universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

"Ergodicity of the Geodesic Flow on Quotient Surfaces of the Hyperbolic Plane"

verfasst von / submitted by

Isabel Gomes Gerber, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2020 / Vienna, 2020

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 821
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Mathematik
Betreut von / Supervisor:	Assoz. Prof. Mag. Dr. Roland Zweimüller, Privatdoz.

In memory of Grace Brown Gerber

Abstract

The goal of my thesis is to show the ergodicity of the geodesic flow on quotient spaces of the hyperbolic plane $\Gamma \setminus \mathbb{H}$, where Γ is a lattice. This statement is presented and proven in the last section of chapter 3.

To be able to understand all the concepts needed, we start by introducing the hyperbolic plane $\mathbb{H}$ in chapter 1 and point out how its geometry differs from Euclidean geometry. In particular, we demonstrate how hyperbolic distance is defined and show its consequences. For instance, we will see that geodesics in the hyperbolic plane consist of vertical lines and semicircles with centre on $\mathbb{R}$. We will also be interested in studying Möbius transformations, which do not alter hyperbolic distances, angles or hyperbolic areas. Furthermore some fundamental properties of hyperbolic geometry will be shown, such as the Gauss-Bonnet Theorem.

Chapter 2 starts by showing various characteristics of the projective special linear group, $PSL_2(\mathbb{R})$, such as the identification between $PSL_2(\mathbb{R})$ and $T^1\mathbb{H}$, the unit tangent bundle of $\mathbb{H}$, or the fact that $PSL_2(\mathbb{R})$ is a closed linear group. The reason why this is useful is that since the geodesic flow on the hyperbolic plane is a function on $T^1\mathbb{H}$ we can also define the geodesic flow as a function on $PSL_2(\mathbb{R})$, which will be done in chapter 3. We will also derive a metric on $PSL_2(\mathbb{R})$. This will be done in a more general way by defining a metric on closed linear groups G. Afterwards we consider properties of Fuchsian groups and introduce the notion of fundamental regions. This will be important since we want Γ to be a Fuchsian group whose fundamental domains have finite measure.

As mentioned before we start chapter 3 by defining the geodesic flow on $T^1\mathbb{H}$ as well as on $PSL_2(\mathbb{R})$. The same can be done for the horocycle flow. In order to define the geodesic flow on the quotient space $\Gamma \setminus PSL_2(\mathbb{R})$ we first demonstrate the identifications $T^1(\Gamma \setminus \mathbb{H}) \cong \Gamma \setminus (T^1\mathbb{H}) \cong \Gamma \setminus PSL_2(\mathbb{R})$ and use the definition of the geodesic flow on $T^1(\Gamma \setminus \mathbb{H})$. Our last step before examining the ergodicity of the geodesic flow on $\Gamma \setminus PSL_2(\mathbb{R})$ will be the definition of a measure and a metric on $\Gamma \setminus G$.

I mainly followed the book [5] by Einsiedler and Ward and the paper [8] by Katok. The others sources were used for additional information on the topics.

Zusammenfassung

Das Ziel meiner Masterarbeit ist die Ergodizität des geodätischen Flusses auf Quotientenräumen der hyperbolischen Ebene $\Gamma \setminus \mathbb{H}$ zu zeigen, wobei Γ ein Gitter ist. Diese Aussage wird im letzten Abschnitt von Kapitel 3 bewiesen.

Um alle benötigten Konzepte verstehen zu können, führen wir zunächst die hyperbolische Ebene $\mathbb{H}$ im ersten Kapitel ein und zeigen auf, wie sich ihre Geometrie von der euklidischen Geometrie unterscheidet. Insbesondere zeigen wir, wie die hyperbolische Distanz definiert ist und welche Konsequenzen dies hat. Zum Beispiel werden wir sehen, dass Geodäten in der hyperbolischen Ebene aus vertikalen Linien und Halbkreisen bestehen, deren Mittelpunkt auf $\mathbb{R}$ liegt. Wir werden auch Möbius-Transformationen untersuchen. Diese verändern keine hyperbolischen Abstände, Winkel oder hyperbolischen Flächen. Außerdem werden einige grundlegende Eigenschaften der hyperbolischen Geometrie aufgezeigt, wie zum Beispiel der Satz von Gauß-Bonnet.

Wir beginnen das zweite Kapitel indem wir verschiedene Merkmale der projektiven speziellen linearen Gruppe $PSL_2(\mathbb{R})$ zeigen, wie zum Beispiel die Identifikation zwischen $PSL_2(\mathbb{R})$ und $T^1\mathbb{H}$, dem Einheits-Tangentialbündel von $\mathbb{H}$, oder die Tatsache, dass $PSL_2(\mathbb{R})$ eine geschlossene lineare Gruppe ist. Dies ist nützlich, weil der geodätische Fluss auf der hyperbolischen Ebene eine Funktion auf $T^1\mathbb{H}$ ist. Daher können wir in Kapitel 3 den geodätischen Fluss auch als Funktion auf $PSL_2(\mathbb{R})$ betrachten. Wir werden auch eine Metrik auf $PSL_2(\mathbb{R})$ herleiten. Dazu definieren wir allgemeiner eine Metrik auf geschlossenen linearen Gruppen G. Anschließend betrachten wir Eigenschaften von Fuchsschen Gruppen und führen den Begriff der Fundamentalregion ein. Dies wird wichtig sein, da Γ eine Fuchssche Gruppe sein soll, deren Fundamentalregionen endliches Maß haben.

Wie bereits erwähnt, beginnen wir das dritte Kapitel mit der Definition des geodätischen Flusses auf $T^1\mathbb{H}$ sowie auf $PSL_2(\mathbb{R})$. Dasselbe kann für den horozyklischen Fluss gemacht werden. Um den geodätischen Fluss auf dem Quotientenraum $\Gamma \setminus PSL_2(\mathbb{R})$ zu definieren, werden zunächst die Identifikationen $T^1(\Gamma \setminus \mathbb{H}) \cong \Gamma \setminus (T^1\mathbb{H}) \cong \Gamma \setminus PSL_2(\mathbb{R})$ gezeigt, um danach den geodätischen Fluss auf $T^1(\Gamma \setminus \mathbb{H})$ zu definieren. Bevor wir die Ergodizität des geodätischen Flusses auf $\Gamma \setminus PSL_2(\mathbb{R})$ untersuchen, definieren wir noch ein Maß und eine Metrik auf $\Gamma \setminus G$.

Ich bin hauptsächlich dem Buch [5] von Einsiedler und Ward und der wissenschaftlichen Arbeit [8] von Katok gefolgt. Die anderen Quellen wurden für zusätzliche Informationen zu den jeweiligen Themen verwendet.

Acknowledgements

I would like to thank my advisor, Prof. Roland Zweimüller, for his support during the process of structuring and during the writing process of my thesis.

Contents

1 Hyperbolic geometry

In this chapter I will introduce properties of the hyperbolic plane which will be relevant for the following chapters. Also some differences between Euclidean geometry and hyperbolic geometry will be studied.

1.1 Hyperbolic length and distance

Our first goal is to define a metric on the hyperbolic plane. This will be done by constructing a Riemannian metric on the hyperbolic plane.

Definition 1.1. The *hyperbolic plane* is defined as the upper half-plane $\mathbb{H} = \{z \in \mathbb{C} : \Im(z) > 0\}$ of the complex plane $\mathbb{C}$. Its *boundary* is given by $\partial\mathbb{H} = \mathbb{R} \cup \infty = \{z \in \mathbb{C} : \Im(z) = 0\} \cup \{\infty\}$.

Definition 1.2. A *path* in $\mathbb{H}$ is a piecewise $\mathcal{C}^1$ curve $\gamma : I \to \mathbb{H}$, where I is the unit interval $[0, 1]$.

Before we define a metric on the hyperbolic plane we will briefly remember the definition of a smooth manifold. Since the hyperbolic plane is a smooth manifold we can define a Riemannian metric on it (by Proposition 13.3 in [10]). For a more detailed discussion see for example [4] or [10].

Definition 1.3. A topological space M is called *topological manifold* if for every $p \in M$ there is an open set $U \subseteq M$ containing p which is homeomorphic to $\mathbb{R}^n$. That is, M is a topological manifold if it is locally homeomorphic to $\mathbb{R}^n$.

Definition 1.4. Let M be a topological manifold and let $\mathcal{A}$ be a family of homeomorphisms $X_\alpha : \mathbb{R}^n \to M$, $U_\alpha \mapsto X_\alpha(U_\alpha)$, where the sets U_α are open in $\mathbb{R}^n$ and α is in some index set A. Then a *smooth manifold* $(M, \mathcal{A})$ is a pair satisfying

(i) $M = \cup_\alpha X_\alpha(U_\alpha)$;

(ii) Let $\alpha, \beta \in A$ such that $X_\alpha(U_\alpha) \cap X_\beta(U_\beta) = V \neq \emptyset$. Then the sets $X_\alpha^{-1}(V)$, $X_\beta^{-1}(V)$ are open in $\mathbb{R}^n$. Additionally the composition

$$\begin{aligned} X_\beta^{-1} \circ X_\alpha : \mathbb{R}^n &\to \mathbb{R}^n \\ X_\alpha^{-1}(X_\beta(U_\beta)) &\mapsto U_\beta \end{aligned}$$

is smooth (see Figure 1.1). By interchanging α and β we get that also the inverse map is smooth;

(iii) The family $\mathcal{A}$ is maximal with respect to (i) and (ii).

Remark 1.5. The family $\mathcal{A}$ is called an *atlas* and (X_α, U_α) is a *coordinate system* for $p \in M$ if $p \in X_\alpha(U_\alpha)$.

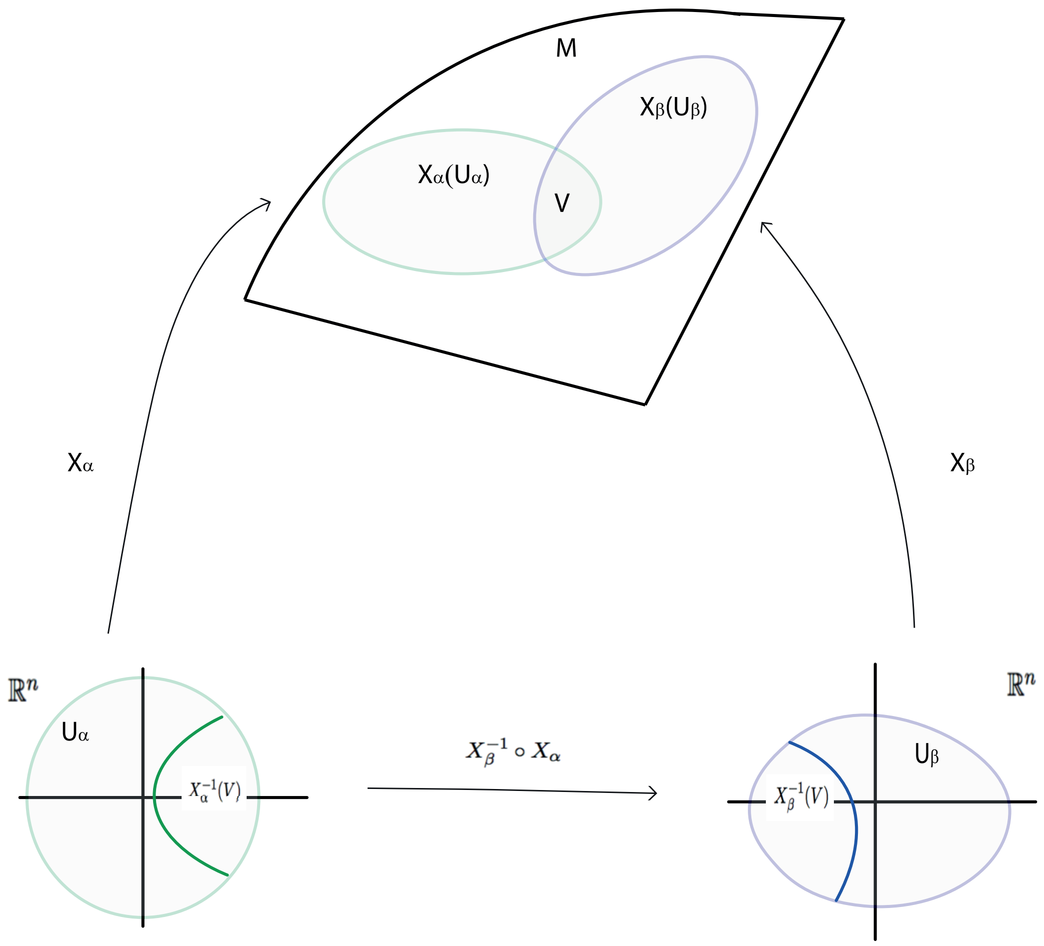

Figure 1.1: Representation of coordinates of a manifold M.

Definition 1.6. Let M be a smooth manifold and let $<,>_p: T_pM \times T_pM \to \mathbb{R}$ be an inner product on T_pM for $p \in M$. If $<,>_p$ varies smoothly from point to point on M, then the collection $(<,>_p)_{p \in M}$ is called a *Riemannian metric* on M.

We define a Riemannian metric on $\mathbb{H}$ as follows.

Definition 1.7. Let $z = x + iy$ be in $\mathbb{H}$ and the vectors (z, u) and (z, v) in $T_z\mathbb{H}$, which is the tangent space of $\mathbb{H}$ at z. Since $T_z\mathbb{H} = \{z\} \times \mathbb{C} \cong \{z\} \times \mathbb{R}^2$ the vectors u and v are in $\{z\} \times \mathbb{R}^2$. Thus we can define an inner product on $\mathbb{R}^2$:

$$<,>_z: \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$$
$$u, v \mapsto < u, v >_z := \frac{(u, v)}{\Im(z)^2} = \frac{(u, v)}{y^2},$$

where (u, v) is the usual inner product in $\mathbb{C} \cong \mathbb{R}^2$.
Then we use the same symbol to define an inner product on $T_z\mathbb{H}$:

$$\begin{aligned} <,>_z &: T_z\mathbb{H} \times T_z\mathbb{H} \to \mathbb{R} \\ & (z,u),(z,v) \mapsto <(z,u),(z,v)>_z := <u,v>_z = \frac{(u,v)}{\Im(z)^2} = \frac{(u,v)}{y^2}. \end{aligned}$$

By Definition 1.6 the *Riemannian metric on* $\mathbb{H}$ or *hyperbolic Riemannian metric* is the collection of the inner products for all z in $\mathbb{H}$.

Now we can define the hyperbolic distance $d(.,.)$ induced by the hyperbolic Riemannian metric and show that it is a metric.

Definition 1.8. Let $\gamma : I \to \mathbb{H}$ be a path in $\mathbb{H}$ and $D\gamma(t) = (\gamma(t), \gamma'(t))$ its derivative at time $t \in [0,1]$. $D\gamma(t)$ is a vector in $T_{\gamma(t)}\mathbb{H}$ with norm $\|D\gamma(t))\|_{\gamma(t)}$. The (hyperbolic) *length* of a path γ is given by

$$\begin{aligned} L(\gamma) &= \int_0^1 \|D\gamma(t))\|_{\gamma(t)} dt = \int_0^1 \sqrt{< D\gamma(t), D\gamma(t) >_{\gamma(t)}} dt \\ &= \int_0^1 \sqrt{\frac{(\gamma'(t),\gamma'(t))}{\Im(\gamma(t))^2}} dt, \end{aligned} \tag{1.1}$$

where the second equation follows by the definition of a norm induced by an inner product and the third equation by Definition 1.7.
Now let z_0 and z_1 be points in $\mathbb{H}$ and consider paths γ in $\mathbb{H}$ connecting these two points. Then we define the *hyperbolic distance* between z_0 and z_1 as

$$\begin{aligned} d &: \mathbb{H} \times \mathbb{H} \to \mathbb{R} \\ & (z_0, z_1) \mapsto d(z_0, z_1) := \inf_{\gamma} L(\gamma), \end{aligned}$$

the infimum over all paths starting at z_0 and ending at z_1.

Proposition 1.9. The hyperbolic distance function is a metric.

Proof. Let $\gamma(t) = x(t) + iy(t)$ for $t \in [0,1]$ be a path in $\mathbb{H}$ going from $z_0 = \gamma(0)$ to $z_1 = \gamma(1)$.
If $z_0 = z_1$ then we get a shortest path if $x(t)$ and $y(t)$ are constant functions and thus

$$L(\gamma) = \int_0^1 \sqrt{\frac{(\frac{dx}{dt})^2 + (\frac{dy}{dt})^2}{y(t)^2}} dt = \int_0^1 \sqrt{\frac{0}{y(t)^2}} dt = 0. \tag{1.2}$$

This shows $d(z_0, z_1) = 0$.
For $z_0 \neq z_1$ at least one the functions $x(t)$, $y(t)$ cannot be constant. Therefore the numerator in the first integral of equation (1.2) is strictly positive on some

non-degenerate subinterval. Since the denominator is positive in any case, the integrand is also positive. It follows that $d(z_0, z_1) > 0$, so the distance is strictly positive.
To show the symmetry of the distance function consider the path

$$\gamma(1-t) = x(1-t) + iy(1-t),$$

$t \in [0,1]$, going in the reverse direction of $\gamma(t)$, form z_1 to z_0. Then

$$\begin{aligned} L(\gamma(1-t)) &= \int_0^1 \sqrt{\frac{(\frac{dx(1-t)}{dt})^2 + (\frac{dy(1-t)}{dt})^2}{y(1-t)^2}} dt = \int_1^0 \sqrt{\frac{(\frac{dx(u)}{-du})^2 + (\frac{dy(u)}{-du})^2}{y(u)^2}} (-du) \\ &= \int_0^1 \sqrt{\frac{(\frac{dx(u)}{du})^2 + (\frac{dy(u)}{du})^2}{y(u)^2}} du = L(\gamma(t)). \end{aligned} \tag{1.3}$$

The second equation follows by the substitution $1-t = u$, $-dt = du$. Hence the length of the path is independent of the direction of the path and so we get

$$d(z_0, z_1) = \inf_{\gamma(t)} L(\gamma(t)) = \inf_{\gamma(1-t)} L(\gamma(1-t)) = d(z_1, z_0).$$

For the triangle inequality consider an additional point z_2 in $\mathbb{H}$. Let γ_1 be a path from z_0 to z_1 and γ_2 a path from z_1 to z_2. Let γ_3 be the path composed from γ_1 and γ_2 going from z_0 to z_2. By our construction the length of γ_3 is $L(\gamma_3) = L(\gamma_1) + L(\gamma_2)$ and by the definition of the hyperbolic distance function we get $d(z_0, z_2) \leq L(\gamma_3) = L(\gamma_1) + L(\gamma_2)$. Then taking the infimum over $L(\gamma_1)$ and $L(\gamma_2)$ gives $d(z_0, z_2) \leq d(z_0, z_1) + d(z_1, z_2)$. □

The notions of hyperbolic length and distance can be extended to the boundary of $\mathbb{H}$.

Definition 1.10. We define the *length* of a path $\gamma : I \to \mathbb{H} \cup \partial\mathbb{H}$ with $\gamma(t) \in \mathbb{H}$ for $t \in (0,1)$ by

$$L(\gamma) := \int_0^1 \sqrt{\frac{(\gamma'(t), \gamma'(t))}{\Im(\gamma(t))^2}} dt,$$

and call the infimum over all such paths the *hyperbolic distance* $d(\gamma(0), \gamma(1))$ between $\gamma(0)$ and $\gamma(1)$.

Lemma 1.11. The hyperbolic distance between any two points $z_0 \in \mathbb{H}$, $z_1 \in \partial\mathbb{H}$ is infinite.

Proof. Let $z_1 = a \in \mathbb{R}$ and $z_0 = b + ci$, where $b \in \mathbb{R}$ and $c \in \mathbb{R}_{>0}$. Then $\gamma(t) = a + (b-a)t + cit$ is a path from z_1 to z_0 for $t \in [0,1]$. Its length is

$$\begin{aligned} L(\gamma) &= \int_0^1 \sqrt{\frac{(\frac{d(a+(b-a)t)}{dt})^2 + (\frac{d(ct)}{dt})^2}{(ct)^2}}dt \geq \int_0^1 \sqrt{\frac{(\frac{d(ct)}{dt})^2}{(ct)^2}}dt = \int_0^1 \frac{1}{t}dt = \ln(t)|_0^1 \\ &= \lim_{k\to 0} \ln(\frac{1}{k}) = \infty. \end{aligned}$$

So $d(z_0, z_1) = \infty$ follows.
If $z_1 = \infty$ then by letting $k \in \mathbb{R}$ go to infinity $\gamma(t) = b + tk + (1-t)ci$ is a path from z_0 to z_1 for $t \in [0,1]$. Its length is

$$\begin{aligned} L(\gamma) &= \lim_{k\to\infty} \int_0^1 \sqrt{\frac{k^2+c^2}{((1-t)c)^2}}dt = \lim_{k\to\infty} \sqrt{k^2+c^2}\frac{1}{-c}\int_0^1 \frac{-c}{c(1-t)}dt \\ &= \lim_{k\to\infty} \sqrt{k^2+c^2}\frac{1}{-c}ln(1-t)|_0^1 = \infty, \end{aligned}$$

and again $d(z_0, z_1) = \infty$ follows. □

Remark 1.12. (i) The previous Lemma shows a considerable difference between hyperbolic length and Euclidean length: For example, if we take the two points $z_0 = 1 + i$ and $z_1 = 1$, then $\gamma(t) = 1 + (1-t)i$, $t \in [0,1]$, is a path from z_0 to z_1. Its Euclidean length is 1, whereas its hyperbolic length is ∞.
(ii) We will see at the end of the next subsection (Remark 1.33 (iv)) that the hyperbolic distance of two points on $\partial\mathbb{H}$ is also infinite.

1.2 Möbius transformations and geodesics

The last section enabled us to measure distances in the hyperbolic plane. Now we are interested in finding functions on the hyperbolic plane which do not alter distances. We will show that Möbius transformations fulfill this requirement. Afterwards we consider paths of shortest lengths.

Definition 1.13. Let $SL_2(\mathbb{R}) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2} : ad - bc = 1\}$ be the *special linear group* and define the action of $SL_2(\mathbb{R})$ on $\mathbb{H}$ by

$$\begin{aligned} T : SL_2(\mathbb{R}) \times \mathbb{H} &\to \mathbb{H} \\ (g, z) &\mapsto T_g(z) := \frac{az+b}{cz+d}. \end{aligned} \tag{1.4}$$

We call T_g a *Möbius transformation.*

Now that we have defined Möbius transformations, let us continue with some basic properties.

