are submultiples of π. (This means that if H_i and H_j are 2 walls of P, then either they are disjoint or at any intersection point their outward-pointing normals form an angle of measure $\pi - \pi/m_{ij}$ where $m_{ij} \in \mathbb{N} \cup \{\infty\}$). The point of this condition is that the reflections across the walls of such a polyhedron generate a discrete group of isometries of X, for which P is a fundamental polyhedron. More precisely, we have the following result which is due to Poincaré when $X = H^2$, to Coxeter when $X = S^n$ or E^n, and in general follows from the more general version of what is now called the Poincaré polyhedron theorem (see section 2.7.5 and [EP]).

Figure 2.1: Coxeter diagram for H_5, a simplex in $\mathrm{H}^4_{\mathbb{R}}$.
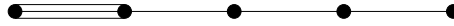
Theorem 2.6.1 *Let P be a Coxeter polyhedron in $X = \mathrm{S}^n$, E^n or H^n, denote $H_1, ..., H_k$ the walls of P, $r_1, ..., r_k$ the reflections across these walls and π/m_{ij} the dihedral angle between the walls H_i and H_j (where $m_{ij} \in \mathbb{N} \cup \{\infty\}$, with the convention that $m_{ii} = 2$ for all i). Then:*
(1) $\Gamma = \langle r_1, ..., r_k \rangle$ is a discrete subgroup of $\mathrm{Isom}(X)$,
(2) P is a fundamental domain for the action of Γ on X, and
(3) Γ admits the presentation $\langle r_1, ..., r_k | (r_i r_j)^{m_{ij}} = 1 \ (\text{for } 1 \leqslant i \leqslant j \leqslant k) \rangle$.

An abstract group given by a presentation of the form in part (3) is called a *Coxeter group*.

A convenient way to encode the information about such a Coxeter group Γ is by means of the associated *Coxeter diagram* D_Γ ([Co]) which is a graph with labeled edges, constructed as follows. The vertices of D_Γ correspond to the generating reflections $r_1, ..., r_k$; two distinct vertices corresponding to r_i and r_j are joined by an edge if (a) $m_{ij} \geqslant 3$, in which case the edge is labeled m_{ij} (alternatively, an edge of multiplicity $m_{ij} - 2$ is drawn), or (b) by a bold (resp. dotted) edge if the walls of r_i and r_j are parallel (resp. ultraparallel), in the hyperbolic case. The point of this convention is that connected components of D_Γ correspond to irreducible factors of Γ.

The numbers m_{ij} also determine a quadratic form via a symmetric matrix called the *Gram matrix* of the polyhedron or group. Namely, if Γ is generated by the k reflections $r_1, ..., r_k$ with m_{ij} as above, then the Gram matrix of Γ is the symmetric $k \times k$ matrix with 1's on the diagonal and (i,j)-entry equal to $-\cos(\pi/m_{ij})$ (resp. $-\cosh d(H_i, H_j)$ if the walls H_i and H_j are parallel or ultraparallel). This is the Gram matrix of the outward-pointing normal vectors to the walls of P. One can then define invariants such as the *determinant, rank, signature* of P (or Γ or D_Γ) as the corresponding invariant of its Gram matrix. Accordingly, we will say that D_Γ is *hyperbolic* if its Gram matrix is of signature $(n, 1)$ for some $n \geqslant 1$, *parabolic* if it is positive semidefinite (but not definite), and *elliptic* if it is positive definite.

The Gram matrix represents a quadratic form on $\mathbb{R}^k$ which is preserved by the so-called "geometric representation" (see e.g. [H]). More relevant from our point of view is the representation that we started with (of Γ in $\mathrm{Isom}(\mathrm{H}^n) < \mathrm{PGL}(n+1, \mathbb{R})$ arising from the polyhedron P in H^n). These representations coincide only when the polyhedron is a simplex, that is when the Gram matrix has maximal rank.

Elliptic and parabolic Coxeter diagrams were classified by Coxeter in 1934 ([Co], see pp. 202–203 of [V1] and [H]), whereas no such classification is known for hyperbolic diagrams, except in dimensions 2 (Poincaré) and 3 by Andreev's theorem ([An]). We now state some partial results, see [V1] for more details. Vinberg has proved the following general existence result for acute-angled hyperbolic polyhedra (Theorem 2.1 of [V2]):

Theorem 2.6.2 (Vinberg) *Let G be an indecomposable symmetric matrix of signature $(n, 1)$ with 1's on the diagonal and non-positive entries off it. Then there exists a convex polyhedron in H^n whose Gram matrix is G, and this polyhedron is unique up to isometry.*

The following results are in striking contrast to the Euclidean and spherical cases, for which compact Coxeter polyhedra exist in all dimensions:

Figure 2.2: Coxeter diagram for Bugaenko's compact polyhedron in $H_{\mathbb{R}}^8$.
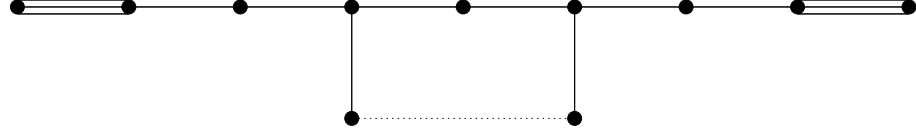
Theorem 2.6.3 (Vinberg) *There do not exist compact Coxeter polyhedra in H^n when $n \geqslant 30$.*

Theorem 2.6.4 (Prokhorov-Khovanskij) *There do not exist finite-volume Coxeter polyhedra in H^n when $n \geqslant 996$.*

There is a large gap between the dimension bounds given in these theorems and the dimensions where compact/finite-volume Coxeter polyhedra are known to exist. The largest dimension of a known compact hyperbolic Coxeter polyhedron is 8 (due to Bugaenko, [Bu]), whose Coxeter diagram is pictured in Figure 2.2; the largest dimension of a known finite-volume hyperbolic Coxeter polyhedron is 21 (due to Borcherds, [Borch]). Both of these are arithmetic (i.e. the corresponding reflection group is an arithmetic lattice). Non-arithmetic compact Coxeter polyhedra are known to exist in H^n for $n \leqslant 5$ (the first examples were due to Makarov [Mak] for $n = 3$, others can be found among the simplex groups); non-arithmetic finite-volume Coxeter polyhedra are known to exist in H^n for $n \leqslant 10$ (Ruzmanov, [Ru]).

