UNIVERSITY OF ZURICH
INSTITUTE OF MATHEMATICS

MASTER THESIS

Flows on the Modular Surface and Continued Fractions

AUTHOR:
Enea Fioroni

SUPERVISOR:
Prof. Dr. Alexander Gorodnik

July 2024

Acknowledgements

I would like to express my deepest gratitude to Professor Dr. Alexander Gorodnik, whose support and guidance have been crucial to the completion of this thesis. His passion for mathematics and dedication to teaching have significantly impacted my academic development, not only for this thesis, but for my entire university journey here in Zürich. His availability enabled me to face and overcome the challenges encountered during the research process. I am immensely thankful for the trust he placed in me and for suggesting such an interesting topic that connected the mathematical fields I liked most during my academic studies.

A big thank you goes to my parents, my brother and sister, my friends, my half-marathon squad and my football teams Ac Canobbio, Ffs Scarsenal, Fc BelloFigo and As Matici who kept my greatest passions alive during my studies.

Grazie

Abstract

In this thesis, we study ergodic properties of geodesic flows on $\mathrm{PSL}(2,\mathbb{Z})\backslash\mathrm{PSL}(2,\mathbb{R})$, that we will see is the unit tangent bundle of the Modular surface $M = \mathrm{PSL}(2,\mathbb{Z})\backslash\mathbb{H}$, and of the Gauss map with respect to the Gauss measure developing a proof using ergodicity of the geodesic flow on T^1M. Moreover, we present some applications of ergodicity of the Gauss Map, as for example the Gauss-Kuzmin Theorem. Before reaching these results we present a preliminary section regarding hyperbolic geometry and discrete spaces. In particular we include an introductive discussion about fundametal domains and Dirichlet regions, arriving to the famous Gauss-Bonnet formula. These first three sections have as main reference chapters 9 and 11 of the book by Einsiedler and Ward "*Ergodic Theory: with a view towards Number Theory*"[7]. We then build a bridge between geodesics in the hyperbolic plane with a special tesselation called *Farey tesselation* and continued fractions, reaching interesting results such as density of geodesics trajectories on the Modular surface and properties of (purely) periodic continued fractions. This chapter borrows Caroline Series's work "The Modular Surface and Continued Fractions" [24]. Lastly, we close this thesis speaking about the old and famous topic *Diphantine Approximation* looking at Dal'Bo's book "*Geodesic and Horocyclic Trajectories*" [5]. Here we will prove results about speed and approximation of irrational numbers by rational numbers in a hyperbolic way, applying settings and results about geodesics and hyperbolic geometry that we developed in the previous chapters.

Contents

1 Introduction **1**

2 Preliminaries: Hyperbolic Geometry **9**
2.1 Hyperbolic Plane, the Action of $\mathrm{PSL}(2,\mathbb{R})$ on $\mathbb{H}$, $\mathrm{T}^1\mathbb{H} \cong \mathrm{PSL}(2,\mathbb{R})$ 9
2.2 Geodesics in $\mathbb{H}$ 13

3 M = $\mathbf{PSL}(2,\mathbb{Z})\backslash\mathbb{H}$, $\mathbf{PSL}(2,\mathbb{Z})\backslash\mathbf{PSL}(2,\mathbb{R})$ and Gauss-Bonnet Formula **17**
3.1 Distance for Closed Linear Groups 17
3.2 Fuchsian Groups, Lattices and Fundamnetal Domains 19
3.3 Hyperbolic Area and Hyperbolic Volume 21
3.4 A $\mathrm{PSL}(2,\mathbb{R})$-Invariant Measure for $\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$ 22
3.5 The main Example of a Fundamental Domain in $\mathbb{H}$ for the lattice $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$ 24
3.6 Dirichlet Regions and Gauss-Bonnet Formula 25

4 Ergodicity of Geodesic Flows on $\mathbf{PSL}(2,\mathbb{Z})\backslash\mathbf{PSL}(2,\mathbb{R})$ and of the Gauss Map **31**
4.1 Geodesic Flows and Horocycle Flows 31
4.2 Ergodicity of Geodesic Flows on $\mathrm{T}^1\mathrm{M}$ 34
4.3 Ergodicity of the Gauss map 39
4.4 Gauss–Kuzmin Theorem and other Applications 51

5 Continued Fractions and Geodesics on $\mathbb{H}$ **57**
5.1 Farey Tesselation, Continued Fractions and Cutting Sequences 57
5.2 Convergents of a number 63
5.3 Dense Trajectories of Geodesics on $\Gamma\backslash\mathbb{H}$ 70
5.4 Periodic Continued Fractions 72

6 Diopanthine Approximation **83**

7 Conclusion **93**

Bibliography **94**

Chapter 1

Introduction

In this thesis we analyze and discuss dynamical properties of geodesics in $\mathbb{H}$ and of geodesic flows on T^1M, the unit tangent bundle of the Modular surface $M = \Gamma\backslash\mathbb{H}$, where $\mathbb{H}$ denotes the hyperbolic plane, while $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$ is a subgroup of $\mathrm{PSL}(2,\mathbb{R})$. We have two main puroposes: firstly, to prove ergodicity of the geodesic flow on T^1M and derive ergodicity of the Gauss Map from it (Chapter 4), and then to construct the continued fraction form of a number from the intersection of geodesics in $\mathbb{H}$ equipped with a particular tesselation (Chapter 5). From here we also prove some applications such as the Gauss-Kuzmin theorem (Chapter 4), state some properties of (purely) periodic continued fractions and of geodesics on $M = \Gamma\backslash\mathbb{H}$ (Chapter 5) and explain some results of Diophantine approximation such as Hurwitz's Theorem from a hyperbolic point of view (Chapter 6).

Philosophically speaking, the most important purpose of this thesis is to show that mathematics is not divided into areas that don't communicate with each other. Indeed, the majority of the results that we prove in this work were originally faced with a more algebraic approach, while here a more geometric and diversified approach gives a tangible proof of how big and intricate mathematics can be, connecting fields that superficially can seem to be not directly linked.

Initially we present an exhaustive first chapter (Chapter 2) that introduces the reader to hyperbolic geometry and the group action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{H}$. We see that $\mathbb{H} = \{x + iy \in \mathbb{C} \mid y > 0\}$ can be equipped with a specific Riemannian metric such that the group action

$$\mathrm{SL}(2,\mathbb{R}) \times \mathbb{H} \to \mathbb{H}$$

$$\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}, z\right) \mapsto \frac{az+b}{cz+d}$$

is isometric with respect to the distance between two points in the hyperbolic plane defined from the norm induced by our Riemannian metric. Furthermore, after having defined the derivative of a curve on $\mathrm{T}_z\mathbb{H} = \{z\} \times \mathbb{C}$ as $\mathrm{D}\phi(t) = (\phi(t), \phi'(t))$, it is central also the role of the action of $\mathrm{PSL}(2,\mathbb{R})$ on $\mathrm{T}\mathbb{H}$ defined as $\mathrm{D}g(z,v) = (g(z), g'(z)v) = \left(\frac{az+b}{cz+d}, \frac{v}{(cz+d)^2}\right)$, that also preserves the hyperbolic Riemannian metric. After proving

some lemmas and introducing the *unit tangent bundle* $\mathrm{T}^1\mathbb{H} = \{(z, v) \in \mathrm{TH} \mid \|v\|_z = 1\}$, we figure out why $\mathrm{T}^1\mathbb{H} \cong \mathrm{PSL}(2, \mathbb{R})$. At the end we exhibit and clarify the concept of geodesic in the hyperbolic plane $\mathbb{H}$. Geodesics are in general curves that minimize the distance between two points in a surface. For example in the usual Cartesian plane $\mathbb{R}^2$, those are represented by straight lines. In the hyperbolic plane, always endowed with the hyperbolic Riemannian metric, it turns out that geodesics are vertical lines and semicircles perpendicular with the real axis, and $\mathrm{SL}(2, \mathbb{R})$ maps thus vertical lines and semicircle to themselves, since it preserves distances. The main reference here is Chapter 9 of Thomas Ward and Manfred Einsiedler's book *"Ergodic Theory: with a view towards Number Theory"*[7].

Then we make a small digression to talk about discrete subspaces in Chapter 3.1. After having defined a distance for a general closed linear group, we introduce the concepts of Fuchsian group, lattice and fundamental domain. Briefly, a Fuchsian group is simply a discrete subgroup of $\mathrm{PSL}(2, \mathbb{R})$, while considering a group action on a space, like the one that we introduced on $\mathbb{H}$, and a Fuchsian group Γ, a fundamental domain F for the Modular surface $\Gamma\backslash\mathbb{H}$ is a connected subset of $\mathbb{H}$ such that orbits ΓF cover entirely $\mathbb{H}$ with the orbits intersecting themselves only on measure zero sets, i.e. only the boundaries are shared. We can see a finite fundamental domain for the action of $\mathrm{PSL}(2, \mathbb{R})$ on $\mathbb{H}$ in the following Figure 1.1.

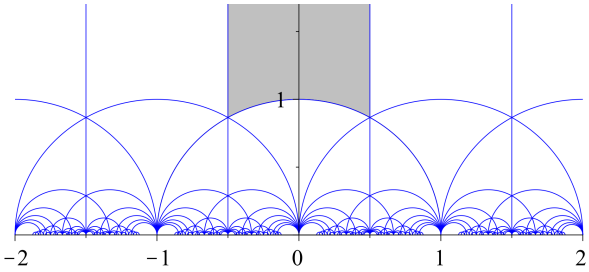

Figure 1.1: Fundamental domain for the Möbius action on $\mathbb{H}$ and $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ [32].

However, using the same definition as before, we can also define a fundamental domain for $\mathrm{T}^1\mathbb{H}$. Clearly we can take the same fundamental domain as before but associating to each point unit tangent vectors. We denote the fundamental domain for the action of $\mathrm{PSL}(2, \mathbb{R})$ on $\mathrm{T}^1\mathbb{H}$ as T^1F, where F denotes the previous fundamental domain. Now since $\mathrm{T}^1\mathbb{H} \cong \mathrm{PSL}(2, \mathbb{R})$, to each element of $(z, v) \in \mathrm{T}^1\mathbb{H}$ corresponds uniquely a $g \in \mathrm{PSL}(2, \mathbb{R})$ such that $Dg(i, i) = (z, v)$, so we can view T^1F as a subset of $\mathrm{PSL}(2, \mathbb{R})$ such that $\forall g \in \mathrm{PSL}(2, \mathbb{R}), card\ (\mathrm{T}^{\mathring{1}}\mathrm{F} \cap \Gamma g) = 1$. So the fundamental domain F for $\Gamma\backslash\mathbb{H}$ is in a sense comparable to the fundamental domain T^1F for $\Gamma\backslash\mathrm{PSL}(2, \mathbb{R})$ defined as $\mathrm{T}^1F = \{g \in \mathrm{PSL}(2, \mathbb{R})$ such that $\mathrm{Dg}(i, i) \in \mathrm{T}^1F\}$. For more details see Proposition 3.15.
Γ is a lattice if there exists a fundamental domain of finite measure. After that, we discuss what an invariant measure for $\Gamma\backslash\mathrm{PSL}(2, \mathbb{R})$ is (even in general for a closed linear group and a lattice).

We finally talk about Dirichlet regions, that are defined as $D(p) = \{z \in \mathbb{H} \mid \mathrm{d}(z,p) < \mathrm{d}(z,\gamma p)$ for all $\gamma \in \Gamma\backslash\{e\}\}$ such that for every $g \in \Gamma \setminus \{e\}$, $p \in \mathbb{H}$ satisfies $g(p) \neq p$. So, after stated but not proved that any Dirichlet region for a lattice in $\mathrm{PSL}(2,\mathbb{R})$ has finitely many sides that are geodesic segments (so is a polygon in $\mathbb{H}$), we finally reach and inductively prove the famous Gauss-Bonnet formula for computing area of hyperbolic polygons (see Fig. 1.2).

Result 1.1. *Let P be a hyperbolic n sided convex polygon in $\mathbb{H}$ with $n \geqslant 3$ vertices in $\overline{\mathbb{H}}$, with angles $\alpha_1, \ldots, \alpha_n$ at the n vertices. Then the hyperbolic area of P is*

$$(n-2)\pi - (\alpha_1 + \cdots + \alpha_n)\,.$$

For a vertex in $\partial\mathbb{H}$ we set the angle to be zero since the circles are tangential there.

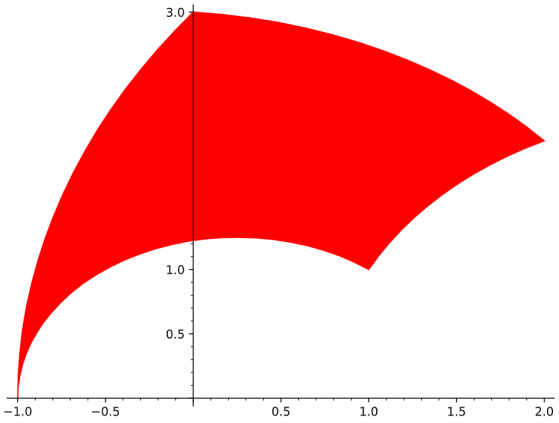

Figure 1.2: Area of an hyperbolic polygon [19].

Here we also refer to Thomas Ward and Manfred Einsiedler's book, partially in chapter 9 again, and in chapter 11 for Dirichlet domains and the Gauss-Bonnet formula [7].

In Chapter 4 we arrive at the heart of this thesis, where we describe the connection between *geodesic flows* on the unit tangent bundle of the Modular surface and the *Gauss map*. We shall demonstrate that the ergodicity of the Gauss map can be inferred from that of the geodesic flow on T^1M. There are several definitions of what it means for a map to be *ergodic* (see Proposition 4.13). The main purpose to prove this is that if a map possesses this property, then we can apply the Birkoff-Ergodic Theorem (Theorem 4.16) to obtain some interesting and often surprising results about convergence at infinity of sums, products, ...But what is a geodesic flow? We call *geodesic flow* the function $\Gamma_t : \mathrm{T}^1\mathbb{H} \to \mathrm{T}^1\mathbb{H}$ that follows the uniquely defined parametrization of a geodesic determined by (z,v) for time t. Since we proved in Chapter 2 that $\mathrm{T}^1\mathbb{H} \cong \mathrm{PSL}(2,\mathbb{R})$, we can also describe it as the right multiplication $R_{a_t}(g) = ga_t^{-1}$ in $\mathrm{PSL}(2,\mathbb{R})$ by the matrix $a_t^{-1} = \begin{bmatrix} \mathrm{e}^{t/2} & 0 \\ 0 & \mathrm{e}^{-t/2} \end{bmatrix}$. The following picture is a nice example of what is happening now (see Fig. 1.3).

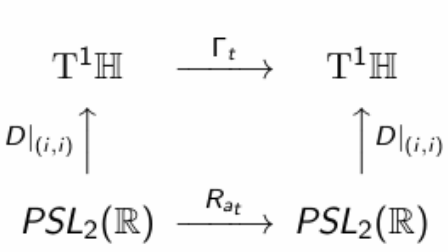

Figure 1.3: From $\mathrm{T}^1\mathbb{H}$ to $\mathrm{PSL}(2,\mathbb{R})$ [27].

After proving ergodicity of the geodesic flow on M, we use this result to prove ergodicity of the Gauss Map with respect to the Gauss measure $\mu(A) = \frac{1}{\log 2}\int_A \frac{1}{1+x}\,\mathrm{d}x$, where A is any Borel measurable set in $\subseteq [0,1]$. The Gauss Map is the map T defined on $Y = [0,1] \setminus \mathbb{Q}$, such that $T(y) = \left\{\frac{1}{y}\right\} = \frac{1}{x} - \lfloor \frac{1}{y} \rfloor$ (see Fig. 1.4).

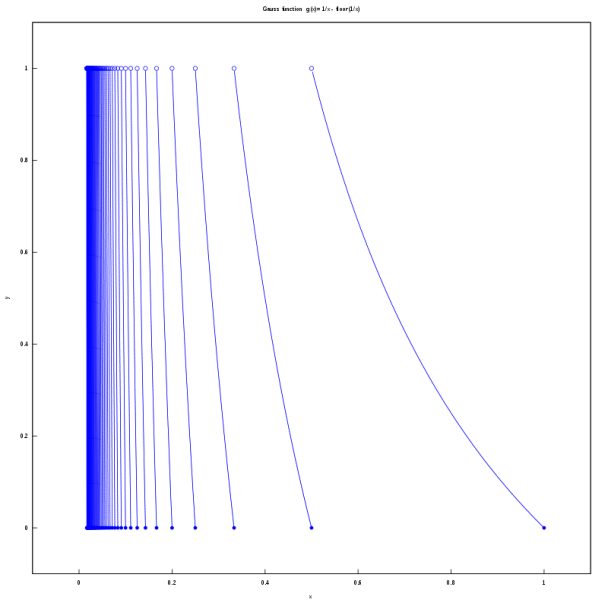

Figure 1.4: Gauss Map T on $Y = (0,1)$ [31].

The proof is intricate, involving new flows and new lemmas and defining a so called *cross section* C (see Definition 4.21).

Finally, in order to apply what we have proved, we provide a proof of the Gauss-Kuzmin Theorem using the Birkoff Ergodic Theorem and ergodicity of the Gauss map. Moreover, we include some interesting results about continued fractions that are also related to the ergodicity of the Gauss map. Gauss-Kutzmin Theorem sates that for any irrational number, which as seen in Chapter 5 has an infinite continued fraction form, it can be found a limit frequency of the presence of any positive natural number in the continued fraction.

Result 1.2 (Gauss-Kuzmin Theorem). *For almost every real number* $x = [a_1, a_2, \ldots] \in (0, 1)$*, i.e. for all* $x \in [0, 1] \setminus \mathbb{Q}$*, the positive natural number* k *appears in the continued fraction with frequency*

$$\frac{1}{\log 2} \log \left(\frac{(k+1)^2}{k(k+2)} \right).$$

In order to accomplish what we said in this short introduction of Chapter 4, we look again to chapter 9 of the book *"Ergodic Theory: with a view towards Number Theory"*[7].

The second main goal that we want to understand is about the continued fraction form of a real number. We see that any real number x can be expressed through a unique continued fraction form (actually two for rational numbers, but closely linked to each other) obtained through an iterative process of representing x as the sum of its integer part and the reciprocal of another number, then writing this other number as the sum of its integer part and another reciprocal, and continuing doing this finitely many times for rational numbers and infinitely many times for irrational numbers. In symbols:

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \frac{1}{a_3 + \cdots}}},$$

where $a_0 \in \mathbb{N}$ and $a_i \in \mathbb{N}_{>0}$ for $i \geq 1$. In particular, Chapter 5 show us that it can be constructed from the intersection of geodesics with a specific tessellation of the hyperbolic plane, called *Farey tesselation* (see Fig. 1.5).

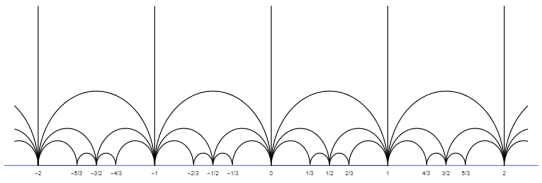

Figure 1.5: Farey tesselation of the hyperbolic plane [24].

We introduce the concept of *cutting sequence*. Basically, for a geodesic ending at $x \in \mathbb{R}$ and crossing the imaginary axis, we construct a binary code made of L and R, that denote, going from the imaginary axis to the real axis, if at the crossing moment of the geodesic with a *Farey side* we create a hyperbolic triangle to the left "L" or to the right "R". The number of consecutive Ls/Rs denotes the natural numbers that compose the continued fraction of x (see Fig. 1.6).

From there we reach interesting results introducing the notion of *convergents of a continued fraction*, which are defined as $\frac{p_n}{q_n} = [a_0; a_1, a_2, \ldots, a_n]$. For example, we discover that $\frac{p_{2n}}{q_{2n}} \leq \frac{p_{2n+2}}{q_{2n+2}} \leq x \leq \frac{p_{2n+1}}{q_{2n+1}} \leq \frac{p_{2n-1}}{q_{2n-1}}$ for all $n \geq 0$ and $\left| x - \frac{p_n}{q_n} \right| \leq \frac{1}{q_n q_{n+1}}$.

Furthermore, we then move to the Modular surface M, and we give a sketch of proof of density of geodeisc trajectories on M.