Remark 1.14. (i) For any matrix $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$ with $ad - bc > 0$ there exists some matrix $\tilde{g} = \begin{pmatrix} \tilde{a} & \tilde{b} \\ \tilde{c} & \tilde{d} \end{pmatrix} \in SL_2(\mathbb{R})$ such that

$$T_g(z) = T_{\tilde{g}}(z)$$

for any $z \in \mathbb{H}$. This can be seen since for any $g \in \mathbb{R}^{2\times 2}$ with $ad - bc > 0$ we can choose $\nu = \frac{1}{\sqrt{ad-bc}}$ and set $\tilde{g} = \begin{pmatrix} \nu a & \nu b \\ \nu c & \nu d \end{pmatrix} = \begin{pmatrix} \tilde{a} & \tilde{b} \\ \tilde{c} & \tilde{d} \end{pmatrix}$.
Then $T_{\tilde{g}}(z) = \frac{\nu az+\nu b}{\nu cz+\nu d} = \frac{az+b}{cz+d} = T_g(z)$ and $\tilde{a}\tilde{d} - \tilde{b}\tilde{c} = \nu^2(ad - bc) = 1$ for any $z \in \mathbb{H}$. Thus a Möbius transformation can be defined as an action of the set $\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2} : ad - bc > 0\}$ on $\mathbb{H}$. But we will from now on assume that $\det(g) = 1$, i.e. $g \in SL_2(\mathbb{R})$.

(ii) The Möbius transformations can also define an action of the *projective special linear group*

$$PSL_2(\mathbb{R}) := SL_2(\mathbb{R})/\{\pm\mathbb{I}_2\}$$

on $\mathbb{H}$ since $T_{-g}(z) = \frac{-az-b}{-cz-d} = \frac{az+b}{cz+d} = T_g(z)$ for $g \in SL_2(\mathbb{R})$ and $z \in \mathbb{H}$.

Remark 1.15. Definition 1.13 really defines an action since:

(i) If $z \in \mathbb{H}$, then $\frac{az+b}{cz+d}$ is defined for all $a, b, c, d \in \mathbb{R}$ such that $ad - bc = 1$. Assume otherwise that $cz + d = 0$. Then $z = -\frac{d}{c} \in \mathbb{R}$, which means z is not in $\mathbb{H}$. Also if $c = 0$ then $d \neq 0$ and if $d = 0$ then $c \neq 0$ since the determinant of g must be equal to 1.

(ii) If we write

$$T_g(z) = \frac{az+b}{cz+d} = \frac{(az+b)(c\bar{z}+d)}{(cz+d)(c\bar{z}+d)} = \frac{ac|z|^2 + adz + bc\bar{z} + bc}{|cz+d|^2},$$

the imaginary part

$$\Im(T_g(z)) = \frac{T_g(z) - \overline{T_g(z)}}{2i} = \frac{(ad-bc)(z-\bar{z})}{2i|cz+d|^2} = \frac{(ad-bc)\Im(z)}{|cz+d|^2} = \frac{\Im(z)}{|cz+d|^2}$$

is strictly positive because the denominator and the numerator are positive.

(iii) The composition of two Möbius transformations is a Möbius transformation: Let T_g and $T_{\tilde{g}}$ be two Möbius transformations with $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\tilde{g} =$

$\begin{pmatrix} \tilde{a} & \tilde{b} \\ \tilde{c} & \tilde{d} \end{pmatrix}$. Then a simple calculation shows that

$$T_{\tilde{g}} \circ T_g(z) = \frac{\tilde{a}\frac{az+b}{cz+d} + \tilde{b}}{\tilde{c}\frac{az+b}{cz+d} + \tilde{d}} = T_{\tilde{g}g}(z)$$

and $\det(\tilde{g}g) = 1$, where $\tilde{g}g$ denotes matrix multiplication. The same holds for $T_g \circ T_{\tilde{g}}(z)$.

Remark 1.16. (i) Remember that for all $g \in SL_2(\mathbb{R})$ we can write the action in (1.4) as a bijective map

$$\begin{aligned} T_g : \mathbb{H} &\to \mathbb{H} \\ z &\mapsto g \cdot z = T_g(z) = \frac{az+b}{cz+d} \end{aligned}$$

with inverse transformation

$$T_g^{-1}(z) = \frac{dz-b}{-cz+a}.$$

We can easily check $T_g^{-1}(T_g(z)) = T_g(T_g^{-1}(z)) = z$.

(ii) $T_g(z)$ is differentiable and the derivative is given by

$$T_g'(z) = \frac{a(cz+d) - c(az+b)}{(cz+d)^2} = \frac{ad-bc}{(cz+d)^2} = \frac{1}{(cz+d)^2}.$$

(iii) We can extend T_g to the boundary of $\mathbb{H}$ as follows. For $z \in \mathbb{R} \setminus \{\frac{-d}{c}\}$ the transformation maps $\mathbb{R}$ to $\mathbb{R}$. If $z = \frac{-d}{c}$ we set $T_g(z) = \infty$ and for $z = \infty$ we define $T_g(\infty) = \inf_{z\to\infty} T_g(z) = \inf_{z\to\infty} \frac{a+\frac{b}{z}}{c+\frac{d}{z}} = \frac{a}{c}$, which is again in $\mathbb{R}$. If $z = \infty$ and $c = 0$ both a and d cannot be 0 because we require $ad - bc = 1$. In that case we set $\frac{a}{c} = \infty$ and $\frac{-d}{c} = \infty$ and obtain $T_g(\infty) = \infty$.

Theorem 1.17. Möbius transformations are homeomorphisms of $\mathbb{H} \cup \partial\mathbb{H}$.

Proof. We have seen in Remark 1.16 that T_g is a bijection on $\mathbb{H}$ and can be extended to $\partial\mathbb{H}$. A simple calculation also shows that T_g is a bijection for points on $\partial\mathbb{H}$. Since for all $z \in \mathbb{H} \cup \partial\mathbb{H}$ both $T_g(z)$ and $T_g^{-1}(z)$ are rational functions, they are both continuous. Hence T_g is a homeomorphism for all $g \in SL_2(\mathbb{R})$. □

Proposition 1.18. The set of all Möbius transformations is a group under composition, which we will call *Möbius group.*

Proof. (i) The closure follows by Remark 1.15(iii).

(ii) The existence of an inverse for all g is given in Remark 1.16(i).

(iii) If we choose $a = d = 1$ and $b = c = 0$, then clearly $\det(g) = 1$ and T_g is the identity element.
(iv) The associativity follows by the associativity of matrix multiplication and point (i).

$\square$

Remark 1.19. Let $Aut(\mathbb{H})$ be the group of homeomorphisms of $\mathbb{H}$. Then we can identify the Möbius group with $Aut(\mathbb{H})$ by Theorem 1.17 and Proposition 1.18. Thus we can write the action defined in (1.4) as a group homomorphism

$$\begin{aligned} SL_2(\mathbb{R}) &\to Aut(\mathbb{H}) \\ g &\mapsto T_g. \end{aligned}$$

Remark 1.20. (i) The derivative of T_g is given by the following map

$$\begin{aligned} DT_g : T\mathbb{H} &\to T\mathbb{H} \\ (z, v) &\mapsto (T_g(z), T_g'(z)v) = (\frac{az+b}{cz+d}, \frac{v}{(cz+d)^2}) \end{aligned}$$

from the tangent bundle $T\mathbb{H} = \dot{\cup}_{z\in\mathbb{H}} T_z\mathbb{H} = \mathbb{H} \times \mathbb{C}$ to itself. It sends $z \in \mathbb{H}$ to $T_g(z) \in \mathbb{H}$ and the vector component v of $(z, v) \in T_z\mathbb{H}$ to $T_g'(z)v \in \mathbb{C}$.

(ii) If $z \in \mathbb{H}$ is fixed we can identify the derivative of T_g at z

$$(DT_g)_z : T_z\mathbb{H} \to T_{T_g(z)}\mathbb{H}$$

with

$$v \mapsto \frac{v}{(cz+d)^2} =: (DT_g)_z v.$$

(iii) For any $z \in \mathbb{H}$, u, v in $T_z\mathbb{H}$ and $(DT_g)_z u, (DT_g)_z v \in T_{T_g(z)}\mathbb{H}$ the calculation

$$\begin{aligned} &< (DT_g)_z u, (DT_g)_z v >_{T_g(z)} \overset{Def.\ 1.7}{=} \frac{((DT_g)_z u, (DT_g)_z v)}{\Im(T_g(z))^2} \\ &\overset{Remark\ 1.15(ii)}{=} (\frac{\Im(z)}{|cz+d|^2})^{-2}(\frac{u}{(cz+d)^2}, \frac{v}{(cz+d)^2}) \\ &= (\frac{\Im(z)}{|cz+d|^2})^{-2}\frac{1}{|cz+d|^4}(u, v) = \frac{1}{\Im(z)^2}(u, v) =< u, v >_z \end{aligned}$$

shows that DT_g preserves the hyperbolic Riemannian metric.

Lemma 1.21. The Möbius transformations T_g are isometries, that is for any z_0, z_1 in $\mathbb{H}$ and for any g in $\in \mathbb{R}^{2\times 2}$ with $\det(g) = 1$ the hyperbolic distance is invariant under T_g.

Proof. By Remark 1.20(iii) we know that the hyperbolic Riemannian metric is invariant under Möbius transformations. Let $\gamma : I \to \mathbb{H}$ be a path in $\mathbb{H}$ connecting z_0 with z_1 and let T_g be a Möbius transformation. If $\gamma(t) = z$ for some t, then $D\gamma(t)$ is a vector in $T_z\mathbb{H}$. By equation (1.1) we get

$$\begin{aligned} L(\gamma) &= \int_0^1 \sqrt{< D\gamma(t), D\gamma(t) >_{\gamma(t)}}dt \overset{D\gamma(t)=:(\gamma(t),u(t))}{=} \int_0^1 \sqrt{< u(t), u(t) >_{\gamma(t)}}dt \\ &= \int_0^1 \sqrt{< (DT_g)_{\gamma(t)}u(t), (DT_g)_{\gamma(t)}u(t) >_{T_g(\gamma(t))}}dt = L(T_g \circ \gamma), \end{aligned}$$

where $T_g \circ \gamma$ is a path from $T_g(z_0)$ to $T_g(z_1)$. Now taking the infimum over all possible paths γ from z_0 to z_1 yields $d(z_0, z_1) = d(T_g(z_0), T_g(z_1))$. □

Definition 1.22. A path of shortest length between two points is called *geodesic*.

This means that the hyperbolic distance between two points is the distance of a geodesic joining these two points.

Definition 1.23. The *angle* between two geodesics at their intersection point $z \in \mathbb{H}$ is defined as the angle between their tangent vectors at z in $T_z\mathbb{H}$.

Remark 1.24. Let (z, u) and (z, v) with $u = (u_0, u_1), v = (v_0, v_1)$ be the tangent vectors of two geodesics at their intersection point z and let θ be the angle between u and v. Since the tangent space $T_z\mathbb{H}$ can be identified with $\mathbb{R}^2$ (see Definition 1.7) we can use the cosine formula to define θ

$$\cos(\theta) = \frac{< u, v >_z}{\|u\|\|v\|}.$$

Note that

$$\frac{< u, v >_z}{\|u\|\|v\|} \overset{Def.\ 1.7}{=} \frac{\frac{(u,v)}{\Im(z)^2}}{\sqrt{\frac{(u,u)}{\Im(z)^2}}\sqrt{\frac{(v,v)}{\Im(z)^2}}} = \frac{(u, v)}{|u||v|},$$

where $|.|$ is the Euclidean norm.

Remark 1.24 shows that even though the distance in hyperbolic space is defined differently from the distance in Euclidean space, the measure of angles coincide.

Now we will determine that the geodesics are the semicircles with centre on the real axis and the vertical lines, as shown in Figure 1.2.

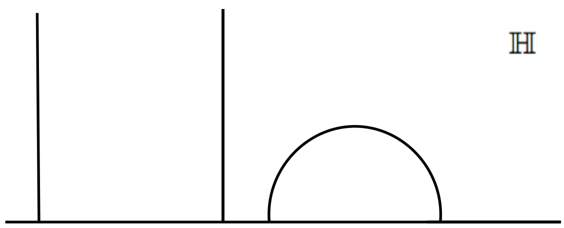

Figure 1.2: Semicircle and vertical lines in $\mathbb{H}$.

Proposition 1.25. The vertical lines in $\mathbb{C}$ and circles in $\mathbb{C}$ with centre in $\mathbb{R}$ can be expressed by the equation

$$\alpha z\bar{z} + \beta z + \beta\bar{z} + \gamma = 0 \tag{1.5}$$

for $\alpha, \beta, \gamma \in \mathbb{R}$ and $z \in \mathbb{C}$.

Proof. By choosing $\alpha = 0$ the equation above becomes $\beta z + \beta\bar{z} + \gamma = 0$. This defines a vertical line $z = x + ik$ for $k \in \mathbb{R}$, since by letting $\frac{-\gamma}{\beta} = 2x$ we get

$$\beta(x + ik) + \beta(x - ik) - 2\beta x = 0.$$

A circle in $\mathbb{C}$ with centre in $z_0 \in \mathbb{R}$ and radius r has the equation

$$|z - z_0| - r^2 = (z - z_0)(\overline{z - z_0}) - r^2 \overset{\beta := -z_0}{=} z\bar{z} + \beta z + \beta\bar{z} + \beta^2 - r^2 = 0.$$

And by choosing $\alpha = 1$ and $\gamma = \beta^2 - r^2$ we get the required equation. $\square$

Definition 1.26. We denote by $\mathcal{H}$ the set of all vertical lines in $\mathbb{H}$ and semicircles in $\mathbb{H}$ with centres in $\mathbb{R}$.

Proposition 1.27. Let H be in $\mathcal{H}$ and let T_g be a Möbius transformation. Then $T_g(H)$ is again in $\mathcal{H}$.

Proof. Claim: A Möbius transformation T_g maps vertical lines in $\mathbb{C}$ and circles in $\mathbb{C}$ with centres in $\mathbb{R}$ to vertical lines in $\mathbb{C}$ or circles in $\mathbb{C}$ with centres in $\mathbb{R}$.
We only need to prove the claim since by Remark 1.16 we already know that T_g is a bijective map from the hyperbolic plane to itself. Thus if the claim is true T_g will map $H \in \mathcal{H}$ bijectively into $\mathcal{H}$
Proof of the claim: By Proposition 1.25 a vertical line or a circle with centre in $\mathbb{R}$ is given by the equation (1.5). Let H be of the form (1.5). Then

$$\begin{aligned} T_g(H) &= \alpha T_g(z)T_g(\bar{z}) + \beta T_g(z) + \beta T_g(\bar{z}) + \gamma \\ &= z\bar{z}(\alpha a^2 + 2\beta ac + \gamma c^2) + z(\alpha ab + \beta ad + \beta bc + \gamma cd) \\ &\quad + \bar{z}(\alpha ab + \beta bc + \beta ad + \gamma cd) + (\alpha b^2 + 2\beta bd + \gamma d^2) = 0. \end{aligned}$$

Since the terms in the brackets are in $\mathbb{R}$ the Möbius transformation of H is also of the form (1.5) and the claim follows. $\square$

Proposition 1.28. The imaginary axis in $\mathbb{H}$ is a geodesic.

Proof. Let $z_0, z_1 \in \mathbb{H}$ be on the imaginary axis, $z_0 = iy_0$, $z_1 = iy_1$, where w.l.o.g. $y_0 < y_1$. Then for $t \in [0, 1]$ the path $\gamma(t) = z_0(1 - t) + z_1 t$ goes from z_0 to z_1 along the imaginary axis. The length of γ is

$$\begin{aligned} L(\gamma) &= \int_0^1 \sqrt{< D\gamma(t), D\gamma(t) >_{\gamma(t)}} = \int_0^1 \sqrt{\frac{(\Im(-z_0 + z_1), \Im(-z_0 + z_1))}{\Im(\gamma(t))^2}} dt \\ &= \int_0^1 \frac{-y_0 + y_1}{y_0(1 - t) + y_1 t} dt = \ln(\frac{y_1}{y_0}). \end{aligned}$$

If we take any other path $\alpha(t) = x(t) + iy(t)$ joining z_0 to z_1 for $t \in [0,1]$, then we can estimate its length by

$$L(\alpha) = \int_0^1 \sqrt{\frac{(\frac{dx}{dt})^2 + (\frac{dy}{dt})^2}{y^2}}dt \geq \int_0^1 \frac{|\frac{dy}{dt}|}{y}dt = \ln(\frac{y_1}{y_0}).$$

Therefore the path of shortest length between any two points iy_0, iy_1 on the imaginary axis is given by the vertical line segment γ and γ is the unique geodesic joining iy_0 and iy_1. Thus the Proposition follows. □

Lemma 1.29. For every $H \in \mathcal{H}$ there exists a Möbius transformation mapping H bijectively to the imaginary axis of $\mathbb{H}$.

Proof. For a vertical line $H = b + ik$, $b \in \mathbb{R}$, $k \in \mathbb{R}_{>0}$, the translation T_g mapping z to $z - b$, with $z \in H$, $g = \begin{pmatrix} 1 & -b \\ 0 & 1 \end{pmatrix}$ and $\det(g) = 1$ is a Möbius transformation mapping H to imaginary axis of $\mathbb{H}$.
If H is a semicircle with endpoints ζ_- and ζ_+ in $\mathbb{R}$ such that $\zeta_- < \zeta_+$, we define the transformation $T_g(z) = \frac{z-\zeta_+}{z-\zeta_-}$. Since $g = \begin{pmatrix} 1 & -\zeta_+ \\ 1 & -\zeta_- \end{pmatrix}$ has determinant $-\zeta_- + \zeta_+ > 0$, the transformation T_g is a Möbius transformation. As shown in Figure 1.3 T_g maps ζ_- to infinity and ζ_+ to zero. Therefore H is mapped to the imaginary axis of $\mathbb{H}$. □

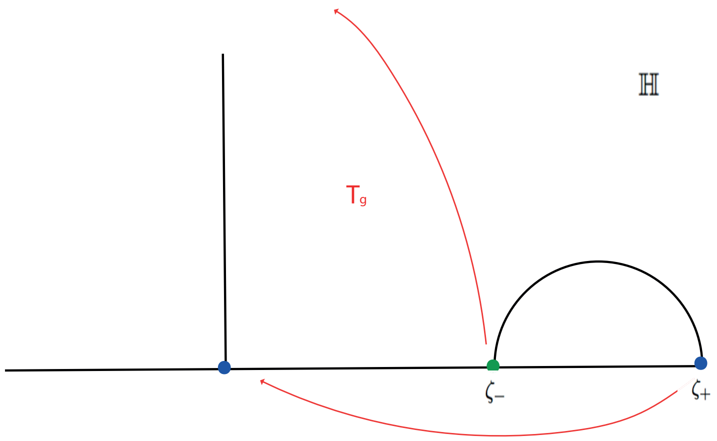

Figure 1.3: Möbius transformation T_g mapping a semicircle to the imaginary axis.

Remark 1.30. It follows from Lemma 1.29 that for any two elements H_1 and H_2 in $\mathcal{H}$ there exists a Möbius transformation T_g such that $T_g H_1 = H_2$. To see this let T_{g_1} be the Möbius transformation mapping H_1 to the imaginary axis in $\mathbb{H}$ and T_{g_2} the Möbius transformation mapping H_2 to the imaginary axis in $\mathbb{H}$. Then $T_g = T_{g_2}^{-1} \circ T_{g_1}$.

Lemma 1.31. For any $H \in \mathcal{H}$ and any $z_0 \in H$ there exists a Möbius transformation T_g mapping H to the imaginary axis in $\mathbb{H}$, such that $T_g(z_0) = i$.

Proof. Let $T_{g'}$ be a Möbius transformation which by Lemma 1.29 maps H to the imaginary axis of $\mathbb{H}$. Also let $T_{\hat{g}}$ be the Möbius transformation mapping z to kz, for $k \in \mathbb{R}_{>0}$. Then $T_{\hat{g}}$ maps the imaginary axis of $\mathbb{H}$ to itself. Thus for a specific k the composition $T_{\hat{g}} \circ T_{g'}$ is the wanted Möbius transformation. □

Theorem 1.32. The elements of $\mathcal{H}$ are the geodesics in $\mathbb{H}$ and for any two points in $\mathbb{H}$ there exists a unique geodesic joining them.

Proof. For any z_0 and z_1 in $\mathbb{H}$ there exists a unique H in $\mathcal{H}$ passing through z_0 and z_1. By Lemma 1.29 there exists a Möbius transformation T_g mapping H to the positive imaginary axis in $\mathbb{H}$. It follows from Proposition 1.28 that the unique geodesic going through the points $T_g z_0$ and $T_g z_1$ is the positive imaginary axis. Since by Lemma 1.21 T_g is an isometry, applying T_g^{-1} to the positive imaginary axis shows that H is the unique geodesic and the segment of H between z_0 and z_1 is the unique geodesic from z_0 to z_1. □

Remark 1.33. (i) From Theorem 1.32 we can conclude that for any point there exists a geodesic in any direction.
(ii) The segment of the geodesic between two points z_0, z_1 is denoted by $[z_0, z_1]$.
(iii) By Proposition 1.27 and Theorem 1.32 Möbius transformations map geodesic to geodesic.
(iv) By Lemma 1.28 and Theorem 1.32 the hyperbolic distance between two points at $\mathbb{R}$ is infinity.

Euclidean geometry can be defined by using the five postulates of Euclid (see [17], chapter 1.7). The fifth postulate is equivalent to the so-called *parallel postulate*. It says that for any straight line k of infinite length and any point x not on that straight line, there exists a unique straight line l of infinite length going through x, which is parallel to the first straight line. We have seen that in hyperbolic space the geodesics are the "straight lines" because they are the paths of minimum length. But in hyperbolic space for any geodesic H we are able to find points $z \notin H$, such that there are infinitely many geodesics going through z which do not intersect H. Both situations are depicted in Figure 1.4. Therefore hyperbolic geometry is a non-Euclidean geometry.