We now review some results concerning finiteness of finite-volume Coxeter polyhedra/groups in any given dimension. Allcock has proved the following ([Al]):

Theorem 2.6.5 (Allcock) *For all $3 \leqslant n \leqslant 19$, there exist infinitely many non-isometric finite-volume Coxeter polyhedra in H^n.*

The idea of Allcock's construction is to start with a given Coxeter polyhedron, such that all the dihedral faces along a given face are right angles; such a polyhedron is called *doublable*: one can reflect it along that face to obtain a new Coxeter polyhedron. Allcock finds polyhedra which are *redoublable* in the sense that they satisfy the right-angled condition along 2 disjoint faces, which allows to double it successively to produce an infinite sequence of pairwise distinct Coxeter polyhedra. Note that by construction, all of the doubled/redoubled polyhedra are commensurable to each other, and it is still unknown whether or not there exist infinitely many distinct commensurability classes of finite-volume Coxeter polyhedra in H^n for $n \geqslant 4$ (for $n = 2$ triangle groups provide infinitely many such groups; for $n = 3$ it follows from Andreev's theorem [An] that there exist infinitely many such groups). However, it is known that the answer is negative for arithmetic Coxeter groups; the following result is due to Borel for $n = 2$ with bounded genus ([Borel]), Agol for $n = 3$ ([Ag]) and independently Agol–Belolipetsky–Storm–Whyte ([ABSW]) and Nikulin ([N]) for $n \geqslant 4$:

Theorem 2.6.6 *For any $n \geqslant 2$, there exist only finitely many maximal arithmetic finite-volume Coxeter polyhedra in H^n.*

2.7 Some non-arithmetic lattices in $\mathrm{Isom}(\mathrm{H}_\mathbb{C}^2)$

In this section we give a brief overview of Mostow's construction of non-arithmetic lattices in $\mathrm{SU}(2,1)$ from [Mos1], followed by a related construction by Deraux, Parker and the author from [ParPau], [DPP1] and [DPP2].

2.7.1 Mostow's lattices

Notation: $\Gamma(p,t) < \mathrm{SU}(2,1)$, where $p = 3, 4$ or 5 and t is a real parameter.

The $\Gamma(p,t)$ are *symmetric complex reflection triangle groups*, i.e.:

- $\Gamma = \langle R_1, R_2, R_3 \rangle$ where each R_i is a complex reflection of order p.

- *symmetric* means that there exists an isometry J of order 3 such that $JR_iJ^{-1} = R_{i+1}$, or equivalently $J(L_i) = L_{i+1}$ where $L_i = \mathrm{Fix}(R_i)$.

Moreover Mostow imposes the *braid relation* $R_iR_jR_i = R_jR_iR_j$.

Facts:

- For fixed p there is a 1-dimensional family of such groups (hence the t).

- Only finitely many of the $\Gamma(p,t)$ are discrete; the discrete ones are lattices.

The second statement is the hardest, and is obtained by a numerical analysis of Dirichlet domains for the groups $\Gamma(p,t)$ acting on complex hyperbolic space. See [De] for a discussion of the complexity of the corresponding Ddirichlet domains, and see [DFP] for a simpler construction of fundamental domains for Mostow's lattices. This produces 15 lattices in $\mathrm{SU}(2,1)$, 7 of which are non-arithmetic (see the Vinberg–Mostow arithmeticity criterion below). For concreteness, the following are presentations for the Mostow lattices (more precisely, of the group $\langle J, R_1 \rangle$, see below) obtained in [DFP]:

$$\tilde\Gamma(p,t) = \langle J, R_1, R_2 \mid J^3 = R_1^p = R_2^p = J^{-1}R_2JR_1^{-1} = R_1R_2R_1R_2^{-1}R_1^{-1}R_2^{-1}$$
$$= (R_2R_1J)^k = ((R_1R_2)^{-1}J)^l = I \rangle,$$

where $p = 3, 4$ or 5, $k = (\frac{1}{4} - \frac{1}{2p} + \frac{t}{2})^{-1}$ and $l = (\frac{1}{4} - \frac{1}{2p} - \frac{t}{2})^{-1}$.

2.7.2 Configuration spaces of symmetric complex reflection triangle groups

We now work with $\tilde\Gamma = \langle R_1, J \rangle$ which contains Γ with index 1 or 3. We now drop the braid relation.

Fact: (For fixed p) the space of such groups has dimension 2. More precisely:

Proposition 2.7.1 *Let R_1 be a complex reflection and J a regular elliptic isometry of order 3 in $\mathrm{PU}(2,1)$. Then $\tilde\Gamma = \langle R_1, J \rangle$ is determined up to conjugacy by the conjugacy class of R_1J.*

In concrete terms, we parametrize the conjugacy class of R_1J by either:

- $\tau := \mathrm{Tr}R_1J$ (good for arithmetic), or

- the angle pair $\{\theta_1, \theta_2\}$ of R_1J when elliptic (good for geometry).