Finally, we prove the fact that periodic continued fractions are solutions of quadratic equations with integral coefficients and a similar fact for purely periodic continued

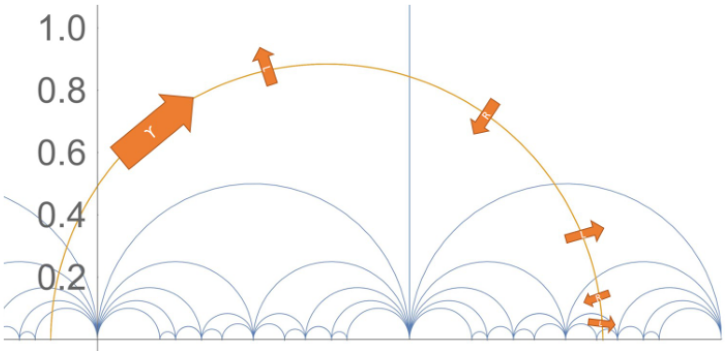

Figure 1.6: Cutting sequence of a geodesic ending up in $x = \frac{81}{50}$ [8].

fractions. These topics are faced in two ways: firstly we provide algebraic proofs like Lagrange and Galois did (following Prof. Ikenaga's work [13][14]), and then we prove them geometrically, through the construction built in this chapter in the hyperbolic plane (see [25]).

Result 1.3. *Let $x \in \mathbb{R} \setminus \mathbb{Q}$. The continued fraction of x is eventually periodic if and only if x is a solution of a quadratic equation of integral coefficients.*

From our construction with the group action of Möbius transformations it is understandable to expect such a result, but for people who face this topic for the first time can be surprising the connection between quadratic equations and continued fractions.

This fifth chapter is entirely based on Caroline Series work *"The Modular Surface and Continued Fraction"* [24], but the main reference for us is a simpler version of [24] published by Loughborough LMS Summer School titled *"Continued Fractions and Hyperbolic Geometry"* [25].

As a last chapter, we go deeply into *Diophantine Approximation* following the third section of chapter 7 of Dal'Bo's book *Geodesic and Horocyclic trajectories* [5]. Diophantine approximation deals with convergence and speed of convergence of approximation of irrational numbers by rational numbers. Here Dal'Bo proposes a geometrical approach to main results about this topic, using what we have proved and discovered in Chapter 5. A famous result that we prove here is the Hurwitz's Theorem.

Result 1.4. *For every irrational number x,we define*

$$\nu(x) = \inf \left\{ c > 0 \mid \exists \quad 1 \leq \left(\frac{p_n}{q_n}\right)_n \in \mathbb{Q}, \left| x - \frac{p_n}{q_n} \right| \leqslant \frac{c}{q_n^2} \text{ and } \lim_{n \to +\infty} |q_n| = +\infty \right\},$$

and one has

$$\nu(x) \leqslant \frac{1}{\sqrt{5}}.$$

Also $\nu(x) = \frac{1}{\sqrt{5}}$ if and only if there exist a, b, c, d in $\mathbb{Z}$ such that $ac - bd = 1$ and $x = \frac{a\mathcal{N}+b}{c\mathcal{N}+d}$, where $\mathcal{N} = \frac{1+\sqrt{5}}{2}$ is the golden ratio .

We try to face and go into detail about the results found in these three references in the most honest and precise way possible. Other references, which are cited in the most transparent way, facilitated our understanding of proofs and concepts within this thesis.

Speaking about history of hyperbolic geometry [4], the first sign of it was born with Euclid's five postulate: indeed the fifth one states that *"If a line segment intersects two straight lines forming two interior angles on the same side that are less than two right angles, then the two lines, if extended indefinitely, meet on that side on which the angles sum to less than two right angles."*, and it was the only one that Euclid did not prove. Over two thousand years later, Gauss, Lobachevsky, Taurinus, Bolyai and others extended the concept of hyperbolic geometry. Later on Poincaré, Klein and many other mathematicians helped to develop hyperbolic geometry further [3].

Regarding the proof of the ergodicity of the geodesic flow on the unit tangent bundle of the Modular surface, the argument was made by the German mathematician Eberhard Hopf in the twentieth century.

Finally, as mentioned in the work of Caroline Series [24], the connection between continued fractions and geodesics in the hyperbolic plane was noted and used by the renowned Austrian mathematician Emil Artin in the twentieth century to establish the dense trajectory of geodesic flows on Riemann surfaces.

Diopanthine Approximationit is named after Diophantus of Alexandria, a Greek mathematician of the third century AD who was famous for his studies about fractions and polygonal numbers and his main work *Arithmetica*, which survived only in part throught the centuries [2].

Chapter 2

Preliminaries: Hyperbolic Geometry

In this preliminary section we provide firstly an understandable view of the hyperbolic plane endowed with its hyperbolic metric. Then we focus on the action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{H}$, discovering $\mathrm{T}^1\mathbb{H} \cong \mathrm{PSL}(2,\mathbb{R})$, that will be a central result in the whole thesis. Finally we introduce the concept of geodesic. The main reference for this section is chapter 9 of Thomas Ward and Manfred Einsiedler's book *"Ergodic Theory: with a view towards Number Theory"*[7].

2.1 Hyperbolic Plane, the Action of $\mathrm{PSL}(2,\mathbb{R})$ on $\mathbb{H}$, $\mathrm{T}^1\mathbb{H} \cong \mathrm{PSL}(2,\mathbb{R})$

We start introducing the hyperbolic plane and its metric and then we focus on the action of $\mathrm{PSL}(2,\mathbb{R})$ firstly on $\mathbb{H}$ and then on $\mathrm{T}^1\mathbb{H}$.

Definition 2.1 (Hyperbolic Plane)**.** The *hyperbolic plane* $\mathbb{H}$ is defined as the upper-half of the complex plane endowed with a Riemaniann metric called hyperbolic Riemannian metric (see Definition 2.3). In symbols,

$$\mathbb{H} = \{x + iy \in \mathbb{C} \mid y > 0\}.$$

Definition 2.2 (Tangent Bundle)**.** $T\mathbb{H} = \mathbb{H} \times \mathbb{C}$ is called *tangent bundel* of the hyperbolic plane. It can be viewed also as the (disjoint) union of $\mathrm{T}_w\mathbb{H} = \{w\} \times \mathbb{C}, \forall w \in \mathbb{H}$.

Definition 2.3 (Hyperbolic Riemannian Metric)**.** The *hyperbolic Riemannian metric* is defined as the conjugate linear bilinear form

$$\langle v, w\rangle_z = \frac{1}{y^2}\Re(\alpha\overline{\beta}) = \frac{1}{y^2}(\alpha \cdot \beta)$$

where $(v \cdot w)$ denotes the usual inner product in $\mathbb{C}$ viewed as $\mathbb{R}^2$, for $z = x + \mathrm{i}y \in \mathbb{H}$ and $v = (z, \alpha), w = (z, \beta) \in \mathrm{T}_z\mathbb{H} = \{z\} \times \mathbb{C}$.

Definition 2.4 (Derivative of a curve)**.** If $\phi : [0,1] \to \mathbb{H}$ is a differentiable curve at $t \in [0,1]$ with $\phi(t) = z$, we define its derivative at t by

$$\mathrm{D}\phi(t) = \big(\phi(t), \phi'(t)\big) \in \mathrm{T}_z\mathbb{H}.$$

Definition 2.5 (Length of a curve)**.** Let $\phi : [0,1] \to \mathbb{H}$ be a continuous piecewise differentiable curve, then its length is defined by

$$\mathrm{L}(\phi) = \int_0^1 \|\mathrm{D}\phi(t)\|_{\phi(t)} \mathrm{d}t,$$

with the induced norm by the hyperbolic Riemannian metric $\|\mathrm{D}\phi(t)\|_{\phi(t)} = \frac{1}{\Im(\phi(t))}|\phi'(t)| = \frac{\sqrt{x'(t)^2+y'(t)^2}}{y(t)}$ for $\phi(t) = x(t) + iy(t)$ and $\phi'(t) = x'(t) + y'(t)i$.

Definition 2.6 (Hyperbolic Distance)**.** We define the *hyperbolic distance* or *hyperbolic metric* as

$$\mathrm{d}\left(z_0, z_1\right) = \inf_{\phi} \mathrm{L}(\phi),$$

taking the infimum over all continuous piecewise differentiable curve going from z_0 to z_1. It is easy to check that d defines indeed a metric.

It could seem strange to pick such a Riemaniann metric, but in the next pages we prove that it gives a metric with respect to which the Möbius action is isometric (Lemma 2.10), so it is particularily useful.

Definition 2.7 (Group Action)**.** Let G be a group and X be a set. A *group action* of G on X is a map $(\cdot) : G \times X \to X$ such that for every $g, h \in G$ and every $x \in X$, it satisfies the following properties:

1. $e \cdot x = x$ for every $x \in X$, where e is the identity element of G,
2. $(gh) \cdot x = g \cdot (h \cdot x)$ for every $g, h \in G$ and $x \in X$.

A group action is *transitive* if for any $x_1, x_2 \in X$ there is a $g \in G$ with $g \cdot x_1 = x_2$, and is *simply transitive* if there is a unique $g \in G$ with $g \cdot x_1 = x_2$.

Definition 2.8 (Möbius Transformations)**.** Consider $\mathrm{SL}(2,\mathbb{R})$ as the group of 2×2 matrices with real entries of determinant one. We define the group action of $\mathrm{SL}(2,\mathbb{R})$ on $\mathbb{H}$ as the map

$$\begin{gathered} \mathrm{SL}(2,\mathbb{R}) \times \mathbb{H} \to \mathbb{H} \\ \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}, z \right) \mapsto \frac{az+b}{cz+d}. \end{gathered} \tag{2.1}$$

This action is called *Möbius transformation or fractional linear transformation*.

The two condition of Definition 2.7 are easy to check. It remains to check that the action ends in $\mathbb{H}$:

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

therefore

$$\Im(g(z)) = \frac{\Im(z)}{|cz+d|^2} \tag{2.2}$$

since $ad - bc = 1$. From the moment that $\Im(z) > 0$, we proved that $\Im(g(z)) > 0$.

Definition 2.9 (Projective Special Linear Group)**.** Since the matrix $(-I_2)$ acts trivially on $\mathbb{H}$, from now on we will work with the action of the *projective special linear group* $\mathrm{PSL}(2,\mathbb{R})$ on $\mathbb{H}$, where

$$\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R})/\{\pm I_2\}$$

We now state an important lemma that will be used a lot in upcoming proofs.

Lemma 2.10. *The action of* $\mathrm{PSL}(2,\mathbb{R})$ *on* $\mathbb{H}$ *defined by (2.1) has the following properties:*

1. $\forall z_1, z_2 \in \mathbb{H}, \forall g \in \mathrm{PSL}(2,\mathbb{R})$,

$$\mathrm{d}\left(g\left(z_1\right), g\left(z_2\right)\right) = \mathrm{d}\left(z_1, z_2\right).$$

   This means that the action is isometric*.*

2. *The action of* $\mathrm{PSL}(2,\mathbb{R})$ *on* $T\mathbb{H}$ *defined by*

$$\mathrm{D}g(z,v) = \big(g(z), g'(z)v\big) = \left(\frac{az+b}{cz+d}, \frac{v}{(cz+d)^2}\right) \tag{2.3}$$

   preserves the Riemannian metric.

3. *The action is transitive on* $\mathbb{H}$*: given any two points* $z_1, z_2 \in \mathbb{H}$*, there is a matrix* $g \in \mathrm{PSL}(2,\mathbb{R})$ *with* $g\left(z_1\right) = z_2$*.*

4. $$\mathrm{Stab}_{\mathrm{PSL}(2,\mathbb{R})}(\mathrm{i}) = \mathrm{PSO}(2)$$

   where,

$$\mathrm{Stab}_{\mathrm{PSL}(2,\mathbb{R})}(\mathrm{i}) = \{g \in \mathrm{PSL}(2,\mathbb{R}) \mid g(\mathrm{i}) = \mathrm{i}\}$$

   and

$$\mathrm{PSO}(2) = \mathrm{SO}(2)/\{\pm I_2\}$$

   with

$$\mathrm{SO}(2) = \left\{ \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \middle| \; \theta \in \mathbb{R} \right\}.$$

Notation 2.11. From now on for $w = (z,v) \in T_z\mathbb{H}$, we simply write v to represent w since the first coordinate is obviously z. In the same way, we write $(\mathrm{D}g)_z v$ to represent the second coordinate of $\mathrm{D}g(w)$ since the first coordinate is simply $g(z)$. So we will abuse the notation writing $v \in \mathrm{T}_z\mathbb{H}$ instead of $w = (z,v) \in \mathrm{T}_z\mathbb{H}$.

Proof. 1. From point 2, we deduce that $\|\mathrm{D}\phi(t)\|_{\phi(t)} = \|\mathrm{D}(g \circ \phi(t))\|_{g(\phi(t))}$. Therefore, length is preserved by the action of $\mathrm{PSL}(2,\mathbb{R})$ on $\mathbb{H}$, and therefore distances are too.

2. For v, w in $\mathrm{T}_z\mathbb{H}$ we have $(\mathrm{D}g)_z v, (\mathrm{D}g)_z w \in \mathrm{T}_{g(z)}\mathbb{H}$ and using (2.2) we show that

$$\begin{aligned}\langle (\mathrm{D}g)_z v, (\mathrm{D}g)_z w \rangle_{g(z)} &= \left(\frac{y}{|cz+d|^2}\right)^{-2} \left(\frac{1}{(cz+d)^2} v \cdot \frac{1}{(cz+d)^2} w\right) \\ &= \frac{1}{y^2}(v \cdot w) = \langle v, w \rangle_z,\end{aligned}$$

by basic properties of inner product in $\mathbb{C}$ and complex numbers, such as that for $x \in \mathbb{C}$, $x\overline{x} = |x|^2$ and for $a, x, y \in \mathbb{C}$, $(x, ay) = \overline{a}(x, y)$.

3. For any $z = x + \mathrm{i}y \in \mathbb{H}$, the matrix $\begin{bmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{bmatrix}$ sends i to $x + \mathrm{i}y$. Thus, if g_1 goes from i to z_1 and if g_2 goes from i to z_2, there exists $h = g_2 \circ g_1^{-1} \in \mathbb{H}$ such that $h(z_1) = z_2$.

4. In the stabilizer we must have $g(\mathrm{i}) = \mathrm{i}$, therefore $|c + \mathrm{i}d| = 1$ by (2.2). Thus there exists a $\theta \in \mathbb{R}$ with $c = \sin\theta, d = \cos\theta$, i.e.

$$g(\mathrm{i}) = \mathrm{i} \Rightarrow \frac{a\mathrm{i} + b}{\sin\theta \mathrm{i} + \cos\theta} = \mathrm{i}. \tag{2.4}$$

Solving the equation, $a = \cos\theta$ and $b = -\sin\theta$. Thus the stabilizer of i is the subgroup $\mathrm{PSO}(2)$ of $\mathrm{PSL}(2, \mathbb{R})$.

□

Corollary 2.12. *From point 4 of Lemma 2.10, we deduce*

$$\mathbb{H} \cong \mathrm{PSL}(2, \mathbb{R}) / \mathrm{PSO}(2)$$

Proof. Since the action is transitive, $\mathrm{PSL}(2, \mathbb{R})(i) = \mathbb{H}$. Moreover $\mathrm{PSL}(2, \mathbb{R})$ is locally and σ-compact since is a subset of $\mathbb{R}^4$. From point 4 of Lemma 2.10, we can conclude using a basic theorem of topological groups theory. □

Definition 2.13 (Unit Tangent Bundle)**.**

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

consists in pairs (z, v) with all unit vectors v attached to all possible $z \in \mathbb{H}$ and is called *unit tangent bundle*.

Lemma 2.14. *A transitive action is* simply transitive *if and only if it has all trivial stabilizers.*

Proof. The left implication is trivial. Now assume that all stabilizers are trivial, but assume that there are $g_1, g_2 \in G$ and $x, y \in X$ such that $g_1 \cdot x = y$ and $g_2 \cdot x = y$. Then $\left(g_2^{-1} g_1\right) \cdot x = x$ and $g_2^{-1} g_1 = e$ because all stabilizers are trivial. Thus $g_1 = g_2$ and so the action is simply transitive. □

Lemma 2.15. *The action of* $\mathrm{PSL}(2, \mathbb{R})$ *on* $\mathrm{T}^1\mathbb{H}$ *described in (2.3) is simply transitive.*

Proof. Since we already know that the action on $\mathbb{H}$ is transitive, we can simplify the whole proof analyzing only the case of vectors $v \in \mathrm{T_i}\mathbb{H}$ with base point i, since we can reach any other point in $\mathbb{H}$ applying any g in $\mathrm{PSL}(2,\mathbb{R})$. From Lemma 2.10, applying a matrix in $\mathrm{PSO}_2(\mathbb{R})$ we stay at i: therefore we have for $g \in \mathrm{PSO}_2(\mathbb{R})$

$$\mathrm{D}g(\mathrm{i}, v) = \left(i, \frac{1}{(\mathrm{i}\sin\theta + \cos\theta)^2} v\right) = (i, (\cos(2\theta) - \mathrm{i}\sin(2\theta))v)$$

Changing θ, $(\mathrm{D}g)_{\mathrm{i}}(v)$ could be any vector of modulus one. Now we have proved transitivity. To prove *simply transitivity*, note that if $(\mathrm{D}g)_{\mathrm{i}}(v) = v$, then

$$\theta \equiv 0 \pmod{2\pi}.$$

So $\theta \in \mathbb{Z}\pi$, and $g = \pm I_2$. This means that all the stabilizers are trivial. We conclude from Lemma 2.14. □

Corollary 2.16. *It follows from Lemma 2.15 that*

$$\mathrm{T}^1\mathbb{H} \cong \mathrm{PSL}(2,\mathbb{R})$$

Proof. This follows directly from the fact that the action (2.3) is simply transitive. The one-to-one correspondence starts with $g = \pm I_2$ and (i, i). Indeed, $\mathrm{D}g(i,i) = (g(i), g'(i)v) = \left(i, \frac{i}{(0\cdot i \pm 1)^2}\right) = (i,i)$, and then the identification is given by $g \mapsto \mathrm{D}g(i,i)$. □

2.2 Geodesics in $\mathbb{H}$

Now we want to better understand what is the minimal distance between two points in $\mathbb{H}$ and discover interesting properties of these minimal distance segments.

Lemma 2.17. *Let $z_0 = y_0\mathrm{i}$ and $z_1 = y_1\mathrm{i}$ with $0 < y_0 < y_1$. Then*

$$\mathrm{d}(z_0, z_1) = \log y_1 - \log y_0$$

and

$$\phi(t) = y_0 \mathrm{e}^t \mathrm{i}$$

is a function from $[0, \mathrm{d}(z_0,z_1)]$ to $\mathbb{H}$, defining a unit speed curve from z_0 to z_1 of length $\mathrm{L}(\phi) = \log y_1 - \log y_0$. We claim also that the curve ϕ is uniquely determined, i.e. if

$$\psi : [0, \mathrm{d}(z_0,z_1)] \to \mathbb{H}$$

is any path from z_0 to z_1 such that $\mathrm{L}(\psi) = \mathrm{d}(z_0,z_1)$, then there is some increasing piecewise differentiable map $f : [0,1] \to [0, \mathrm{d}(z_0,z_1)]$ with $\psi = \phi \circ f$.