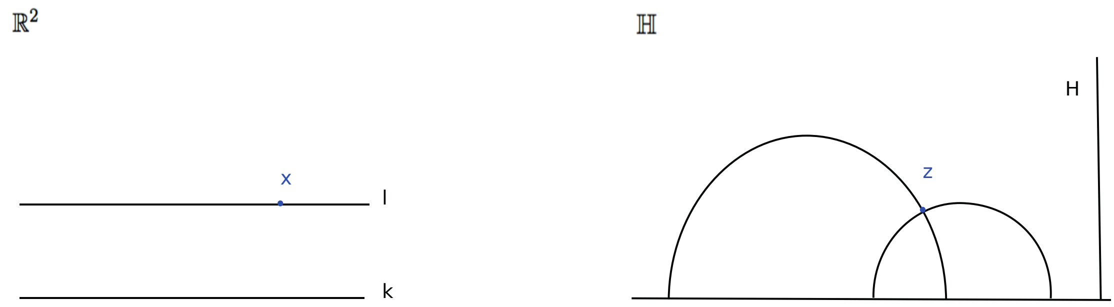

Figure 1.4: Parallel postulate in $\mathbb{R}^2$ and $\mathbb{H}$.

1.3 Hyperbolic area

After examining Möbius transformations of lines and paths in the hyperbolic plane in the last section, we are now interested in studying the influence of Möbius transformations on the hyperbolic area. Additionally we prove the Gauss-Bonnet Theorem, which gives us another example of the difference between Euclidean geometry and hyperbolic geometry. We start by defining the hyperbolic area.

Definition 1.34. Let A be a Borel subset of $\mathbb{H}$. The *hyperbolic area* of A is given by

$$\mu(A) = \int_A \frac{dxdy}{y^2}.$$

Definition 1.35. A map from $\mathbb{H}$ to $\mathbb{H}$ that preserves angles is called *conformal*.

This means that if f is a conformal map and γ_1, γ_2 are two paths in $\mathbb{H}$ intersecting with angle θ at the point z, then the paths $f \circ \gamma_1$, $f \circ \gamma_2$ intersect with the same angle θ at the point $f(z)$.

Proposition 1.36. Möbius transformations are conformal.

Proof. Let T_g be a Möbius transformation and γ_1, γ_2 two paths in $\mathbb{H}$. Assume the paths intersect at a point $\gamma_1(0) = \gamma_2(0) =: z$ with tangent vectors $\gamma_1'(0)$, $\gamma_2'(0)$. Then the tangent vectors of $T_g \circ \gamma_1$, $T_g \circ \gamma_2$ at their intersection point $T_g(z)$ are $(DT_g)_z\gamma_1'(0)$, $(DT_g)_z\gamma_2'(0)$.
To check if the angle between the vectors $\gamma_1'(0)$, $\gamma_2'(0)$ and $(DT_g)_z\gamma_1'(0)$, $(DT_g)_z\gamma_2'(0)$ is the same it suffices to check if the cosine formula for both vector pairs is the same. The cosine formula for the angle θ between $(DT_g)_z\gamma_1'(0)$, $(DT_g)_z\gamma_2'(0)$ is

$$\cos(\theta) = \frac{< (DT_g)_z\gamma_1'(0), (DT_g)_z\gamma_2'(0) >_{T_g(z)}}{\|(DT_g)_z\gamma_1'(0)\|_{T_g(z)}\|(DT_g)_z\gamma_2'(0)\|_{T_g(z)}}.$$

We have seen in Remark 1.20 (iii) that

$$< (DT_g)_z\gamma_1'(0), (DT_g)_z\gamma_2'(0) >_{T_g(z)} = < \gamma_1'(0), \gamma_2'(0) >_z .$$

Thus it follows that

$$\begin{aligned}\|(DT_g)_z\gamma_1'(0)\|_{T_g(z)} &= \sqrt{< (DT_g)_z\gamma_1'(0), (DT_g)_z\gamma_1'(0) >_{T_g(z)}}\\ &= \sqrt{< \gamma_1'(0), \gamma_1'(0) >_z} = \|\gamma_1'(0)\|_z\end{aligned}$$

and

$$\|(DT_g)_z\gamma_2'(0)\|_{T_g(z)} = \|\gamma_2'(0)\|_z.$$

Therefore we can show

$$\cos(\theta) = \frac{< (DT_g)_z\gamma_1'(0), (DT_g)_z\gamma_2'(0) >_{T_g(z)}}{\|(DT_g)_z\gamma_1'(0)\|_{T_g(z)}\|(DT_g)_z\gamma_2'(0)\|_{T_g(z)}} = \frac{< \gamma_1'(0), \gamma_2'(0) >_z}{\|\gamma_1'(0)\|_z\|\gamma_2'(0)\|_z},$$

which implies that Möbius transformations are conformal. □

Theorem 1.37. Möbius transformations preserve hyperbolic area. That is, for any Borel subset A of $\mathbb{H}$ and any Möbius transformation T_g we have $\mu(A) = \mu(T_g(A))$.

Proof. Let T_g be a Möbius transformation and let $z = x + iy$, $T_g(z) = u + iv$ and A be a set in $\mathbb{H}$. Then

$$\begin{aligned}\mu(T_g(A)) &= \int_{T_g(A)} \frac{1}{\Im(T_g(A))^2} dudv \overset{f(u,v):=\frac{1}{\Im(T_g(A))^2}}{=} \int_{T_g(A)} f(u,v)dudv\\ &= \int_A (f\circ T_g)(x,y)|\det(J_{\mathbb{R}}(T_g)(x,y))|dxdy,\end{aligned} \tag{1.6}$$

where $\det(J_{\mathbb{R}}(T_g))$ is the real Jacobian determinant and the last equation follows by the change of variables formula.

Note the following:

(i) The real Jacobian determinant can be obtained from the complex Jacobian determinant because of the fact that $\det(J_{\mathbb{R}}(T_g)) = |\det(J_{\mathbb{C}}(T_g))|^2$;

(ii) By Remark 1.15 (ii) we have $(f\circ T_g)(x,y) = f(\frac{az+b}{cz+d}) = (\frac{\Im(z)}{|cz+d|^2})^{-2}$;

(iii) $|\det(J_{\mathbb{R}}(T_g))| = \frac{|1|^2}{|cz+d|^4}$, since $J_{\mathbb{C}}(T_g) = \frac{1}{(cz+d)^2}$ by Remark 1.16 (ii).

Thus by using the comment above the last integral of equation (1.6) becomes

$$\int_A (\frac{\Im(z)}{|cz+d|^2})^{-2}\frac{|1|^2}{|cz+d|^4}dxdy = \int_A \frac{1}{y^2}dxdy = \mu(A),$$

which finishes our proof. □

We would like to determine the hyperbolic area of specific geometric subsets of $\mathbb{H}$. The Gauss-Bonnet Theorem below gives us a formula for the hyperbolic area of hyperbolic triangles (which correspond to hyperbolic 3-gons).

Definition 1.38. A *hyperbolic n-gon* is an area of $\mathbb{H}$ bounded by n geodesic segments, that is for n vertices $z_0, z_1, ..., z_{n-1}$ in $\mathbb{H} \cup \partial\mathbb{H}$ it is bounded by the segments $[z_0, z_1], [z_1, z_2], ...[z_{n-1}, z_0]$.

Remark 1.39. (i) Remember that geodesics meet $\partial\mathbb{H}$ at right angles, since they are semicircles with centre in $\partial\mathbb{H}$ or vertical lines. Therefore if two geodesic segments meet at a vertex on $\partial\mathbb{H}$ the angle between the two geodesic segments at this vertex is 0.

(ii) Hyperbolic triangles can be distinguished by the number of vertices on $\partial\mathbb{H}$.

Theorem 1.40. (*Gauss-Bonnet*)
A hyperbolic triangle $\triangle$ with angles α, β and γ has hyperbolic area

$$\mu(\triangle) = \pi - (\alpha + \beta + \gamma). \tag{1.7}$$

Proof. Assume we have a triangle $\triangle$ with one vertex on $\partial\mathbb{H}$. By Proposition 1.36 and Theorem 1.37 Möbius transformations do not change the angles or the area of a triangle. Therefore we can map the vertex on $\partial\mathbb{H}$ to ∞ by a Möbius transformation, where the denominator of the transformation is $z - \zeta$ (like we have seen in Lemma 1.29). Then the angle at ∞ is 0 by Theorem 1.37. If we apply the Möbius trasformations $z \mapsto z + c$ and $z \mapsto kz$ for suitable c and k, we get that the geodesic segment joining the other two vertices belongs to a geodesic with radius 1 and centre $(0, 0)$. Let us call the angles at these vertices α and β and let the vertical geodesic at angle α be at $x = a$ and the vertical geodesic at angle β be at $x = b$, just as in Figure 1.5.

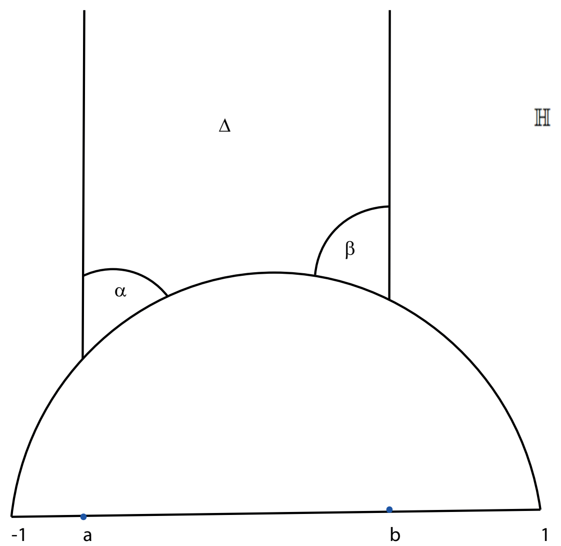

Figure 1.5: Hyperbolic triangle $\triangle$.

Then we can calculate the hyperbolic area

$$\mu(\triangle) = \int_\triangle \frac{1}{y^2} dxdy = \int_a^b dx \int_{\sqrt{1-x^2}}^\infty \frac{1}{y^2} dy = \int_a^b \frac{1}{\sqrt{1-x^2}} dx$$
$$= \int_{\pi-\alpha}^{\beta} \frac{-\sin\theta}{\sin\theta} d\theta = \pi - (\alpha + \beta),$$

where the second last equation follows from the substitution $x = \cos\theta$. Now assume our triangle $\triangle$ has vertices A, B and C, where none of them is in $\partial\mathbb{H}$. We then follow the geodesic containing the geodesic segment joining the vertices A and B until we reach $\partial\mathbb{H}$ and call this new point $D \in \partial\mathbb{H}$. By Theorem 1.32 we can connect the points C and D by a geodesic segment. This method is shown in Figure 1.6.

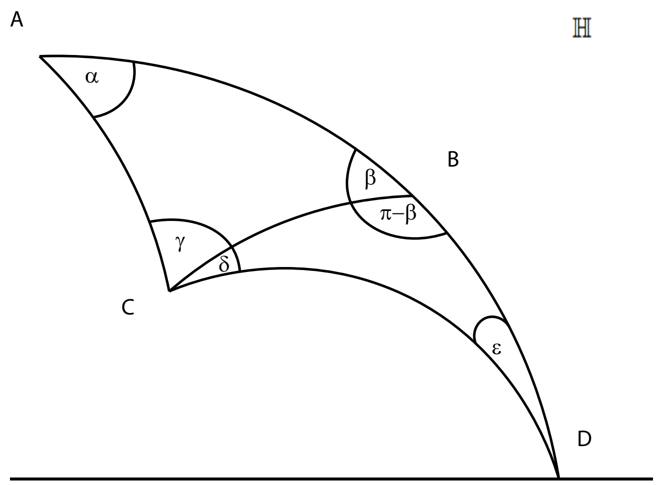

Figure 1.6: Triangles $\triangle$, $\triangle_1$ and $\triangle_2$.

Thus we get a triangle $\triangle_1$ with vertices A, C and D plus another triangle $\triangle_2$ with vertices B, C and D. Since both triangles $\triangle_1, \triangle_2$ have one vertex on $\partial\mathbb{H}$ we can calculate

$$\mu(\triangle) = \mu(\triangle_1) - \mu(\triangle_2) = (\pi - (\alpha+\gamma+\delta+\epsilon)) - (\pi - (\pi-\beta+\delta+\epsilon)) = \pi - (\alpha+\beta+\gamma).$$

$\square$

Remark 1.41. From the Gauss-Bonnet Theorem some significant differences between hyperbolic and Euclidean geometry follow:

(i) The sum of the angles of a hyperbolic triangle is strictly less than π, whereas in Euclidean geometry the sum is always equal to π.

(ii) The area of an arbitrary hyperbolic triangle is at most π. It is equal to π if and only if all the angles are zero. But in Euclidean geometry we can construct triangles with area greater than π and no triangle has area equal to zero.

(iii) The formula (1.7) depends on the angles of the triangle, while in Euclidean geometry the area of a triangle clearly does not depend on its angles.

2 Fuchsian groups and fundamental regions

We start this chapter by showing different properties of $PSL_2(\mathbb{R})$, such as identifications with $T^1\mathbb{H}$ and the Möbius group. We also prove the facts that $PSL_2(\mathbb{R})$ is a topological space as well as a closed linear group and that the hyperbolic volume is invariant under the action of $PSL_2(\mathbb{R})$ by DT_g. At the end of the first section we develop a method to define a metric on $PSL_2(\mathbb{R})$. In the second section we show that Fuchsian groups act properly discontinuously on $PSL_2(\mathbb{R})$ and in the last section we look at some properties of fundamental regions of Fuchsian groups.

2.1 The projective special linear grop

Definition 2.1. The *unit tangent bundle* of $\mathbb{H}$ is defined as

$$T^1\mathbb{H} = \{(z, v) \in T\mathbb{H} : \|v\|_z = 1\},$$

that is the collection of all unit vectors v at $z \in \mathbb{H}$ from all the tangent spaces $T_z\mathbb{H}$.

We have seen in Remark 1.14(ii) the action $PSL_2(\mathbb{R}) \times \mathbb{H} \to \mathbb{H}$, $(g, z) \mapsto \frac{az+b}{cz+d}$, by Möbius transformations. Thus by restricting the derivative DT_g on vectors $v \in T\mathbb{H}$ with $\|v\|_z = 1$, DT_g defines an action of $PSL_2(\mathbb{R})$ on $T^1\mathbb{H}$ by Remark 1.20.
It will prove useful to show that $T^1\mathbb{H} \cong PSL_2(\mathbb{R})$. For this, let us first remember some properties of group actions and the first isomorphism theorem.

Definition 2.2. Let

$$\begin{aligned} G \times X &\to X \\ (g, x) &\mapsto g \cdot x \end{aligned}$$

be an action of a group G on a nonempty set X.

(i) The action is called *transitive* if for all $x, y \in X$ there is some $g \in G$ such that $g \cdot x = y$.

(ii) The action is called *simply transitive* or *regular* if for all $x, y \in X$ there exists a unique $g \in G$ such that $g \cdot x = y$. This means that if $g \cdot x = x$ we must have that g is the identity element of G.

(iii) For every $x \in X$ the *orbit of* x (or G-*orbit* of x) is defined as

$$G \cdot x := \{g \cdot x : g \in G\}.$$

(iv) For every $x \in X$ the *stabilizer* of x is defined as

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

Proposition 2.3. Let G act on X. For any $x \in X$ there exists a bijective map

$$\begin{aligned} f : G/Stab_G(x) &\to G \cdot x \\ yStab_G(x) &\mapsto y \cdot x, \end{aligned}$$

where $y \in G$.

The proof can be found for example in [1] Proposition 6.8.4.

Theorem 2.4. (First Isomorphism Theorem)
Let G and G' be groups, let $\Phi : G \to G'$ be a surjective group homomorphism and let N be the kernel of Φ. If $\pi : G \to G \setminus N := \bar{G}$ is the quotient map, then there exists a unique isomorphism $\bar{\Phi} : \bar{G} \to G'$, where $\bar{\Phi} = \Phi \circ \pi^{-1}$. That is, $\bar{G}$ is isomorphic to G'.

$$\begin{tikzcd} G \arrow[r, "\Phi"] \arrow[d, "\pi"] & G' \\ \bar{G} \arrow[ur, "\exists!\bar{\phi}"] & \end{tikzcd}$$

Figure 2.1: First Isomorphism Theorem.

The First Isomorphism Theorem is proved in [1] (Theorem 2.12.10).

Lemma 2.5. The action $PSL_2(\mathbb{R}) \times \mathbb{H} \to \mathbb{H}$ given by Möbius transformations is transitive.

Proof. Let $z_0 = x_0 + iy_0$ and $z_1 = x_1 + iy_1$ be in $\mathbb{H}$. From Lemma 1.31 we know that there exists a Möbius transformation T_g mapping z_0 to i. We need to make sure that g is in $PSL_2(\mathbb{R})$. For that choose $g = \begin{pmatrix} \frac{1}{\sqrt{y_0}} & -\frac{x_0}{\sqrt{y_0}} \\ 0 & \sqrt{y_0} \end{pmatrix}$. Clearly $\det(g) = 1$ and $T_g(z_0) = i$. Now we want to send i to z_1 via the Möbius transformation $T_{\tilde{g}}$, with $\tilde{g} = \begin{pmatrix} \sqrt{y_1} & \frac{x_1}{\sqrt{y_1}} \\ 0 & \frac{1}{\sqrt{y_1}} \end{pmatrix}$. Again we can check that $\det(\tilde{g}) = 1$ and $T_{\tilde{g}}(i) = z_1$. Therefore the Möbius transformation mapping z_0 to z_1 is given by $T_{\tilde{g}g}$, where $\tilde{g}g \in PSL_2(\mathbb{R})$. □

Corollary 2.6. $PSL_2(\mathbb{R})$ can be identified with the Möbius group.

Proof. It follows Lemma 2.5 that the action $SL_2(\mathbb{R}) \times \mathbb{H} \to \mathbb{H}$ is also transitive. Thus using Remark 1.19 the group homomorphism $f : SL_2(\mathbb{R}) \to Aut(\mathbb{H})$ is surjective. By Theorem 2.4 we can identify

$$SL_2(\mathbb{R}) \setminus ker(f) = SL_2(\mathbb{R}) \setminus \{\pm\mathbb{I}_2\} = PSL_2(\mathbb{R})$$

with $Aut(\mathbb{H}) \overset{Remark\ 1.19}{\cong}$ Möbius group. □

Example 2.7. For the point $i \in \mathbb{H}$ we can calculate its stabilizer with respect to the action $PSL_2(\mathbb{R}) \times \mathbb{H} \to \mathbb{H}$ given by Möbius transformations:

$$\begin{aligned} Stab_{PSL_2(\mathbb{R})}(i) &= \{g \in PSL_2(\mathbb{R}) : T_g(i) = i\} = \{\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R}\}/\{\pm\mathbb{I}_2\} \\ &=: PSO(2). \end{aligned}$$

This follows because for $T_g(i) = i$ to be true we need $\Im(i) = \Im(T_g(i))$ and thus by Remark 1.15(ii) we need $|ci + d|^2 = 1$. This holds if and only if $c = \sin\theta$ and $d = \cos\theta$ for some $\theta \in \mathbb{R}$. Thus we obtain

$$T_g(i) = \frac{ai + b}{\sin\theta i + \cos\theta} = i$$

if and only if $a = \cos\theta$ and $b = -\sin\theta$.
We call $PSO(2)$ the *projective special orthogonal group.*
Since the action is transitive the orbit of i is $PSL_2(\mathbb{R}) \cdot i = \mathbb{H}$. And by using Proposition 2.3 we can identify $\mathbb{H}$ with $PSL_2(\mathbb{R})/PSO(2)$.

Lemma 2.8. The action

$$\begin{aligned} DT : PSL_2(\mathbb{R}) \times T^1\mathbb{H} &\to T^1\mathbb{H} \\ (g, (z, v)) &\mapsto (T_g(z), T_g'(z)v) \end{aligned}$$

is simply transitive.

Proof. By Lemma 2.5 there exists some T_g with $g \in PSL_g(\mathbb{R})$ mapping any element z_0 of $\mathbb{H}$ to any element z_1 of $\mathbb{H}$. Thus we can choose $z_0 := i$ with unit vector $v \in T_i^1\mathbb{H}$. To prove that DT is simply transitive we need to show that

$$\begin{aligned} DT : PSL_2(\mathbb{R}) \times T^1\mathbb{H} &\to T^1\mathbb{H}, \\ (g, (i, v)) &\mapsto (T_g(i), T_g'(i)v) \stackrel{!}{=} (i, v) \end{aligned}$$

holds if and only if $g = \{\pm\mathbb{I}_2\} \in PSL_2\mathbb{R}$ (by the second description of a simply transitive action in Definition 2.2(ii)).
Let us assume that $T_g(i) = i$. By Example 2.7 we know that if $T_g(i) = i$ then $g = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in PSO(2)$ and we can calculate

$$T_g'(i)v = \frac{v}{(\sin\theta i + \cos\theta)^2} = (\cos(2\theta) - i\sin(2\theta))v. \tag{2.1}$$

Therefore $T_g'(i)v$ is a unit vector for any angle θ. This shows that DT really maps unit vector to unit vector and thus DT is transitive. From equation (2.1) it can be seen that $T_g'(i)v = v$ if and only if $\theta = n\pi$ for $n \in \mathbb{Z}$ and thus

$$g = \begin{pmatrix} \cos(n\pi) & -\sin(n\pi) \\ \sin(n\pi) & \cos(n\pi) \end{pmatrix} = \{\pm\mathbb{I}_2\},$$

so DT is simply transitive. $\square$

Remark 2.9. We have seen in Theorem 1.32 that any geodesic γ is uniquely determined by any two points lying on γ. It is also possible to uniquely determine any geodesic γ by any point z on γ and a corresponding unit vector $v \in T^1_z\mathbb{H}$, such that γ' at z has the same slope as v.

Theorem 2.10. There is an identification between $PSL_2(\mathbb{R})$ and $T^1\mathbb{H}$.