Recall that the *angle pair* of an elliptic element of $\mathrm{PU}(2,1)$ is the pair of angles of the eigenvalues of the correpsonding element of $\mathrm{U}(2)$.

Figure 2.3 illustrates the admissible angle pairs $\{\theta_1, \theta_2\}$ for elliptic products R_1J when R_1 is a complex reflection through angle $2\pi/p$ and J a regular elliptic isometry of order 3, for $p = 3, .., 10$. For each fixed value of p the admissible region is the polygon bounded by the blue line segments; the angle pairs corresponding to Mostow's lattices (resp. to normal subgroups of Mostow's lattices) comprise the green (resp. red) segments. The crosses indicate the sporadic groups, see below.

Notation: We denote $\Gamma(2\pi/p, \tau) = \langle R_1, J \rangle$, where R_1 is a complex reflection through angle $2\pi/p$, J a regular elliptic isometry of order 3, and $\tau := \mathrm{Tr}R_1J$. For concreteness we give explicit matrices for the generators and Hermitian form, denoting $\psi = 2\pi/p$:

$$J = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{2.7.1}$$

$$R_1 = \begin{bmatrix} e^{2i\psi/3} & \tau & -e^{i\psi/3}\,\overline{\tau} \\ 0 & e^{-i\psi/3} & 0 \\ 0 & 0 & e^{-i\psi/3} \end{bmatrix} \tag{2.7.2}$$

These preserve the Hermitian form $\langle \mathbf{z}, \mathbf{w} \rangle = \mathbf{w}^* H_\tau \mathbf{z}$ where

$$H_\tau = \begin{bmatrix} 2\sin(\psi/2) & -ie^{-i\psi/6}\tau & ie^{i\psi/6}\overline{\tau} \\ ie^{i\psi/6}\overline{\tau} & 2\sin(\psi/2) & -ie^{-i\psi/6}\tau \\ -ie^{-i\psi/6}\tau & ie^{i\psi/6}\overline{\tau} & 2\sin(\psi/2) \end{bmatrix}. \tag{2.7.3}$$
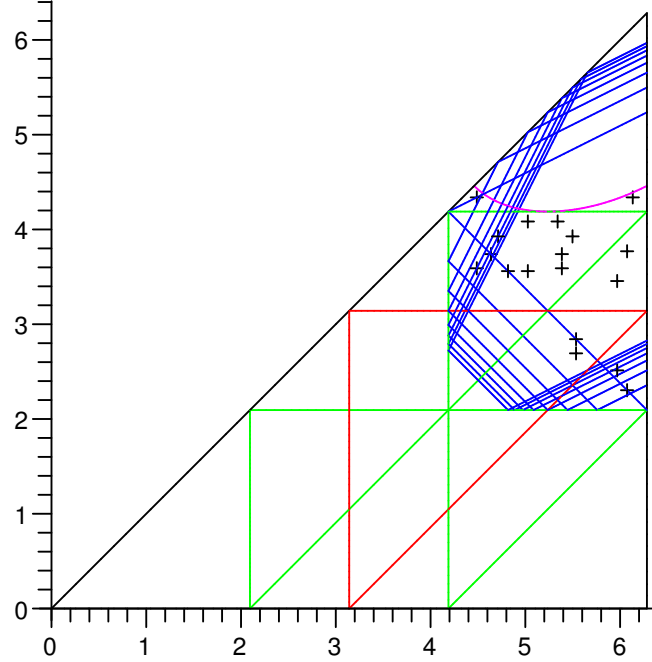
2.7.3   Sporadic groups

The following result ([ParPau]) gives a *necessary* condition for dicscreteness, under the assumption that 2 given words in the group are elliptic or parabolic (see [Sc] for a discussion of why these words are relevant), and provides the motivation to study the so-called sporadic groups.

Theorem 2.7.1 *Let R_1 be a complex reflection and J a regular elliptic isometry of order 3 in $\mathrm{PU}(2,1)$. Suppose that R_1J and $R_1R_2 = R_1JR_1J^{-1}$ are elliptic. If the group $\Gamma = \langle R_1, J \rangle$ is discrete then one of the following is true:*

- Γ *is one of Mostow's lattices.*

- Γ *is a subgroup of one of Mostow's lattices.*

- Γ *is one of the sporadic groups listed below.*

Figure 2.3: Polygons of configurations of $\Gamma(2\pi/p, \tau)$ for $p = 3, ..., 10$

Mostow's lattices correspond to $\tau = e^{i\phi}$ for some angle ϕ; subgroups of Mostow's lattices to $\tau = e^{2i\phi} + e^{-i\phi}$ for some angle ϕ, and sporadic groups are those for which τ takes one of the 18 values $\{\sigma_1, \overline{\sigma_1}, ..., \sigma_9, \overline{\sigma_9}\}$ where the σ_i are given in the following list:

$$\sigma_1 := e^{i\pi/3} + e^{-i\pi/6} 2\cos(\pi/4) \qquad \sigma_2 := e^{i\pi/3} + e^{-i\pi/6} 2\cos(\pi/5)$$
$$\sigma_3 := e^{i\pi/3} + e^{-i\pi/6} 2\cos(2\pi/5) \qquad \sigma_4 := e^{2\pi i/7} + e^{4\pi i/7} + e^{8\pi i/7}$$
$$\sigma_5 := e^{2\pi i/9} + e^{-i\pi/9} 2\cos(2\pi/5) \qquad \sigma_6 := e^{2\pi i/9} + e^{-\pi i/9} 2\cos(4\pi/5)$$
$$\sigma_7 := e^{2\pi i/9} + e^{-i\pi/9} 2\cos(2\pi/7) \qquad \sigma_8 := e^{2\pi i/9} + e^{-i\pi/9} 2\cos(4\pi/7)$$
$$\sigma_9 := e^{2\pi i/9} + e^{-i\pi/9} 2\cos(6\pi/7).$$