Proof. Firstly, let's introduce $\alpha : [y_0, y_1] \to \mathbb{H}$, $t \mapsto t \cdot i$. Then,

$$\mathrm{L}(\alpha) = \int_{y_0}^{y_1} \frac{|\alpha'(t)|}{\alpha_y(t)} \mathrm{d}t = \int_{y_0}^{y_1} \frac{1}{t} \mathrm{d}t = \log y_1 - \log y_0.$$

Suppose that $\eta : [0,1] \to \mathbb{H}$ is another path joining z_0 to z_1, and write $\eta(t) = \eta_x(t) + \mathrm{i}\eta_y(t)$ with $\eta_x(t), \eta_y(t) \in \mathbb{R}$. Then

$$\mathrm{L}(\eta) = \int_0^1 \frac{\|\eta'(t)\|}{\eta_y(t)}\mathrm{d}t \geqslant \int_0^1 \frac{\left|\eta_y'(t)\right|}{\eta_y(t)}\mathrm{d}t \geqslant \int_0^1 \frac{\eta_y'(t)}{\eta_y(t)}\mathrm{d}t = \log y_1 - \log y_0.$$

Thus $\log y_1 - \log y_0$ is the distance between z_0 and z_1. Equality holds in the first inequality if and only if $\eta_x'(t) = \eta_x(t) = 0$ for all $t \in [0,1]$, and in the second if and only if $\eta_y'(t) \geqslant 0$ for all $t \in [0,1]$.
Thus if both inequalities are equalities, ϕ is indeed uniquely determined. Now we prove that ϕ is a unit speed curve of length $\log y_1 - \log y_0$.

$$\begin{aligned}
\mathrm{L}(\phi) &= \int_0^{\mathrm{d}(z_0,z_1)} \|\mathrm{D}\phi(t)\|_{\phi(t)}\mathrm{d}t \\
&= \int_0^{\mathrm{d}(z_0,z_1)} \frac{|\phi'(t)|}{\phi_y(t)}\mathrm{d}t \\
&= \int_0^{\mathrm{d}(z_0,z_1)} \frac{e^t}{e^t}\mathrm{d}t \\
&= \int_0^{\mathrm{d}(z_0,z_1)} \mathrm{d}t = \mathrm{d}\left(z_0, z_1\right) \\
&= \log y_1 - \log y_0.
\end{aligned}$$

□

Definition 2.18 (Geodesic)**.** A *geodesic* is the path of minimal length joining two points. The set $\{y\mathrm{i} \mid y > 0\}$ is called *geodesic curve* for $\mathbb{H}$.
Since from point 1 of Lemma 2.10 we know that $g \in \mathrm{PSL}(2,\mathbb{R})$ preserves distances, g sends geodesics to geodesics in $\mathbb{H}$.

Theorem 2.19. *For any two points $z_0, z_1 \in \mathbb{H}$, there is a unique path*

$$\phi : [0,\ \mathrm{d}\left(z_0, z_1\right)] \to \mathbb{H}$$

of unit speed with $\phi(0) = z_0$ and $\phi\left(\mathrm{d}\left(z_0, z_1\right)\right) = z_1$. Moreover, there is a unique isometry $g \in \mathrm{PSL}(2,\mathbb{R})$ such that $\phi(t) = g\left(\mathrm{e}^t\mathrm{i}\right)$.

Proof. We first claim that there exists $g \in \mathrm{PSL}(2,\mathbb{R})$ such that $g^{-1}(z_0) = i$ and $g^{-1}(z_1) = iy_1$ for some $y_1 > 1$. Referring to Lecture notes of B. Solomyak [27], by transitivity, we can find $\widetilde{g} \in \mathrm{PSL}(2,\mathbb{R})$ with $\widetilde{g}^{-1}\left(z_0\right) = i$. Then also $h\widetilde{g}^{-1}\left(z_0\right) = i$ for any $h \in \mathrm{PSO}(2)$, since we proved that $\mathrm{PSO}(2)$ is the stabilizer of i. Let $h = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ and $\widetilde{g}^{-1}\left(z_1\right) = \widetilde{x}_1 + i\widetilde{y}_1$. Then $g^{-1} = h\widetilde{g}^{-1}$ satisfies

$$\begin{aligned}
\Re(g^{-1})\left(z_1\right) = \Re(h\left(\widetilde{z}_1\right)) &= \frac{\frac{1}{2}\sin(2\theta)\left(\left|\widetilde{z}_1\right|^2 - 1\right) + \cos(2\theta)\widetilde{x}_1}{\left(\widetilde{x}_1\sin(\theta) + \cos(\theta)\right)^2 + \left(\widetilde{y}_1\sin(\theta)\right)^2} = 0 \text{ (see Eq. 2.2)} \\
\iff \tan(2\theta) &= \frac{2\widetilde{x_1}}{1 - \left|\widetilde{z}_1\right|^2},
\end{aligned}$$

so we get two solutions θ_1 and $\theta_2 = \theta_1 + \frac{\pi}{2}, \theta_1 \in (-\frac{\pi}{4}, \frac{\pi}{4})$.
One of them corresponds to $y_1 > 1$, since if it is smaller than 1, the matrix $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$, that belongs to $\mathrm{PSO}(2)$, would increase the imaginary part. By uniqueness of the unit speed geodesic $\phi_0(t) = \mathrm{e}^t\mathrm{i}$ connecting i to iy in Lemma 2.17, so $\phi(t) = g\left(\mathrm{e}^t\mathrm{i}\right)$ is the unique geodesic path of unit speed connecting z_0 and z_1.
To prove uniqueness of $g \in \mathrm{PSL}(2,\mathbb{R})$, suppose that $t \mapsto g_1\left(\mathrm{e}^t\mathrm{i}\right)$ is any geodesic path from z_0 to z_1. Then $t \mapsto g^{-1}g_1\left(\mathrm{e}^t\mathrm{i}\right)$ is a geodesic path from i to $y_1\mathrm{i}$ of unit speed, and by what we have just proved, it must be equal to the path $t \mapsto \mathrm{e}^t\mathrm{i}$. Taking the derivative at $t = 0$ we see that $\mathrm{D}\left(g^{-1}g_1\right)(\mathrm{i},\mathrm{i}) = (\mathrm{i},\mathrm{i})$, which shows that $g^{-1}g_1 = I_2$ by Lemma 2.15, that implies $g_1 = g$. □

The following lemma is taken from Stewart's book *"Complex Analysis"* [28], adapting it to the upper half plane of the complex plane.

Lemma 2.20. *We call* basic transformations *the following möbius transformations:*

1. *translation:* $f(z) = z + b, b \in \mathbb{R}$,
2. *dilation:* $f(z) = az, a \in \mathbb{R} \setminus \{0\}$,
3. *inversion:* $f(z) = -\frac{1}{z}$.

Any Möbius transformation is a product of basic ones.

Proof.

$$\text{- c = 0:} \quad f(z) = \frac{a}{d}z + \frac{b}{d}$$

$$\text{- } c \neq 0: \quad f(z) = \frac{1}{c} \cdot \frac{az+b}{z+\frac{d}{c}} = \frac{1}{c} \cdot \frac{a\left(z+\frac{d}{c}\right) - \frac{ad}{c} + b}{z+\frac{d}{c}}$$
$$= \frac{a}{c} - \frac{1}{c^2} \cdot \frac{1}{z+\frac{d}{c}}.$$

□

Definition 2.21 (Conformal map)**.** A map that preserves angles between two curves is said to be *conformal.*

Proposition 2.22. *Möbius transformations are conformal maps.*

Proof. The proof can be found at page 54 in J.W. Anderson's book *Hyperbolic Geometry* [1]. The proof is really straight forward and basically consists in proving that for each basic transformation angles are preserved. □

Theorem 2.23. $\mathrm{PSL}(2,\mathbb{R})$ *maps the geodesic curve* $\{y\mathrm{i} \mid y > 0\}$ *either to another vertical line or to an upper half of a circle with center in the real axis (i.e, perpendicular to the real axis). In other words, geodesics of* $\mathbb{H}$ *are vertical lines and upper half of circles perpendicular to the real axis. Combining the last two theorems, we can conclude that for any two different points in the hyperbolic plane, there is a unique geodesic through them.*

Proof. We want to prove that for each basic transformation, vertical lines and half-circles perpendicular to the real line are mapped to themselves. Consider the upper half of a circle

$$(x-f)^2+y^2=r^2$$

with center in the real axis. Firstly note that a transformation in $\mathrm{PSL}(2,\mathbb{R})$ of the form $z \mapsto az+b$ maps the vertical line $\{yi \mid y>0\}$ to the vertical line $\Re(z)=b$ and maps the upper half of the circle $(x-f)^2+y^2=r^2$ to the upper half of the circle $(x-(af+b))^2+y^2=a^2r^2$. Consider now the transformation $w: z \mapsto -\frac{1}{z}$. We express in polar coordinates (r,θ) the transformation w, that becomes $(r,\theta) \mapsto \left(\frac{1}{r}, \pi-\theta\right)$. Indeed expressing the complex number z in its Euler form $z=re^{i\theta}$, w becomes $re^{i\theta} \mapsto \frac{-e^{i(-\theta)}}{r}$. So a half circle perpendicular to the real line is mapped onto a circle of radius $1/r$ and perpendicular to the real line. The claim follows since both vertical lines and circles with real centers can be defined by equations of the form

$$\alpha r^2+\beta r\cos\theta+\gamma=0$$

with $(\alpha,\beta,\gamma) \neq (0,0,0)$. Indeed, we know that the equation of a line L in $\mathbb{R}^2$ has the form

$$ax+by+c=0$$

for some $a,b,c \in \mathbb{R}$. In the complex plane we can write x and y in terms of z as $x=\frac{1}{2}(z+\bar{z}), y=\frac{1}{2i}(z-\bar{z})$. If we substitute these expressions in our equation for the line, we get

$$\frac{1}{2}(a-ib)z+\frac{1}{2}(a+ib)\bar{z}+c=0.$$

So if we let $\beta=(a-ib)/2$, then the equation of L is

$$\beta z+\bar{\beta}\bar{z}+c=0.$$

Now we do the same thing for the circle. The equation for a circle in $\mathbb{R}^2$ is

$$(x-x_0)^2+(y-y_0)^2=r^2.$$

Let $z=x+iy$ and $z_0=x_0+iy_0$. Then we get $|z-z_0|^2=r^2$, which we can write as

$$(z-z_0)\,(\overline{z-z_0})=r^2.$$

Let $\beta=-\overline{z_0}$ and $\gamma=z_0\overline{z_0}-r^2$, then the equation for the circle becomes

$$z\bar{z}+\beta z+\bar{\beta}\bar{z}+\gamma=0,$$

where with $\beta \in \mathbb{R}$ we have half circles perpendicular to $\mathbb{R}$ and if we also have $\alpha=0$ we have vertical lines [30]. Since we know that $\mathrm{PSL}(2,\mathbb{R})$ maps $\mathbb{H}$ to itself and since each Möbius transformation can be expressed as composition of basic ones, we have finished with the proof. □

Chapter 3

M = $\mathbf{PSL}(2,\mathbb{Z})\backslash\mathbb{H}$, $\mathbf{PSL}(2,\mathbb{Z})\backslash\mathbf{PSL}(2,\mathbb{R})$ and Gauss-Bonnet Formula

In this chapter we start constructing a Riemaniann metric (and afterwards a distance) for a general closed linear group in order to define also a measure on the unit tangent bundle of the Modular surface defined as $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}(2,\mathbb{R})$ thanks to Corollary 2.16. We explain new concepts like Fuchsian groups, lattices and fundamental domains, which are also needed to define such a measure. Then we prove (left) invariance of areas and volumes in $\mathbb{H}$ with respect to $\mathrm{PSL}(2,\mathbb{Z})$. We end the chapter introducing Dirichlet regions in order to state and prove the famous Gauss Bonnet formula. The principal reference to accomplish our goals is again Chapter 9 and Chapter 11 of Thomas Ward and Manfred Einsiedlers book *"Ergodic Theory: with a view towards Number Theory"*[7].

3.1 Distance for Closed Linear Groups

Firstly we introduce some general concepts to define a Riemaniann metric for a general closed linear group. This is useful because we proved that $\mathrm{T}^1\mathbb{H} \cong \mathrm{PSL}(2,\mathbb{R})$. So we need this short discussion to define a distance on $\mathrm{PSL}(2,\mathbb{R})$.

Definition 3.1 (Closed Linear Group)**.** A group G is called *closed linear group* if it can be embedded into $\mathrm{GL}_d(\mathbb{R})$ for some $d \geqslant 1$ such that the image is a closed set in $\mathrm{GL}_d(\mathbb{R})$, viewed with the topology of $\mathbb{R}^{d^2}$.

Definition 3.2 (Lie Algebra)**.** Let G be a closed linear group such that $G \subseteq \mathrm{GL}_d(\mathbb{R})$. We call $\mathfrak{g} = \{v \in \mathrm{Mat}_{dd}(\mathbb{R}) \mid \exp(tv) \in G \text{ for all } t \in \mathbb{R}\}$ the *lie algebra* of G.

Proposition 3.3. $\mathfrak{g}$ *is the subspace consisting of all derivatives* $\phi'(0)$ *of paths*

$$\phi : [a,b] \to G \subseteq \mathrm{GL}_d(\mathbb{R})$$

with $a < 0$ *and* $b > 0$ *such that* $\phi(0) = I$.

Proof. The proof is taken from Hall's book *"Lie Groups, Lie Algebras, and Representations"* [9]. Assume X is in $\mathfrak{g}$, then by definition of Lie algebra, $\gamma(t) = e^{tX}$ ends in G and then $\gamma(0) = I$ and $\frac{d\gamma}{dt}|_{t=0} = X$. Conversely, suppose that $\gamma(t)$ is a smooth curve in G with $\gamma(0) = I$. For a sufficiently small t, we can write $\gamma(t) = e^{\delta(t)}$, where δ is a smooth curve in $\mathfrak{g}$ since the exponential map is a bijection in a neighborhoud of zero. Now, the derivative of $\delta(t)$ at $t = 0$ is the same as the derivative of $t \mapsto t\delta'(0)$ at $t = 0$. Thus, by the chain rule, we conclude

$$\gamma'(0) = \frac{d}{dt}e^{\delta(t)}\bigg|_{t=0} = \frac{d}{dt}e^{t\delta'(0)}\bigg|_{t=0} = \delta'(0).$$

Since $\delta(t)$ belongs to $\mathfrak{g}$ for all sufficiently small t, that $\delta'(0) = \gamma'(0)$ belongs to $\mathfrak{g}$. □

Definition 3.4 (Tangent Bundel of a Closed Linear Group)**.** We define the *tangent bundle* $\mathrm{T}G$ of the closed linear group G to be $G \times \mathfrak{g}$. For $g \in G$ we define the tangent space of G at g by $\mathrm{T}_g G = \{g\} \times \mathfrak{g}$.

Definition 3.5 (Derivative of a curve in G)**.** For a path $\phi : [0,1] \to G$ which is differentiable at $t_0 \in [0,1]$ the corresponding tangent vector is

$$\mathrm{D}\phi\left(t_0\right) = \left(\phi\left(t_0\right), \phi\left(t_0\right)^{-1}\phi'\left(t_0\right)\right).$$

Note that the derivative $\phi'\left(t_0\right)$ is meant in the usual sense:

$$\phi : [0,1] \to G \subseteq \mathrm{Mat}_{dd}(\mathbb{R})$$

is viewed as a vector-valued function of dimension d^2. We need to check that

$$\mathrm{D}\phi\left(t_0\right) \in \mathrm{T}G$$

that is, that $\phi\left(t_0\right)^{-1}\phi'\left(t_0\right) \in \mathfrak{g}$. The curve defined by

$$\phi\left(t_0\right)^{-1}\phi(t)$$

has values in G and has $\phi\left(t_0\right)^{-1}\phi(t_0) = I$. By Proposition 3.3,

$$\phi\left(t_0\right)^{-1}\phi'\left(t_0\right) \in \mathfrak{g}$$

as required.

Proposition 3.6. *Let $\phi : [0,1] \to G$ be a continuous curve which is differentiable at $t_0 \in [0,1]$, and let $g \in G$. Then the curves $(g\phi)(t) = g\phi(t)$ and $\left(\phi g^{-1}\right)(t) = \phi(t)g^{-1}$ are also differentiable at t_0, and if $\mathrm{D}\phi\left(t_0\right) = \left(\phi\left(t_0\right), v\right)$ then*

$$\mathrm{D}(g\phi)\left(t_0\right) = \left(g\phi\left(t_0\right), v\right);$$
$$\mathrm{D}\left(\phi g^{-1}\right)\left(t_0\right) = \left(\phi\left(t_0\right)g^{-1}, gvg^{-1}\right).$$

Proof. By definition,

$$\mathrm{D}(g\phi)(t_0) = \left(g\phi(t_0), (g\phi(t_0))^{-1} g\phi'(t_0)\right) = \left(g\phi(t_0), \phi(t_0)^{-1}\phi'(t_0)\right)$$

which shows the first claim since the $\mathfrak{g}$ component is unchanged by g. Similarly,

$$\begin{aligned}\mathrm{D}\left(\phi g^{-1}\right)(t_0) &= \left(\phi(t_0) g^{-1}, \left(\phi(t_0) g^{-1}\right)^{-1} \phi'(t_0) g^{-1}\right) \\ &= \left(\phi(t_0) g^{-1}, g\left(\phi(t_0)^{-1}\phi'(t_0)\right) g^{-1}\right).\end{aligned}$$

□

Definition 3.7 (Riemannian Metric for Closed Linear Groups)**.** Choosing an inner product $\langle\cdot,\cdot\rangle$ on $\mathfrak{g}$, we define a Riemannian metric by letting

$$\langle x, y\rangle_g = \langle u, v\rangle$$

for $u, v \in \mathfrak{g}$ and $x = (g, u), y = (g, v) \in \mathrm{T}_g G$.

Definition 3.8 (Distance for Closed Linear Groups)**.** Let G be a closed linear group. The Riemannian metric defined in Definition 3.7 defines a left-invariant metric on G^0, the connected component of G containing the identity element e. Left invariance comes from Proposition 3.6. For a piecewise smooth curve $\phi : [0,1] \to G$ such that with $\phi(0) = g_0$ and $\phi(1) = g_1$, we define its length with

$$\mathrm{L}(\phi) = \int_0^1 \|\mathrm{D}\phi(t)\|_{\phi(t)} \mathrm{d}t.$$

Now we define a distance with

$$\mathrm{d}_G(g_0, g_1) = \inf_{\phi} \mathrm{L}(\phi),$$

where

$$\mathrm{d}_G(hg_0, hg_1) = \mathrm{d}_G(g_0, g_1)$$

for all $h, g_0, g_1 \in G^0$.

3.2 Fuchsian Groups, Lattices and Fundamnetal Domains

Definition 3.9 (Discrete Subgroup)**.** Consider a group X. A subgroup $Y \subseteq X$ is *discrete* if for every point $z \in Y$ there exists a neighborhood intersecting Y only in the point z.

Definition 3.10 (Fuchsian Group)**.** A *Fuchsian group* is a discrete subgroup $\Gamma \leqslant \mathrm{PSL}(2,\mathbb{R})$.

Definition 3.11 (Lattice)**.** A *lattice* in $\mathrm{PSL}(2,\mathbb{R})$ is a discrete subgroup $\Gamma \leqslant \mathrm{PSL}(2,\mathbb{R})$ such that a fundamental domain (see Definition 3.13) for the quotient space $\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$ has finite measure. Equivalently, applying the same reasoning that we used in the introduction, $\Gamma \leqslant \mathrm{PSL}(2,\mathbb{R})$ is a lattice in $\mathbb{H}$ if it is a discrete subgroup and there is a fundamental domain subset of $\mathbb{H}$ with finite measure (see Definition 3.14).

Definition 3.12 (Uniform Lattice)**.** A lattice is said to be *uniform* if the quotient space $\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$ is compact.

Definition 3.13 (Fundamental Domain for $\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$)**.** Consider the quotient group $\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$. A *fundamental domain* F for $\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$ with Γ a Fuchsian group is a measurable subset of $\mathrm{PSL}(2,\mathbb{R})$ such that:

$$\forall g \in \mathrm{PSL}(2,\mathbb{R}), card\ (\mathring{F} \cap \Gamma g) = 1.$$

Definition 3.14 (Fundamental Domain for $\Gamma\backslash\mathbb{H}$)**.** This definition of *fundamental domain* is taken from Dal'Bo's book *"Geodesic and Horocyclic Trajectories"* [5]: F is a *fundamental domain* for $\Gamma\backslash\mathbb{H}$ with Γ that is a Fuchsian group, if the following properties hold:

1. F is a closed connected subset of $\mathbb{H}$ with non-empty interior,
2. $\bigcup_{\gamma\in\Gamma} \gamma F = \mathbb{H}$,
3. $\mathring{F} \cap \gamma\mathring{F} = \emptyset,\ \forall\gamma \in \Gamma \setminus \{e\}$.