Proof. Since the action DT of $PSL_2(\mathbb{R})$ is simply transitive on $T^1\mathbb{H}$, its stabilizer is the singleton $\{\pm\mathbb{I}_2\}$ and its orbit $PSL_2(\mathbb{R}) \cdot (z,v)$, for any $(z,v) \in T^1\mathbb{H}$, is given by $T^1\mathbb{H}$. By Proposition 2.3 we can identify $SL_2(\mathbb{R})/\{\pm\mathbb{I}_2\} = PSL_2(\mathbb{R})$ with $T^1\mathbb{H}$, where the identification is given by

$$h \in PSL_2(\mathbb{R}) \mapsto DT_h(z,v) \in T^1\mathbb{H}$$

for some fixed $(z,v) \in T^1\mathbb{H}$.
To make this identification more explicit consider the elements $(i,i),(z',v') \in T^1\mathbb{H}$, where (z',v') is arbitrary and (i,i) is the point i together with the vector based at i with unit length pointing upwards. Now if we consider the geodesic going through z' in the direction of v' there is a Möbius transformation T_g mapping i to $T_g(i) = z'$, by Lemma 1.29 and Lemma 1.31, such that $DT_g(i) = v'$. This is shown in Figure 2.2. Thus we can identify $g \in PSL_2(\mathbb{R})$ with $DT_g(i,i) = (z',v') \in T^1\mathbb{H}$. $\square$

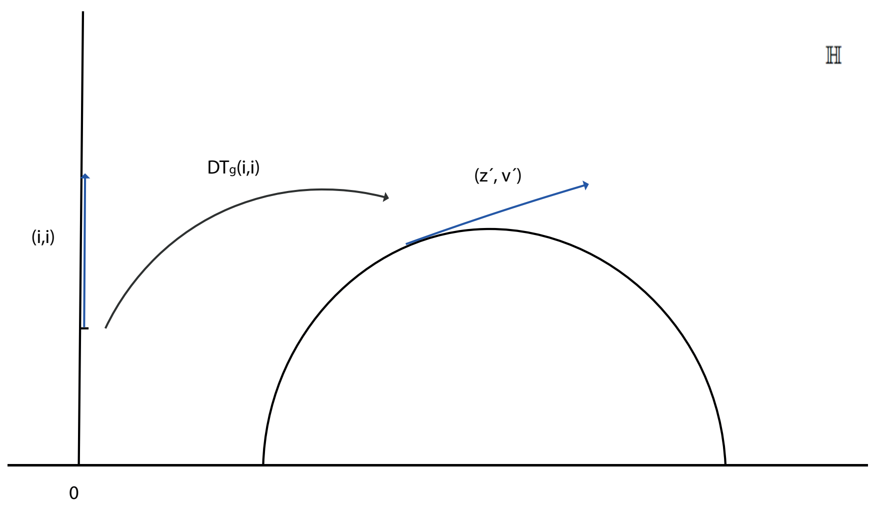

Figure 2.2: DT_g maps (i,i) to (z',v').

Corollary 2.11. The action

$$\begin{aligned} PSL_2(\mathbb{R}) \times PSL_2(\mathbb{R}) &\to PSL_2(\mathbb{R}) \\ (g,h) &\mapsto g \cdot h, \end{aligned}$$

where $g \cdot h$ denotes matrix multiplication, corresponds to the action

$$DT : PSL_2(\mathbb{R}) \times T^1\mathbb{H} \to T^1\mathbb{H}$$
$$(g, DT_h(i,i)) = (g, (T_h(i), T'_h(i)i)) \mapsto DT_{gh}(i,i) = (T_{gh}(i), T'_{gh}(i)i).$$

Proof. We can identify h with $DT_h(i,i)$ by Theorem 2.10. By the definition of DT in Remark 1.20 we know DT maps $(g, T_h(i))$ to $T_g \circ T_h(i) \overset{Proposition\ 1.18}{=} T_{gh}(i)$. So it remains to show that $T'_g \circ T'_h(i)i = T'_{gh}(i)i$:

Let $h = \begin{pmatrix} \tilde{a} & \tilde{b} \\ \tilde{c} & \tilde{d} \end{pmatrix}$, $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, and $g \cdot h = gh = \begin{pmatrix} a\tilde{a} + b\tilde{c} & a\tilde{b} + b\tilde{d} \\ c\tilde{a} + d\tilde{c} & c\tilde{b} + d\tilde{d} \end{pmatrix}$. Then

$$\begin{aligned} T'_g \circ T'_h(i)i &= T'_g(\frac{\tilde{a}i + \tilde{b}}{\tilde{c}i + \tilde{d}})\frac{i}{(\tilde{c}i + \tilde{d})^2} = \frac{1}{(c\frac{\tilde{a}i+\tilde{b}}{\tilde{c}i+\tilde{d}} + d)^2}\frac{i}{(\tilde{c}i + \tilde{d})^2} \\ &= \frac{i}{((c\tilde{a} + d\tilde{c})i + c\tilde{b} + d\tilde{d})^2} = T'_{gh}(i)i. \end{aligned}$$

□

Remark 2.12. It follows from Lemma 1.21 that the group of all isometries on $\mathbb{H}$, $Isom(\mathbb{H})$, contains $PSL_2(\mathbb{R})$.

Now let us show some topological properties of $PSL_2(\mathbb{R})$.

Definition 2.13. (i) We can identify any matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2}$ with the vector $(a,b,c,d) \in \mathbb{R}^4$, where $\mathbb{R}^4$ carries the natural topology. Therefore we can identify $SL_2(\mathbb{R})$ with the subspace

$$X = \{(a,b,c,d) \in \mathbb{R}^4 : ad - bc = 1\}$$

equipped with the subspace topology.

(ii) By defining the equivalence relation $\sim$ on X, given by $(a,b,c,d) \sim (a',b',c',d')$ if and only if $(a,b,c,d) = \pm(a',b',c',d')$, we are able to identify $PSL_2(\mathbb{R})$ with the quotient space $X/\sim$ carrying the quotient topology.

Definition 2.14. We call a subgroup of the *general linear group*

$$GL_2(\mathbb{R}) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2} : ad - bc \neq 0\}$$

a *linear group.* A topological group G is called a *closed linear group* if there exists a map $f : G \to GL_2(\mathbb{R})$ such that f is a homeomorphism from G to $f(G)$, i.e. f is an embedding, and $f(G)$ is closed in $GL_2(\mathbb{R})$.

Remark 2.15. $PSL_2(\mathbb{R})$ is a topological group. The idea of the proof is the following:
First we show that the general linear group $GL_2(\mathbb{R})$ is a topological group. Since $SL_2(\mathbb{R})$ is a subgroup of $GL_2(\mathbb{R})$ we can conclude that $SL_2(\mathbb{R})$ is also a topological group. Next we can show that if N is a normal subgroup of $SL_2(\mathbb{R})$, then the quotient $SL_2(\mathbb{R})/N$ equipped with the quotient topology is a topological group. Because $\{\pm\mathbb{I}_2\}$ is a normal subgroup of $SL_2(\mathbb{R})$ the quotient $SL_2(\mathbb{R})/\{\pm\mathbb{I}_2\} = PSL_2(\mathbb{R})$ is a topological group.
The details are in [12], more specifically in Theorem 5.1.1 and Theorem 5.1.4.

Example 2.16. The special linear group $SL_2(\mathbb{R})$ is a closed linear group:
Consider the maps

$$SL_2(\mathbb{R}) \overset{id}{\to} GL_2(\mathbb{R}) \overset{\det}{\to} \mathbb{R}.$$

The identity map is a homeomorphism from $SL_2(\mathbb{R})$ to $id(SL_2(\mathbb{R})) = SL_2(\mathbb{R})$. Since the determinant map is continuous and $\{1\}$ is closed in $\mathbb{R}$ the inverse $\det^{-1}(\{1\}) = SL_2(\mathbb{R})$ is closed in $GL_2(\mathbb{R})$.

Corollary 2.17. $PSL_2(\mathbb{R})$ is a closed linear group.

Proof. Consider the following conjugation map:

$$\begin{gathered} \phi : SL_2(\mathbb{R}) \to GL(Mat_{22}(\mathbb{R})) \cong GL_2(\mathbb{R}) \\ g \mapsto \phi_g(m) = gmg^{-1} \end{gathered}$$

from the special linear group to the group of invertible linear automorphisms of 2×2 matrices over $\mathbb{R}$, where $m \in Mat_{22}(\mathbb{R})$. ϕ is a homomorphism since

$$\phi_{gh}(m) = ghmh^{-1}g^{-1} = g\phi_h(m)g^{-1} = \phi_g(\phi_h(m))$$

for all $g, h \in SL_2(\mathbb{R})$, $m \in Mat_{22}(\mathbb{R})$. The kernel of ϕ is the set $\{\pm\mathbb{I}_2\}$ because

$$\phi_{\mathbb{I}_2}(m) = \phi_{-\mathbb{I}_2}(m) = m.$$

Thus by the quotient group mapping property for a projection

$$\pi : SL_2(\mathbb{R}) \to SL_2(\mathbb{R})/\{\pm\mathbb{I}_2\} = PSL_2(\mathbb{R})$$

there is a unique homomorphism

$$\tilde{\phi} : PSL_2(\mathbb{R}) \to GL(Mat_{22}(\mathbb{R}))$$

such that $\tilde{\phi} = \phi \circ \pi^{-1}$.

$$\begin{tikzcd} SL_2(\mathbb{R}) \arrow[r, "\phi"] \arrow[d, "\pi"] & GL_2(\mathbb{R}) \\ PSL_2(\mathbb{R}) \arrow[ur, "\tilde{\phi}"] & \end{tikzcd}$$

Figure 2.3: Quotient group mapping property.

Our goal is to show that the image of $PSL_2(\mathbb{R})$ is a closed subset of $GL(Mat_{22}(\mathbb{R}))$. Therefore we need to make sure that $\tilde{\phi}$ is continuous, injective and has a closed image:

(i) It is clear that $\tilde{\phi}$ is continuous and it is injective as $\tilde{\phi}_g(m) = \tilde{\phi}_g(n)$ if and only if $m = n$, for $n, m \in Mat_{22}(\mathbb{R})$.

(ii) Claim: $\tilde{\phi}$ is a proper map. That is, for any compact set K in $GL_2(\mathbb{R})$, $\tilde{\phi}^{-1}(K)$ is compact in $PSL_2(\mathbb{R})$.

Proof of the claim: Consider the two basis vectors $m = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ and $\tilde{m} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$ of $Mat_{22}(\mathbb{R})$ and calculate for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $PSL_2(\mathbb{R})$

$$gmg^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} -ac & a^2 \\ -c^2 & ac \end{pmatrix},$$

$$g\tilde{m}g^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} bd & -b^2 \\ d^2 & -bd \end{pmatrix}.$$

Thus if the images are bounded, then a, b, c, d are bounded and $\tilde{\phi}$ is a proper map. It follows that $\tilde{\phi}(PSL_2(\mathbb{R}))$ is closed since the image of a proper map is closed (see Proposition 5.2.17 in [9]).

□

We have seen in Theorem 1.37 that the hyperbolic area is invariant under the action of $PSL_2(\mathbb{R})$ by Möbius transformations T_g. Now we want to define a hyperbolic volume on $T^1\mathbb{H}$ and show that it is also invariant under the action of $PSL_2(\mathbb{R})$ by the derivative of the Möbius transformations DT_g.

Definition 2.18. Let m be a measure on the measurable space $(T^1\mathbb{H}, \mathcal{B}_{T^1\mathbb{H}})$, where $\mathcal{B}_{T^1\mathbb{H}}$ is the Borel σ-algebra on $T^1\mathbb{H}$. Let $V \in \mathcal{B}_{T^1\mathbb{H}}$ and let $\theta \in [0, 2\pi i)$ be the angle of the unit tangent vector $(z, v) = (z, e^\theta)$ at $z = x + iy \in \mathbb{H}$. The *hyperbolic volume* of V is defined by

$$m(V) = \int_V d\mu d\theta \overset{Def.\ 1.34}{=} \int_V \frac{1}{y^2} dx dy d\theta.$$

Theorem 2.19. The hyperbolic volume is invariant under the action of $PSL_2(\mathbb{R})$ on $T^1\mathbb{H}$.

Proof. Let $(z, e^\theta) \in T^1\mathbb{H}$ with $\theta \in [0, 2\pi i)$, $z = x + iy$, and let (z, e^θ) be in $V \in \mathcal{B}_{T^1\mathbb{H}}$. Remember that the action of $PSL_2(\mathbb{R})$ on $T^1\mathbb{H}$ is given by

$$DT_g(z, e^\theta) = (T_g(z), \frac{e^\theta}{(cz + d)^2})$$

(see Remark 1.20 (i)) and define

$$e^{\theta'} := \frac{e^{\theta}}{(cz+d)^2}$$

with $\theta' \in [0, 2\pi i)$. Let $T_g(z) = u + iv$. Then the hyperbolic volume of $DT_g(V)$ is given by

$$\begin{aligned} m(DT_g(V)) &= \int_{DT_g(V)} \frac{1}{v^2} dudvd\theta' \overset{f(u,v):=\frac{1}{v^2}}{=} \int_{DT_g(V)} f(u,v) dudvd\theta' \\ &= \int_V (f \circ DT_g)(x,y,\theta) |\det(J_{\mathbb{R}}(DT_g)(x,y,\theta))| dxdyd\theta, \end{aligned} \tag{2.2}$$

where we used the change of variables formula in the last equation. As explained in the proof of Theorem 1.37 we calculate the absolute value of the Jacobian determinant

$$\begin{aligned} |\det(J_{\mathbb{R}}(DT_g))| = |\det(J_{\mathbb{C}}(DT_g))|^2 &= |\det \begin{pmatrix} \frac{d}{dz}T_g(z) & \frac{d}{d\theta}T_g(z) \\ \frac{d}{dz}e^{\theta'} & \frac{d}{d\theta}e^{\theta'} \end{pmatrix} |^2 \\ &= |\frac{e^{\theta'}}{(cz+d)^2}|^2 \overset{|e^{\theta'}|=1}{=} \frac{1}{|cz+d|^4} \end{aligned}$$

and the composition of f with DT_g

$$(f \circ DT_g)(x,y,\theta) = \frac{1}{\Im(T_g(z))^2} = (\frac{|cz+d|^2}{\Im(z)})^2 = \frac{|cz+d|^4}{\Im(z)^2},$$

so that equation (2.2) becomes

$$\int_V \frac{|cz+d|^4}{\Im(z)^2} \frac{1}{|cz+d|^4} dxdyd\theta = \int_V \frac{1}{y^2} dxdyd\theta = m(V)$$

as claimed. □

Definition 2.20. The hyperbolic volume m on $T^1\mathbb{H}$ is also called *Liouville measure.*

We now would like to define a metric on $PSL_2(\mathbb{R})$ or more precisely on any closed linear group G. This will be done by deriving a left-invariant metric from a left-invariant Riemannian metric. For that we first need to introduce the notion of Lie algebras. A more detailed introduction can be found for example in [5].

Definition 2.21. The map defined by the absolutely convergent power series

$$\begin{aligned} \exp : Mat_{dd}(\mathbb{C}) &\to Mat_{dd}(\mathbb{C}) \\ m &\mapsto \sum_{n=0}^{\infty} \frac{1}{n!} m^n \end{aligned}$$

is called *(matrix) exponential map.*

Remark 2.22. We can consider the exponential map $\exp : Mat_{dd}(\mathbb{R}) \to GL_d(\mathbb{R})$ since $\exp(Mat_{dd}(\mathbb{R})) \subseteq GL_d(\mathbb{R})$. At $0 \in Mat_{dd}(\mathbb{R})$ it is locally invertible.

Definition 2.23. The *logarithm map*

$$\begin{aligned} \log : GL_d(\mathbb{R}) &\to Mat_{dd}(\mathbb{R}) \\ g \mapsto \log(g) &= \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n}(g - \mathbb{I}_d)^n \end{aligned}$$

is the inverse of the exponential map. If g is close enough to $\mathbb{I}_d$ then $\log(g)$ is convergent.

Proposition 2.24. Let G be a closed linear group contained in $GL_d(\mathbb{R})$. Then there exists a neighbourhood B of $\mathbb{I}_d$ in G such that $\log(B) \subseteq Mat_{dd}(\mathbb{R})$. Additionally $\log(B)$ is a neighbourhood of 0 contained in a linear subspace $\mathcal{G}$ of $Mat_{dd}(\mathbb{R})$.

To prove Proposition 2.24 we need an exact definition of the subspace $\mathcal{G}$ and an additional Lemma.

Definition 2.25. We call the subspace $\mathcal{G}$ mentioned in Proposition 2.24 the *Lie algebra* of G. It can be characterised in the following equivalent ways

(i) $\mathcal{G} := \{m \in Mat_{dd}(\mathbb{R}) : \exp(tm) \in G, \forall t \in \mathbb{R}\} = T_{\mathbb{I}_d}G$.
(ii) Let $\Phi : [a,b] \to G$ be a path in G such that $\Phi(t) = \mathbb{I}_d$ for $t \in [a,b]$. Then $\mathcal{G}$ consists of all derivatives $\Phi'(t)$ of all paths $\Phi(t)$ at $t \in [a,b]$.

Figure 2.4 depicts one path $\Phi(t) \in G$ with $\Phi(t^\circ) = \mathbb{I}_d$, $t^\circ \in [a,b]$, and its derivative at t° in $\mathcal{G}$ as well as how to get from G to $\mathcal{G}$ and vice versa with the functions log and exp, respectively.

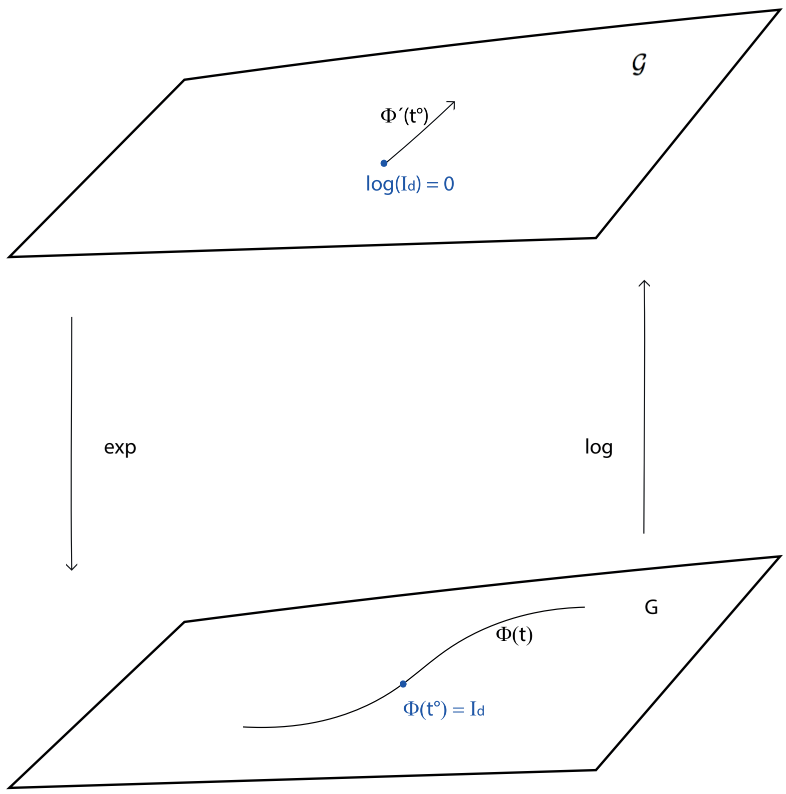

Figure 2.4: The exponential map and the logarithm map.

Lemma 2.26. There exists a neighbourhood A of $0 \in Mat_{dd}(\mathbb{R})$ with the property that for any sequence m_j converging to $m \in A$, as $j \to \infty$, the following holds

$$(\mathbb{I}_d + \frac{m_j}{j})^j \stackrel{j\to\infty}{\to} \exp(m).$$

Proof. Let m be sufficiently small and j be sufficiently large. Then

$$\begin{aligned} j\log(\mathbb{I}_d + \frac{m_j}{j}) &= j\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}(\mathbb{I}_d + \frac{m_j}{j} - \mathbb{I}_d)^n = j\sum_{n=1}^{\infty}\frac{(-1)^{n+1}}{n}(\frac{m_j}{j})^n \\ &= j(\frac{m_j}{j} - (\frac{m_j}{2j})^2 + (\frac{m_j}{3j})^3 - +...) = m_j + O(\frac{1}{j}), \end{aligned}$$

implies that $j\log(\mathbb{I}_d + \frac{m_j}{j}) \stackrel{j\to\infty}{\to} m$. Therefore

$$\exp(j\log(\mathbb{I}_d + \frac{m_j}{j})) = (\mathbb{I}_d + \frac{m_j}{j})^j \stackrel{j\to\infty}{\to} \exp(m).$$

□

Proof. (of Proposition 2.24)
We will first show that $\mathcal{G}$ is a linear subspace of $Mat_{dd}(\mathbb{R})$ and then that there exists some neighbourhood B of $\mathbb{I}_d$ in G such that $\log(B)$ is a neighbourhood of 0 contained in $\mathcal{G}$.

(i) Let $k \in \mathbb{R}$ and $v \in \mathcal{G} = \{m \in Mat_{dd}(\mathbb{R}) : \exp(tm) \in G, \forall t \in \mathbb{R}\}$. Then kv is also in $\mathcal{G}$, since $\exp(tkv) \in G$ for all $k, t \in \mathbb{R}$. To show that $\mathcal{G}$ is closed under addition let $v, w \in \mathcal{G}$ and $t > 0$ such that $t(v+w)$ is an element of the neighbourhood A of $0 \in Mat_{dd}(\mathbb{R})$ from Lemma 2.26. Let

$$(g_n)_{n\geq 1} = ((\exp(\frac{t}{n}v)\exp(\frac{t}{n}w))^n)_{n\geq 1}$$

be a sequence in G. Then if we use the approximation

$$\exp(\frac{t}{n}v) = \mathbb{I}_d + \frac{t}{n}v + \mathcal{O}(\frac{1}{n^2})$$

we can write g_n as

$$g_n = ((\mathbb{I}_d + \frac{t}{n}v + \mathcal{O}(\frac{1}{n^2}))(\mathbb{I}_d + \frac{t}{n}w + \mathcal{O}(\frac{1}{n^2}))) = (\mathbb{I}_d + \frac{1}{n}(t(v+w) + \mathcal{O}(\frac{1}{n})))^n.$$

Observe that $t(v+w) + \mathcal{O}(\frac{1}{n})$ converges to $t(v+w)$ as $n \to \infty$. Therefore we can use Lemma 2.26 to conclude that g_n converges to $\exp(t(v+w))$ as $n \to \infty$. Since G is a closed linear group the limit $\exp(t(v+w))$ is in G. Thus by the definition of $\mathcal{G}$, $v+w$ is in $\mathcal{G}$.

(ii) Consider a linear complement $V \subseteq Mat_{dd}(\mathbb{R})$ of $\mathcal{G}$ and define the map

$$\begin{aligned} \phi : \mathcal{G} \times V &\to GL_d(\mathbb{R}) \\ (tu, tv) &\mapsto (\exp(tu))(\exp(tv)), \end{aligned}$$

for $t \in \mathbb{R}$. Since

$$\frac{d}{dt}(\exp(tu)\exp(tv)) = (u+v)\exp(tu)\exp(tv),$$

the derivative of ϕ at $t=0$ is given by $\phi'(tu,tv)|_{t=0} = u+v$. Thus

$$\phi' : \mathcal{G} \times V \to Mat_{dd}(\mathbb{R})$$

is invertible at $(0,0) \in \mathcal{G} \times V$. Now, by the inverse function theorem there exists some neighbourhood U of $(0,0) \in \mathcal{G} \times V$ as well as some neighbourhood B_1 of $\phi(0,0) = \mathbb{I}_d \in GL_d(\mathbb{R})$ such that the map

$$\phi|_U : U \to B_1$$

is a diffeomorphism. Therefore every $g \in B_1$ can be written as $g = \exp(u)\exp(v)$ such that $u \in \mathcal{G}$ and $v \in V$.