Therefore, for each value of $p \geqslant 3$, we have a finite number of groups to study, the $\Gamma(2\pi/p, \sigma_i)$ and $\Gamma(2\pi/p, \overline{\sigma_i})$ which are hyperbolic (i.e. preserve a form of signature (2,1)). We determined exactly which sporadic groups are hyperbolic; notably these exist for all values of p, and more precisely:

Proposition 2.7.2 *For $p \geqslant 4$ and $\tau = \sigma_1, \sigma_2, \sigma_3, \overline{\sigma_4}, \sigma_5, \sigma_6, \sigma_7, \overline{\sigma_8}$ or σ_9, $\Gamma(2\pi/p, \tau)$ is hyperbolic.*

2.7.4 Arithmeticity

The symmetric complex reflection triangle groups are all integral groups preserving a form, in the following sense (recall that $\tau = \text{Tr}(R_1 J)$):

Proposition 2.7.3 *The generators R_1, R_2 and R_3, as well as the Hermitian form, may be conjugated within* $\mathrm{SU}(2,1)$ *and scaled so that their matrix entries lie in the ring* $\mathbb{Z}[\tau, \overline{\tau}, e^{\pm 2i\pi/p}]$.

The following result gives a concrete criterion to determine whether or not a given integral group is contained in an arithmetic lattice. This formulation is essentially due to Mostow (Lemma 4.1 of [Mos1]), following the Vinberg criterion from [V2] (see p. 226 of [V1]).

Proposition 2.7.4 (Vinberg, Mostow) *Let E be a purely imaginary quadratic extension of a totally real field F, and H a Hermitian form of signature (2,1) defined over E.*

1. $\mathrm{SU}(H; \mathcal{O}_E)$ *is a lattice in* $SU(H)$ *if and only if for all* $\varphi \in \mathrm{Gal}(F)$ *not inducing the identity on F, the form* $^{\varphi}H$ *is definite. Moreover, in that case,* $\mathrm{SU}(H; \mathcal{O}_E)$ *is an arithmetic lattice.*

2. *Suppose* $\Gamma \subset \mathrm{SU}(H; \mathcal{O}_E)$ *is a lattice. Then* Γ *is arithmetic if and only if for all* $\varphi \in \mathrm{Gal}(F)$ *not inducing the identity on F, the form* $^{\varphi}H$ *is definite.*

Note that when the group Γ as in the Proposition is non-arithmetic, it necessarily has infinite index in $SU(H, \mathcal{O}_K)$ (which is non-discrete in $SU(H)$). Using this criterion we get the following (Theorem 3.1 of [Pau]):

Theorem 2.7.2 *For any $p \geqslant 3$ and $\tau \in \{\sigma_1, \overline{\sigma_1}, ..., \sigma, \overline{\sigma_9}\}$, $\Gamma(2\pi/p, \tau)$ is contained in an arithmetic lattice in* $\mathrm{SU}(2,1)$ *if and only if $p = 3$ and $\tau = \overline{\sigma_4}$.*

2.7.5   Commensurability classes

Proposition 2.7.5 (Deligne-Mostow) $\mathbb{Q}[\mathrm{TrAd}\Gamma]$ *is a commensurability invariant.*

This gives a concrete tool to distinguish commensurability classes; for unitary groups (in fact, of any signature), traces under the adjoint representation are easily computed by the following (see "Lemma 4.2" of [Mos1] and [Pau]):

Lemma 2.7.1 *For $\gamma \in \mathrm{SU}(n,1)$, $\mathrm{Tr}(\mathrm{Ad}\gamma) = |\mathrm{Tr}(\gamma)|^2$.*

Proposition 2.7.6 *For $\Gamma = \Gamma(2\pi/p, \overline{\sigma_4})$, $\mathbb{Q}[\mathrm{TrAd}\Gamma] = \mathbb{Q}[\cos\frac{2\pi}{p}, \sqrt{7}\sin\frac{2\pi}{p}]$.*

Corollary 2.7.3   *(1) The 6 groups $\Gamma = \Gamma(2\pi/p, \bar{\sigma}_4)$ with $p = 3, 4, 5, 6, 8, 12$ lie in different commensurability classes.*

(2) The 6 groups $\Gamma = \Gamma(2\pi/p, \bar{\sigma}_4)$ with $p = 3, 4, 5, 6, 8, 12$ are not commensurable to any Mostow or Picard lattice.

2.7.6 Discreteness and fundamental domains

The following is the main result of [DPP2]:

Theorem 2.7.4 *Let $p \geqslant 3$, $R_1 \in \mathrm{SU}(2,1)$ be a complex reflection through angle $2\pi/p$ and $J \in \mathrm{SU}(2,1)$ be a regular elliptic map of order 3 Suppose that $\tau = \overline{\sigma_4} = \mathrm{Tr}(R_1 J) = -(1+i\sqrt{7})/2$.*
   Define $c = 2p/(p-4)$ and $d = 2p/(p-6)$.
   The group $\langle R_1, J \rangle$ is a lattice whenever c and d are both integers, possibly infinity, that is when $p = 3, 4, 5, 6, 8, 12$.
   Moreover, writing $R_2 = JR_1J^{-1}$ and $R_3 = JR_2J^{-1} = J^{-1}R_1J$, this group has presentation

$$
\left\langle \; R_1, R_2, R_3, J \; \middle| \; \begin{array}{c} R_1^p = J^3 = (R_1J)^7 = id, \\ R_2 = JR_1J^{-1}, \; R_3 = J^{-1}R_1J, \\ (R_1R_2)^2 = (R_2R_1)^2, \\ (R_1R_2)^{2c} = (R_1R_2R_3R_2^{-1})^{3d} = id \end{array} \right\rangle.
$$

Note that the first of these groups is the arithmetic lattice from Theorem 2.7.2; the other 5 groups are non-arithmetic lattices, not commensurable to each other or to any Mostow or Picard lattice by Corollary 2.7.3.