Proposition 3.15. *For $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$, if F is a fundamental domain for $\Gamma\backslash\mathbb{H}$ (Definition 3.14), then $\mathrm{T}^1F \cong \{g \in \mathrm{PSL}(2,\mathbb{R})$ such that $Dg(i,i) \in \mathrm{T}^1F\}$ is a fundamental domain for $\Gamma\backslash\mathrm{T}^1\mathbb{H} \cong \Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$ and satisfies $\forall g \in \mathrm{PSL}(2,\mathbb{R})$, $card(\mathring{\mathrm{T}^1}F \cap \Gamma g) = 1$ (Definition 3.13).*

Proof. The key of the proof comes from Corollary 2.16, i.e. $\mathrm{PSL}(2,\mathbb{R}) \cong \mathrm{T}^1\mathbb{H}$. We prove firstly that T^1F is a fundamental domain for $\Gamma\backslash\mathrm{T}^1\mathbb{H}$. All the properties that we have to satisfy (only 2 and 3 since 1 is trivially satisfied) in Definition 3.14 hold simply passing to the unit tangent bundle because of simply transitivity of $\mathrm{PSL}(2,\mathbb{R})$ on $\mathrm{T}^1\mathbb{H}$ (Lemma 2.14) and because the associated unit vector v is invariant under the derivative action on $\mathrm{T}^1\mathbb{H}$. More precisely for v such that $\|v\|_z = \frac{|v|}{y} = 1$ for $z = x + iy$, $\mathrm{Dg}(z,v) = \left(g(z), \frac{v}{(cz+d)^2}\right)$ with $\|\frac{v}{(cz+d)^2}\|_{g(z)} = \frac{|v|}{|(cz+d)^2|}\frac{|(cz+d)^2|}{y} = \frac{|v|}{y} = 1$. Now to each point $(x,v) \in \mathrm{T}^1F$ we know there exists $g \in \mathrm{PSL}(2,\mathbb{R})$ which satisfies $\mathrm{Dg}(i,i) = (g(i), g'(i)i) = (x,v)$. This g exists and is unique from $\mathrm{T}^1\mathbb{H} \cong \mathrm{PSL}(2,\mathbb{R})$ (Corollary 2.16). So we can correspond to each element in T^1F an element in $\mathrm{PSL}(2,\mathbb{R})$, and thus, translating the conditions 2 and 3 for $\mathrm{PSL}(2,\mathbb{R})$ instead of $\mathrm{T}^1\mathbb{H}$, we get that

1. $\bigcup_{\gamma\in\mathrm{PSL}(2,\mathbb{Z})} \gamma\mathrm{T}^1F = \mathrm{PSL}(2,\mathbb{R})$,
2. $\mathring{\mathrm{T}^1}\mathrm{F} \cap \gamma\mathring{\mathrm{T}^1}\mathrm{F} = \emptyset,\ \forall\gamma \in \Gamma \setminus \{e\}$.

is equivalent to say that for each $g \in \mathrm{PSL}(2,\mathbb{R})$, $card\ (\mathring{\mathrm{T}^1}\mathrm{F} \cap \Gamma g) = 1$. □

Example 3.16 (Modular Group)**.** $\mathrm{PSL}(2,\mathbb{Z}) = \mathrm{SL}(2,\mathbb{Z})/\{\pm I_2\}$ is called *Modular group*.

3.3 Hyperbolic Area and Hyperbolic Volume

In this subsection we prove invariant properties of areas and volumes on $\mathbb{H}$ and $\mathrm{T}^1\mathbb{H}$ under the respective actions of $\mathrm{PSL}(2,\mathbb{R})$.

Definition 3.17 (Hyperbolic Area and Hyperbolic Volume)**.** We define the *hyperbolic area form* as $\mathrm{d}A = \frac{1}{y^2}\,\mathrm{d}x\,\mathrm{d}y$ on $\mathbb{H}$, and the *hyperbolic volume form* as

$$\mathrm{d}m = \frac{1}{y^2}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\theta$$

on $\mathrm{T}^1\mathbb{H}$, where θ is the angle of the unit tangent vector at $z = x + \mathrm{i}y$.

Lemma 3.18. *Hyperbolic volume and area forms are both invariant under the respective actions of* $\mathrm{PSL}(2,\mathbb{R})$.

Proof. The proof follows the one of J.W.Anderson's book "*Hyperbolic Geometry*" [1]. Let's do the proof for the area. We just have to apply the *change of variables theorem* to a Möbius transformation f. We first rewrite f in terms of real and imaginary parts.

$$\begin{aligned} f(z) = \frac{az+b}{cz+d} &= \frac{(az+b)(c\bar{z}+d)}{(cz+d)(c\bar{z}+d)} \\ &= \frac{acx^2+acy^2+bd+bcx+adx}{(cx+d)^2+c^2y^2} + i\frac{y}{(cx+d)^2+c^2y^2} \end{aligned}$$

with $ad-bc=1$ and $a,b,c,d \in \mathbb{R}$. Consider the function $f : \mathbb{H} \to \mathbb{H}$ given by

$$f(x,y) = \left(\frac{acx^2+acy^2+bd+bcx+adx}{(cx+d)^2+c^2y^2}, \frac{y}{(cx+d)^2+c^2y^2}\right).$$

We see that

$$Df(x,y) = \begin{bmatrix} \frac{(cx+d)^2-c^2y^2}{((cx+d)^2+c^2y^2)^2} & \frac{2cy(cx+d)}{((cx+d)^2+c^2y^2)^2} \\ \frac{-2cy(cx+d)}{((cx+d)^2+c^2y^2)^2} & \frac{(cx+d)^2-c^2y^2}{((cx+d)^2+c^2y^2)^2} \end{bmatrix}.$$

In particular, we have that

$$\det(Df(x,y)) = \frac{1}{((cx+d)^2+c^2y^2)^2}.$$

For the computation of an hyperbolic area X in $\mathbb{H}$, we integrate the function $h(x,y) = \frac{1}{y^2}$, so we need to calculate the composition

$$h \circ f(x,y) = \frac{\left((cx+d)^2+c^2y^2\right)^2}{y^2}. \tag{3.1}$$

To conclude,

$$\begin{aligned}\text{Area}_{\mathbb{H}}(f(X)) &= \int_{f(X)} \mathrm{d}A \\ &= \int_{f(X)} \frac{1}{y'^2}\,\mathrm{d}x'\mathrm{d}y' \\ &= \int_X h(f(z))\frac{1}{\left((cx+d)^2+c^2y^2\right)^2}\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_X \frac{1}{y^2}\,\mathrm{d}x\mathrm{d}y \\ &= \int_X \mathrm{d}A = \text{Area}_{\mathbb{H}}(X).\end{aligned}$$

by Equation (3.1), where $z = x + \mathrm{i}y$ and $z' = f(z) = x' + \mathrm{i}y'$. To prove invariance of the volume we have to do similar computations to find the Jacobian for the derivative action on $\mathrm{T}^1\mathbb{H}$, and then argue in the same way we did for the area. More details can be found at pages 307-308 of Ward-Einsiedler's book [7]. □

3.4 A $\mathrm{PSL}(2,\mathbb{R})$-Invariant Measure for $\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$

Now we present a $\mathrm{PSL}(2,\mathbb{R})$-invariant measure for $\Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$.

Proposition 3.19. *If $\Gamma \leqslant \mathrm{PSL}(2,\mathbb{R})$ is a lattice, then $\mathrm{d}m = \frac{1}{y^2}\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}\theta$ (see Lemma 3.17) induces a finite $\mathrm{PSL}(2,\mathbb{R})$-right invariant measure m_X on $X = \Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$.*
Consider the usual quotient map

$$\pi : \mathrm{PSL}(2,\mathbb{R}) \to X$$

such that $\pi(g) = \Gamma g$ for $g \in \mathrm{PSL}(2,\mathbb{R})$ and let's call F a finite volume fundamental domain. Then

$$m_X(B) = m\left(F \cap \pi^{-1}(B)\right)$$

defines the $\mathrm{PSL}(2,\mathbb{R})$ right invariant measure on X for $B \subseteq X$ measurable .

For the proof it is enough to prove the next proposition, which is a generalization of Proposition 3.19.

Proposition 3.20. *Let G be a closed linear group, and let $\Gamma \leqslant G$ be a lattice in the sense that Γ is discrete and that there is a fundamental domain F for $X = \Gamma\backslash G$ with finite left invariant measure m_G. Then any fundamental domain has the same measure as F, G is unimodular (and so also right invariant), and the left invariant measure m_G induces a finite measure m_X on X via*

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

for all measurable $B \subseteq X$. Moreover, the right G-action $R_g(x) = xg^{-1}$ for $x \in X$ and $g \in G$ leaves the measure m_X invariant.

Proof. We first show that any two fundamental domains $F, F' \subseteq G$ for $\Gamma\backslash G$ have the same volume. In fact we want to prove that if $B, B' \subseteq G$ are measurable sets with the property that $\pi|_B$ and $\pi|_{B'}$ are injective, and $\pi(B) = \pi(B')$, then $m_G(B) = m_G(B')$. By assumptions on $\pi|_B$ and $\pi|_{B'}$, for every $g \in B$ there is a unique $\gamma \in \Gamma$ with $g \in \gamma B'$, so

$$B = \bigsqcup_{\gamma\in\Gamma} B \cap \gamma B'$$

and similarly

$$B' = \bigsqcup_{\gamma'\in\Gamma} B' \cap \gamma' B.$$

Indeed, by contradiction we can assume that there exists γ_1 and $\gamma_2 \in \Gamma$ such that $g \in \gamma_1 B'$ and $g \in \gamma_2 B'$. This means that there exist g_1' and g_2' such that $g = \gamma_1 g_1' = \gamma_2 g_2'$. This implies that $g \in \Gamma g_2'$ and $g \in \Gamma g_2'$, but since cosets are disjoint, these two cosets are the same, but by injectivity of π we conclude that $g_2' = g_1'$, that implies $\gamma_1 = \gamma_2$. However, these two decompositions are equivalent in the sense that one can be used to derive the other: given $\gamma \in \Gamma$ and a chosen set $B \cap \gamma B'$ we get

$$\gamma^{-1}\left(B \cap \gamma B'\right) = B' \cap \gamma^{-1} B.$$

For the left invariant measure m_G we therefore have

$$m_G(B) = \sum_{\gamma\in\Gamma} m_G\left(B \cap \gamma B'\right) = \sum_{\gamma\in\Gamma} m_G\left(B' \cap \gamma^{-1} B\right) = m_G\left(B'\right).$$

This proves the claim, and in particular $m_G(F) = m_G(F')$ for any two fundamental domains F and F', since $\pi(F) = \pi(F')$. Now notice that for any $g \in G$ the set $F' = Fg$ is another fundamental domain (see Definition 3.13) whose measure satisfies

$$m_G(F) = m_G\left(F'\right) = \mathrm{mod}(g) m_G(F).$$

The equality holds since left invariant measures on locally compact groups are unique up to scalar. Since our assumption is that $m_G(F) < \infty$, and $m_G(F) > 0$ since Γ is discrete, we deduce that $\mathrm{mod}(G) = 1$ and that G is unimodular, and so also right-invariant. For more details about unimodularity see Appendix C in [7]. Now let $B \subseteq X$ be a measurable set. We define $m_X(B) = m_G\left(\pi^{-1}(B) \cap F\right)$ and note that this definition is independent of the choice of fundamental domain F by the claim above. Now write $C = \pi^{-1}(B) \cap F$ and note that

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

Indeed,

$$\begin{aligned}\pi^{-1}(Bg) = \{h \in G \text{ such that } \Gamma h \in Bg\} = \\ \{h \in G \text{ such that } \Gamma h g^{-1} \in B\} = \\ \{hg \in G \text{ such that } \Gamma h \in B\} = \pi^{-1}(B)g.\end{aligned}$$

Then by above $m_G(C) = m_G(Cg)$, and

$$m_X(Bg) = m_G(Cg) = m_G(C) = m_X(B),$$

so $m_X(B) = m_X\left(R_g^{-1}(B)\right)$ as claimed. □

3.5 The main Example of a Fundamental Domain in $\mathbb{H}$ for the lattice $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$

Here we introduce and prove the most famous example of a fundamental domain in $\mathbb{H}$ for the Möbius action.

Example 3.21. Consider $E = \left\{z \in \mathbb{H} \text{ such that } |z| \geqslant 1, |Re(z)\,\big| \leqslant \frac{1}{2}\right\}$ (see Fig.3.1) and the Modular group $\mathrm{PSL}(2,\mathbb{Z})$. We claim that $\mathrm{PSL}(2,\mathbb{Z})$ is a lattice with E a fundamental domain of finite measure. In particular we claim that

$$A(\gamma E \cap E) = 0 \tag{3.2}$$

for $\gamma \in \mathrm{PSL}_2(\mathbb{Z})\backslash\{I_2\}$, and

$$\mathbb{H} = \bigcup_{\gamma \in \mathrm{PSL}(2,\mathbb{Z})} \gamma E. \tag{3.3}$$

Now we want to prove it.
Consider $\tau = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ and $\sigma = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ in $\mathrm{PSL}_2(\mathbb{Z})$ that act on the set E. We have $\sigma(z) = -\frac{1}{z}, \tau(z) = z + 1$, and

$$\sigma^2 = (\sigma\tau)^3 = I_2.$$

Firstly we prove (3.3).
Since one inclusion is obvious, we want to prove $\mathbb{H} \subseteq \bigcup_{\gamma \in \mathrm{PSL}_2(\mathbb{Z})} \gamma E$. Let $z \in \mathbb{H}$. We want to find $\gamma^{-1} \in \mathrm{PSL}_2(\mathbb{Z})$ such that $z \in \gamma^{-1}E$, or equivalently that $\gamma(z) \in E$. Recall from Equation (2.2), that for $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$,

$$\Im(\gamma(z)) = \frac{\Im(z)}{|cz + d|^2}.$$

Since we are in $\mathrm{PSL}(2,\mathbb{Z})$ (so a discrete space), there exists a matrix γ such that $\Im(\gamma(z))$ is maximal. Choose $k \in \mathbb{Z}$ such that $\left|\Re\left(\tau^k \gamma z\right)\right| \leqslant \frac{1}{2}$. We claim that $w = \tau^k \gamma z \in E$: if

$|w| < 1$ then $\Im\left(-\frac{1}{w}\right) > \Im(z)$, contradicting the maximal property that gamma gives to $\Im(\gamma(z))$. Therefore $|w| \geqslant 1$ and $w \in E$ as wanted.

Now let $z, w \in E$ and $\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{PSL}(2,\mathbb{Z})$ such that $\gamma z = w$. We want to show that w ends up to be a point on the boundary of E, i.e either $|\Re(z)| = \frac{1}{2}$ (and $z = w \pm 1$), or $|z| = 1$ (and $w = -1/z$), such that Equation (3.2) is proved (since the area of lines is zero). Before that, note that if $\Im(\gamma z) < \Im(z)$ we can replace the pair (z, γ) by $\left(\gamma(z), \gamma^{-1}\right)$. Doing this we may assume without loss of generality that $\Im(\gamma z) \geqslant \Im(z)$. Therefore we get $|cz + d| \leqslant 1$ by Equation (2.2). Since $z \in E$, $|\Re(z)| \leq \frac{1}{2}$ and $|z| \geq 1$, that imply $\Im(z) > \frac{1}{2}$. Hence, we conclude that $|c| < 2$, otherwise $|cz + d| > 1$. So $c = 0, \pm 1$. Let's analyze all the cases. If $c = 0$, then $d = \pm 1$ because of Equation (2.2) and since $d \in \mathbb{Z}$. Thus the map γ is a translation map by $\pm b$. By assumption, $|\Re(z)| \leqslant \frac{1}{2}$ and $|\Re(\gamma z)| \leqslant \frac{1}{2}$, so this implies that $b = 0$ and $\gamma = I_2$ (trivial case) or that $b = \pm 1$ and $\{\Re(z), \Re(\gamma z)\} = \left\{\frac{1}{2}, -\frac{1}{2}\right\}$, as wanted.

Now consider $\kappa = -\frac{1}{2} + \mathrm{i}\frac{\sqrt{3}}{2}$.
If $c = 1$, the conditions $z \in E$ and $|z + d| \leqslant 1$ imply that $d = 0$ unless $z = \kappa$ or $z = -\bar{\kappa}$. For $d = 0$, we will have $|z| \leqslant 1$ and so $|z| = 1$. If $c = -1$, the argument is the same. This shows (3.2), and thus that E is a fundamental domain. Lastly, we want to show that our fundamental domain has finite measure, i.e. that $\mathrm{PSL}(2,\mathbb{Z})$ is a lattice in $\mathrm{PSL}(2,\mathbb{R})$. Here what it follows:

$$\begin{aligned} \mathrm{Area}(E) = \int_{z\in E} \mathrm{d}A &\leqslant \int_{\sqrt{3}/2}^{\infty} \int_{-1/2}^{1/2} \frac{\mathrm{d}x\,\mathrm{d}y}{y^2} \\ &= \int_{\sqrt{3}/2}^{\infty} \frac{1}{y^2}\,\mathrm{d}y = \frac{2}{\sqrt{3}} < \infty. \end{aligned}$$

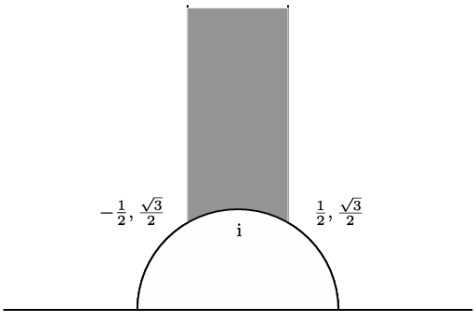

Figure 3.1: Set E in the Example 3.21 [7].

3.6 Dirichlet Regions and Gauss-Bonnet Formula

Now we will refer to Chapter 11 of *"Ergodic Theory with a view towards Number Theory"* [7], to speak about *Dirchlet domains* and to state and prove the famous Gauss-Bonnet formula.

Definition 3.22 (Properly Discontinuous Action)**.** Let K be a locally compact metric space and let Γ be a countable group. Suppose that there is an action $\phi : \Gamma \times K \to K$ that is also an homeomorphism. ϕ is said to be *properly discontinuous* if for any compact subset $U \subseteq K$ the set $\{\gamma \in \Gamma \mid \gamma U \cap U \neq \varnothing\}$ is finite.

Definition 3.23 (Open Fundamental Domain)**.** Assume the same settings that we have in Definition 3.22 on ϕ, K and Γ. An open set $F \subseteq K$ is called an open fundamental domain for ϕ if

1. $g_1 \neq g_2 \in \Gamma \Rightarrow g_1 F \cap g_2 F = \varnothing$,
2. $\bigcup_{g\in G} g\bar{F} = K$.

Lemma 3.24. *Let $\Gamma \subseteq \mathrm{PSL}(2,\mathbb{R})$ be an infinite subgroup. Then, Γ is a Fuchsian group if and only if the action $(\cdot) : \Gamma \times \mathbb{H} \to \mathbb{H}$ is properly discontinuous.*

Proof. Firstly we prove the reverse implication by contrapposition. So assume that Γ is not discrete. It means that we can reach the identity element $e = \pm I_2$ choosing a sequence of elements (g_n) with $g_n \neq e$ for all $n \geqslant 1$ and $g_n \to e$ as $n \to \infty$. Now let $P \subseteq \mathbb{H}$ be a compact set containing an open set (since we are in a locally compact metric space), then $g_n P \cap P \neq \varnothing$ for all large n, therefore $\{\gamma \in \Gamma \mid \gamma P \cap P \neq \varnothing\}$ is infinite, showing that the action of Γ is not properly discontinuous.
Conversely, assume that Γ is discrete and again let P be a compact set. We claim that $\{g \in \Gamma \mid gP \cap P \neq \varnothing\}$ is finite if the set $B = \{g \in \mathrm{SL}_2(\mathbb{R}) \mid gP \cap P \neq \varnothing\}$ is compact. To prove that, note that $B = \{g \in \mathrm{SL}_2(\mathbb{R}) \mid gP \cap P \neq \varnothing\}$ is closed. To see this we look at the limit of sequences in B. For those sequences (g_n) such that $g_n P \cap P \neq \varnothing$, there exists a $p \in P$ such that $(g_n)p$ is in P for every n. Since the set P is compact, for every sequence there exists a convergence subsequence such that $(g_{n_k})p \in P$ for $k \to \infty$, so for all sequences in B, also the limit is in B, therefore B is closed. We need to show that B is a bounded set in $\mathrm{SL}_2(\mathbb{R})$ (as a subset of $\mathbb{R}^4$). By compactness of P, there exist $R, \varepsilon > 0$ such that for every $w \in P$ holds $|w| \leqslant R$ and $\Im(w) \geqslant \varepsilon$. For $g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in B$, $gz \in P$ for some $z \in P$, and thus $\left|\frac{az+b}{cz+d}\right| \leqslant R$ and $\Im\left(\frac{az+b}{cz+d}\right) = \frac{\Im(z)}{|cz+d|^2} \geqslant \varepsilon$. Thus

$$|cz + d|^2 \leqslant \frac{1}{\varepsilon}\Im(z) \leqslant \frac{R}{\varepsilon}$$

and

$$|az + b|^2 \leqslant R^2|cz + d|^2 \leqslant \frac{R^3}{\varepsilon}.$$

Since z belongs to some compact subset P of $\mathbb{H}$, this implies that the coefficients in the matrices of B are in a bounded subset of $\mathbb{R}^4$. □

Definition 3.25 (Dirchlet Region)**.** Let Γ be an infinite Fuchsian group, and let $p \in \mathbb{H}$ such that for every element g in $\Gamma \setminus \{e\}$, $g(p) \neq p$. The set

$$D = D(p) = \{z \in \mathbb{H} \mid \mathrm{d}(z,p) < \mathrm{d}(z,\gamma p) \text{ for all } \gamma \in \Gamma\backslash\{e\}\}$$

is called a *Dirichlet region* for Γ.