To show that $\log(B)$ is a neighbourhood of 0 contained in $\mathcal{G}$, we will show that $B \subseteq B_1$ is a neighourhood of $\mathbb{I}_d$ such that $\log(B \cap G) \subseteq \mathcal{G}$. Assume by contradiction that $\log(B \cap G) \subseteq V$. Then there exists a sequence v_m in $V \setminus \{0\}$ converging to 0 as $m \to \infty$. By Lemma 2.26

$$(\mathbb{I}_d + \frac{v_m}{m})^m \overset{m \to \infty}{\to} \exp(0) \in G$$

and $\exp(v_m) \in G$. Since the unit ball in V is compact we can choose a subsequence $\frac{v_m}{\|v_m\|}$ of v_m in the unit ball converging to $w \in V$. It can be shown that $\exp(\mathbb{Z}v_m)$ is a subgroup of G and that $\mathbb{Z}v_m$ is a discrete subgroup of V. Then $\mathbb{Z}v_m$ converges to the subspace $\mathbb{R}w$ of V for $m \to \infty$. Thus $\exp(\mathbb{R}w) \subseteq G$, which implies $w \in \mathcal{G}$. This is a contradiction to $w \in V$.

□

Remark 2.27. It can be shown that the Lie algebra $\mathcal{G}$ of any closed linear group $G \subseteq GL_d(\mathbb{R})$ uniquely determines $G^\circ := \exp(\mathcal{G})$, which is the maximal path-connected component of $\mathbb{I}_d \in G$ and also a normal, open, closed subgroup of G.

Definition 2.28. Let G be a closed linear group. For any $g \in G$ the tangent space of G at g is defined as $T_gG := \{g\} \times \mathcal{G}$. Therefore the tangent bundle to G is defined as $TG := G \times \mathcal{G}$.

Definition 2.29. Let $\Phi : [0,1] \to G$ be a path in G such that Φ is differentiable at $t_0 \in [0,1]$. Then the tangent vector at $\Phi(t_0)$ is defined by

$$D\Phi(t_0) := (\Phi(t_0), \Phi(t_0)^{-1}\Phi'(t_0)).$$

Remark 2.30. We need to check that $D\Phi(t_0)$ really lies in $G \times \mathcal{G}$. The first component $\Phi(t_0)$ is in G by definition. For the second component of $D\Phi(t_0)$ consider the curve $\alpha(t) := \Phi(t_0)^{-1}\Phi(t)$ with values in G. Then $\alpha(t_0) = \mathbb{I}_d \in G$ and $\frac{d\alpha}{dt}(t_0) = \Phi(t_0)^{-1}\frac{d\Phi}{dt}(t_0) \in \mathcal{G}$ by Definition 2.25 (ii).

Proposition 2.31. Consider a continuous path $\Phi : [0,1] \to G$ that at $t_0 \in [0,1]$ is differentiable. Then $(g\Phi)(t) = g\Phi(t)$ and $(\Phi g^{-1})(t) = \Phi(t)g^{-1}$ are curves which are differentiable at t_0 for $g \in G$. Additionally $D(g\Phi)(t_0) = (g\Phi(t_0), v)$ and $D(\Phi g^{-1}) = (\Phi(t_0)g^{-1}, gvg^{-1})$ if $D\Phi(t_0) = (\Phi(t_0), v)$.

Proof. Note that

$$\begin{aligned} D(g\Phi)(t_0) &\overset{Def.}{=} (g\Phi(t_0), (g\Phi(t_0))^{-1}g\Phi'(t_0))) = (g\Phi(t_0), \Phi(t_0)^{-1}g^{-1}g\Phi'(t_0)) \\ &= (g\Phi(t_0), \Phi(t_0)^{-1}\Phi'(t_0)), \end{aligned}$$

so $D(g\Phi)(t_0) = (g\Phi(t_0), v)$ follows for $v = \Phi(t_0)^{-1}\Phi'(t_0)$. The other equation

$$\begin{aligned} D(\Phi g^{-1})(t_0) &= (\Phi(t_0)g^{-1}, (\Phi(t_0)g^{-1})^{-1}\Phi'(t_0)g^{-1}) \\ &= (\Phi(t_0)g^{-1}, g(\Phi(t_0)^{-1}\Phi'(t_0))g^{-1}) = (\Phi(t_0)g^{-1}, gvg^{-1}) \end{aligned}$$

also follows by definition. □

Remark 2.32. We can interpret the equation $D(g\Phi)(t_0) = (g\Phi(t_0), v)$ in Proposition 2.31 in the following way. The left translation

$$\begin{aligned} L_g : G &\to G \\ h &\mapsto gh \end{aligned}$$

has the derivative

$$\begin{aligned} D(L_g)_h : T_hG &\to T_{gh}G \\ (h, v) &\mapsto (gh, v). \end{aligned}$$

So it can be seen that $D(L_g)_h$ moves the base point h to gh but leaves v unchanged. If we choose an inner product on $\mathcal{G}$, denoted by $<,>$, we can define a Riemannian metric on G as the collection of inner products

$$< (g, u), (g, v) >_g := < u, v >_g = < u, v >, \tag{2.3}$$

where $u, v \in T_gG$, $g \in G$.
Notice that we defined $TG \overset{Def.\ 2.28}{=} G \times \mathcal{G} \overset{Def.\ 2.25}{=} G \times T_{\mathbb{I}_d}G$, where $\mathcal{G}$ contains all derivatives of paths going through the identity $\mathbb{I}_d \in G$. Then to get derivatives of curves going through $g \in G$ we can use

$$\begin{aligned} D(L_g)_{\mathbb{I}_d} : T_{\mathbb{I}_d}G &\to T_gG \\ (\mathbb{I}_d, v) &\mapsto (g, v). \end{aligned}$$

Just as in the first chapter we define a metric $d_G(.,.)$ induced by the Riemannian metric on G. We start again by defining the length of paths on G.

Definition 2.33. Let $\Phi : [0, 1] \to G$ be a path in G. The length of Φ is given by

$$L(\Phi) = \int_0^1 \|D\Phi(t)\|_{\Phi(t)} dt \overset{(2.3)}{=} \int_0^1 \sqrt{< D\Phi(t), D\Phi(t) >_{\Phi(t)}} dt.$$

Then for paths starting at $\Phi(0) = g_0 \in G^\circ$ and ending at $\Phi(1) = g_1 \in G^\circ$ the metric on G° is defined (as in the Definition 1.8) as

$$\begin{aligned} d_G : G^\circ \times G^\circ &\to \mathbb{R} \\ (g_0, g_1) &\mapsto d_G(g_0, g_1) := \inf_\Phi L(\Phi), \end{aligned}$$

where the infimum is taken over all such paths.

Corollary 2.34. The metric d_G is left-invariant on G°. That is, for all $h, g_0, g_1 \in G^\circ$ we have $d_G(hg_0, hg_1) = d_G(g_0, g_1)$.

Proof. By

$$\begin{aligned}
L(h\Phi) &= \int_0^1 \sqrt{< D(h\Phi)(t), D(h\Phi)(t) >_{h\Phi(t)}}dt \\
&\overset{Proposition\ 2.31}{=} \int_0^1 \sqrt{< (h\Phi(t), v), (h\Phi(t), v) >_{h\Phi(t)}}dt = \int_0^1 \sqrt{< v, v >_{h\Phi(t)}}dt \\
&\overset{Remark\ 2.32}{=} \int_0^1 \sqrt{< v, v >_{\Phi(t)}}dt = \int_0^1 \sqrt{< (\Phi(t), v), (\Phi(t), v) >_{\Phi(t)}}dt = L(\Phi)
\end{aligned}$$

the length of Φ is left invariant for $h \in G^\circ$. Thus it follows that the metric d_G is also left-invariant on G°. □

2.2 Discrete subgroups of closed linear groups

We want to show that a subgroup of $PSL_2(\mathbb{R})$ is a Fuchsian group if and only if it acts properly discontinuously on $\mathbb{H}$.

Definition 2.35. We call a subgroup of a closed linear group *discrete* if it has a discrete topology as subspace topology.

Remark 2.36. Γ is discrete if and only if it has the following property: A sequence $\{g_n\}$ of elements of a subgroup Γ converges to the identity element if and only if for a sufficiently large n the element g_n is equal to the identity element.

Example 2.37. (i) $SL_2(\mathbb{Z}) = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{Z}^{2\times 2} : ad - bc = 1\}$ is a discrete subgroup of $SL_2(\mathbb{R})$.
(ii) The *modular group* $PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z})/\{\pm\mathbb{I}_2\}$ is a discrete subgroup of $PSL_2(\mathbb{R})$.

Definition 2.38. A discrete subgroup of $PSL_2(\mathbb{R})$ is called a *Fuchsian group*.

Remark 2.39. (i) Remember that a *discrete set* is a set $A \subseteq X$, where every point $x \in A$ has a neighbourhood in X containing only x.
(ii) We will later need the fact that the intersection of a discrete set A and a compact set K is finite. This is true because if we assume by contradiction that $A \cap K$ is infinite, then since K is compact there exists a limit point $x \in A \cap K$. This would mean that there are infinitely many neighbourhoods of x intersecting $A \cap K$. But this cannot be true since A is a discrete set.

Definition 2.40. Let $\{M_\alpha : \alpha \in A\}$ be a family of subsets of a locally compact metric space X. If for every compact set $K \subseteq X$, $M_\alpha \cap K \neq \emptyset$ only for finitely many $\alpha \in A$, then $\{M_\alpha : \alpha \in A\}$ is called *locally finite*.

Definition 2.41. Let G be a group which acts on X, a locally compact metric space. If for every $x \in X$ the family of singletons $\{\{gx\} : g \in G\}$ is locally finite then we say that G acts *properly discontinuously* on X.

Corollary 2.42. The group G acts properly discontinuously on the locally compact metric space X if and only if the order of $Stab_G(x)$ is finite for every $x \in X$ and the orbit of x has no accumulation points for every $x \in X$.

Proof. "$\Rightarrow$": Assume G acts properly discontinuously on X. By Definitions 2.40 and 2.41 for every $x \in X$ there exists a compact set $K \subseteq X$ such that $\{gx\} \cap K \neq \emptyset$ only for finitely many $g \in G$. Therefore for every $x \in X$ each orbit of x has no accumulation points. Now assume that $Stab_G(x)$ has infinite order and $x \in K$. Then $\{x\} = \{gx\} \cap K \neq \emptyset$ for infinitely many $g \in G$.
"$\Leftarrow$": Let $K \subseteq X$ be compact and define the set $K' := K \cap Gx$. We know that the orbit Gx of x is a discrete set (by Remark 2.43). Thus by Remark 2.39 K' is finite, which implies $\{gx\} \cap K \neq \emptyset$ only for finitely many $g \in G$ $\square$

Remark 2.43. The condition that the orbit of x has no accumulation points is equivalent to the condition of the orbit of x being a discrete set. Assume that G acts on X by Möbius transformations. Then one direction of this equivalence can be shown as follows:
Assume by contradiction that the orbit of x has an accumulation point $s \in X$. That is, there exists a sequence $(g_n)_n \in G$ such that $T_{g_n}(x)$ converges to s. But this means that for any $\epsilon > 0$ the distance $d_X(T_{g_n}(x), T_{g_{n+1}}(x)) < \epsilon$. We have seen that Möbius transformations are isometries, and so

$$d_X(T_{g_n}(x), T_{g_{n+1}}(x)) = d_X(x, T_{g_n^{-1}} T_{g_{n+1}}(x)) < \epsilon.$$

Thus x is an accumulation point for its orbit, which implies that the orbit of x is not a discrete set.

Now we can turn to the main statement of this section.

Theorem 2.44. A subgroup Γ of $PSL_2(\mathbb{R})$ acts properly discontinuously on $\mathbb{H}$ if and only if Γ is a Fuchsian group.

To prove Theorem 2.44 we need the following auxiliary Lemma.

Lemma 2.45. The set $E = \{g \in PSL_2(\mathbb{R}) : T_g(z) \in K\}$ is compact for a compact subset K of $\mathbb{H}$ and a point z in $\mathbb{H}$.

Proof. Consider the set $E_1 = \{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R}) : \frac{az+b}{cz+d} \in K\}$ and the projection map $\pi : SL_2(\mathbb{R}) \to PSL_2(\mathbb{R})$. If E_1 is compact then $\pi(E_1) = E$ is compact, since the image of any compact set under a continuous map is compact. Define

$$\begin{aligned} &\beta : SL_2(\mathbb{R}) \to \mathbb{H} \\ &\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \beta(\begin{pmatrix} a & b \\ c & d \end{pmatrix}) := \frac{az+b}{cz+d}. \end{aligned}$$

If we identify E_1 with a subset of $\mathbb{R}^4$, as done in Definition 2.13, we show E_1 is compact by showing it is closed and bounded:

(i) $E_1 = \beta^{-1}(K)$ is closed as K is closed and β is continuous.

(ii) Since K is bounded there exists some constant $M_1 > 0$ such that

$$|\frac{az+b}{cz+d}| < M_1$$

for all $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in E_1$. And since K is a subset of $\mathbb{H}$ there is a constant $M_2 > 0$ such that

$$\Im(\frac{az+b}{cz+d}) \overset{Remark\ 1.15(ii)}{=} \frac{\Im(z)}{|cz+d|^2} \geq M_2.$$

Thus $|cz+d| \leq \sqrt{\frac{\Im(z)}{M_2}}$ and $|az+b| < M_1\sqrt{\frac{\Im(z)}{M_2}}$. This shows that a, b, c, d are bounded and consequently E_1 is bounded.

□

Proof. (of Theorem 2.44)
Assume first that Γ is a Fuchsian group and let K be a compact subset of $\mathbb{H}$ and $z \in \mathbb{H}$. Showing that Γ acts properly discontinuously on $\mathbb{H}$ is the same as showing that the set $\{g \in \Gamma : T_g(z) \in K\}$ is finite. If we write

$$\{g \in \Gamma : T_g(z) \in K\} = \{g \in PSL_2(\mathbb{R}) : T_g(z) \in K\} \cap \Gamma$$

and notice that the first term on the right hand side is compact by Lemma 2.45 the claim follows, because the intersection of a compact and a discrete set is finite (Remark 2.39).
On the other hand let $\{g \in \Gamma : T_g(z) \in K\}$ be finite for every $z \in \mathbb{H}$ and compact $K \subseteq \mathbb{H}$, but assume that Γ is not discrete. Since Γ is not discrete there exists a sequence $\{g_k\}$ of elements in Γ, where the g_k are distinct and not equal to the identity element, such that g_k converges to the identity element for $k \to \infty$. Thus for any point $s \in \mathbb{H}$ that is not fixed by any g_k the sequence $\{T_{g_k}(s)\}$ does not contain s, consists of distinct points, and converges to s for $k \to \infty$. Hence any compact set K in $\mathbb{H}$ containing s in its interior contains infinitely many points of the s orbit, i.e. $\{g_k \in \Gamma : T_{g_k}(s) \in K\}$ is an infinite set. This is a contradiction to our assumption that Γ acts properly discontinuously on $\mathbb{H}$. □

Corollary 2.46. For a subgroup Γ of $PSL_2(\mathbb{R})$ and for any $z \in \mathbb{H}$ the orbit $\Gamma z = \{T_g(z) : g \in \Gamma\}$ of z is a discrete subset of $\mathbb{H}$ if and only if the action of Γ on $\mathbb{H}$ is properly discontinuous.

Proof. If Γ acts properly discontinuously on $\mathbb{H}$ then by Corollary 2.42 for all $z \in \mathbb{H}$ the orbit Γz has no accumulation points. Then it follows by Remark 2.43 that Γz is discrete.
If Γ does not act properly discontinuously on $\mathbb{H}$ then Γ is not discrete by Theorem 2.44. By using the argumentation in the proof of Theorem 2.44 we find for any $s \in \mathbb{H}$ a sequence $\{T_{g_k}(s)\}$ of distinct points converging to s. Thus the orbit Γs of s is not a discrete subset of $\mathbb{H}$. $\square$

2.3 Fundamental domains

Let a subgroup Γ of $PSL_2(\mathbb{R})$ act properly discontinuously on $\mathbb{H}$.

Definition 2.47. A closed set F in X is called *fundamental domain* or *fundamental region* of Γ if the following are satisfied:

(i) $X = \cup_{g\in\Gamma} T_g(F)$
(ii) $\overset{\circ}{F} \cap T_g(\overset{\circ}{F}) = \emptyset$ for all $g \in \Gamma$ not equal to the identity element.

Note that $\overset{\circ}{F}$ denotes the interior of F and $\partial F = F \setminus \overset{\circ}{F}$ the boundary of F.
The *tessellation* of X is the family $\{T_g(F) : g \in \Gamma\}$.

We will now consider two fundamental domains.

Example 2.48. The subgroup $\Gamma = \{g_n : T_{g_n}(z) = z + n, n \in \mathbb{Z}\}$ of the group of all Möbius transformations is a Fuchsian group.
The closed set $F = \{z \in \mathbb{H} : 0 \leq \Re(z) \leq 1\}$ is a fundamental domain of Γ since:

(i) $\cup_{n\in\mathbb{Z}} T_{g_n}(F) = \cup_{n\in\mathbb{Z}}\{z \in \mathbb{H} : n \leq \Re(z) \leq n + 1\} = \mathbb{H}$ and
(ii) For $n, m \in \mathbb{Z}$ we have that $T_{g_n}(\overset{\circ}{F}) \cap T_{g_m}(\overset{\circ}{F}) \neq \emptyset$ if and only if $m = n$.

Proposition 2.49. A fundamental domain of $PSL_2(\mathbb{Z})$ is given by the set

$$F = \{z \in \mathbb{H} : |\Re(z)| \leq \frac{1}{2}, |z| \geq 1\},$$

which is shown in Figure 2.5.

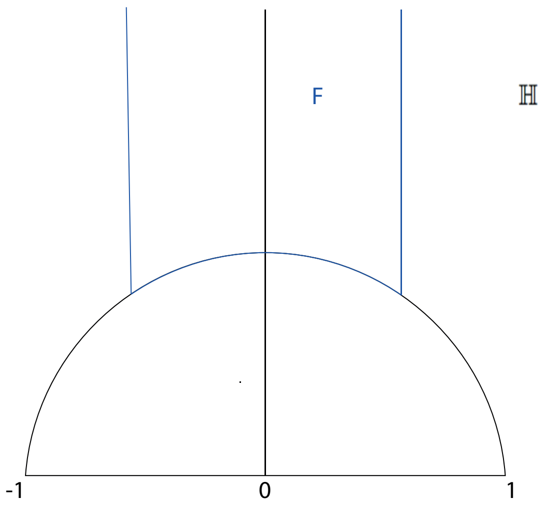

Figure 2.5: A fundamental domain F of $PSL_2(\mathbb{Z})$.

Proof. (i) To get the first property in Definition 2.47 we will show that for any $z \in \mathbb{H}$ there exists some $g \in PSL_2(\mathbb{Z})$ such that $T_g(z) \in F$.
Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PSL_2(\mathbb{Z})$, then for any $z \in \mathbb{H}$ we have $\Im(T_g(z)) = \frac{\Im(z)}{|cz+d|^2}$ by Remark 1.15. Let m be any positive real number. Since $|cz+d| < m$ only for finitely many pairs $c, d \in \mathbb{Z}$ there must exist a matrix $g \in PSL_2(\mathbb{Z})$ such that $T_g(z)$ has maximal imaginary part. That is,

$$\Im(T_g(z)) = \max\{\Im(T_h(z)) : h \in PSL_2(\mathbb{Z})\}.$$

Let $\tau = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $k \in \mathbb{Z}$ such that $|\Re(T_{\tau^k} T_g(z))| \leq \frac{1}{2}$.
Claim: $w := T_{\tau^k g}(z) \in F$
Proof of the claim: Assume $|w| = |T_{\tau^k g}(z)| = |T_g(z) + k| < 1$. Since $\Im(\frac{-1}{w}) = \frac{w-\bar{w}}{2i|w|^2} = \frac{\Im(w)}{|w|^2}$ and $\Im(w) = \Im(T_g(z))$ it follows that $\Im(\frac{-1}{w}) > \Im(T_g(z))$, contradicting the maximality of the imaginary part of $T_g(z)$. Thus $|w| \geq 1$ and $w \in F$.

(ii) To get the second property in Definition 2.47 we will show that the boundary of F gets mapped to itself. For that let $z, w \in F$ and set $w = T_g(z)$.
Claim: Either $|\Re(z)| = \frac{1}{2}$ and $w = z \pm 1$, or $|z| = 1$ and $w = \frac{-1}{z}$.
Proof of the claim: Assume w.l.o.g. that $\Im(w) \geq \Im(z)$. This implies together with Remark 1.15 that $|cz + d| \leq 1$. Since $z \in F$ it follows that $c = 0, \pm 1$.

By assuming $c = 0$ it follows that $d = \pm 1$ and so $\Im(w) = \Im(z)$, which means that $w = z \pm b$. Since $z, w \in F$ we know that $|\Re(z)| \leq \frac{1}{2}$ and $|\Re(w)| \leq \frac{1}{2}$ and thus either $b = 0$ and $g = \ I_2$ or $b = \pm 1$ and $\{\Re(z), \Re(w)\} = \{\frac{-1}{2}, \frac{1}{2}\}$. If we assume that $c = 1$, then since $z \in F$ and $|cz + d| \leq 1$ either $d = 0$ and thus $|z| = 1$ or $d = \pm 1$ in the case of $z = \frac{\sqrt{3}}{2} i \mp \frac{1}{2}$. In the case $c = -1$ by replacing g with $-g$ we get the same results as for $c = 1$ as T_g and T_{-g} define the same Möbius transformations.

$\square$

Proposition 2.50. $PSL_2(\mathbb{Z})$ is generated by $\tau = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\sigma = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

Notice that for $\tau, \sigma \in PSL_2(\mathbb{Z})$ we get the Möbius transformations

$$T_\tau(z) = z + 1, T_{\tau^{-1}}(z) = z - 1$$

and

$$T_\sigma(z) = -\frac{1}{z} = T_{\sigma^{-1}}(z)$$

for all $z \in \mathbb{H}$.