   Strategy: Construct a polyhedron in $\mathrm{H}^2_{\mathbb{C}}$ and use the Poincaré Polyhedron Theorem (Theorem 2.7.5 below) to prove that it is a fundamental domain for the action of Γ.

Key ingredients:

- a *polyhedron D* in $\mathrm{H}^2_{\mathbb{C}}$ with *side-pairings*;

- for each orbit of 2-faces, local tessellation conditions around these 2-faces must be satisfied.

We now give a slightly simplified exposition of the Poincaré Polyhedron Theorem (Theorem 2.7.5 below) following [DFP] in the compact case. For a more general class of polyhedra, including noncompact ones, see section 3.2 of [DPP2]; see also [EP] for background on this theorem and an exposition in the constant curvature case.

   Definition: A *polyhedron* is a cellular space homeomorphic to a compact polytope. In particular, each codimension two cell is contained in exactly two codimension one cells. Its realization as a cell complex in a manifold X is also referred to as a polyhedron. We will say a polyhedron is *smooth* if its faces are smooth.

   Definition: A *Poincaré polyhedron* is a smooth polyhedron D in X with codimension one faces T_i such that

1. The codimension one faces are paired by a set Δ of isometries of X which respect the cell structure (the side-pairing transformations). We assume that if $\gamma \in \Delta$ then $\gamma^{-1} \in \Delta$.

2. For every $\gamma_{ij} \in \Delta$ such that $T_i = \gamma_{ij}T_j$ then $\gamma_{ij}D \cap D = T_i$.

Remark: If $T_i = T_j$, that is if a side-pairing maps one side to itself then we impose, moreover, that γ_{ij} be of order two and call it a reflection. We refer to the relation $\gamma_{ij}^2 = 1$ as a reflection relation.

Cycles: Let T_1 be an $(n-1)$-face and F_1 be an $(n-2)$-face contained in T_1. Let T_1' be the other $(n-1)$-face containing F_1. Let T_2 be the $(n-1)$-face paired to T_1' by $g_1 \in \Delta$ and $F_2 = g_1(F_1)$. Again, there exists only one $(n-1)$-face containing F_2 which we call T_2'. We define recursively g_i and F_i, so that $g_{i-1} \circ \cdots \circ g_1(F_1) = F_i$.

Definition: *Cyclic* is the condition that for each pair (F_1, T_1)(an $(n-2)$-face contained in an $(n-1)$-face), there exists $r \geqslant 1$ such that, in the construction above, $g_r \circ \cdots g_1(T_1) = T_1$ and $g_r \circ \cdots g_1$ restricted to F_1 is the identity. Moreover, calling $g = g_r \circ \cdots \circ g_1$, there exists a positive integer m such that $g_1^{-1}(P) \cup (g_2 \circ g_1)^{-1}(P) \cup \cdots \cup g^{-1}(P) \cup (g_1 \circ g)^{-1}(P) \cup (g_2 \circ g_1 \circ g)^{-1}(P) \cup \cdots \cup (g^m)^{-1}(P)$ is a cover of a closed neighborhood of the interior of F_1 by polyhedra with disjoint interiors.

The relation $g^m = (g_r \circ \cdots g_1)^m = \mathrm{Id}$ is called a *cycle relation*.

Theorem 2.7.5 (Poincaré Polyhedron Theorem) *Let D be a compact Poincaré polyhedron in $\mathrm{H}_{\mathbb{C}}^n$ with side-pairing transformations Δ satisfying condition* **Cyclic**. *Let Γ be the group generated by Δ. Then Γ is a discrete subgroup of $\mathrm{Isom}(\mathrm{H}_{\mathbb{C}}^n)$, D is a fundamental domain for Γ and Γ has presentation:*

$$\Gamma = \langle \Delta \mid \text{cycle relations, reflection relations} \rangle$$

Description of the domains D and E: We construct 2 related polyhedra in $\mathrm{H}_{\mathbb{C}}^2$. D will be a fundamental domain for the lattice Γ, and E will be a fundamental domain for the action of Γ modulo $\langle P \rangle$, where $P = R_1 J$ has order 7.

E is constructed as follows: start with 4 bisectors $\mathcal{R}^{\pm}$ and $\mathcal{S}^{\pm}$, with $R_1(\mathcal{R}^+) = \mathcal{R}^-$ and $S_1(\mathcal{S}^+) = \mathcal{S}^-$. ($S_1$ is a special element in Γ - namely $S_1 = P^2 R_1 P^{-2} R_1 P^{-2}$ - and is related to an obvious complex reflection in Γ by $P^2 S_1 = R_2 R_3^{-1} R_2^{-1}$).

E is then defined as the intersection of the 28 half-spaces bounded by $P^k(\mathcal{R}^{\pm})$, $P^k(\mathcal{S}^{\pm})$ $(k = 0, ..., 6)$ and containing O_P, the isolated fixed point of P.