Proposition 3.26 (Existence of Dirichlet Region)**.** *Dirichlet Region always exists.*

Proof. Since a Fuchsian group is countable and any non-trivial element can only fix at most two points in $\mathbb{C}$ (indeed, at most one in $\mathbb{H}$), there must be points in $\mathbb{H}$ not fixed by any non-identity element. Moreover, a Dirichlet region is the intersection of the hyperbolic half planes

$$\{z \in \mathbb{H} \mid \mathrm{d}(z,p) < \mathrm{d}(z,\gamma p)\}$$

for each $\gamma \in \Gamma\backslash\{e\}$. □

Lemma 3.27. *Let's take $\gamma \in \mathrm{PSL}(2,\mathbb{R})$ and define L_γ as the geodesic in $\mathbb{H}$ such that $L_\gamma := \{z \in \mathbb{H} \mid \mathrm{d}(z,p) = \mathrm{d}(z,\gamma p)\}$. Then the open set*

$$D_\gamma = \{z \in \mathbb{H} \mid \mathrm{d}(z,p) < \mathrm{d}(z,\gamma p)\}$$

is the connected component of $\mathbb{H}\backslash L_\gamma$ containing p. It follows that a Dirichlet region is connected and convex, where convex here means that a geodesic path joining two points in the Dirichlet region D is in D.

Proof. We may apply an isometry $g \in \mathrm{PSL}(2,\mathbb{R})$ to map p and γp to $-r+\mathrm{i}$ and $r+\mathrm{i}$ respectively, such that with this chosen r we can ensure that $\mathrm{d}(-r+\mathrm{i}, r+\mathrm{i}) = \mathrm{d}(p,\gamma p)$, and therefore the geodesic L_γ will be the upper-half imaginary axis. Now D_γ is the left side of the hyperbolic plane, thus it is convex. □

Lemma 3.28. *Let Γ be an infinite Fuchsian group. Then, any Dirichlet region for Γ is an open fundamental domain for the action of Γ on $\mathbb{H}$.*

Proof. Let $D = D(p)$ be a Dirichlet region for a point p. Firstly let us prove that the Dirichlet region is open. The action of Γ is properly discontinuous (see Lemma 3.24), so we have that for any $z \in \mathbb{H}$, picking a compact set $\{w \in \mathbb{H} \mid \mathrm{d}(w,p) \leq \mathrm{d}(z,p)+1\}$ there are only finitely many $\gamma \in \Gamma$ with

$$\mathrm{d}(\gamma z, p) \leqslant \mathrm{d}(z,p) + 1,$$

say $\gamma_1, \ldots, \gamma_n$. Note that for $z' \in B_{\frac{1}{2}}(z) = \{w \in \mathbb{H} \text{ s.t. } d(z,w) < \frac{1}{2}\}$, this list of elements will include those $\gamma \in \Gamma$ for which $\mathrm{d}(\gamma z', p) \leqslant \mathrm{d}(z', p)$ since $\mathrm{d}(z', p) < \mathrm{d}(z,p) + 1$. In particular, there is some $w \in \Gamma z$ with

$$\mathrm{d}(w,p) \leqslant \mathrm{d}(\gamma z, p) = \mathrm{d}\left(z, \gamma^{-1} p\right)$$

for all $\gamma \in \Gamma$. Note that $\mathrm{d}\left(z, \gamma^{-1}p\right)$ is the bound in the inequality defining a Dirchlet region. Without loss of generality, assume that $z = w$. If $z \in D$, then a point z' close to z belongs to $D_{\gamma_1} \cap \cdots \cap D_{\gamma_n}$ by Lemma 3.27, so that $z' \in D$ and hence every point in D has an open neighbourhood in D (in this case $D_{\gamma_1} \cap \cdots \cap D_{\gamma_n} \cap B_{1/2}(z)$), so D is open.

Now we want to prove $\bigcup_{\gamma\in\Gamma} \gamma \cdot \bar{D} = \mathbb{H}$. To do that it is enough to show that $\forall z \in \mathbb{H}$, there exists a $\gamma \in \Gamma$ such that $\gamma^{-1} z \in \bar{D}$.
If $z \in D$, then $\gamma = Id_2$. If $z \notin D$, then z belongs to some of the boundaries of the sets $D_{\gamma_1}, \ldots, D_{\gamma_n}$. Assume that

$$z \in L_{\gamma_1} \cap \cdots \cap L_{\gamma_m} \cap D_{\gamma_{m+1}} \cap \cdots \cap D_{\gamma_n}.$$

Then by Lemma 3.27, in particular from hyperbolic convexity of Dirichlet regions, the geodesic path joining z to p will be in D entirely except at z, so $z \in \bar{D}$ and therefore we are concluded this part of the proof.

Lastly we want to prove that having $g_1 \neq g_2$, $g_1 D \cap g_2 D = \varnothing$. Assume that points z and w in D have $z = \gamma w$ for some $\gamma \in \Gamma\backslash\{e\}$. Then

$$\mathrm{d}(w,p) < \mathrm{d}\left(w, \gamma^{-1}p\right) = \mathrm{d}(z,p)$$

and

$$\mathrm{d}(z,p) < \mathrm{d}(z,\gamma p) = \mathrm{d}(w,p)$$

which is a contradiction. In this last part of the proof, we have chosen g_2 as the identity, while the two inequalities follow from the definition of a Dirichlet domain.

□

Corollary 3.29. *The boundary of a Dirichlet region is made up of geodesic segments contained in geodesics defined by*

$$L_\gamma = \{z \in \mathbb{H} \mid \mathrm{d}(z,p) = \mathrm{d}(z,\gamma p)\}$$

for $\gamma \in \Gamma\backslash\{e\}$.

Proof. The proof is included in the second part of the proof of the previous lemma. □

Definition 3.30 (Hyperbolic Plane with Boundary)**.** We write

$$\overline{\mathbb{H}} = \mathbb{H} \cup \partial\mathbb{H}$$

for the union of $\mathbb{H}$ with its boundary $\partial\mathbb{H} = \mathbb{R} \cup \{\infty\}$.

Definition 3.31 (Complex Polygon)**.** We define a convex polygon $K \subseteq \mathbb{H}$ by taking points on geodesic paths connecting two vertices. Moreover, any point of $\bar{K}^{\overline{\mathbb{H}}} \cap \partial\mathbb{H}$ is also a vertex, where we denote with $\bar{K}^{\overline{\mathbb{H}}}$ boundaries and closures taken in the set $\overline{\mathbb{H}}$.

Theorem 3.32. *A Dirichlet region for a lattice in* $\mathrm{PSL}(2,\mathbb{R})$ *has finitely many sides. That is, it is a convex hyperbolic polygon.*

Proof. The proof is very long and complicated, and not so relevant for this paper. It can be found at page 353 in Einsiedler's book [7]. □

We finally can state and prove the famous *Gauss - Bonnet* formula.

Proposition 3.33 (Gauss - Bonnet Formula)**.** *Let P be a hyperbolic n sided convex polygon in $\mathbb{H}$ with $n \geqslant 3$ vertices in $\overline{\mathbb{H}}$, with angles $\alpha_1, \ldots, \alpha_n$ at the n vertices. Then the hyperbolic area of P is*

$$(n-2)\pi - (\alpha_1 + \cdots + \alpha_n)\,.$$

For a vertex in $\partial\mathbb{H}$ we set the angle to be zero since the circles are tangential in that point.

Proof. Let's argue by induction.

1. Induction case: take $n = 3$. In $\mathbb{H}$ we can have many types of triangles, since geodesics can be vertical lines or semicircles perpendicular to the real axis. Let us analyze all the cases. Firstly let us analyze the triangle $T = BPC$ with one vertex $P = \infty$ (see Fig. 3.2), and through a Möbius transformation we place the other two points on a semicircle with the origin where the two dashed segments meet. Remember that area is preserved by the action of $\mathrm{PSL}(2,\mathbb{R})$ (see Lemma 3.18). In this case we easily compute the area of the triangle, that is

$$\mathrm{Area}(T) = \int_{\cos(\pi-\beta)}^{\cos(\gamma)} \left(\int_{\sqrt{1-x^2}}^{\infty} \frac{\mathrm{d}y}{y^2} \right) \mathrm{d}x = \pi - (\beta + \gamma)\,.$$

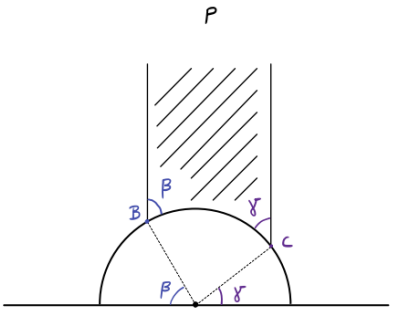

Figure 3.2: Step 1 of the Induction proof [30].

Now referring to Fig. 3.3, let's assume that all three vertices lie in $\mathbb{H}$ and assume without loss of generality that z_1 and z_2 do not have the same real part. It follows that the geodesic through z_1 and z_2 meets the real axis at w (a point at infinity). Then the triangle $z_1 z_3 w$ has interior angles $\pi - \alpha_1, \beta, 0$ while the triangle $z_2 z_3 w$ has interior angles $\alpha_2, \alpha_3 + \beta, 0$. From the fact that Möbius maps are conformal (see Proposition 2.22), we can move hyperbolic triangles $z_2 z_3 w$ and $z_1 z_3 w$ in a situation like the one in Figure 3.2, and thus we can compute their areas.

Thus,

$$\begin{aligned} \mathrm{Area}\,(z_1 z_2 z_3) &= \mathrm{Area}\,(z_2 z_3 w) - \mathrm{Area}\,(z_1 z_3 w) \\ &= \pi - (\alpha_2 + \alpha_3 + \beta) - (\pi - (\pi - \alpha_1 + \beta)) \\ &= \pi - (\alpha_1 + \alpha_2 + \alpha_3)\,. \end{aligned}$$

2. Induction assumption: assume that the formula holds for all polygons with $(n-1)$ sides.

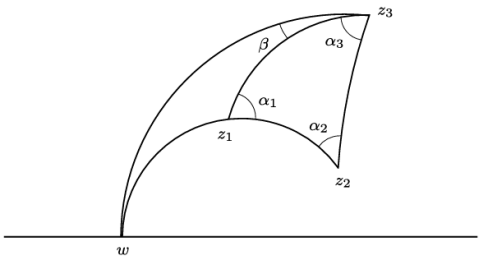

Figure 3.3: Triangle in $\mathbb{H}$ [7].

3. Induction proof: Cutting our n-polygon P in one triangle T and one $(n-1)$-polygon Q, we will have

$$\begin{aligned}\text{Area}(P) &= \text{Area}(Q) + \text{Area}(T)\\ &= ((n-1)-2)\pi - (\alpha_1 + \cdots + \alpha_{n-3} + (\alpha_{n-2} - \beta_2) + (\alpha_{n-1} - \beta_1))\\ &\quad + \pi - (\beta_1 + \beta_2 + \alpha_n)\\ &= (n-2)\pi - (\alpha_1 + \cdots + \alpha_n) \text{ (see Fig. 3.4).}\end{aligned}$$

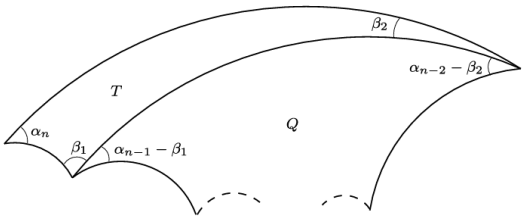

Figure 3.4: Step 3 of Induction proof [7].

□

Chapter 4

Ergodicity of Geodesic Flows on $\mathrm{PSL}(2,\mathbb{Z})\backslash\mathrm{PSL}(2,\mathbb{R})$ and of the Gauss Map

Let's introduce two types of flows on $\mathrm{T}^1\mathbb{H}$: geodesic and horocyle flows. After that we prove ergodicity of geodesic flows on $\mathrm{PSL}(2,\mathbb{Z})\backslash\mathrm{T}^1\mathbb{H} \cong \mathrm{PSL}(2,\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{R})$, that is the unit tangent bundle of the Modular surface. Thanks to that we will prove ergodicty of the Gauss map with respect to the Gauss measure and several results related to that. We always follow chapter 9 of "*Ergodic Theory: with a view towards Number Theory*" [7] and we pick applications of the ergodicity of the Gauss map from chapter 3 of the same book.

4.1 Geodesic Flows and Horocycle Flows

A geodesic can be uniquely determined by two points in $\mathbb{H}$ (see Theorem 2.19), or by a base point z and a unit vector v in $\mathrm{T}^1_z\mathbb{H}$ such that the geodesic passes through z in direction v. In other words we can say that travelling at unit speed on a path of minimal distance, given a point and a direction, is called geodesic flow.

Definition 4.1 (Geodesic Flow)**.** We call *geodesic flow* the function $\Gamma_t : \mathrm{T}^1\mathbb{H} \to \mathrm{T}^1\mathbb{H}$ that follows the uniquely defined parametrization of a geodesic determined by (z, v) for time t.

Example 4.2. For the vector (i, i),

$$\Gamma_t((\mathrm{i},\mathrm{i})) = \left(\mathrm{e}^t\mathrm{i}, \mathrm{e}^t\mathrm{i}\right).$$

Otherwise, we can see $\left(e^t\mathrm{i}, \mathrm{e}^t\mathrm{i}\right)$ as the image of (i,i) under the derivative action of the matrix $a_t^{-1} = \begin{bmatrix} \mathrm{e}^{t/2} & 0 \\ 0 & \mathrm{e}^{-t/2} \end{bmatrix}$,

$$\Gamma_t((\mathrm{i},\mathrm{i})) = \mathrm{D}\begin{bmatrix} \mathrm{e}^{t/2} & 0 \\ 0 & \mathrm{e}^{-t/2} \end{bmatrix}(\mathrm{i},\mathrm{i}) = \left(\mathrm{e}^t\mathrm{i}, \mathrm{e}^t\mathrm{i}\right),$$

since $a_t^{-1}(z) = e^t z$ and $(a_t^{-1})'(z) = e^t$, and therefore we can express $\Gamma_t(i,i) = \left(\mathrm{D}a_t^{-1}\right)(i,i)$.

Remark 4.3. Any other geodesic flow is a path in $\mathrm{T}^1\mathbb{H}$ made by a unique transformation g in $\mathrm{PSL}(2,\mathbb{R})$ (see Corollary 2.16) of $e^t i$, so the geodesic flow described by $(z,v) = \mathrm{D}g(\mathrm{i},\mathrm{i}) \in \mathrm{T}^1\mathbb{H}$ is given by

$$\Gamma_t((z,\mathrm{v})) = \mathrm{D}g\left(\Gamma_t(i,i)\right) = \mathrm{D}g\left(e^t i, e^t i\right) = (\mathrm{D}g)\left(\mathrm{D}a_t^{-1}\right)(i,i) = \left(\mathrm{D}g a_t^{-1}\right)(i,i).$$

Thus $\Gamma_t((z,v))$ in $\mathrm{T}^1\mathbb{H}$ corresponds to $R_{a_t}(g) = g a_t^{-1}$ in $\mathrm{PSL}(2,\mathbb{R})$. The Figure 4.1 taken from B. Solomyak's lecture notes [27] shows how to switch from $\mathrm{T}^1\mathbb{H}$ to $\mathrm{PSL}(2,\mathbb{R})$.

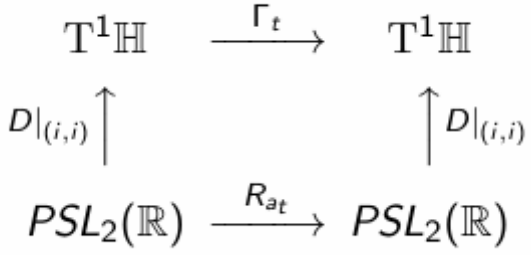

Figure 4.1: From $\mathrm{T}^1\mathbb{H}$ to $\mathrm{PSL}(2,\mathbb{R})$ [27].

Another flow type that we want to mention is the *horocycle flow*. The starting question from which we start is: "Which two pairs $(z_0,v_0),(z_1,v_1)$ have the geodesic flows that converge towards one another"?

Example 4.4. Consider our reference vector (i,i) and a second vector $(x+\mathrm{i},\mathrm{i})$. The geodesic flows of these two vectors are both vertical lines and are parallel to each other, with

$$\Gamma_t((\mathrm{i},\mathrm{i})) = \left(\mathrm{e}^t\mathrm{i}, \mathrm{e}^t\mathrm{i}\right),$$
$$\Gamma_t((x+\mathrm{i},\mathrm{i})) = (\mathrm{D}g a_t^{-1})(i,i) = \left(x+\mathrm{e}^t\mathrm{i}, \mathrm{e}^t\mathrm{i}\right)$$

for $g = \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix}$. The distance between their base points $\mathrm{e}^t\mathrm{i}$ and $x+\mathrm{e}^t\mathrm{i}$ for t that goes to infinity tends to zero, indeed

$$d(\left(\mathrm{e}^t\mathrm{i}, x+\mathrm{e}^t\mathrm{i}\right) \leq L([\mathrm{e}^t\mathrm{i}, x+\mathrm{e}^t\mathrm{i}]) = \int_0^{|x|} \frac{1}{e^t} ds = \frac{|x|}{e^t}.$$

Definition 4.5 (Stable Horocylce flow)**.** The *stable horocylce flow* on $\mathrm{T}^1\mathbb{H}$ for a point $(z,v) = \mathrm{D}g(\mathrm{i},\mathrm{i})$ is

$$u^-(s)((z,v)) := \mathrm{D}\left(g \begin{bmatrix} 1 & -s \\ 0 & 1 \end{bmatrix}\right)(\mathrm{i},\mathrm{i}),$$

that corresponds on $\mathrm{PSL}(2,\mathbb{R})$ by

$$R_{u^-(s)}(g) = gu^-(-s) = g(u^-(s))^{-1}$$

for $g \in \mathrm{PSL}(2,\mathbb{R})$ and $u^-(s) = \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \in U^- = \left\{ \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle| \, s \in \mathbb{R} \right\}$.