Proof. We want to show that any element $h \in PSL_2(\mathbb{Z})$ can be written as

$$\tau^{n_1} \sigma \tau^{n_2} \sigma ... \sigma \tau^{n_j},$$

for $n_i \in \mathbb{Z}$, $1 \leq i \leq j$. Let F be the fundamental domain of $PSL_2(\mathbb{Z})$ and let z_0 be in the interior of F. Let h be any element of $PSL_2(\mathbb{Z})$ such that $z = T_h(z_0) \in \mathbb{H}$. Since $\tau, \sigma \in PSL_2(\mathbb{Z})$ there is a word

$$g := \tau^{n_1} \sigma \tau^{n_2} \sigma ... \sigma \tau^{n_j}$$

generated by τ and σ which is in $PSL_2(\mathbb{Z})$. By the same argumentation as in the proof of Proposition 2.49 there is such a g where $\Im(T_g(z))$ is maximal. Then using the first claim of the proof of Proposition 2.49 shows that there is a $k \in \mathbb{Z}$ such that $T_{\tau^k g}(z) \in F$. Then

$$\tilde{g} := \tau^k g$$

is also generated by τ and σ and we can write $T_{\tilde{g}}(z) = T_{\tilde{g}h}(z_0)$. Since $z_0 \in \mathring{F}$ and $T_{\tilde{g}}(z) \in F$ it follows by using the second claim in the proof of Proposition 2.49 that $\tilde{g}h$ must be the identity matrix in $PSL_2(\mathbb{Z})$. Thus $h = \tilde{g}^{-1}$ is also generated by σ and τ. $\square$

Remark 2.51. Analogously to the proof of Proposition 2.50 it can be shown that the sets $U^+ := \{\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} : s \in \mathbb{R}\}$ and $U^- := \{\begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} : s \in \mathbb{R}\}$ generate $SL_2(\mathbb{R})$.

Even though fundamental regions are not uniquely determined the following Theorem states that the hyperbolic area of any two fundamental regions for a Fuchsian group is the same.

Theorem 2.52. Let two fundamental regions F_1, F_2 of a Fuchsian group Γ be given. Assume additionally that $\mu(F_1) < \infty$ and $\mu(\partial F_1) = \mu(\partial F_2) = 0$. Then it follows that $\mu(F_1) = \mu(F_2)$.

Proof. Since F_2 is a fundamental region $\cup_{g\in\Gamma} T_g(\mathring{F_2}) \subset \mathbb{H}$. Therefore we can write

$$F_1 = F_1 \cap \mathbb{H} \supseteq F_1 \cap (\cup_{g\in\Gamma} T_g(\mathring{F_2})) = \cup_{g\in\Gamma}(F_1 \cap T_g(\mathring{F_2})).$$

Then the hyperbolic area of F_1 is given by

$$\mu(F_1) \geq \mu(\cup_{g\in\Gamma}(F_1 \cap T_g(\mathring{F_2}))) = \sum_{g\in\Gamma} \mu(F_1 \cap T_g(\mathring{F_2})), \tag{2.4}$$

where the equation follows since the sets $F_1 \cap T_g(\mathring{F_2})$, $g \in \Gamma$, are disjoint. By Theorem 1.37 hyperbolic area is preserved by Möbius transformations. Thus equation (2.4) becomes

$$\begin{aligned}\mu(F_1) &\geq \sum_{g\in\Gamma} \mu(T_g^{-1}(F_1) \cap \mathring{F_2}) \overset{Remark\ 1.16(i)}{=} \sum_{g\in\Gamma} \mu(T_{g^{-1}}(F_1) \cap \mathring{F_2}) \\ &= \sum_{g\in\Gamma} \mu(T_g(F_1) \cap \mathring{F_2}).\end{aligned} \tag{2.5}$$

The last equation follows by the fact that the sum is taken over all $g \in \Gamma$. The sets $T_g(F_1) \cap \mathring{F_2}$ are not disjoint and so equation (2.5) becomes

$$\mu(F_1) \geq \mu(\cup_{g\in\Gamma}(T_g(F_1) \cap \mathring{F_2})) = \mu(\mathring{F_2}) = \mu(F_2),$$

using the assumptions that F_1 is a fundamental region and $\mu(\partial F_2) = 0$. We obtain $\mu(F_2) \geq \mu(F_1)$ by swapping F_1 with F_2. Thus $\mu(F_1) = \mu(F_2)$. □

Definition 2.53. A Fuchsian group Γ is called a *lattice* in $PSL_2(\mathbb{R})$ if its fundamental domain F has finite measure

Example 2.54. We have seen that

$$F = \{z \in \mathbb{H} : |\Re(z)| \leq \frac{1}{2}, |z| \geq 1\}$$

is a fundamental domain of $PSL_2(\mathbb{Z})$. Therefore $\Im(z) \geq \frac{\sqrt{3}}{2}$ for any $z \in F$ and $PSL_2(\mathbb{Z})$ is a lattice in $PSL_2(\mathbb{R})$ since

$$\mu(F) = \int_{z\in F} dA \leq \int_{\frac{\sqrt{3}}{2}}^{\infty} \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{dxdy}{y^2} = \int_{\frac{\sqrt{3}}{2}}^{\infty} \frac{dy}{y^2} = \frac{2}{\sqrt{3}} < \infty.$$

3 Dynamics of the geodesic flow

3.1 The geodesic flow

We have seen in Remark 2.9 that any geodesic γ of unit speed is uniquely determined by a point z on γ and the unit vector v in the direction of γ with base point z.

Definition 3.1. The *geodesic flow* on $\mathbb{H}$ is given by

$$\begin{aligned} g_t : T^1\mathbb{H} &\to T^1\mathbb{H} \\ (z,v) &\mapsto g_t(z,v) = (\gamma(t), \gamma'(t)), \end{aligned}$$

for the geodesic $\gamma(t)$ going through $z \in \mathbb{H}$ at time 0 in the direction of $v = \gamma'(0)$.

Remark 3.2. (i) The geodesic flow is a usual flow since:

1) g_0 is equal to the identity map because

$$g_0(z,v) = (\gamma(0), \gamma'(0)) = (z,v).$$

2)

$$\begin{aligned} g_s(g_t(z,v)) &= g_s(g_t(\gamma(0), \gamma'(0)) = g_s(\gamma(t), \gamma'(t)) \\ &= (\gamma(s+t), \gamma'(s+t)) = g_{s+t}(\gamma(0), \gamma'(0)) = g_{s+t}(z,v), \end{aligned}$$

for all $s, t \in \mathbb{R}$.

(ii) By Proposition 1.28 the imaginary axis is a geodesic. The vector i with unit length pointing upwards and base point i determines the imaginary axis. We can parameterize the geodesic by

$$\gamma(t) = ie^t$$

for $t \in \mathbb{R}$. Since $\gamma(0) = i$, $\gamma'(0) = i$ and $\|\gamma'(t)\|_{\gamma(t)} = \sqrt{\frac{e^{2t}}{e^{2t}}} = 1$ we can define the geodesic flow along the imaginary axis as

$$g_t(i,i) = (ie^t, ie^t).$$

(iii) Let us consider the matrix

$$a_t^{-1} := \begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{\frac{-t}{2}} \end{pmatrix} \in PSL_2(\mathbb{R}).$$

Then the derivative of the Möbius transfrmation $T_{a_t^{-1}}$ of the point $(i,i) \in T^1\mathbb{H}$ is given by

$$DT_{a_t^{-1}}(i,i) = (T_{a_t^{-1}}(i), T'_{a_t^{-1}}(i)i) = (\frac{e^{\frac{t}{2}}i + 0}{0 + e^{\frac{-t}{2}}}, \frac{i}{(0 + e^{\frac{-t}{2}})^2}) = (ie^t, ie^t) = g_t(i,i).$$

(iv) Remember that for elements g in $PSL_2(\mathbb{R})$ the Möbius transformation T_g bijectively maps geodesic to geodesic (Lemma 1.29 and Remark 1.33). Thus for any point $(z,v) \in T^1\mathbb{H}$ determining a geodesic γ_1, there exists a unique element g of $PSL_2(\mathbb{R})$ such that DT_g maps the parametrization of the imaginary axis to the parametrization of γ_1, i.e.

$$g_t(z,v) = DT_g(g_t(i,i)).$$

Corollary 3.3. The geodesic flow on $\mathbb{H}$ is described by right multiplication by a_t^{-1}, that is the geodesic flow

$$\begin{aligned} g_t : T^1\mathbb{H} &\to T^1\mathbb{H} \\ (z,v) = DT_g(i,i) &\mapsto DT_{ga_t^{-1}}(i,i) \end{aligned}$$

corresponds to the right multiplication by a_t^{-1},

$$\begin{aligned} R_{a_t} : PSL_2(\mathbb{R}) &\to PSL_2(\mathbb{R}) \\ g &\mapsto R_{a_t}(g) = ga_t^{-1}. \end{aligned}$$

Proof. By Theorem 2.10 we already know we can identify g and $ga_t^{-1} \in PSL_2(\mathbb{R})$ with $DT_g(i,i)$ and $DT_{ga_t^{-1}}(i,i) \in T^1\mathbb{H}$ respectively. Therefore it remains to show that the equation $g_t(z,v) = DT_{ga_t^{-1}}(i,i)$ holds:
By Remark 3.2 we have

$$g_t(z,v) = DT_g(g_t(i,i)) = DT_g(ie^t, ie^t).$$

Now let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PSL_2(\mathbb{R})$. Then $DT_g(ie^t, ie^t) = (\frac{aie^t+b}{cie^t+d}, \frac{ie^t}{(cie^t+d)^2})$ and by multiplying the first term with $\frac{e^{\frac{-t}{2}}}{e^{\frac{-t}{2}}}$ and the second with $\frac{e^{-t}}{e^{-t}}$ we get

$$\begin{aligned} DT_g(ie^t, ie^t) &= (\frac{aie^{\frac{t}{2}} + be^{\frac{-t}{2}}}{cie^{\frac{t}{2}} + de^{\frac{-t}{2}}}, \frac{i}{(cie^{\frac{t}{2}} + de^{\frac{-t}{2}})^2}) = DT_{\begin{pmatrix} ae^{\frac{t}{2}} & be^{\frac{-t}{2}} \\ ce^{\frac{t}{2}} & de^{\frac{-t}{2}} \end{pmatrix}}(i,i) \\ &= DT_{ga_t^{-1}}(i,i). \end{aligned}$$

Thus $g_t(z,v) = DT_{ga_t^{-1}}(i,i)$. □

Remember that by Corollary 2.11 the derivative action

$$DT : PSL_2(\mathbb{R}) \times T^1\mathbb{H} \to T^1\mathbb{H}$$

corresponds to left multiplication in $PSL_2(\mathbb{R})$.

Let us recall the definition of the stable and unstable manifold as well as the definition of a horocycle.

Definition 3.4. The *stable manifold* of (z, v) for the geodesic flow is given by

$$W^s((z,v)) := \{(z', v') \in T^1\mathbb{H} : d(g_t(z,v), g_t(z',v')) \overset{t\to\infty}{\to} 0\}$$

and the *unstable manifold* of (z, v) by

$$W^u((z,v)) := \{(z', v') \in T^1\mathbb{H} : d(g_t(z,v), g_t(z',v')) \overset{t\to-\infty}{\to} 0\}.$$

Definition 3.5. The *horocycles centered at infinity* are the horizontal lines

$$\{t + ir : t \in \mathbb{R}\}, r \in \mathbb{R}_{>0}.$$

The *horocycles centered at* x, $x \in \mathbb{R}$, are circles which at the point x are tangent to $\mathbb{R}$.

Corollary 3.6. For the geodesic flow through the point $(i, i) \in T^1\mathbb{H}$ the stable manifold is the set of upwards pointing vectors on the horizontal line $\{t+i : t \in \mathbb{R}\}$.

Proof. We know that for any vector $(x + i, i), x \in \mathbb{R}$, the geodesic is a vertical line. Then for $g = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \in PSL_2(\mathbb{R})$ the Möbius transformation T_g satisfies $DT_g(i, i) = (i + x, i)$. Thus

$$\begin{aligned} g_t(x+i,i) &\overset{Corollary\ 3.3}{=} DT_{ga_t^{-1}}(i,i) = DT_{\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{\frac{-t}{2}} \end{pmatrix}}(i,i) \\ &= DT_{\begin{pmatrix} e^{\frac{t}{2}} & xe^{\frac{-t}{2}} \\ 0 & e^{\frac{-t}{2}} \end{pmatrix}}(i,i) = (x + ie^t, ie^t) \end{aligned}$$

and the geodesic trajectories $g_t(x+i, i) = (x+ie^t, ie^t)$ and $g_t(i, i) = (ie^t, ie^t)$ move parallel to each other. Now let

$$h(k) = \frac{x}{e^k} + ie^t, k \in \mathbb{R}_{\geq 0},$$

be a path from $x + ie^t$ to ie^t. The length of $h(k)$ is given by

$$\begin{aligned} L(h(k)) &= \int_0^\infty \frac{\sqrt{(\frac{d}{dk}\frac{x}{e^k})^2 + (\frac{d}{dk}e^t)^2}}{e^t} dk = \frac{1}{e^t}\int_0^\infty \sqrt{(-\frac{x}{e^k})^2} dk \\ &= \frac{|x|}{e^t}\int_0^\infty \frac{1}{e^k} dk = \frac{|x|}{e^t}. \end{aligned}$$

Since $\frac{|x|}{e^t} \to 0$, as $t \to \infty$, the distance between $g_t(i, i)$ and $g_t(x + i, i)$ tends to zero. We now want to show that no other points $(z, v) \in T^1\mathbb{H}$ belong to the stable

manifold.
First consider $(z, v) := (x + iy, i)$ for any $x \in \mathbb{R}$ and $y \in (0, \infty) \setminus \{1\}$. We can calculate

$$\begin{aligned} g_t(x+iy,i) &= DT_{\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} e^{\frac{t}{2}} & 0 \\ 0 & e^{\frac{-t}{2}} \end{pmatrix}}(i,i) = DT_{\begin{pmatrix} ye^{\frac{t}{2}} & xe^{\frac{-t}{2}} \\ 0 & e^{\frac{-t}{2}} \end{pmatrix}}(i,i) \\ &= (x+iye^t, ie^t) \end{aligned}$$

and the length of the path

$$h(k) = x + ike^t,$$

$k \in [1, y]$ (or $k \in [y, 1]$ if $y \in (0, 1)$), between $x + ie^t$ and $x + iye^t$ given by

$$L(h(k)) = \int \sqrt{\frac{e^{2t}}{(ke^t)^2}} dk = |\ln(y)| > 0$$

for all $t \in \mathbb{R}$. This means that the distance between $g_t(x + i, i)$ and $g_t(x + iy, i)$ is constant and positive for all $t \in \mathbb{R}$. By the triangle inequality we have

$$d(g_t(x+iy,i), g_t(i,i)) + d(g_t(i,i), g_t(x+i,i)) \geq d(g_t(x+iy,i), g_t(x+i,i)).$$

By the above

$$d(g_t(i,i), g_t(x+i,i)) \overset{t\to\infty}{\to} 0.$$

Thus

$$d(g_t(x+iy,i), g_t(i,i)) > 0$$

for all $t \in \mathbb{R}$. This means that the distance between $g_t(x + iy, i)$ and $g_t(i, i)$ does not tend to zero as $t \to \infty$.
Now consider $(z, v) \in T^1\mathbb{H}$ with $v \neq i$. The corresponding geodesic of (z, v) is a semicircle with endpoints on $\mathbb{R}$, so $g_t((z, v)) \to u \in \mathbb{R}$ as $t \to \infty$. On the other hand, $g_t(i, i) \to \infty$ as $t \to \infty$. Therefore $d(g_t(z, v), g_t(i, i))$ does not tend to zero as $t \to \infty$.
Hence we can conclude that the set $\{(x + i, i) : x \in \mathbb{R}\}$ is the stable manifold of (i, i) for the geodesic flow. □

Remark 3.7. It can be shown that for the geodesic flow through the point $(i, -i) \in T^1\mathbb{H}$ the unstable manifold is the set of downwards pointing vectors on the line $\{t + i : t \in \mathbb{R}\}$. That is, the geodesic flow $g_t(i, -i)$ is given by

$$DT_{ga_t}(i,i) = (ie^{-t}, -ie^{-t})$$

and

$$g_t(x+i,-i) = DT_{ga_t}(i,-i) = (x+ie^{-t}, -ie^{-t}).$$

By choosing the path

$$h(k) = kx + ie^{-t}, k \in [0,1],$$

from ie^{-t} to $x + ie^{-t}$, we calculate

$$L(h(k)) = \int_0^1 \frac{\sqrt{x^2}}{e^{-t}} dk = |x|e^t \to 0$$

as $t \to -\infty$ and continue analogously to the proof in Corollary 3.6.

Remark 3.8. Notice that we can equivalently define the horocycles as curves whose perpendicular geodesics converge to the same point. The set of vectors on a horocyle defining these geodesics which converge to the same point is called *stable horocycle* and corresponds to the stable manifold for a vector on this horocyle. On the other hand the set of vectors on a horocyle whose distances under the geodesic flow go to infinity is called *unstable horocycle* and corresponds to the unstable manifold for a vector on this horocyle.

Figure 3.1 shows a horocycle centered at infinity and a horocycle centered at x in blue. The corresponding stable horocycles/manifolds are represented by green vectors whereas the unstable horocycles/manifolds are represented by red vectors.

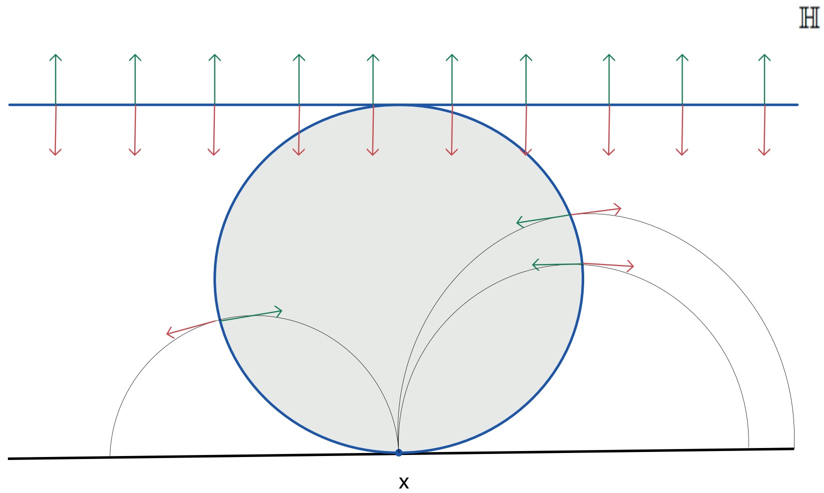

Figure 3.1: Horocycles and stable/unstable manifolds.

Definition 3.9. The *stable horocycle flow* is given by

$$\begin{aligned} h_s &: T^1\mathbb{H} \to T^1\mathbb{H} \\ &(z,v) \mapsto h_s(z,v) = h_s(DT_g(i,i)) = DT_{gu^-(s)^{-1}}(i,i), \end{aligned} \tag{3.1}$$

if the Möbius transformation T_g maps $(i,i) \in T^1\mathbb{H}$ to $(z,v) \in T^1\mathbb{H}$ and

$$u^-(s)^{-1} = u^-(-s) = \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix} \in PSL_2(\mathbb{R}).$$

It sends (z, v) belonging to a stable horocycle to another vector on the same stable horocycle.
Analogously, the *unstable horocycle flow* is given by

$$\begin{aligned} h_s : T^1\mathbb{H} &\to T^1\mathbb{H} \\ (z, v) &\mapsto h_s(z, v) = h_s(DT_g(i, i)) = DT_{gu^+(s)^{-1}}(i, i), \end{aligned} \tag{3.2}$$

with

$$u^+(s)^{-1} = u^+(-s) = \begin{pmatrix} 1 & 0 \\ -s & 1 \end{pmatrix} \in PSL_2(\mathbb{R}).$$

It sends (z, v) belonging to an unstable horocycle to another vector on the same unstable horocycle.

Remark 3.10. Just as in Corollary 3.3, (3.1) corresponds to

$$\begin{aligned} R_{u^-(s)} : PSL_2(\mathbb{R}) &\to PSL_2(\mathbb{R}) \\ g &\mapsto R_{u^-(s)}(g) = gu^-(-s) \end{aligned}$$

and (3.2) corresponds to

$$\begin{aligned} R_{u^+(s)} : PSL_2(\mathbb{R}) &\to PSL_2(\mathbb{R}) \\ g &\mapsto R_{u^+(s)}(g) = gu^+(-s). \end{aligned}$$

3.2 Dynamics on $\Gamma \setminus PSL_2(\mathbb{R})$

Notice that the geodesic flow on $\mathbb{H}$ is not recurrent, since for any $g \in PSL_2(\mathbb{R})$ the orbit leaves any compact set at some time. To get more exciting dynamics we will consider geodesic flows on quotient spaces of $PSL_2(\mathbb{R})$.
Let F be a fundamental domain for the action of a Fuchsian group Γ on $\mathbb{H}$ and let $\pi : \mathbb{H} \to \Gamma \setminus \mathbb{H}$ be the natural projection induced by Γ, where $\Gamma \setminus \mathbb{H}$ consists of Γ-orbits.

Definition 3.11. We call F of Γ *locally finite* if and only if for each compact subset K of $\mathbb{H}$ the set $\{T_g(F) \cap K : g \in \Gamma\}$ is finite.

Theorem 3.12. If F is locally finite then there exists a homeomorphism between $\Gamma \setminus F$ and $\Gamma \setminus \mathbb{H}$.

The proof can be found in [2]. It is included in the proof of Theorem 9.2.4.

Remark 3.13. By Theorem 2.52, we know that if a fundamental region F of Γ has finite hyperbolic area, then $\mu(F) = \mu(\Gamma \setminus F)$. Thus combining this with the last Theorem shows

$$\mu(\Gamma \setminus \mathbb{H}) = \mu(\Gamma \setminus F) = \mu(F).$$

Corollary 3.14. $T^1F := \{(z,v) \in T\mathbb{H} : z \in F, \|v\|_z = 1\}$ is a fundamental domain for the action of Γ on $PSL_2(\mathbb{R})$.