Proposition 2.7.7 $(E, \partial E)$ *is homeomorphic to* $(B^4, \partial B^4)$*(with some vertices removed when Γ is NC).*

Typical results: Given an element $g \in \langle$ side-pairings $\rangle$, one wishes to prove that $g(D) \cap D$ is exactly a certain k-face of D (where $0 \leqslant k \leqslant 3$). One way to do this is to prove that $g(D)$ and D are on opposite sides of well-chosen bisectors. We now illustrate the arguments that come into play.

Bisector intersections: Pairwise intersections of bisectors can be nasty, e.g. disconnected, singular... (Goldman). However, if the bisectors are *coequidistant* (i.e. their complex spines intersect, away from their real spines), then their intersection is nice:

Figure 2.4: A view of the domain E for $p = 12$
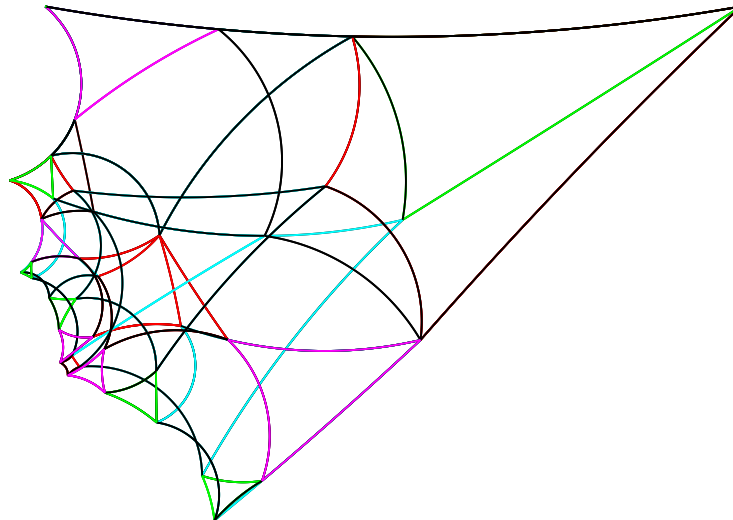
Theorem 2.7.6 (Giraud, 1934) *If B_1 and B_2 are 2 coequidistant bisectors, then $B_1 \cap B_2$ is a (non-totally geodesic) smooth disk. Moreover, there exists a unique bisector $B_3 \neq B_1, B_2$ containing it.*

Proposition 2.7.8 *All 2-faces of E are contained in Giraud disks or complex lines.*

Bad projections of Giraud disks: In order to show that $g(D)$ and D are on opposite sides of a certain bisector $\mathcal{B}$, we argue on the successive k-skeleta of D (for $k = 0, 1, 2, 3$). The bisector is given by a certain equation which can be written as a real quadratic equation in the real and imaginary parts of the ball coordinates of points of $H^2_\mathbb{C}$. Checking that the vertices of D and $g(D)$ lie on the appropriate sides of $\mathcal{B}$ amounts to checking a finite number of such inequalities. For the 1-skeleton, it turns out that all 1-faces of our polyhedra are geodesic segments, which makes matters simpler as the intersection of a geodesic and a bisector is governed by a single quadratic polynomial (of 1 variable); in other words we only need to compute the corresponding discriminant for each 1-face of D and $g(D)$.

The most delicate arguments are along the 2-skeleton; one cannot argue that the 2-skeleton is automatically on the same side of the bisector as the 1-skeleton. While this is true for 2-faces contained in complex lines, it may fail for 2-faces contained in Giraud disks, as Figures 2.5 and 2.6 illustrate. Recall that the bisector $\mathcal{B}$ is the preimage of its real spine under projection to its complex spine, therefore the fact that a subset of $H^2_\mathbb{C}$ lies on one side or the other of $\mathcal{B}$ can be seen by projecting to the complex spine. Figure 2.5 shows a Giraud disk whose projection to a complex line has 2 self-intersections; Figure 2.6 shows a Giraud disk whose projection to a complex line "spills over" the projection of its boundary. (Both figures courtesy of Martin

Figure 2.5: Projection of a Giraud disk with pinching
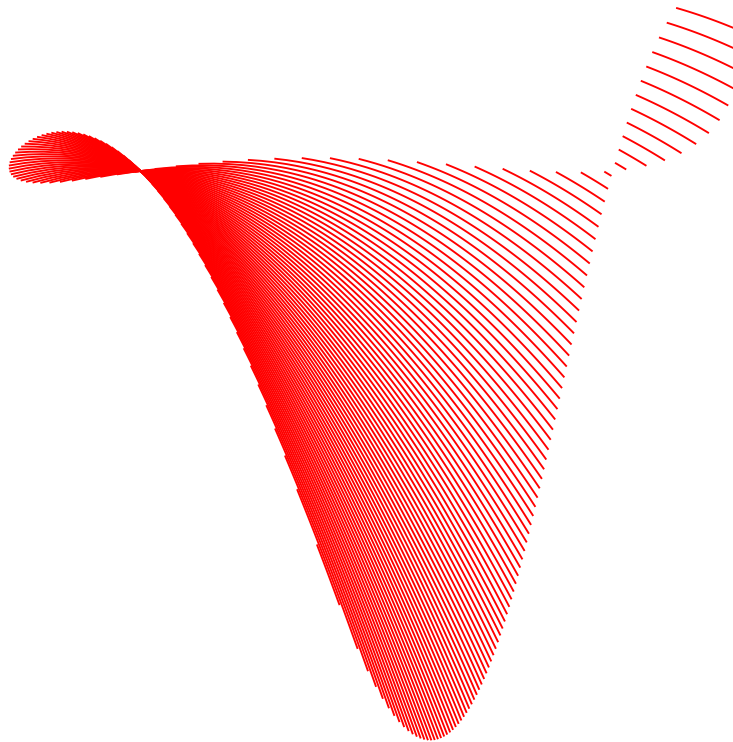
Deraux). In particular, if $\mathcal{B}$ is the bisector whose real spine is the horizontal axis in Figure 2.6, then there are points of the interior of the face which lie on the opposite side of the bisector than the edges of the face (if the face is large enough so that its edges are close to the boundary of the Giraud disk). We must argue that this does not happen in the cases that we consider; in [DPP2] we argue by computing critical points of the distance functions for which the corresponding bisectors are level sets, and showing that these critical points do not lie in the interior of the corresponding Giraud face.