Proposition 4.6. *The stable horocycle flow of a point $(z,v) = Dg(i,i)$ is the set containing those points such that their associated geodesic flow converge (for $t \to \infty$) to the geodesic flow of (z,v).*

Proof. For any two pairs $(z_1,v_1) = \mathrm{D}g_1(i,i), (z_2,v_2) = \mathrm{D}g_2(i,i)$, we want to show that

$$d_{\mathrm{PSL}(2,\mathbb{R})}\left(\Gamma_t\left(z_0,v_0\right),\Gamma_t\left(z_1,v_1\right)\right) \longrightarrow 0$$

as $t \to \infty$ if and only if $\mathrm{D}g_2(i,i) = (z_2,v_2) = \mathrm{D}(g_1(u^-(y)^{-1}))(i,i)$ for some $y \in \mathbb{R}$. By Definition 3.8, Remark 4.3 and Corollary 2.16,

$$d_{\mathrm{PSL}(2,\mathbb{R})}\left(\Gamma_t\left(z_0,v_0\right),\Gamma_t\left(z_1,v_1\right)\right) \longrightarrow 0 \iff d_{\mathrm{PSL}(2,\mathbb{R})}\left(g_1a_t^{-1}, g_2a_t^{-1}\right) \longrightarrow 0.$$

By Definition 3.8, the metric on $\mathrm{PSL}(2,\mathbb{R})$ is left invariant, so

$$d_{\mathrm{PSL}(2,\mathbb{R})}\left(g_1a_t^{-1}, g_2a_t^{-1}\right) \longrightarrow 0 \iff a_tg_1^{-1}g_2a_t^{-1} \to I_2$$

as $t \to \infty$. Calling $g_1^{-1}g_2 = \begin{bmatrix} x & y \\ z & w \end{bmatrix}$, $a_tg_1^{-1}g_2a_t^{-1} = \begin{bmatrix} x & e^{-t}y \\ e^tz & w \end{bmatrix}$, and so

$$\begin{bmatrix} x & e^{-t}y \\ e^tz & w \end{bmatrix} \to I_2 \iff g_1^{-1}g_2 = \begin{bmatrix} 1 & y \\ 0 & 1 \end{bmatrix} \iff g_2 = g_1\begin{bmatrix} 1 & y \\ 0 & 1 \end{bmatrix} \text{ for } y \in \mathbb{R}.$$

By Corollary 2.16, this is true if and only if $(z_2,v_2) = Dg_2(i,i) = D(g_1(u^-(y))^{-1})(i,i)$. □

From Example 4.4 and Proposition 4.6, it is easy to see that for a point with vertical direction, so with a vertical associated geodesic flow, the associated horocycle flow is the horizontal line passing through z with vertical unit vector v at each point. For all other points (z,v) in $\mathrm{T}^1\mathbb{H}$, the horocycle flow is described by all the points on the circle tangent to $\mathbb{R}$ in the point where the geodesic flow associated to (z,v) with vector v pointing towards the inside of the circle ends. All points on the circle have associated orthogonal vectors v to the circle where they lay, with v pointing towards the inside of the circle. Indeed, seeing Proposition 4.6, these are the points such that the distance between their associated geodesic flows goes to zero, that translates in converging to the same point on the real axis (see Fig. 4.2).

Definition 4.7 (Unstable Horocylce Flow)**.** The *unstable horocylce flow* on $\mathrm{T}^1\mathbb{H}$ for a point $(z,v) = \mathrm{D}g(\mathrm{i},\mathrm{i})$ is

$$u^+(s)((z,v)) = \mathrm{D}\left(g\begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix}\right)(\mathrm{i},\mathrm{i}),$$

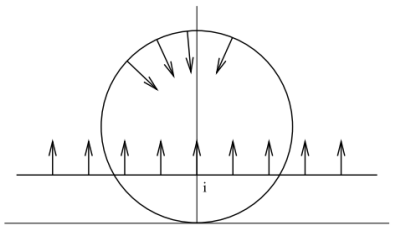

Figure 4.2: Two horocylce flows in $\mathrm{T}^1\mathbb{H}$ [7].

that corresponds on $\mathrm{PSL}(2,\mathbb{R})$ by

$$R_{u^+(s)}(h) = hu^+(-s) = h(u^+(s))^{-1}$$

for $h \in \mathrm{PSL}(2,\mathbb{R})$ and $u^+(s) = \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix} \in U^+ = \left\{ \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix} \middle| \; s \in \mathbb{R} \right\}$.

Proposition 4.8. *The unstable horocycle flow of a point $(z,v) = Dg(i,i)$ is the set containing those points such that their associated geodesic flow converge for $t \to -\infty$ to the geodesic flow of (z,v).*

Proof. The proof is almost identical to the one of Proposition 4.6. □

The points on $\mathbb{H}$ associated to the unstable horocylce flow of $(z,v) = Dg(i,i)$ are the same as those of the stable horocycle flow, but the associated orthogonal vectors v to the circle tangent to $\mathbb{R}$ are pointing outside the circle (or pointing towards the real axis in case of horizontal horocycle flow).

4.2 Ergodicity of Geodesic Flows on $\mathrm{T}^1\mathrm{M}$

In this section we prove ergodicity of geodesic flows on T^1M and we use this result to also prove ergodicity of the Gauss map. The reference is always chapter 9 of Ward and Einsiedler's book, in particular chapters 9.4 and 9.6 [7]. The argument was developed by E. Hopf in more general settings in his paper from 1971 [12].

Proposition 4.9 (Generators of $\mathrm{SL}_2(\mathbb{R})$)**.** *For any $g \in \mathrm{SL}_2(\mathbb{R})$,*

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

for $s_1, s_2, s_3, s_4 \in \mathbb{R}$, $U^- = \left\{ \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle| \; s \in \mathbb{R} \right\}$ and $U^+ = \left\{ \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix} \middle| \; s \in \mathbb{R} \right\}$.

Proof. Take a general element γ in $\mathrm{SL}(2,\mathbb{R})$

$$\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ such that } ad - bc = 1.$$

Note,

$$\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{c}{a} & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & \frac{1}{a} \end{bmatrix} = \begin{bmatrix} a & b \\ c & \frac{bc+1}{a} \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ by } ad - bc = 1.$$

Now remark that for any $\delta \in \mathbb{R}$ such that $b - \delta a \neq 0$ and for $a \neq 0$,

$$\gamma = \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{c}{a} + \frac{1-a}{a(b-\delta a)} & 1 \end{bmatrix} \begin{bmatrix} 1 & b - \delta a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{a-1}{b-\delta a} & 1 \end{bmatrix} \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix},$$

since

$$\begin{bmatrix} 1 & 0 \\ \frac{c}{a} & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{1-a}{a(b-\delta a)} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{c}{a} + \frac{1-a}{a(b-\delta a)} & 1 \end{bmatrix}$$

and because

$$\begin{bmatrix} a & b \\ 0 & \frac{1}{a} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \frac{1-a}{a(b-\delta a)} & 1 \end{bmatrix} \begin{bmatrix} 1 & b - \delta a \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{a-1}{b-\delta a} & 1 \end{bmatrix} \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix}.$$

For $a = 0$, automatically we have $c \neq 0$ and we get

$$\gamma = \begin{bmatrix} 0 & \frac{-1}{c} \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ (c\delta + 1 - d)c & 1 \end{bmatrix} \begin{bmatrix} 1 & \frac{-1}{c} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix} \begin{bmatrix} 1 & \delta \\ 0 & 1 \end{bmatrix}.$$

□

Now we recall some definitions and results of ergodic theory that are helpful in order to reach our aims.

Definition 4.10 (Measure-preserving transformation)**.** Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{C}, \nu)$ be probability spaces. A map ϕ from X to Y is measurable if $\phi^{-1}(A) \in \mathcal{B}$ for any $A \in \mathcal{C}$, and is measure-preserving if it is measurable and $\mu\left(\phi^{-1}(B)\right) = \nu(B)$ for all $B \in \mathcal{C}$.

Lemma 4.11. *For a function $T : X \to X$, a measure μ on X is T-invariant (or T is measure-presrving) if and only if*

$$\int f \, \mathrm{d}\mu = \int f \circ T \, \mathrm{d}\mu$$

for all $f \in L^1_\mu$.

Definition 4.12 (Ergodic transformation)**.** A measure-preserving transformation $T : X \to X$ of a probability space $(X, \mathcal{B}, \mu)$ is ergodic if for any $B \in \mathcal{B}$,

$$T^{-1}(B) = B \Longrightarrow \mu(B) = 0 \text{ or } \mu(B) = 1.$$

Proposition 4.13 (Ergodicity)**.** *The following are equivalent properties for a measure-preserving transformation T of $(X, \mathcal{B}, \mu)$.*

1. *T is ergodic.*
2. *For any $B \in \mathcal{B}$, $\mu\left(T^{-1}B \triangle B\right) = 0$ implies that $\mu(B) = 0$ or $\mu(B) = 1$.*

3. *For $A \in \mathcal{B}, \mu(A) > 0$ implies that $\mu\left(\bigcup_{n=1}^{\infty} T^{-n}A\right) = 1$.*

4. *For $A, B \in \mathcal{B}, \mu(A)\mu(B) > 0$ implies that there exists $n \geqslant 1$ with*

$$\mu\left(T^{-n}A \cap B\right) > 0.$$

5. *For $f : X \to \mathbb{C}$ measurable such that $f \circ T = f$ almost everywhere, then f is equal to a constant almost everywhere.*

The fifth point of the last proposition is actually the only one that we will use to prove ergodicty.

Theorem 4.14 (Lusin's Theorem)**.** *Let $(X, \mathcal{B}, \mu)$ be a Borel probability space and let $f : X \to \mathbb{R}$ be a measurable function. Then, for any $\varepsilon > 0$, there is a continuous function $g : X \to \mathbb{R}$ with the property that*

$$\mu(\{x \in X \mid f(x) \neq g(x)\}) < \varepsilon.$$

Note that this directly implies that $f|_K$ is continuous on $K = \{x \in X \mid f(x) = g(x)\}$, and K is a compact set.

Theorem 4.15. *Let G be a σ-locally metrizable compact group, and let m_G denote a left invariant measure on G. $\mathcal{B}_G$ is the Borel σ-algebra of G. Then for any two Borel sets $B_1, B_2 \in \mathcal{B}_G$ with $m_G(B_1)\, m_G(B_2) > 0$, the set*

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

is open and non-empty. The same holds for $\{g \in G \mid m_G(B_1 g \cap B_2) > 0\}$. Moreover, $m_G(B) > 0$ if and only if $m_G\left(B^{-1}\right) > 0$ for $B \in \mathcal{B}_G$.

Theorem 4.16 (Birkoff's Pointwise Ergodic Theorem)**.** *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system. If $f \in L^1_\mu$, then*

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

converges almost everywhere and in L^1_μ to a T-invariant function $f^* \in L^1_\mu$, and

$$\int f^* \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu.$$

If T is ergodic, then

$$f^*(x) = \int f \, \mathrm{d}\mu \text{ a.e. }.$$

Theorem 4.17 (Heine-Cantor Theorem)**.** *If $f : X_1 \to X_2$ is a continuous function and X_1 is compact, then f is uniformly continuous.*

Theorem 4.18 (Ergodicity of Geodesic Flow on $\mathrm{T}^1 M$)**.** *Let $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$. It follows that any non-trivial element of the geodesic flow (that is, the map R_{a_t} for some $t \neq 0$) is an ergodic transformation on $X = \Gamma\backslash\mathrm{PSL}(2,\mathbb{R})$ with respect to m_X.*

Proof. Remember from before that $U^- = \left\{ \begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix} \middle| \ s \in \mathbb{R} \right\}, U^+ = \left\{ \begin{bmatrix} 1 & 0 \\ s & 1 \end{bmatrix} \middle| \ s \in \mathbb{R} \right\}$ and $a_t = \begin{bmatrix} \mathrm{e}^{-t/2} & 0 \\ 0 & \mathrm{e}^{t/2} \end{bmatrix}$. Recall also that $R_g(h) = hg^{-1}$. The main idea for this proof is to prove that a measurable R_{a_t} - invariant function $f : X \to \mathbb{R}$ is constant, and this implies that R_{a_t} is ergodic by point 5 of Theorem 4.13. To prove that the function is constant we will prove that $f(xg) = f(x)\ \forall g \in \mathrm{PSL}(2,\mathbb{R})$. Moreover we know from Proposition 4.9 that any $g \in \mathrm{PSL}(2,\mathbb{R})$ can be written as product of four elements in U^- and U^+. This will be useful later.
First of all let us normalize the measure m_X such that $m_X(X) = 1$. m_X is finite since it is equal to the volume derived from the area found in Example 3.21.
Take a function $f : X \to \mathbb{R}$ to be measurable and R_{a_t}-invariant.
Fix $\varepsilon > 0$. By Lusin's theorem we are allowed to choose a compact set $K \subseteq X$ of measure $m(K) > 1 - \varepsilon$ such that $f|_K$ is continuous.
Now let's define

$$B = \left\{ x \in X \middle| \lim_{n\to\infty} \frac{1}{n} \sum_{\ell=0}^{n-1} \chi_K \left(R_{a_t}^{\ell} x \right) > \frac{1}{2} \right\}.$$

Define now

$$g^*(x) = \lim_{n\to\infty} \frac{1}{n} \sum_{\ell=0}^{n-1} \chi_K \left(R_{a_t}^{\ell} x \right) \in [0,1].$$

Since R_{a_t} is measure preserving (see Proposition 3.19), by Birkoff ergodic theorem (See Theorem 4.16) we get

$$\int_X g^* \, \mathrm{d}m_X = \int_X \chi_K \, \mathrm{d}m_X = m_X(K) \geqslant 1 - \varepsilon$$

and we can reach this inequality:

$$\begin{aligned} 1 - \varepsilon \leqslant \int_X g^* \, \mathrm{d}m_X &= \int_B g^* \, \mathrm{d}m_X + \int_{X\backslash B} g^* \, \mathrm{d}m_X \\ &\leqslant m_X(B) + \frac{1}{2} m_X(X\backslash B) \\ &= m_X(B) + \frac{1}{2} \left(1 - m_X(B)\right) \\ &= \frac{1}{2} m_X(B) + \frac{1}{2}, \end{aligned}$$

where $\int_{X\backslash B} g^* \, \mathrm{d}m_X \leqslant \frac{1}{2} m_X(X\backslash B)$ comes from the definition of B.
Thus $m_X(B) \geqslant 2 - 2\varepsilon > 1 - 2\varepsilon$. So for a.e. $x, y \in X$, we can find a sequence $(R_{a_t}^{\ell})_{\ell\in\mathbb{N}_{>0}}$ such that $R_{a_t}^{\ell}$ maps x and y into K.
Suppose now that x and $y = R_{u^-(s)}x$ for some $s \in \mathbb{R}$ both belong to the set B.
By R_{a_t}-invariance we have

$$f(x) = f\left(R_{a_t}^{\ell}(x) \right), f(y) = f\left(R_{a_t}^{\ell}(y) \right)$$

for all $\ell \geq 1$, and since $d(\Gamma x_1, \Gamma x_2) = \min_{\gamma\in\Gamma} d(x_1, \gamma x_2)$, we have the inequality below:

$$\begin{aligned}\mathrm{d}_X\left(R^\ell_{a_t}(x), R^\ell_{a_t}(y)\right) &= \mathrm{d}_X\left(xa_t^{-\ell}, xu^-(-s)a_t^{-\ell}\right)\\ &\leqslant \mathrm{d}_{\mathrm{PSL}(2,\mathbb{R})}\left(a_t^{-\ell}, u^-(-s)a_t^{-\ell}\right)\\ &= \mathrm{d}_{\mathrm{PSL}(2,\mathbb{R})}\left(I_2, a_t^\ell u^-(-s)a_t^{-\ell}\right) \to 0\end{aligned}$$

as $\ell \to \infty$. The limit goes to zero since $a_t^\ell u^-(-s)a_t^{-\ell} = \begin{bmatrix} 1 & -s\mathrm{e}^{-t\ell} \\ 0 & 1 \end{bmatrix}$.
Since asymptotically both x and y spend more than half of their future in K, there exists a convergent common sub-sequence ℓ_n such that

$$R^{\ell_n}_{a_t}(x), R^{\ell_n}_{a_t}(y) \in K$$

for $n \to \infty$.
By Heine-Cantor Theorem, since K is compact, $f|_K$ is uniformly continuous, therefore as $R^{\ell_n}_{a_t}(x)$ and $R^{\ell_n}_{a_t}(y)$ become closer, $f\left(R^{\ell_n}_{a_t}(x)\right)$ and $f\left(R^{\ell_n}_{a_t}(y)\right)$ also become closer. For that reason $f(x) = f\left(R^{\ell_n}_{a_t}(x)\right) = f\left(R^{\ell_n}_{a_t}(y)\right) = f(y) = f\left(R_{u^-(s)}(x)\right)$ as n goes to ∞ and for $x, y \in B$.

If $\varepsilon_1 < \varepsilon$, then we can choose a compact subset $K_1 \subseteq X$ with $m_X(K_1) > 1 - \varepsilon_1$ for which $f|_{K_1}$ is continuous. We may assume that $K \subseteq K_1$, since continuity of $f|_K$ and $f|_{K_1}$ implies continuity of $f|_{K\cup K_1}$. Then we define the set B_1 defined from K_1 just as we did for B, and thus $B \subseteq B_1$. By repeating this process, for example choosing $\epsilon = 1/n$ and letting n go to ∞, we will be able to conclude the existence of a set X' with $m_X(X') = 1$ with the property that for $x, y = R_{u^-(s)}(x) \in X'$ we have $f(x) = f(y)$. Applying the same argument to $R^{-1}_{a_t}$, that is defined as $R^{-1}_{a_t}(g) = ga_t$, we come to the same conclusion for points x and $y = R_{u^+(s)}(x)$ in X'', where X'' is an other set of full measure. Taking the intersection $X_1 = X' \cap X''$ we get a full measure set X_1.

Now our proof technique is to proceed by contradiction. Assume that $f : X \to \mathbb{R}$ is not constant almost everywhere with respect to m_X. Then there exist disjoint intervals $I_1, I_2 \subseteq \mathbb{R}$ for which

$$C_j = \{h \in \mathrm{PSL}(2,\mathbb{R}) \mid f(\Gamma h) \in I_j\} \text{ for } j = 1, 2$$

have measure $m_{\mathrm{PSL}(2,\mathbb{R})}(C_j) \in (0,1)$, since f is assumed to be non-constant. By Theorem 4.15,

$$\exists g \in \mathrm{PSL}(2,\mathbb{R}) \text{ such that } m_{\mathrm{PSL}(2,\mathbb{R})}\left(C_1 \cap C_2 g^{-1}\right) > 0. \tag{4.1}$$

Define now for the same g of (4.1) the set X_g in the following way:
for $g = u^+(s_4)\,u^-(s_3)\,u^+(s_2)\,u^-(s_1)$, with $s_1, s_2, s_3, s_4 \in \mathbb{R}$ (see Proposition 4.9),

$$\begin{aligned}X_g := X_1 &\cap R^{-1}_{u^-(s_1)}(X_1) \cap R^{-1}_{u^+(s_2)u^-(s_1)}(X_1)\\ &\cap R^{-1}_{u^-(s_3)u^+(s_2)u^-(s_1)}(X_1) \cap R^{-1}_g(X_1).\end{aligned}$$

X_g has full measure since it is the intersection of full measure sets, which is insured by invariance of m_X under $\mathrm{PSL}(2,\mathbb{R})$.
Now,

$$x \in X_1 \cap R^{-1}_{u^-(s_1)}(X_1) \Longrightarrow f(x) = f\left(R_{u^-(s_1)}(x)\right),$$
$$y = R_{u^-(s_1)}(x) \in X_1 \cap R^{-1}_{u^+(s_2)}(X_1) \Longrightarrow f(x) = f(y) = f\left(R_{u^+(s_2)}(y)\right),$$

and continuing in this way we prove

$$f(x) = f\left(R_g(x)\right)$$

for all $x \in X_g$. However, consider the set

$$D_g = \{h \in \mathrm{PSL}(2,\mathbb{R}) \mid \Gamma h \in X_g\}.$$

D_g is a full measure set with respect to $m_{\mathrm{PSL}(2,\mathbb{R})}$.
Indeed, $1 = m_X(X_g) = m_{\mathrm{PSL}(2,\mathbb{R})}(\pi^{-1}(X_g) \cap F) = m_{\mathrm{PSL}(2,\mathbb{R})}(D_g \cap F)$.
Therefore,

$$\exists h \in C_1 \cap C_2 g \cap D_g.$$

This generates a contradiction, since

$$f(\Gamma h) = f\left(\Gamma h g^{-1}\right)$$

by definition of D_g and X_g, but $f(\Gamma h) \in I_1$ and $f\left(\Gamma h g^{-1}\right) \in I_2$ by definition of C_1 and C_2. Thus we conclude that f is constant, and so a geodesic flow is an ergodic transformation on X with respect to m_X.

□

4.3 Ergodicity of the Gauss map

Now we want to prove the ergodicity of the Gauss map. To derive this result we apply the ergodicity of the geodesic flow proved in the previous section.

Definition 4.19 (Gauss Map)**.** Let $Y = (0,1)$. Then we define the Gauss map as $T : Y \to Y$ such that $T(y) = \left\{\frac{1}{y}\right\} = \frac{1}{y} - \lfloor\frac{1}{y}\rfloor$.

Remark 4.20. As previously said for any semi-circle geodesic, note that there are two uniquely associated positive real numbers $\tilde{y}$, y, i.e. the limit points of the flow on the real axis. We will write $-\tilde{y}$ and y (respectively $\tilde{y}$ and $-y$) for these points according to the orientation of the geodesic.