Proof. By Theorem 2.10, $PSL_2(\mathbb{R})$ can be identified with $T^1\mathbb{H}$. Thus we can consider the action

$$DT : \Gamma \times T^1\mathbb{H} \to T^1\mathbb{H}.$$

It follows that

$$T^1\mathbb{H} = \cup_{g\in\Gamma} DT_g(T^1F),$$

since we assume F to be a fundamental region of Γ on $\mathbb{H}$. Let $T_g(z) = \tilde{z}$, $z \in F$, then $\tilde{z} \in F$ if and only if g is the identity. Therefore

$$(T^1\overset{\circ}{F}) \cap DT_g(T^1\overset{\circ}{F}) = \emptyset$$

follows for all g not equal to the identity. $\square$

Remark 3.15. Let Γ be a Fuchsian group which does not contain elliptic elements, i.e. it does not contain fixed points in $\mathbb{H}$. Then it follows that $T^1(\Gamma\setminus\mathbb{H}) = \Gamma\setminus T^1\mathbb{H}$ and $\Gamma \setminus PSL_2(\mathbb{R})$ are homeomorphic. To get an idea of why this is true consider the following maps.

$$\begin{array}{ccc} T^1\mathbb{H} & \xrightarrow{\phi} & PSL_2(\mathbb{R}) \\ \downarrow{\scriptstyle \pi'} & & \downarrow{\scriptstyle \pi} \\ \Gamma \setminus T^1\mathbb{H} & & \Gamma \setminus PSL_2(\mathbb{R}) \end{array}$$

Figure 3.2: Identification of $\Gamma \setminus T^1\mathbb{H}$ with $\Gamma \setminus PSL_2(\mathbb{R})$.

Let

$$\begin{aligned} \phi : T^1\mathbb{H} &\to PSL_2(\mathbb{R}) \\ (z,v) = DT_g(i,i) &\mapsto g \end{aligned}$$

be the homeomorphism from Theorem 2.10, and let

$$\begin{aligned} \pi' : T^1\mathbb{H} &\to \Gamma \setminus T^1\mathbb{H} \\ (z,v) &\mapsto \Gamma(z,v) = \{(z',v') := DT_h(z,v) : h \in \Gamma\}, \\ \pi : PSL_2(\mathbb{R}) &\to \Gamma \setminus PSL_2(\mathbb{R}) \\ g &\mapsto \Gamma g = \{hg : h \in \Gamma\} \end{aligned}$$

be the corresponding projections. Since

$$(z',v') = DT_h(z,v) = DT_h(DT_g(i,i)) \overset{Cor.\ 2.11}{=} DT_{hg}(i,i),$$

we know by Theorem 2.10 that we can identify $(z',v') \in \Gamma \setminus T^1\mathbb{H}$ with $hg \in \Gamma \setminus PSL_2(\mathbb{R})$. Thus $\Gamma \setminus T^1\mathbb{H} \cong \Gamma \setminus PSL_2(\mathbb{R})$.

Remark 3.16. By using the projection $\pi' : T^1\mathbb{H} \to T^1(\Gamma \setminus \mathbb{H})$ we can define the geodesic flow on $\Gamma \setminus \mathbb{H}$

$$\begin{aligned} g_t : T^1(\Gamma \setminus \mathbb{H}) &\to T^1(\Gamma \setminus \mathbb{H}) \\ y := \Gamma(z, v) &\mapsto g_t(y). \end{aligned}$$

Remark 3.17. As in Corollary 3.3 the geodesic flow on $\Gamma \setminus \mathbb{H}$ corresponds to right multiplication by a_t^{-1}

$$\begin{aligned} R_{a_t} : \Gamma \setminus PSL_2(\mathbb{R}) &\to \Gamma \setminus PSL_2(\mathbb{R}) \\ x := \Gamma g &\mapsto R_{a_t}(x) = x a_t^{-1}. \end{aligned}$$

Example 3.18. Let $\Gamma = PSL_2(\mathbb{Z})$. We have seen in Example 2.54 that a fundamental domain of the action of $PSL_2(\mathbb{Z})$ on $\mathbb{H}$ is given by

$$F = \{z \in \mathbb{H} : |\Re(z)| \leq \frac{1}{2}, |z| \geq 1\}.$$

Since F is locally finite we can identify $PSL_2(\mathbb{Z}) \setminus \mathbb{H}$ with $PSL_2(\mathbb{Z}) \setminus F$ by Theorem 3.12 and thus we can regard the geodesic flow on $PSL_2(\mathbb{Z}) \setminus F$ given by

$$\begin{aligned} g_t : T^1(PSL_2(\mathbb{Z}) \setminus F) &\to T^1(PSL_2(\mathbb{Z}) \setminus F) \\ (z, v) = DT_g(i, i) &\mapsto g_t(z, v) = DT_{g a_t^{-1}}(i, i) \end{aligned}$$

as the map

$$\begin{aligned} R_{a_t} : PSL_2(\mathbb{Z}) \setminus PSL_2(\mathbb{R}) &\to PSL_2(\mathbb{Z}) \setminus PSL_2(\mathbb{R}) \\ x := PSL_2(\mathbb{Z}) g &\mapsto R_{a_t}(x) = x a_t^{-1}. \end{aligned}$$

Thus we can identify x with $(z, v) = DT_g(i, i) \in T^1(PSL_2(\mathbb{Z}) \setminus F)$ such that $g \in PSL_2(\mathbb{R})$ and $z = T_g(i) \in F$. If the geodesic is a vertical line, then the geodesic flow $R_{a_t}(x)$ follows the vertical line to infinity (represented by the black arrow in Figure 3.3). If we assume that the geodesic is not a vertical line, then the geodesic flow $R_{a_t}(x)$ follows the geodesic uniquely determined by (z, v) until the boundary of F is reached. That point $\tilde{z}$ on the boundary of F has a corresponding unit vector $\tilde{v}$ pointing outside of F. By applying $\tau^{\pm}$ or $\sigma^{\pm}$ to $(\tilde{z}, \tilde{v})$ we obtain a point $\bar{z}$ on the boundary of F with a corresponding vector $\bar{v}$ pointing inwards of F. Again $(\bar{z}, \bar{v})$ determines a new geodesic which is followed by the geodesic flow until the boundary of F is reached, where we repeat the former process. A possible geodesic trajectory is illustrated in Figure 3.3, were the blue arrow represents the vector (z, v), the green arrow the vector $(\tilde{z}, \tilde{v})$ and the red arrow the vector $(\bar{z}, \bar{v})$.

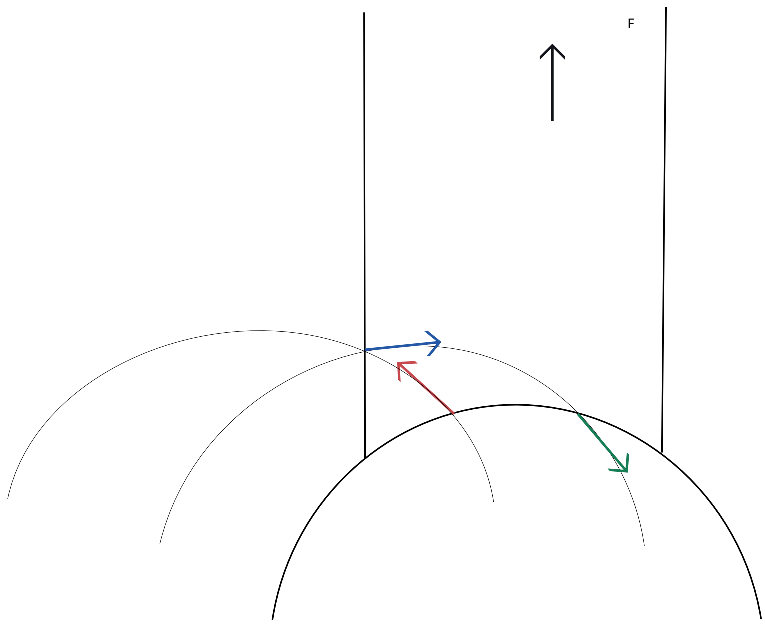

Figure 3.3: Geodesic flow on $PSL_2(\mathbb{Z}) \setminus F$.

Our next goal is to define a measure and a metric on the quotient space $\Gamma \setminus G$. We will need both in our last proof. Let us start by defining a left Haar measure on a locally compact topological group G.

Definition 3.19. Consider a locally compact topological group G with Borel-σ-algebra $\mathcal{B}_G$. We call a measure m_G on the Borel subsets of G a *left Haar measure* if m_G satisfies the following properties:

(i) m_G is left translation invariant, that is

$$m_G(B) = m_G(gB)$$

for all Borel subsets $B \in \mathcal{B}_G$ and for all $g \in G$;

(ii) The measure $m_G(O)$ of any open subset $O \subseteq G$ is positive;

(iii) The measure $m_G(K)$ of any compact subset $K \subseteq G$ is finite.

Remark 3.20. (i) We can analogously define a right Haar measure.

(ii) It can be shown (Corollary 8.8 in [5]) that m_G is unique up to a constant C. This means that for any left Haar measure μ and for any Borel set $B \in \mathcal{B}_G$ there exists a constant $C \in \mathbb{R}_{>0}$ such that $\mu(B) = Cm_G(B)$. Notice that $m_G(Bg)$ is also a left Haar measure for any $g \in G$ if m_G is. Thus there must exist a unique continuous homomorphism $\mod$ form the group G into the multiplicative group $(\mathbb{R}_{>0}, \cdot)$ such that

$$\mod (g) m_G(B) = m_G(Bg).$$

Definition 3.21. We call the function $\mod : G \to \mathbb{R}_{>0}$ described above *modular function* or *modular character*.

Definition 3.22. G is called *unimodular* if m_G is also a right Haar measure, or equivalently, if $\mod(G) = \{1\}$.

Theorem 3.23. Let Γ be a discrete subgroup of the closed linear group G, let

$$\pi : G \to X := \Gamma \setminus G$$

be the natural projection and let $F \subseteq G$ be a fundamental domain of Γ with finite left Haar measure. Then the following hold:

(i) $m_G(F') = m_G(F)$ for any other fundamental region F' of Γ;
(ii) G is unimodular;
(iii) The measure

$$m_X(B) := m_G(\pi^{-1}(B) \cap F)$$

on X defined for all measurable subsets B of X is finite;
(iv) For all $x \in X$ and $g \in G$ the measure m_X is invariant under the action $R_g(x) = xg^{-1}$.

Proof. (i) We prove a more general fact: Let $A, A' \subseteq G$ be two measurable sets and let $\pi|_A, \pi|_{A'}$ be injective such that $\pi(A) = \pi(A')$. Then A and A' have the same left Haar measure. Since projections are surjective and we assume $\pi|_A, \pi|_{A'}$ to be injective it follows that $\pi|_A, \pi|_{A'}$ are bijective as maps into $\pi(A)$. Thus for every $g \in A$ there exists a unique $\gamma \in \Gamma$ and $g' \in A'$ such that $g = \gamma g' \in \gamma A'$. Thus we can write

$$A = \dot{\bigcup}_{\gamma \in \Gamma} A \cap \gamma A' \tag{3.3}$$

and

$$A' = \dot{\bigcup}_{\gamma' \in \Gamma} A' \cap \gamma' A. \tag{3.4}$$

We can relate $A \cap \gamma A'$ with $A' \cap \gamma' A$ in the following way

$$\gamma^{-1}(A \cap \gamma A') = \gamma^{-1} A \cap A' \overset{\gamma^{-1} =: \gamma'}{=} \gamma' A \cap A', \tag{3.5}$$

for $\gamma \in \Gamma$. To prove the claim we just need to gather the points above:

$$\begin{aligned} m_G(A) &\overset{(3.3)}{=} \sum_{\gamma \in \Gamma} m_G(A \cap \gamma A') = \sum_{\gamma \in \Gamma} m_G(A' \cap \gamma^{-1} A) \\ &\overset{(3.5)}{=} \sum_{\gamma' \in \Gamma} m_G(A' \cap \gamma' A) \overset{(3.4)}{=} m_G(A'), \end{aligned} \tag{3.6}$$

where the second equation follows from (3.5) and the fact that m_G is a left Haar measure. Now if we consider two fundamental domains $F, F' \subseteq G$ we can use equation (3.6) to show that $m_G(F) = m_G(F')$.

(ii) If F is a fundamental domain then $F' := Fg$ is also a fundamental domain for any $g \in G$. By Remark 3.20 we have $m_G(Fg) = \mod(g) m_G(F)$. Then using our result form (i) yields

$$m_G(F) = m_G(F') = m_G(Fg) = \mod(g) m_G(F)$$

for any $g \in G$. On the one hand $m_G(F)$ is positive because Γ is discrete and on the other hand $m_G(F)$ is finite by assumption. Thus $\mod(G) = \{1\}$ follows.

(iii) Since $m_G(F)$ is finite the measure

$$m_X(B) = m_G(\pi^{-1}(B) \cap F) \leq m_G(F)$$

is also finite for any measurable set $B \subseteq X$.

(iv) Since $\pi(F) = \pi(F') = X$ for any two fundamental domains F and F' it follows that $B \cap \pi(F) = B \cap \pi(F')$ for any measurable set $B \subseteq X$. Let $A = \pi^{-1}(B) \cap F$ and $A' = \pi^{-1}(B) \cap F'$, then by applying the Haar measure to these sets we get

$$m_G(\pi^{-1}(B) \cap F) = m_G(A) \overset{(3.6)}{=} m_G(A') = m_G(\pi^{-1}(B) \cap F'), \tag{3.7}$$

which shows the independence of m_X of the fundamental regions. If we define $D := \pi^{-1}(B) \cap F$ then

$$Dg = \pi^{-1}(Bg) \cap F' \subseteq F' := Fg$$

and

$$m_G(D) = m_G(Dg)$$

by the unimodularity of G. Then the equation

$$\begin{aligned} m_X(R_g^{-1}(B)) = m_X(Bg) &\overset{Def.}{=} m_G(\pi^{-1}(Bg) \cap F') = m_G(Dg) = m_G(D) \\ &= m_G(\pi^{-1}(B) \cap F) \overset{Def.}{=} m_X(B) \end{aligned}$$

proves the claim that m_X is right translation invariant, that is

$$m_X(B) = m_X(Bg)$$

for all measurable sets $B \subseteq X$ and for all $g \in G$.

$\square$

Remark 3.24. (i) $m_X(X) = m_G(F) < \infty$,

(ii) $m_X(O) > 0$ for any open set $O \subseteq X$.
To see this recall that $\pi^{-1}(O)$ is open in G, so that $m_G(\pi^{-1}(O)) > 0$. Now we can write

$$\pi^{-1}(O) = \pi^{-1}(O) \cap G = \cup_{\gamma \in \Gamma}(\pi^{-1}(O) \cap F\gamma),$$

hence there exists a $\gamma \in \Gamma$ such that

$$m_G(\pi^{-1}(O) \cap F\gamma) > 0.$$

But $F' := F\gamma$ is also a fundamental region, and by the independence of m_X of the fundamental regions (equation (3.7))

$$m_X(O) = m_G(\pi^{-1}(O) \cap F) = m_G(\pi^{-1}(O) \cap F') > 0.$$

(iii) Since m_G is left translation invariant, so is m_X.

By the points above and the right translation invariance shown in the last proof the measure m_X is called a Haar measure, but X may not be a group.

Now let us define a metric on X.

Definition 3.25. The metric on $X = \Gamma \setminus G$ is defined by

$$\begin{aligned} d_X(\Gamma g_1, \Gamma g_2) &:= \inf_{\gamma_1, \gamma_2 \in \Gamma} d_G(\gamma_1 g_1, \gamma_2 g_2) \overset{Corollary\ 2.34}{=} \inf_{\gamma_1, \gamma_2 \in \Gamma} d_G(g_1, (\gamma_1)^{-1}\gamma_2 g_2) \\ &\overset{\gamma := (\gamma_1)^{-1}\gamma_2}{=} \inf_{\gamma \in \Gamma} d_G(g_1, \gamma g_2), \end{aligned}$$

for $g_1, g_2 \in G$.

Remark 3.26. (i) On X the map

$$\begin{aligned} R_g : X &\to X \\ \Gamma h &\mapsto R_g(\Gamma h) = \Gamma h g^{-1} \end{aligned}$$

is well-defined.

(ii) Let

$$\begin{aligned} \pi : G &\to X \\ g &\mapsto \Gamma g \end{aligned}$$

be the quotient map. Then for $g_1, g_2 \in G$ we have

$$d_X(\pi(g_1), \pi(g_2)) = d_X(\Gamma g_1, \Gamma g_2) \overset{Definition\ 3.25}{\leq} d_G(g_1, g_2).$$

3.3 Ergodicity of the geodesic flow

Since $PSL_2(\mathbb{R})$ is a closed linear group we can apply Theorem 3.23 to $PSL_2(\mathbb{R})$ and define $X := \Gamma \setminus PSL_2(\mathbb{R})$. Point (iv) in Theorem 3.23 has shown us that for any $g \in PSL_2(\mathbb{R})$ the map $R_g : X \to X$ is a measure preserving transformation with respect to the Haar measure m_X. Therefore the time-t-map of the geodesic flow $R_{a_t} : X \to X$, $a_t \in PSL_2(\mathbb{R})$, is also a measure preserving transformation with respect to m_X. The last theorem in this section will show that for $t \neq 0$ the geodesic flow R_{a_t} is ergodic with respect to m_X. But first we need to remember some useful theory.

Definition 3.27. Let $(X, \mathcal{B}_X, \mu)$ be a measure space and let $T : X \to X$ be a measure preserving transformation, that is $\mu(A) = \mu(T^{-1}(A))$ for any $A \in \mathcal{B}_X$. T is called *ergodic* with respect to μ if $A = T^{-1}A$ implies either $\mu(A) = 0$ or $\mu(A^c) = 0$, for any $A \in \mathcal{B}_X$.

Proposition 3.28. Let T be a measure preserving transformation on the measure space $(X, \mathcal{B}_X, \mu)$. To say that T is ergodic is equivalent to the following statement: Let $f : X \to \mathbb{R}$ be a measurable function such that $f \circ T = f$ μ-almost everywhere, that is f is T-invariant μ-almost everywhere. Then f is constant μ-almost everywhere.

Proof. Assume first that f is T-invariant μ-almost everywhere but f is not constant $\mu-$almost everywhere. Then there exists some $a \in \mathbb{R}$ such that the sets $A := f^{-1}((-\infty, a])$ and $A^c := f^{-1}((a, \infty))$ have positive measure. Thus we get

$$\begin{aligned} A = f^{-1}((-\infty, a]) &\overset{T-invariance}{=} (f \circ T)^{-1}((-\infty, a]) \\ &= T^{-1} \circ f^{-1}((-\infty, a]) = T^{-1}(A) \end{aligned}$$

and similarly

$$\begin{aligned} A^c = f^{-1}((a, \infty)) &\overset{T-invariance}{=} (f \circ T)^{-1}((a, \infty)) \\ &= T^{-1} \circ f^{-1}((a, \infty)) = T^{-1}(A^c). \end{aligned}$$

Therefore T is a measure preserving transformation with $A = T^{-1}(A)$, $A^c = T^{-1}(A^c)$ and $\mu(A) > 0, \mu(A^c) > 0$, which implies that T is not ergodic.
Now suppose $A \in \mathcal{B}_X$ such that $\mu(A) > 0$ and $A = T^{-1}(A)$. Define $f := \mathcal{I}_A$ to be the indicator function of A. Then

$$\mu(f \circ T^{-1}(A)) \overset{A=T^{-1}(A)}{=} \mu(f(A))$$

shows T-invariance of f μ-almost everywhere and thus f is constant μ-almost everywhere. Since we assumed $\mu(A) > 0$, it follows that $f = 1$ μ-almost everywhere. Then $f^c := \mathcal{I}_{A^c} = 0$ μ-almost everywhere and

$$\mu(A^c) = \int_X f^c d\mu = 0,$$

which implies ergodicity of T. $\square$

Theorem 3.29. (Lusin's Theorem)
Let $(X, \mathcal{B}_X, \mu)$ be a measure space and let $f : X \to \mathbb{R}$ be a measurable function that is finite μ-almost everywhere. Then for any $\epsilon > 0$ there exists a compact set $K \subseteq X$ such that $f|_K : K \to \mathbb{R}$ is continuous and $\mu(X \setminus K) < \epsilon$.

A proof can be found in [14] or in any standard measure theory book.

Definition 3.30. Let $(f_n)_{n\geq 1}$ be a sequence in $\mathcal{L}^p_\mu = \{f \in \mathcal{L}^0_\mu : (\int |f|^p d\mu)^{\frac{1}{p}} < \infty\}$, $p \in [0, \infty)$. $(f_n)_{n\geq 1}$ *converges in* $\mathcal{L}^p_\mu$ to $f \in \mathcal{L}^p_\mu$ if $\lim_{n\to\infty} \|f_n - f\|_p = 0$.

Theorem 3.31. (Birkhoff's Pointwise Ergodic Theorem)
Let $(X, \mathcal{B}_X, \mu)$ be a measure space and let $T : X \to X$ be a measure preserving transformation. Assume $f \in \mathcal{L}^1_\mu$. Then for any $x \in X$ almost everywhere $\frac{1}{n}\sum_{j=0}^{n-1} f(T^j x)$ converges to $f^*(x)$, where $f^* \in \mathcal{L}^1_\mu$ is a T-invariant function and

$$\int f d\mu = \int f^* d\mu. \tag{3.8}$$

Additionally $f^*(x) = \int f d\mu$ almost everywhere, if T is ergodic.

The proof of Proposition 3.31 can be found in [5] (Theorem 2.30).