Figure 2.6: Projection of a Giraud disk with spilling
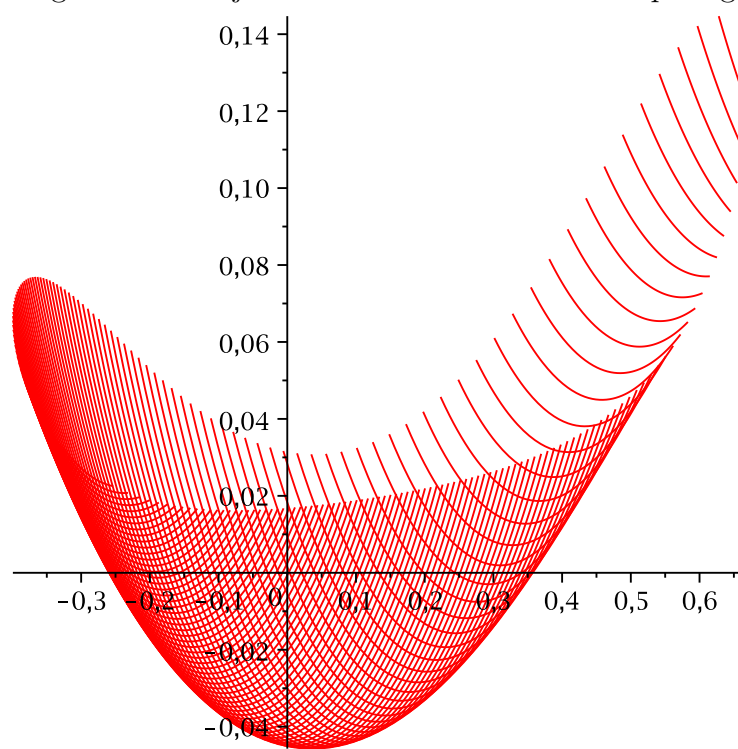
Bibliography

[Ag] I. Agol; *Finiteness of arithmetic Kleinian reflection groups.* International Congress of Mathematicians Vol. II, pp. 951–960, Eur. Math. Soc. , Zürich, 2006.

[Al] D. Allcock; *Infinitely many hyperbolic Coxeter groups through dimension 19.* Geometry and Topology **122** (2006), 159–169.

[An] E.M. Andreev; *On convex polyhedra in Lobachevskij spaces.* Math. USSR Sb. **10** (1970), 413–440.

[ABSW] I. Agol, M. Belolipetsky, P. Storm, K. Whyte; *Finiteness of arithmetic hyperbolic reflection groups.* Groups Geom. Dyn. **2** (2008), 481–498.

[BGS] W. Ballmann, M. Gromov, V. Schroeder; Manifolds of Nonpositive Curvature. Birkhaüser Progress in Math. vol. 61, 1985.

[Be] A. Beardon; The geometry of discrete groups. Graduate Texts in Mathematics Vol. 91. Springer-Verlag, New York (1983).

[BP] R. Benedetti, C. Petronio; Lectures on Hyperbolic Geometry, Universitext, Springer-Verlag (1992).

[Borch] R. Borcherds; *Automorphism groups of Lorentzian lattices.* J. Algebra **111** (1987), 133–153.

[Borel] A. Borel; *Commensurability classes and volumes of hyperbolic 3-manifolds.* Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **8** (1981), no. 1, 1–33.

[Bu] V.O. Bugaenko; *Arithmetic crystallographic groups generated by reflections, and reflective hyperbolic lattices.* Adv. Sov. **8** (1992), 33–56.

[ChGr] S. Chen, L. Greenberg; *Hyperbolic spaces*, in Contributions to Analysis. Academic Press, New York (1974), 49–87.

[Co] H.S.M. Coxeter. *Discrete groups generated by reflections.* Ann. Math. **35** (1934), 588–621.

[dlH] P. de la Harpe; *On Hilbert's metric for simplices.* Geometric group theory, Vol. 1 (Sussex, 1991), London Math. Soc. Lecture Note Ser. **181**, Cambridge Univ. Press, Cambridge (1993), 97–119.

[DM] P. Deligne & G.D. Mostow; *Monodromy of hypergeometric functions and non-lattice integral monodromy.* Publ. Math. I.H.E.S. **63** (1986), 5–89.

[De] M. Deraux; *Dirichlet domains for the Mostow lattices.* Experiment. Math. 14 (2005), 467–490

[DFP] M. Deraux, E. Falbel, J. Paupert; *New constructions of fundamental polyhedra in complex hyperbolic space.* Acta Math. **194** (2005), 155–201.

[DPP1] M. Deraux, J.R. Parker, J. Paupert; *Census of the complex hyperbolic sporadic groups.* Experiment. Math. **20** (2011), 467–486.