Definition 4.21 (C_+)**.**

$$C_+ = \{(\mathrm{i}b, \mathrm{v}) \in \mathrm{T}^1\mathbb{H} \mid \Re\mathrm{v} > 0, \Im\mathrm{v} < 0, y \in [0,1), \tilde{y} \geqslant 1\} \text{ (see Fig. 4.3)}.$$

Definition 4.22 (C_-)**.**

$$C_- = \{(\mathrm{i}b, \mathrm{v}) \in \mathrm{T}^1\mathbb{H} \mid \Re\mathrm{v} < 0, \Im\mathrm{v} < 0, y \in [0,1), \tilde{y} \geqslant 1\}.$$

We write $C = C_+ \cup C_-$.

For C_+ and C_- we give two coordinates, (y, z), where $z = \frac{1}{y+\bar{y}}$.

Lemma 4.23. *Let $x = (\mathrm{i}b, \mathrm{v})$ be in $C_+$with natural coordinates (y, z). The next visit, if there is one, of the geodesic flow to the set $\pi(C)$ occurs in $\pi(C_-)$ and has coordinates*

$$\bar{T}(y, z) = \left(\left\{ \frac{1}{y} \right\}, y(1 - yz) \right).$$

A similar property holds for points in C_-.

Proof. The isometry $z \mapsto -\frac{1}{z}$ sends the usual fundamental domain E to another fundamental domain (the geodesic triangle with all sides that are a part of a semi-circle inside the radius-one semi-circle with center at the origin). Note that the map σ always reverses the orientation of the geodesic. The image of the geodesic in Fig.4.3 under the isometry $\sigma : z \mapsto -\frac{1}{z}$ is shown in Fig. 4.4.

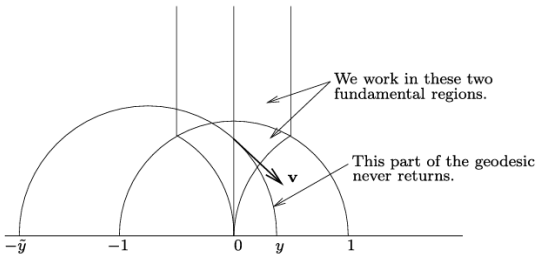

Figure 4.3: Fundamental region E and σE [7].

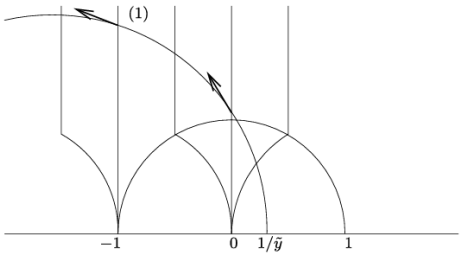

Figure 4.4: The first visit (1) to $\pi(D)$ is not in $\pi(C)$ [7].

We define $D := \left\{(z, \mathrm{v}) \in \mathrm{T}^1\mathbb{H} \mid \Re(z) = 0\right\}$ for the collection of arrows attached to the imaginary axis. The next return of the geodesic flow to $\pi(D)$, which is marked as (1) in Fig. 4.4, is not in $\pi(C)$ properties required for C_+ or for C_- are not satisfied. So we continue the geodesic flow to the next return to $\pi(D)$ and repeat this until the return is to $\pi(C)$. If we apply the isometry $\tau^n(x) = x + n$ for that unique n such that $-\frac{1}{y} + n \in (-1, 0]$ (see Fig. 4.5), then the intersection of the image of the geodesic with $\{z \mid \Re(z) = 0\}$ represents the first return to $\pi(C)$, which in this case is to $\pi(C_-)$ because of the orientation.

The new coordinates

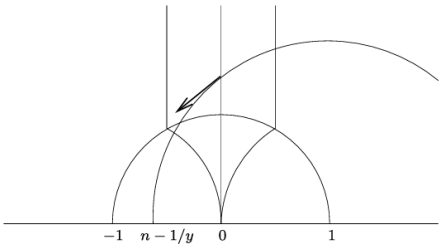

Figure 4.5: The return to $\pi(C)$ [7].

$$\left(\frac{1}{y} - n, \frac{1}{\frac{1}{\tilde{y}} + \frac{1}{y}}\right)$$

are as claimed, since

$$\frac{1}{\frac{1}{\tilde{y}} + \frac{1}{y}} = \frac{y\tilde{y}}{y + \tilde{y}} = y\left(1 - \frac{y}{y + \tilde{y}}\right) = y(1 - yz).$$

□

Definition 4.24 (Return time function)**.** For $x \in \pi(C)$, the function

$$r_C(x) = \min\{t \mid t > 0, R_{a_t}x \in \pi(C)\} \tag{4.2}$$

is called the *return time function* for the geodesic.

Definition 4.25 (Flow [7])**.** A flow is a family $\{T_t \mid t \in \mathbb{R}\}$ of measurable transformations of the probability space $(X, \mathcal{B}, \mu)$ satisfying the identity $T_sT_t = T_{s+t}$ for all $s, t \in \mathbb{R}$, and with $T_0 = e$. The flow is measure-preserving if T_t preserves μ for each $t \in \mathbb{R}$, and is measurable (as a flow) if the map $(x, t) \mapsto T_t(x)$ is a measurable map from $(X \times \mathbb{R}, \mathcal{B} \otimes \mathcal{B}_{\mathbb{R}})$ to $(X, \mathcal{B})$.

Now we introduce some useful theorems for the proof of the ergodicity of the Gauss map (Proposition 4.37) without proving them.

Theorem 4.26. *Let $(X, \mathcal{B}, \mu)$ be a Borel probability space, and $\mathcal{A} \subseteq \mathcal{B}$ a σ-algebra. Then there exists an $\mathcal{A}$-measurable full measure set $X' \subseteq X$ and a system $\{\mu_x^{\mathcal{A}} \mid x \in X'\}$ of measures on X, referred to as conditional measures, with the following property. $\mu_x^{\mathcal{A}}$ is a probability measure on X with*

$$E(f \mid \mathcal{A})(x) = \int f(y)\mathrm{d}\mu_x^{\mathcal{A}}(y)$$

almost everywhere for all $f \in \mathcal{L}^1(X, \mathcal{B}, \mu)$.

Theorem 4.27. *Let $(X, \mathcal{B}, \mu, T)$ be a measure-preserving system and $f \in L^1_\mu$. Then*

$$\frac{1}{M}\sum_{n=0}^{M-1} f \circ T^n \longrightarrow E(f \mid \mathcal{E}) \text{ for } M \to \infty$$

almost everywhere and in L^1, where $\mathcal{E} = \{B \in \mathcal{B} \mid \mu(T^{-1}(B) \triangle B) = 0\}$.

Lemma 4.28. *Let $T : (Y,\mu) \to (Y,\mu)$ be an invertible measure-preserving map on a probability space, and let $r : Y \to \mathbb{R}^+$ be a measurable function. Then a flow may be defined on the space*

$$X_r = \{(y,s) \mid y \in Y, 0 \leqslant s < r(y)\},$$

(with measure m_r defined by the restriction of $\mu \times m_{\mathbb{R}}$ to X_r, with $m_{\mathbb{R}}$ denoting the Lebesgue measure) by

$$T_t(y,s) = \begin{cases} (y, s+t) & \text{if } 0 \leqslant s+t < r(y), \\ (Ty, s+t-r(y)) & \text{if } 0 \leqslant s+t-r(y) < r(Ty), \\ \left(T^2y, s+t-r(y)-r(Ty)\right) & \text{if } 0 \leqslant s+t-r(y)-r(Ty) < r\left(T^2y\right) \\ \vdots \end{cases}$$

Proof. Define the map

$$\widetilde{T}(y,s) = (Ty, s - r(y))$$

on the space $Y \times \mathbb{R}$, which we equip with the infinite measure $\mu \times m_{\mathbb{R}}$. Notice that $\widetilde{T}$ preserves $\mu \times m_{\mathbb{R}}$. We check this using Lemma 4.11. Firstly,

$$\int_{Y\times\mathbb{R}} f(\widetilde{T}(y,s))\, \mathrm{d}m_r(y,s) = \int_Y \left(\int_{\mathbb{R}} f(Ty, s-r(y))\, \mathrm{d}s \right)\, \mathrm{d}\mu(y)$$

by Fubini's theorem. Then, remember that Lebesgue's integral is invariant under translation because the differential does not change applaing the *change of variables theorem*. Therefore now we have

$$\int_Y \left(\int_{\mathbb{R}} f(Ty, s-r(y))\, \mathrm{d}s \right)\, \mathrm{d}\mu(y) = \int_Y \left(\int_{\mathbb{R}} f(Ty, s)\, \mathrm{d}s \right)\, \mathrm{d}\mu(y).$$

Then again by Fubini's theorem we get

$$\int_Y \left(\int_{\mathbb{R}} f(Ty, s)\, \mathrm{d}s \right)\, \mathrm{d}\mu(y) = \int_{\mathbb{R}} \left(\int_Y f(Ty, s)\, \mathrm{d}\mu(y) \right)\, \mathrm{d}s.$$

Finally by $\mu-$invariance of T, we conclude

$$\int_{\mathbb{R}} \left(\int_Y f(Ty, s)\, \mathrm{d}\mu(y) \right)\, \mathrm{d}s = \int_{\mathbb{R}} \left(\int_Y f(y, s)\, \mathrm{d}\mu(y) \right)\, \mathrm{d}s = \int_{Y\times\mathbb{R}} f(y,s)\, \mathrm{d}m_r(y,s).$$

The inverse of $\widetilde{T}$ is

$$\widetilde{T}^{-1}(y,s) = \left(T^{-1}y, s + r\left(T^{-1}y\right)\right)$$

so $\widetilde{T}$ defines a $\mathbb{Z}$-action on $Y \times \mathbb{R}$ (which plays the role of the isometries in Γ). The action is

$$\mathbb{Z} \times (Y \times \mathbb{R}) \to (Y \times \mathbb{R})$$
$$(n, (y, s)) \mapsto \widetilde{T}^n(y, s).$$

Let's verify that it is actually an action following Definition 2.7. The first property is satisfied since taking the identity element $0 \in \mathbb{Z}$, we get $\widetilde{T}^0(y, s) = (y, s)$. The second point to verify is also satisfied: note indeed that

$$\begin{aligned}\widetilde{T}^{n+m}(y, s) &= \left(T^{n+m}s, s - \sum_{k=0}^{n+m-1} r\left(T^k y\right)\right) \\ &= \left(T^n(T^m s), s - \sum_{k=0}^{m-1} r\left(T^k y\right) - \sum_{k=m}^{m+n-1} r\left(T^k y\right)\right) \\ &= \widetilde{T}^n(\widetilde{T}^m(y, s)).\end{aligned}$$

Now we claim that $X_r = \{(y, s) \mid y \in Y, 0 \leqslant s < r(y)\}$ is a fundamental domain. It means that we have to prove the usual properties:

1. $\widetilde{T}^n \mathring{X}_r \cap \mathring{X}_r = \varnothing$ for all $n \neq 0$;
2. for $\mu \times m_{\mathbb{R}}$ a.e. $x = (y, s) \in Y \times \mathbb{R}$, there exists a unique $n \in \mathbb{Z}$ with the property that $\widetilde{T}^n x \in \mathring{X}_r$ (see Definition 3.14).

Let us prove the first property: $\widetilde{T}^n(y, s) = (T^n y, s - \sum_{i=0}^{n-1} r(T^i y))$, for $(y, s) \in \mathring{X}_r$. $T^n y$ belongs to Y, but $(s - \sum_{i=1}^{n-1} r(T^i y))$ is negative because (y, s) belongs to $\mathring{X}_r$, so then $(T^n y, s - \sum_{i=1}^{n} r(T^i y))$ does not belong to $\mathring{X}_r$.

The second property is more complicated to prove: let $\mathcal{E} \subseteq \mathcal{B}_Y$ be the σ-algebra of T-invariant sets; then by Theorems 4.26 and 4.27, we have

$$E(r \mid \mathcal{E})(y) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} r\left(T^k y\right) = \int r \, \mathrm{d}\mu_y^{\mathcal{E}} > 0$$

almost everywhere. Therefore, for almost every $y \in Y$ we have $\sum_{k=0}^{n-1} r\left(T^k y\right) \nearrow \infty$. So if $(y, s) \in Y \times \mathbb{R}$ has $s > 0$ then there exists some n for which

$$s - \sum_{k=0}^{n} r\left(T^k y\right) < 0$$

but

$$s - \sum_{k=0}^{n-1} r\left(T^k y\right) \geqslant 0.$$

Then by definition of X_r we get

$$\widetilde{T}^n(y,s) = \left(T^n s, s - \sum_{k=0}^{n-1} r\left(T^k y\right)\right) \in X_r \text{ since } 0 \leq s - \sum_{k=0}^{n-1} r\left(T^k y\right) < r(T^n y).$$

For later $k > n-1$, $s - \sum_{k=0}^{k} r\left(T^k y\right) \leq 0$, so we have the uniqueness of the n. The case $s < 0$ is similar. We reach that X_r is a fundamental domain. We also claim that $\widetilde{T}$ commutes with our flow $T_t(y,s) = (y, s+t)$. Indeed,

$$T_t \circ \widetilde{T}(y,s) = T_t(Ty, s - r(y)) = (Ty, s + t - r(y)) = \widetilde{T}(y, s+t) = \widetilde{T} \circ T_t(y,s).$$

Now from the definition of fundamental domain, we know that $\widetilde{T}(X_r)$ is also a fundamental domain, but also $T_t(X_r)$ is another fundamental domain since by definition T_t is defined only if $0 \leqslant s + t < r(y)$. Moreover, $T_t(X_r)$ can be decomposed into pieces for which a power of $\widetilde{T}$ brings them back to X_r. The definition of T_t on X_r precisely describes this process. Note that also T_t preserve the measure $\mu \times m_{\mathbb{R}}$ for the same argument of $\widetilde{T}$ presented above and thus since both preserve it on $Y \times \mathbb{R}$, the action on X_r preserves the restriction of the measure. So by Definition 4.25, we finally conclude that T_t is a measure preserving. □

Let's state a useful proposition for later.

Proposition 4.29. *Let $(X, \mathcal{B}, \mu, T)$ a measure preserving system and let G be a σ-compact metric group acting continuously on a compact metric space X preserving a measure $\mu \in \mathcal{M}(X)$ (space of measure on X). Then for $B \in \mathcal{B}$ the following properties are equivalent:*

1. *B is invariant in the sense that $\mu(gB \triangle B) = 0$ for all $g \in G$;*
2. *B is invariant in the sense that there is a set $B' \in \mathcal{B}$ with $\mu(B \triangle B') = 0$ and with $g \cdot B' = B'$ for all $g \in G$.*

Lemma 4.30. *For the special flow in Lemma 4.28,*

1. *the measure m_r is finite if and only if r is integrable,*
2. *assuming m_r is finite, the map T is ergodic if and only if the flow $\{T_t\}$ is ergodic.*

In the space $X_r = \{(y,s) \mid y \in Y, 0 \leqslant s < r(y)\}$, we call a set of the form

$$F_y = \{(y,s) \mid 0 \leqslant s < r(y)\}$$

a (vertical) fiber (see Fig. 4.6).

Proof. Firstly, let's prove point one. It follows immediately from the definition of the measure in Lemma 4.28, again using Fubini,

$$m_r(X_r) = \int_{X_r} \mathrm{d}m_r = \int_Y \int_0^{r(y)} \mathrm{d}s \, \mathrm{d}\mu(y) = \int_Y r(y) \mathrm{d}\mu(y).$$

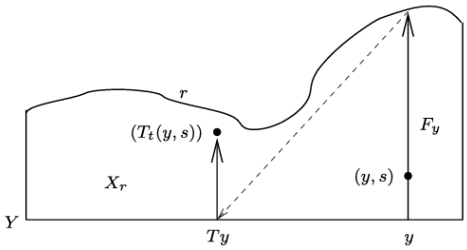

Figure 4.6: Flow under ceiling function r [7].

and $m_r(X_r)$ is finite if and only if $r(y) \in L^1_\mu$, i.e. $\int_Y |r(y)| \, \mathrm{d}\mu(y) = \int_Y r(y) \, \mathrm{d}\mu(y) < \infty$.

Now we want to prove the second statement. Let's assume that the map T is ergodic. Assume that for a measurable function f on X_r we have $f \circ T_t = f$ a.e. for all $t \in \mathbb{R}$. Then, by Proposition 4.13, if we prove that f is constant, we prove that T_t is ergodic. We claim that f can be modified on a null set to ensure that $f \circ T_t = f$ everywhere for all $t \in \mathbb{R}$. The claim can be proved using Proposition 4.29. In particular, $\bar{f}(y) = f((y,0))$ defines a T-invariant function on Y which must be constant μ-almost everywhere by ergodicity of T. This implies that f is m_r-almost everywhere constant, and so the flow is also ergodic.

Assume now that T_t is ergodic for m_r, and let f be a measurable function such that $f = f \circ T$. Define $\bar{f}(y,s) = f(y)$. We obtain a measurable function on X_r that is strictly invariant under the flow. Indeed, $\bar{f} \circ T_t(y,s) = f \circ T^n(y) = f(y) = \bar{f}(y,s)$, for some $n \in \mathbb{N}$. Since T_t is assumed to be ergodic, using Proposition 4.13 again, we conclude that $\bar{f}$ is constant a.e., which implies that f is constant a.e, which implies that T is ergordic. □

Let's again state known general definitions and results that will be useful tools later for the proof of Proposition 4.37.

Definition 4.31 (Topologically conjugate [7])**.** Let $T : X \to X$ and $S : Y \to Y$ be continuous maps of compact metric spaces (that is, topological dynamical systems). Then a homeomorphism $\theta : X \to Y$ with $\theta \circ T = S \circ \theta$ is called a topological conjugacy, and if there is such a conjugacy then T and S are topologically conjugate.

Definition 4.32 (Absolutely continuous)**.** The measure ν is absolutely continuous with respect to μ, written $\nu \ll \mu$, if $\mu(A) = 0 \implies \nu(A) = 0$ for any $A \in \mathcal{B}$. If $\nu \ll \mu$ and $\mu \ll \nu$, then μ and ν are said to be equivalent.

Theorem 4.33 (Radon-Nikodym derivative [7])**.** *For a σ-finite measure space $(X, \mathcal{B}, \mu)$, if $\nu \ll \mu$ then there is a measurable function $f \geqslant 0$ on X with the property that*

$$\nu(A) = \int_A f \, \mathrm{d}\mu$$

for any set $A \in \mathcal{B}$.

Lemma 4.34. *Let $O \subseteq \mathbb{R}^n$ be open, let $\phi : O \to \mathbb{R}^n$ be a smooth injective map with non-vanishing Jacobian $J_\phi = |\det \mathrm{D}\phi|$. Then the pushforward $\phi_* m_O$ of the Lebesgue measure $m_O = m_{\mathbb{R}^n}|_O$ restricted to O is absolutely continuous with respect to $m_{\mathbb{R}^n}$ and is given by*

$$\mathrm{d}\phi_* m_O = J_\phi^{-1} \circ \phi^{-1}\, \mathrm{d}m_{\phi(O)}.$$

Moreover, if we consider a measure $\mathrm{d}\mu = F\, \mathrm{d}m_O$ absolutely continuous with respect to m_O, then similarly

$$\mathrm{d}\phi_*\mu = F \circ \phi^{-1} J_\phi^{-1} \circ \phi^{-1}\, \mathrm{d}m_{\phi(O)}.$$

Proposition 4.35. *(Push forward change of variables formula [7])*
Let $\mathcal{M}(X) = \left\{\mu \in C(X) \mid \int 1\, \mathrm{d}\mu = 1, \int f\, \mathrm{d}\mu \geqslant 0 \text{ for } f \in C(X) \text{ with } f \geqslant 0\right\}$, with $C(X)$ that denotes the space of continuous maps from X to $\mathbb{R}$, and let a continuous map $T : X \to X$ induces a map $T_* : \mathcal{M}(X) \to \mathcal{M}(X)$ defined by $T_*(\mu)(A) = \mu\left(T^{-1}A\right)$ for any Borel set $A \subseteq X$. For $f \geqslant 0$ a measurable map and $\mu \in \mathcal{M}(X)$,

$$\int f\, \mathrm{d}T_*\mu = \int f \circ T\, \mathrm{d}\mu.$$

Lemma 4.36 (Invariance of Gauss measure for the Gauss map [7])**.** *The Gauss map $T(x) = \left\{\frac{1}{x}\right\}$ on $(0,1)$ preserves the Gauss measure μ given by*

$$\mu(A) = \frac{1}{\log 2} \int_A \frac{1}{1+x}\, \mathrm{d}x$$

for any Borel measurable set $A \subseteq [0,1]$.