Proposition 3.32. Let m_G be a left Haar measure on G, where G is a metrizable, σ-locally compact group. Then the sets

$$\{g \in G : m_G(gB_1 \cap B_2) > 0\}$$

and

$$\{g \in G : m_G(B_1 g \cap B_2) > 0\}$$

are non-empty and open, if $B_1, B_2 \in \mathcal{B}_G$ such that $m_G(B_1)m_G(B_2) > 0$.
Additionally, if $B \in \mathcal{B}_G$ then

$$m_G(B) > 0 \iff m_G(B^{-1}) > 0. \tag{3.9}$$

Proof. We know that

$$m_G(gB_1 \cap B_2) = \int \mathcal{I}_{gB_1}(h)\mathcal{I}_{B_2}(h) dm_G(h). \tag{3.10}$$

So if $h \in gB_1$, then there exists a $\tilde{h} \in B_1$ such that $h = g\tilde{h}$. Therefore $g = h\tilde{h}^{-1}$ and $g \in hB_1^{-1}$. Thus we can write

$$\mathcal{I}_{hB_1^{-1}}(g) = \mathcal{I}_{gB_1}(h). \tag{3.11}$$

Then

$$\begin{aligned}
\int m_G(gB_1 \cap B_2)dm_G(g) &\overset{(3.10)}{=} \int (\int \mathcal{I}_{gB_1}(h)\mathcal{I}_{B_2}(h)dm_G(h))dm_G(g)\\
&\overset{(3.11)}{=} \int (\int \mathcal{I}_{hB_1^{-1}}(g)\mathcal{I}_{B_2}(h)dm_G(h))dm_G(g)\\
&\overset{Fubini}{=} \int \mathcal{I}_{B_2}(h)(\int \mathcal{I}_{hB_1^{-1}}(g)dm_G(g))dm_G(h)\\
&= \int \mathcal{I}_{B_2}(h)m_G(hB_1^{-1})dm_G(h)\\
&\overset{(*)}{=} m_G(B_1^{-1})\int \mathcal{I}_{B_2}(h)dm_G(h) = m_G(B_1^{-1})m_G(B_2),
\end{aligned} \tag{3.12}$$

where equation $(*)$ follows because m_G is a left Haar measure. Notice that $\mathcal{I}_{hB_1^{-1}}(g), \mathcal{I}_{B_2}(h)$ are not negative but they might not be integrable. In that case we can exchange B_1 and B_2 with sequences of subsets which have compact closures in order to use Fubini's theorem. Now set $G = B_2$, then

$$\begin{aligned}
m_G(B_1^{-1})m_G(G) &\overset{(3.12)}{=} \int m_G(gB_1 \cap G)dm_G(g)\\
&= m_G(gB_1)\int dm_G(g) = m_G(B_1)m_G(G),
\end{aligned}$$

and so $m_G(B_1) = m_G(B_1^{-1})$, which implies (3.9). It follows that

$$\int m_G(gB_1 \cap B_2)dm_G(g) = m_G(B_1^{-1})m_G(B_2) = m_G(B_1)m_G(B_2) > 0,$$

which implies that $O := \{g \in G : m_G(gB_1 \cap B_2) > 0\}$ is not empty. To show that O is also open we write $B_1 = \cup_{n=1}^{\infty} A_n$ as a countable union of open sets with compact closures. We can do this since we assumed G to be σ-compact. Then we choose $g, g_1 \in O$ such that $m_G(gB_1 \cap B_2) > 0$. Therefore, there must exist some A_n such that $\epsilon := m_G(gA_n \cap B_2) > 0$. Now we want to show that the difference between $m_G(gA_n \cap B_2)$ and $m_G(g_1A_n \cap B_2)$ is smaller than ϵ. For this we write

$$m_G(g_1A_n \cap B_2) = \int \mathcal{I}_{g_1A_n}(h)\mathcal{I}_{B_2}(h)dm_G(h) = \int \mathcal{I}_{A_n}(g_1^{-1}h)\mathcal{I}_{B_2}(h)dm_G(h)$$

and using $f := \mathcal{I}_{A_n}$ we can show that for g, g_1 sufficiently close to each other the second term of

$$|m_G(g_1A_n \cap B_2) - m_G(gA_n \cap B_2)| \leq |\int (f(g_1^{-1}h) - f(g^{-1}h))\mathcal{I}_{B_2}(h)dm_G(h)|$$

is smaller than ϵ. See Lemma 8.7 in [5] for more details on how to prove this.
To show that $O' := \{g \in G : m_G(B_1 g \cap B_2) > 0\}$ is non-empty and open remember that $m_G(B_i) = m_G(B_i^{-1})$ for $i = 1, 2$ and so we get

$$0 < m_G(B_1)m_G(B_2) = m_G(B_1^{-1})m_G(B_2^{-1}),$$

which implies

$$m_G(B_1^{-1}) > 0, m_G(B_2^{-1}) > 0. \tag{3.13}$$

Notice that we can write

$$\{g \in G : m_G(B_1 g \cup B_2) > 0\} = \{h \in G : m_G(hB_1^{-1} \cap B_2^{-1}) > 0\}^{-1}.$$

Then since

$$\int m_G(hB_1^{-1} \cap B_2^{-1}) dm_G(h) = m_G(B_1^{-1})m_G(B_2^{-1}) \overset{(3.13)}{>} 0,$$

it follows that $\{h \in G : m_G(hB_1^{-1} \cap B_2^{-1}) > 0\}$ is not empty and consequently $\{h \in G : m_G(hB_1^{-1} \cap B_2^{-1}) > 0\}^{-1}$ is not empty. To show that O' is open we can repeat the part above, where we showed that O is open. $\square$

Theorem 3.33. (Heine-Cantor)
Let X be a compact metric space and Y be a metric space. Then a continuous function $f : X \to Y$ is uniformly continuous.

The proof of Theorem 3.33 can be found for example in [13] (Theorem 4.19).

Theorem 3.34. Let Γ be a lattice in $PSL_2(\mathbb{R})$ and $X = \Gamma \setminus PSL_2(\mathbb{R})$. Then $R_{a_t} : X \to X$ is ergodic with respect to m_X for any $t \neq 0$.

Proof. Let $f : X \to \mathbb{R}$ be a measurable function such that $f \circ R_{a_t} = f$ m_X-almost everywhere for $t \neq 0$. We want to show f is constant and use Proposition 3.28 to conclude that R_{a_t} is ergodic with respect to m_X.

We start by normalising the measure m_X such that $m_X(X) = 1$. Then by Lusin's Theorem for any $\epsilon > 0$ we can find a compact set K in X with measure $m_X(K) > 1 - \epsilon$ and continuous function $f|_K : K \to \mathbb{R}$.
Claim:

$$m_X(B) > 1 - 2\epsilon \tag{3.14}$$

for the set

$$B := \{x \in X : \lim_{n\to\infty} \frac{1}{n} \sum_{l=0}^{n-1} \mathcal{I}_K(R_{a_t}^l x) > \frac{1}{2}\}, \tag{3.15}$$

containing points that are in K for more than $\frac{1}{2}$ of their future time.
Proof of the claim: Notice that

$$g^*(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{l=0}^{n-1} \mathcal{I}_K(R_{a_t}^l x)$$

exists almost everywhere (by Theorem 3.31) and is in the interval $[0,1]$ by definition. Now by using equation (3.8) we get

$$\int g^* dm_X = \int \mathcal{I}_K dm_X = m_X(K).$$

Thus

$$\begin{aligned} 1-\epsilon < m_X(K) &= \int_X g^* dm_X = \int_B g^* dm_X + \int_{X\setminus B} g^* dm_X \\ &\overset{(1)}{\leq} \int_B dm_X + \frac{1}{2}\int_{X\setminus B} dm_X = m_X(B) + \frac{1}{2} m_X(X\setminus B) \\ &\overset{m_X(X)=1}{=} m_X(B) + \frac{1}{2}(1-m_X(B)), \end{aligned}$$

where inequality (1) follows since $g^* \in (\frac{1}{2}, 1]$ on B and $g^* \in [0, \frac{1}{2})$ in $X \setminus B$. Therefore we obtain $1 - 2\epsilon < m_X(B)$.

Now let the points

$$x, y := R_{u^-(s)} x \in B$$

be connected by a stable manifold (see Remark 3.10 and Remark 3.17), for $u^-(s) := \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$, $s \in \mathbb{R}$. Since we assumed f to be R_{a_t}-invariant we get

$$f(x) = f(R_{a_t}^l(x)), f(y) = f(R_{a_t}^l(y)) \tag{3.16}$$

for all $l \geq 1$. Moreover

$$\begin{aligned} d_X(R_{a_t}^l(x), R_{a_t}^l(y)) &= d_X(R_{a_t}^l(x), R_{a_t}^l(R_{u^-(s)}x)) = d_X(xa_t^{-l}, R_{a_t}^l(xu^-(-s))) \\ &= d_X(xa_t^{-l}, xu^-(-s)a_t^{-l}) \overset{(1)}{\leq} d_{PSL_2(\mathbb{R})}(\mathbb{I}_2, a_t^l u^-(-s) a_t^{-l}) \\ &= d_{PSL_2(\mathbb{R})}(\mathbb{I}_2, \begin{pmatrix} 1 & -se^{-lt} \\ 0 & 1 \end{pmatrix}) \overset{l\to\infty}{\to} 0, \end{aligned} \tag{3.17}$$

i.e. the distance between $R_{a_t}^l(x)$ and $R_{a_t}^l(y)$ goes to 0 as l goes to infinity. Note that (1) follows by Remark 3.24 (iii) and Remark 3.26 (ii).
The points x, y are in K for more than $\frac{1}{2}$ of their future time, since we assumed

$x, y \in B$. Thus there exists a sequence of common return times to K, $(l_n)_{n\geq 0}$ going to infinity as $n \to \infty$, such that

$$R_{a_t}^{l_n}(x), R_{a_t}^{l_n}(y) \in K.$$

We have seen that $f|_K$ is continuous and using the Theorem of Heine-Cantor shows that $f|_K$ is uniformly continuous. This means that for any $\epsilon > 0$ there exists a $\delta > 0$ such that for all $R_{a_t}^{l_n}(x), R_{a_t}^{l_n}(y) \in K$ with $d_X(R_{a_t}^{l_n}(x), R_{a_t}^{l_n}(y)) < \delta$ we get

$$d_{\mathbb{R}}(f(R_{a_t}^{l_n}(x)), f(R_{a_t}^{l_n}(y))) \overset{(3.16)}{=} d_{\mathbb{R}}(f(x), f(y)) < \epsilon. \tag{3.18}$$

Because of (3.17) the distance between $R_{a_t}^{l_n}(x)$ and $R_{a_t}^{l_n}(y)$ also goes to 0 as $n \to \infty$, which by (3.18) implies $f(x) = f(y)$.

We can go through the same procedure as in the beginning of this proof for any other $0 < \epsilon_1 < \epsilon$. That is, for ϵ_1 we can find a compact set $K_1 \subseteq X$ such that $f|_{K_1}$ is continuous and $m_X(K_1) > 1 - \epsilon_1$ by Lusin's Theorem. Since $f|_K$ and $f|_{K_1}$ are continuous, so is $f|_{K\cup K_1}$ and we can assume $K \subseteq K_1$. Again we define the set

$$B_1 := \{x \in X : \lim_{n\to\infty} \sum_{l=0}^{n-1} \mathcal{I}_{K_1}(R_{a_t}^l x) > \frac{1}{2}\}$$

satisfying $B \subseteq B_1$.
Because ϵ was arbitrary we can continue this way (by letting $\epsilon \to 0$) until we find a compact set $X' \subseteq X$ such that $m_X(X') = 1$ and $f|_{X'}$ is continuous. It then follows by the above that since

$$B' := \{x \in X : \lim_{n\to\infty} \sum_{l=0}^{n-1} \mathcal{I}_{K'}(R_{a_t}^l x) > \frac{1}{2}\} \subset X$$

and

$$m_X(B') \overset{(3.14)}{>} 1 - 2\epsilon$$

(with $\epsilon \to 0$) we must have

$$m_X(B') = 1 = m_X(X) = m_X(X'). \tag{3.19}$$

Thus it follows analogously to our discussion above that $f(x) = f(y)$ for $x, y := R_{u^-s}(x) \in B'$ which corresponds to $x, y \in X'$ by (3.19).

We can repeat what we have done so far for $R_{a_t}^{-1}$. This means that for $\epsilon > 0$ we can find a compact set $\tilde{K} \subseteq X$ such that the set

$$\tilde{B} := \{x \in X : \lim_{n\to\infty} \frac{1}{n} \sum_{l=0}^{n-1} \mathcal{I}_{\tilde{K}}(R_{a_t}^{-l} x) > \frac{1}{2}\}$$

has measure $m_X(\tilde{B}) > 1 - 2\epsilon$.
Then for

$$x, y := R_{u^+(s)}x \in \tilde{B},$$

connected by an unstable manifold (see Remark 3.10), where $u^+(s) := \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}$ and $s \in \mathbb{R}$, the inequality (3.17) becomes

$$\begin{aligned} d_X(R_{a_t}^{-l}(x), R_{a_t}^{-l}(y)) &= d_X(xa_t^l, xu^+(-s)a_t^l) \leq d_{PSL_2(\mathbb{R})}(\mathbb{I}_2, a_t^{-l}u^+(-s)a_t^l) \\ &= d_{PSL_2(\mathbb{R})}(\mathbb{I}_2, \begin{pmatrix} 1 & 0 \\ -se^{-lt} & 1 \end{pmatrix}) \overset{l\to\infty}{\to} 0. \end{aligned}$$

In contrast to before, x, y are in $\tilde{K}$ for more than $\frac{1}{2}$ of their past time. But the rest follows analogously. Therefore we eventually find a set X'' with full measure such that $f(x) = f(y)$ for $x, y = R_{u^+(s)}(x) \in X''$.

Now on the set $X_1 := X' \cap X''$ we get $f(x) = f(y)$ for $x, y := R_{u^-(s)}(x)$ as well as $x, y = R_{u^+(s)}(x)$. That is, f is constant on points which are connected by a stable or unstable manifold.
Remember that by Remark 2.51 U^+, U^- generate $SL_2(\mathbb{R})$. Hence it can be shown that any element g of $SL_2(\mathbb{R})$ can be written as

$$g = u^+(s_4)u^-(s_3)u^+(s_2)u^-(s_1),$$

for $s_1, s_2, s_3, s_4 \in \mathbb{R}$. To understand g better let us consider what $R_g(x)$ does. Assume $R_g(x) = y$. $R_g(x)$ sends x first along the stable manifold containing x to a point y_1. It will need time s_1 to get to y_1. Afterwards y_1 is sent during time s_2 to the point y_2 along the unstable manifold containing y_1 and y_2. Then repeat the procedure until you reach the point y. The idea is shown in Figure 3.4, where we represented the stable manifolds as green lines and the unstable manifolds as red lines.

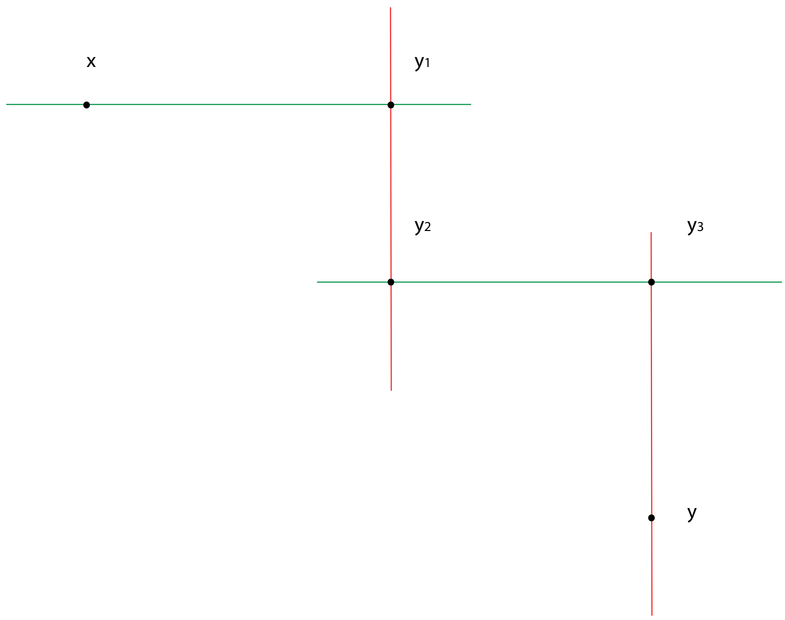

Figure 3.4: Points x, y_1, y_2, y_3 and y on the stable/unstable manifolds.

Now we claim:

$$m_X(X_g) = 1$$

for

$$X_g := X_1 \cap R^{-1}_{u^-(s_1)}(X_1) \cap R^{-1}_{u^+(s_2)u^-(s_1)}(X_1) \cap R^{-1}_{u^-(s_3)u^+(s_2)u^-(s_1)}(X_1) \cap R_g^{-1}(X_1)$$

and

$$f(x) = f(R_g(x)) \tag{3.20}$$

for all $x \in X_g$.
Proof of the claim:

(i) Let $x \in X_1$, then

$$y_1 := R_{u^-(s_1)}(x) \in X_1 \iff x \in R^{-1}_{u^-(s_1)}(X_1).$$

Thus for $x \in X_1 \cap R^{-1}_{u^-(s_1)}(X_1)$ it follows that

$$f(x) = f(R_{u^-(s_1)}(x))$$

by the argument above. Since m_X is invariant under $R_{u^-(s_1)}$, $R^{-1}_{u^-(s_1)}(X_1)$ has full measure.

(ii) Now define $y_2 := R_{u^+(s_2)}(y_1) = R_{u^+(s_2)}R_{u^-(s_1)}(x)$. Then

$$y_2 \in X_1 \iff x \in (R_{u^+(s_2)}R_{u^-(s_1)})^{-1}(X_1) =: R^{-1}_{u^+(s_2)u^-(s_1)}(X_1).$$

Thus for $x \in X_1 \cap R^{-1}_{u^-(s_1)}(X_1) \cap R^{-1}_{u^+(s_2)u^-(s_1)}(X_1)$ we get

$$f(x) = f(R_{u^+(s_2)u^-(s_1)}(x)),$$

again by the above argument. Since m_X is invariant under $R_{u^+(s_2)}$ and $R^{-1}_{u^-(s_1)}(X_1)$ has full measure, we know that $R^{-1}_{u^+(s_2)u^-(s_1)}(X_1)$ has full measure. By continuing this way the claim follows.

Now let us assume $f : X \to \mathbb{R}$ is not constant almost everywhere with respect to m_X. Then we can find disjoint intervals $I_1, I_2 \subseteq \mathbb{R}$, such that $f(\Gamma h)$ is either in I_1 or in I_2 for $h \in PSL_2(\mathbb{R})$. Then since f is not constant almost everywhere the sets

$$\begin{aligned} C_1 &:= \{h \in PSL_2(\mathbb{R}) : f(\Gamma h) \in I_1\}, \\ C_2 &:= \{h \in PSL_2(\mathbb{R}) : f(\Gamma h) \in I_2\} \end{aligned}$$

have neither zero measure nor full measure with respect to $m_{PSL_2(\mathbb{R})}$, so

$$m_{PSL_2(\mathbb{R})}(C_1) m_{PSL_2(\mathbb{R})}(C_2) > 0.$$

Proposition 3.32 then implies that there exist $g \in PSL_2(\mathbb{R})$ such that

$$m_{PSL_2(\mathbb{R})}(C_1 \cap C_2 g) > 0.$$

Now consider the set

$$D_g := \{h \in PSL_2(\mathbb{R}) : \Gamma h \in X_g\}.$$

Since $m_X(X_g) = 1$ it follows that $m_X(X_g^c) = 0$ and therefore $m_{PSL_2(\mathbb{R})}(D_g^c) = 0$. This implies that there is some $h \in PSL_2(\mathbb{R})$ with $h \in C_1 \cap C_2 g \cap D_g$. But $h \in D_g$ implies

$$\Gamma h \in X_g \overset{(3.20)}{\Longrightarrow} f(\Gamma h) = f(\Gamma h g^{-1}), \tag{3.21}$$

whereas

$$f(\Gamma h) \in I_1 \tag{3.22}$$

for $h \in C_1$ and

$$f(\Gamma h g^{-1}) \in I_2 \tag{3.23}$$

for $hg^{-1} \in C_2 \iff h \in C_2 g$. Because we assumed the intervals I_1, I_2 to be disjoint, (3.22) and (3.23) contradict (3.21). Consequently, f is constant almost everywhere with respect to m_X, which is what we needed to show to prove the ergodicity of the geodesic flow. □

Remark 3.35. The last proof uses the so-called Hopf Argument:
Let us use the notation from Theorem 3.31.

(i) By Theorem 3.31 for any $x \in X$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j(x) = f^*(x)$$

almost everywhere.

(ii) If we choose an element y on the stable manifold

$$W^s(x) := \{y \in X : \lim_{n\to\infty} |T^n(x) - T^n(y)| = 0\}$$

of x, then the distance between $f \circ T^j(x)$ and $f \circ T^j(y)$ goes to zero for $j \to \infty$. Thus,

$$f^*(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^j(y) = f^*(y),$$

which implies that f is constant on stable manifolds.

By the T-invariance of f, we can write $f^*(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f \circ T^{-j}(x)$ and conclude that f is also constant on unstable manifolds.

Remark 3.36. (i) Theorem 3.34 in particular implies that the geodesic flow on $PSL_2(\mathbb{Z}) \setminus PSL_2(\mathbb{R})$ is ergodic.

(ii) Using Remark 3.16 and Remark 3.17 we can rephrase Theorem 3.34: The geodesic flow

$$g_t : T^1(\Gamma \setminus \mathbb{H}) \to T^1(\Gamma \setminus \mathbb{H})$$

is ergodic with respect to the Liouville measure (see Theorem 17.4 in [8]).

Bibliography

[1] M. Artin. *Algebra, second edition.* Pearson, 2010.

[2] Alan F. Beardon. *The Geometry of Discrete Groups.* Springer, New York, 1983.

[3] H. Bruin. *Lecture Notes on Ergodic Theory.* University of Vienna, 2014.

[4] M. do Carmo. *Riemannian Geometry.* Birkhäuser, Boston, 1992.

[5] M. Einsiedler and T. Ward. *Ergodic Theory with a view towards Number Theory.* Springer, London, 2011.

[6] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems.* Cambridge University Press, 1995.

[7] S. Katok. *Fuchsian groups.* The University of Chicago Press, Chicago, 1992.

[8] S. Katok. *Fuchsian groups, geodesic flows on surfaces of constant negative curvature and symbolic coding of geodesics.* Clay Mathematics Proceedings, Volume 8, New York, 2008.

[9] M. Kunzinger. *Lecture Notes in General Topology.* University of Vienna, 2016.

[10] J. M. Lee. *Introduction to Smooth Manifolds, Second Edition.* Springer, New York, 2013.

[11] J. R. Munkres. *Topology.* Prentice Hall, New Jersey, 2000.

[12] J. G. Ratcliffe. *Foundation of Hyperbolic Manifolds, Third Edition.* Springer Nature, Switzerland, 2019.

[13] W. Rudin. *Analysis.* R. Oldenbourg Verlag, Munich, 1998.

[14] M. Struwe. *Vorlesung Analysis III Mass und Integral.* ETH Zürich, 2013.

[15] M. Keane T. Bedford and C. Series. *Ergodic Theory, symbolic dynamics and hyperbolic spaces.* Oxford University Press, New York, 1991.

[16] L. W. Tu. *An Introduction to Manifolds, Second Edition.* Springer, 2011.

[17] C. Walkden. *Lecture Notes in Hyperbolic Geometry.* University of Manchester, 2019.