[DPP2] M. Deraux, J.R. Parker, J. Paupert; *New non-arithmetic complex hyperbolic lattices.* Preprint 2013.

[EP] D.B.A. Epstein, C. Petronio; *An exposition of Poincaré's polyhedron theorem.* Enseign. Math. (2) **40** (1994), no. 1-2, 113–170.

[FalPar] E. Falbel, J. Parker; *The geometry of the Eisenstein-Picard modular group.* Duke Math. J. **131** (2006) 249–289.

[FZ] E. Falbel, V. Zocca; *A Poincaré's polyhedron theorem for complex hyperbolic geometry.* J. Reine Angew. Math. **516** (1999), 138–158.

[GHL] S. Gallot, D. Hulin, J. Lafontaine;

[G] W.M. Goldman; Complex Hyperbolic Geometry. Oxford Mathematical Monographs. Oxford University Press (1999).

[GPS] M. Gromov, I. Piatetski-Shapiro; *Non-arithmetic groups in Lobachevsky spaces.* Publ. Math. IHES **66** (1987), 93–103.

[H] J.E. Humphreys; Reflection Groups and Coxeter Groups. Cambridge Studies in advanced mathematics vol. 29. Cambridge University Press (1990).

[KM] A. Kappes, M. Moeller; *Lyapunov spectrum of ball quotients with applications to commensurability questions.* Preprint 2012. arXiv:1207.5433.

[L] M. Lackenby; Hyperbolic Manifolds. Lecture notes (Oxford 2000). Available at: `http://people.maths.ox.ac.uk/lackenby/`

[MR] C. Maclachlan & A.W. Reid; The arithmetic of hyperbolic 3-manifolds. Graduate Texts in Mathematics **219**. Springer-Verlag, New York (2003).

[Mak] V.S. Makarov; *On a class of discrete groups in Lobachevskij space having infinite fundamental domain of finite measure.* Sov. Math. Dokl. **7** (1966), 328–331.

[McR] D.B. McReynolds; Arithmetic lattices in $SU(n, 1)$. Monograph, preliminary version.

[Mos1] G.D. Mostow; *On a remarkable class of polyhedra in complex hyperbolic space.* Pacific J. Maths. **86** (1980), 171–276.

[Mos2] G.D. Mostow; *Generalized Picard lattices arising from half-integral conditions.* Publ. Math. I.H.E.S. **63** (1986), 91–106.

[MSW] D. Mumford, C. Series, D. Wright (with cartoons by Larry Gonick); Indra's pearls : the vision of Felix Klein, Cambridge University Press (2002).

[N] V.V. Nikulin; *Finiteness of the number of arithmetic groups generated by reflections in Lobachevsky spaces.* Izv. Math. **71** (2007), no. 1, 53–56.

[Par1] J.R. Parker; *Unfaithful complex hyperbolic triangle groups I: Involutions.* Pacific J. Maths. **238** (2008), 145–169.

[Par2] J.R. Parker; *Complex hyperbolic lattices*, in Discrete Groups and Geometric Structures. Contemp. Maths. **501** AMS, (2009) 1–42.

[ParPau] J.R. Parker & J. Paupert; *Unfaithful complex hyperbolic triangle groups II: Higher order reflections.* Pacific J. Maths. **239** (2009), 357–389.

[Pau] J. Paupert; *Unfaithful complex hyperbolic triangle groups III: arithmeticity and commensurability.* Pacific J. Maths. **245** (2010), 359–372.

[Pic1] E. Picard; *Sur une classe de groupes discontinus de substitutions linéaires et sur les fonctions de deux variables indpéndantes restant invariables par ces substitutions.* (French) Acta Math. **1** (1882), no. 1, 297–320.

[Pic2] E. Picard; *Sur les fonctions hyperfuchsiennes provenant des séries hypergéométriques de deux variables.* Ann. Sci. École Norm. Sup. (3) , **2** (1885), 357–384.

[Ra] J.G. Ratcliffe; Foundations of Hyperbolic Manifolds, second edition. Graduate Texts in Mathematics Vol.149. Springer (2006).

[Ru] O.P. Ruzmanov; *Examples of non-arithmetic crystallographic Coxeter groups in the n-dimensional Lobachevskij space for $6 \leqslant n \leqslant 10$.* Vopr. Teor. Grupp. Gomologicheskoy Algebry **9** (1989), 138–142 (Russian).

[Sc] R.E. Schwartz; *Complex hyperbolic triangle groups.* Proceedings of the International Congress of Mathematicians Vol II (Beijing, 2002), 339–349. Higher Ed. Press, Beijing, 2002.

[Th] W.P. Thurston; The Geometry and Topology of Three-Manifolds. Lecture notes (Princeton 1980). Available at: `http://library.msri.org/books/gt3m/`

[Ti] J. Tits; *Classification of algebraic semisimple groups.* In: 1966 Algebraic Groups and Discontinuous Subgroups (Proc. Sympos. Pure Math., Boulder, Colo., 1965), Amer. Math. Soc. (1966), 33–62.

[To] D. Toledo; *Representations of surfaces groups in complex hyperbolic space.* J. Differential Geom. **29** (1989), 125–133.

[V1] E.B. Vinberg (ed.); Geometry II, in: Encyclopedia of Mathematical Sciences, vol. 29, Springer–Verlag (1991).

[V2] E. B. Vinberg; *Hyperbolic reflection groups.* Russian Math. Surveys **40:1** (1985), 29–66.

[WM] D. Witte Morris; Introduction to Arithmetic Groups. Available at:
  `http://people.uleth.ca/∼dave.morris/books/IntroArithGroups.html`