Proof. We need to show that $\mu\left(T^{-1}[a,b]\right) = \mu([a,b])$ for every $a, b > 0$. Firstly note the following:

$$\frac{1}{k+1} < x < \frac{1}{k} \Leftrightarrow k < \frac{1}{x} < k+1,\ , T(x) = \frac{1}{x} - k, \text{ and we have } \frac{1}{x} - k = a \Leftrightarrow x = \frac{1}{a+k}.$$

Clearly,

$$T^{-1}([a,b]) = \{x \mid a \leqslant T(x) \leqslant b\} = \bigsqcup_{n=1}^{\infty} \left[\frac{1}{b+n}, \frac{1}{a+n}\right]$$

is a disjoint union. It follows that

$$\begin{aligned}
\mu\left(T^{-1}[a,b]\right) &= \frac{1}{\log 2} \sum_{n=1}^{\infty} \int_{1/(b+n)}^{1/(a+n)} \frac{1}{1+x}\, \mathrm{d}x \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \left(\log\left(1 + \frac{1}{a+n}\right) - \log\left(1 + \frac{1}{b+n}\right)\right) \\
&= \frac{1}{\log 2} \sum_{n=1}^{\infty} \left(\log(a+n+1) - \log(a+n) - (b+n+1) + \log(b+n)\right),
\end{aligned}$$

but

$$\sum_{n=1}^{N} \left(\log(a+n+1) - \log(a+n)\right) = \log(a+N+1) - \log(a+1),$$

so

$$\begin{aligned}
&= \frac{1}{\log 2}\sum_{n=1}^{\infty}\left(\log\left(a+n+1\right)-\log\left(a+n\right)-\left(b+n+1\right)+\log\left(b+n\right)\right)\\
&= \frac{1}{\log 2}\lim_{N\to\infty}\left(\log\left(b+1\right)-\log\left(a+1\right)+\log\left(\frac{a+N+1}{b+N+1}\right)\right)\\
&= \frac{1}{\log 2}\left(\log\left(b+1\right)-\log\left(a+1\right)\right)\\
&= \mu([a,b]).
\end{aligned}$$

□

After a lot of preliminary work, we can finally state and prove the ergodicity of the Gauss Map.

Proposition 4.37. *(Ergodicity of the Gauss Map) The Gauss map $T(y)=\left\{\frac{1}{y}\right\}$ on $(0,1)=Y$ is ergodic with respect to the Gauss measure $\mathrm{d}\mu=\frac{1}{\log 2}\frac{1}{1+x}\,\mathrm{d}x$. Moreover the return time function r_C is integrable.*

Proof. We will prove ergodicity for the invertible extension of the Gauss map, i.e. the map $\bar{T}:\bar{Y}\to\bar{Y}$ given by

$$\bar{T}(y,z)=(Ty,y(1-yz))$$

on the set

$$\bar{Y}=\left\{(y,z)\in Y\times[0,1)\,\middle|\,0\leqslant z\leqslant\frac{1}{1+y}\right\}.$$

From page 92 of Ward and Einsiedler's book [7], we know that $\bar{T}$ preserves Lebesgue's measure.

This will prove the ergodicity of the Gauss map in the following way: we prove ergodicity again using the last point of Proposition 4.13, so we need that if f is an invariant measurable function for T on Y, then f is constant a.e. . Note that if f is an invariant measurable function for T on Y, $\bar{f}(y,z)=f(y)$ is an invariant function for $\bar{T}$ on $\bar{Y}$, indeed $\bar{f}\circ\bar{T}(y,z)=f\circ T(y)=f(y)=\bar{f}(y,z)$. We also have to pay attention to the sign of the cross section C defined before (see Definition 4.21), therefore define $\widetilde{Y}=\bar{Y}\times\{\pm1\}$ and the transformation $\widetilde{T}:\widetilde{Y}\to\widetilde{Y}$ by

$$\widetilde{T}(y,z,\varepsilon)=(\bar{T}(y,z),-\varepsilon),\ \epsilon\in\{-1,1\}.$$

We write $r_C(y,z)=\min\{t\mid t>0,R_{a_t}(y,z)\in\pi(C)\}$ for the return time of a point (see Definition 4.24) in $\pi(C)$ with coordinates (y,z,ε).
We claim that the special flow $\widetilde{T}$ on $X_{r_C}=\{(y,s)\mid y\in Y,0\leqslant s<r_C(y)\}$ (see Lemma 4.28) under r_C can be embedded into

$$\mathrm{T}^1M=\mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}(2,\mathbb{R}).$$

This means that there is an homeomorphism ϕ between the two spaces, where the image of this homeomorphism matches with T^1M. Indeed, for $(y,z,\varepsilon,s)\in\bar{Y}\times\{\pm1\}\times\mathbb{R}$ satisfying $0\leqslant s<r_C(y,z)$, we define

$$\phi(y,z,\varepsilon,s)=R_{a_s}(x(y,z,\varepsilon))$$

with $x(y,z,\varepsilon)\in\pi(C)$. Remember from Corollary 2.16 that $\mathrm{T}^1\mathbb{H}\cong\mathrm{PSL}(2,\mathbb{R})$ and since $C\subseteq\mathrm{T}^1\mathbb{H}$ the definition of ϕ is coherent. Continuity of R_{a_s} and $R_{a_{-s}}$ are satisfied. We still need to show that ϕ is bijective. From the definition of $r_C(y,z)$, it is clear that $\phi(y,z,\varepsilon,s)\notin\pi(C)$ unless $s=0$, since $s<r_C(y,z)$. We claim that $\phi:X_{r_C}\to\mathrm{T}^1M$ is injective. Let's prove it. Suppose

$$w=\phi\left(y_1,z_1,\varepsilon_1,s_1\right)=\phi\left(y_2,z_2,\varepsilon_2,s_2\right).$$

Then $s_1=\min\left\{s\mid R_{a_{-s}}(w)\in\pi(C)\right\}=s_2$ since r_C is defined to be the time of the first return from $\pi(C)$ to $\pi(C)$. This implies $(y_1,z_1,\varepsilon_1)=(y_2,z_2,\varepsilon_2)$ are the natural coordinates of the point $R_{a_{-s_1}}(w)\in\pi(C)$, thus ϕ is injective. We claim now that $\phi\left(T_t(y,z,\varepsilon,s)\right)=R_{a_t}(\phi(y,z,\varepsilon,s))$.
The claim follows from Lemma 4.28. In particular, if $0\leqslant s+t<r_C(y,z)$, then

$$\phi\left(T_t(y,z,\varepsilon,s)\right)=\phi(y,z,\varepsilon,s+t)=R_{a_{s+t}}(x(y,z,\varepsilon))=R_{a_t}(\phi(y,z,\varepsilon,s)),\tag{4.3}$$

from the definition of R_{a_t}.
If $r_C(y,z)\leqslant s+t<r_C(\bar{T}(y,z))$, then

$$\phi\left(T_t(y,z,\varepsilon,s)\right)=\phi\left(\bar{T}(y,z),-\varepsilon,s+t-r_C(y,z)\right)=R_{a_{s+t-r_C(y,z)}}(x(\bar{T}(y,z),-\varepsilon)).$$

However, by Lemma 4.23, in particular by the fact that return times alternate to be in $\pi(C_-)$ or $\pi(C_+)$ and by the definition of r_C as time of first return to $\pi(C)$, we also have

$$R_{a_{r_C(y,z)}}(x(y,z,\varepsilon))=x(\bar{T}(y,z),-\varepsilon).$$

Thus,

$$R_{a_{-r_C(y,z)}}((\bar{T}(y,z),-\varepsilon))=x(y,z,\varepsilon).$$

This gives

$$\phi\left(T_t(y,z,\varepsilon,s)\right)=R_{a_{s+t}}(x(y,z,\varepsilon))=R_{a_t}(\phi(y,z,\varepsilon,s)).$$

Let us also do the next case.
If $r_C(y,z)+r_C(\bar{T}(y,z))\leqslant s+t<r_C(\bar{T}^2(y,z))$, then

$$\begin{aligned}\phi\left(T_t(y,z,\varepsilon,s)\right)&=\phi\left(\bar{T}^2(y,z),\varepsilon,s+t-r_c(y,z)-r_c(\bar{T}(y,z)\right)\\&=R_{a_{s+t-r_C(y,z)-r_c(\bar{T}(y,z)}}(x(\bar{T}^2(y,z),\varepsilon)).\end{aligned}$$

And for the same reason as above, in particular because of

$$R_{a_{-r_C(\bar{T}(y,z))}}((\bar{T}^2(y,z),\varepsilon)) = x(\bar{T}(y,z),-\varepsilon),$$

we conclude

$$\phi\left(T_t(y,z,\varepsilon,s)\right) = R_{a_{s+t}}(x(y,z,\varepsilon)) = R_{a_t}(\phi(y,z,\varepsilon,s)).$$

Other cases follow the same proof of the first cases.

We claim now that $\phi_*\left(m_{r_C}\right) = cm_{\mathrm{T}^1M}$ for some finite $c > 0$. The main idea of the proof is that smoothness of ϕ gives absolute continuity of $\phi_* m_{r_C}$ with respect to the measure m_{T^1M}, and then the ergodicity of the geodesic flow (Theorem 4.18) shows that the Radon-Nikodym derivative (Theorem 4.33) $\frac{\mathrm{d}\phi_*\left(m_{r_C}\right)}{\mathrm{d}m_{\mathrm{T}^1M}} = c$ is constant almost everywhere. Let us focus firstly on absolutely continuity of $\phi_* m_{r_C} = m_{r_C} \circ \phi^{-1}$ with respect to m_{T^1M}. We need absolutely continuity in order to use Radon-Nikodym Theorem 4.33. We want to apply Lemma 4.34. We already proved that ϕ is injective. We still need to check smoothness and non-vanishing Jacobian. Consider $(y,z,+1) \in \tilde{Y}$ corresponding to points in $\pi\left(C_+\right)$. We also assume also that $y > 0$ and $z > 0$, which leads to the endpoints y and $-\widetilde{y} = -\frac{1}{z} + y$ of the geodesic that defines the point $x(y,z,+1) = (\mathrm{i}b, \exp(\mathrm{i}\theta)) \in C_+$, where $\exp(\mathrm{i}\theta)$ is the euler form of the unit vector v that represents the second component of $x(y,z,+1) \in C_+$.
Consider the map

$$\psi : \begin{pmatrix} y \\ z \end{pmatrix} \xrightarrow{\mathrm{f}} \begin{pmatrix} y \\ \widetilde{y} \end{pmatrix} \xrightarrow{\mathrm{g}} \begin{pmatrix} b \\ -\pi/2 + \theta \end{pmatrix} \in \mathbb{R}^+ \times \left(-\frac{\pi}{2}, 0\right).$$

Note that $\theta \in (0, \pi/2)$ since we are in $\pi(C_+)$, and thus $\theta = \arcsin(\frac{b}{\text{radius}})$. We want to check that ψ is smooth and with non-vanishing Jacobian. By the chain rule, the Jacobian of ψ is $J_\psi(y,z) = J_{f\circ g}(y,z) = J_f(g(y,z))J_g((y,z))$.
We have that

$$f(y,z) = (y, \frac{1}{z} - y)$$

Therefore,

$$J_f = \left| det\left(\begin{bmatrix} 1 & 0 \\ -1 & \frac{1}{z^2} \end{bmatrix}\right)\right| = \frac{1}{z^2}.$$

Moreover, the equation C of the geodesic, which is a half circle (for $b > 0$) of radius $(\frac{y+\widetilde{y}}{2})$, is given by

$$C(a,b) : (\frac{y+\widetilde{y}}{2})^2 = (a - (\frac{y-\widetilde{y}}{2}))^2 + b^2$$

And since we are interested in having b, we get for $x = 0$

$$b = \sqrt{2}\sqrt{y\widetilde{y}}.$$

So we can finally define

$$g(y,\widetilde{y}) = \left(\sqrt{2}\sqrt{y\widetilde{y}}, -\pi/2 + \arcsin\left(\frac{\sqrt{2}\sqrt{y\widetilde{y}}}{(\frac{y+\widetilde{y}}{2})}\right)\right).$$

We have $g_y = \sqrt{\frac{1}{2y}}\sqrt{\widetilde{y}}$ and $g_{\widetilde{y}} = \sqrt{\frac{1}{2\widetilde{y}}}\sqrt{y}$. From the fact that $y > 0, z > 0$ and $\widetilde{y} > 0$, and since $(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}$ and our $\arcsin$ is in $(0, \pi/2)$, we deduce easily that $J_\psi > 0$ and that ψ is smooth.

Consider now

$$\phi : \begin{pmatrix} y \\ z \\ s \end{pmatrix} \longmapsto \begin{pmatrix} a' \\ b' \\ \theta' \end{pmatrix} \in \mathbb{R} \times \mathbb{R}_+ \times \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \tag{4.4}$$

$$\phi(y,z,s) = R_{a_s}(x(y,z))$$

for some $y > 0, z > 0$ and $s = 0$. As the application of R_{a_s} is smooth, this Jacobian exists. From the structure of the map, the Jacobian of (4.4) at $s = 0$ is of the form

$$\begin{bmatrix} 0 & 0 & * \\ * & * & * \\ * & * & * \end{bmatrix}.$$

However, as the partial derivative of a' with respect to s is non-zero (since $\theta' \neq -\frac{\pi}{2}$), it follows that the Jacobian is non-zero for $s = 0$. Moreover, the application of R_{a_s} is smooth on $\mathrm{T}^1\mathbb{H}$ with smooth inverse $R_{a_{-s}}$. Therefore, by Lemma 4.34, we have that $\phi_* m_{r_C}$ is absolutely continuous with respect to $\mathrm{d}a\,\mathrm{d}b\,\mathrm{d}\theta$, and hence with respect to $\mathrm{d}m_{\mathrm{T}^1M}$ by Lemma 3.18. It follows that

$$\frac{\mathrm{d}\left(\phi_* m_{r_C}\right)}{\mathrm{d}m_{\mathrm{T}^1M}} = F \tag{4.5}$$

where $F \geqslant 0$ is the Radon-Nikodym derivative. However, as m_{r_C} is invariant under the special flow by construction and $R_{a_t} \circ \phi = \phi \circ T_t$, using the definition of push-forward (see Prop. 4.35) we get

$$\left(R_{a_t}\right)_* \phi_* m_{r_C} = m_{r_C} \circ \phi^{-1} \circ R_{a_t}^{-1} = m_{r_C} \circ \left(R_{a_t} \circ \phi\right)^{-1}.$$

Then by conjugacy property of ϕ (see eq. 4.3)

$$m_{r_C} \circ \left(R_{a_t} \circ \phi\right)^{-1} = \left(R_{a_t}\phi\right)_* m_{r_C} = \left(\phi T_t\right)_* m_{r_C}.$$

And by invariance of ϕ under the special flow T_t, we conclude

$$\left(\phi T_t\right)_* m_{r_C} = \phi_* m_{r_C},$$

showing invariance of $\phi_* m_{r_C}$ under the geodesic flow. Applying this to (4.5) and applying Radon-Nikodin Theorem (Theorem 4.33), we get for any measurable $f \geqslant 0$,

$$(\phi_* m_{r_C})(f) = (R_{a_t})_* \phi_* m_{r_C}(f) = \int f \circ R_{a_t} \mathrm{d}(\phi_* m_{r_C}) = \int f \circ R_{a_t} F \, \mathrm{d}m_{\mathrm{T}^1 M}.$$

The second inequality comes from pushforward change of variables formula (Proposition 4.35). However, by R_{a_t}-invariance of $m_{\mathrm{T}^1 M}$ (see Proposition 3.19), we also have

$$\int fF \circ R_{a_t} \, \mathrm{d}m_{\mathrm{T}^1 M} = \int fF \, \mathrm{d}m_{\mathrm{T}^1 M} = (\phi_* m_{r_C})(f).$$

Therefore we have

$$\int f \circ R_{a_t} F \, \mathrm{d}m_{\mathrm{T}^1 M} = \int fF \, \mathrm{d}m_{\mathrm{T}^1 M} = \int fF \circ R_{a_t} \, \mathrm{d}m_{\mathrm{T}^1 M}.$$

This implies that $F = F \circ R_{a_t}$ almost everywhere for all $t \in \mathbb{R}$, since the second equality holds for every $f \geq 0$. Ergodicity of the geodesic flow (Theorem 4.18) shows that $F = c$ almost everywhere for some constant c. It remains to check that $c > 0$ is finite. This follows from smoothness, which makes it easy to find an open set of finite measure in the special flow (viewed as a subset of $\mathbb{R}^3$) that is sent onto an open set in $\mathrm{T}^1 M$.
It is interesting to see what follows from this claim. It follows that m_{r_C} is a finite measure (and so r_C is integrable by Lemma 4.30), which may be normalized to be a probability measure. Moreover, since ϕ is a homeomorphism, ϕ is a conjugacy between the special flow of $\widetilde{T}$ under r_C and the geodesic flow. The geodesic flow is ergodic by Theorem 4.18, therefore because of the conjugacy, the special flow of $\widetilde{T}$ under r_C is also ergodic. By Lemma 4.30, $\widetilde{T}$ is then ergodic. This gives ergodicity of $\bar{T}$, since an invariant function for $\widetilde{T}$ is an invariant function for $\bar{T}$, following the same reasoning for T and $\bar{T}$ at the beginning of the proof. So we finally conclude that T is ergodic with respect to the Gauss measure. □

4.4 Gauss–Kuzmin Theorem and other Applications

Now that we proved ergodicity of the Gauss Map with respect to the Gauss measure, we can use it to prove a nice application regarding *continued fractions*. More precisely about the limit for $n \to \infty$ of the frequency of a positive natural number in the continued fraction of an irrational number $x \in (0,1)$ (Theorem 4.42). The main reference is chapter 3 from the book *Ergodic Theory: with a view towards Number Theory* [7]. Let's start by introducing what a continued fraction is:

Definition 4.38. For every real number $x \in \mathbb{R}$, we can express x through its *continued fraction*, namely

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \frac{1}{a_3 + \dots}}},$$

where $a_0 \in \mathbb{N}$ and $a_i \in \mathbb{N}$ and for $i \geq 1$. The notation for a continued fraction is $x = [a_0; a_1, a_2, a_3, \dots]$. If $x \in [0,1), x = [a_0, a_1, a_2, a_3, \dots]$.

Proposition 4.39. *For $x \in (0,1)$ such that $x = [a_1(x), a_2(x), \ldots]$, then $T^n(x) = [a_{n+1}(x), a_{n+2}(x), \ldots]$.*

Proof. The proof is really straightforward by induction: as basecase let's take $n = 1$ and call $x' = [a_2(x), a_3(x), \ldots]$, then $x = \frac{1}{a_1 + x'}$ and thus $T(x) = \{a_1 + x'\} = x'$. Assume now that it holds for n. We want to prove it for $n+1$.

$$\begin{aligned} T^{n+1}(x) &= T \circ T^n(x) \\ &= T([a_{n+1}(x), a_{n+2}(x), \ldots]) \\ &= \left\{ \frac{1}{[a_{n+1}(x), a_{n+2}(x), \ldots]} \right\} \\ &= \{a_{n+1} + [a_{n+2}(x), a_{n+3}(x), \ldots]\} = [a_{n+2}(x), a_{n+3}(x), \ldots]\,. \end{aligned}$$

□

Corollary 4.40 (Continued Fraction and the Gauss Map [7])**.** *For any irrational $x \in [0,1]\backslash\mathbb{Q}$ the sequence $(a_n(x))_{n\geq 1} = \left\lfloor \frac{1}{T^{n-1}x} \right\rfloor \in \mathbb{N}_{>0}$ gives the digits of the continued fraction expansion to $x = [a_1(x), a_2(x), \ldots]$. T is the Gauss map that we already introduced, i.e. for $y \in (0,1)$, $T(y) = \left\{ \frac{1}{y} \right\} = \frac{1}{y} - \left\lfloor \frac{1}{y} \right\rfloor$.*

Proof. The proof comes directly from the previous proposition and from the definition of the Gauss map. □

Definition 4.41 (Frequency of digits)**.** Let's consider $k \in \mathbb{N}_{>0} = \{1, 2, \ldots\}$. Then we naturally define the frequency of k in $x = [a_1, a_2, \ldots] \in [0,1] \setminus \mathbb{Q}$ as

$$D_N(x,k) := \frac{\text{card}(i \in \{1,2,\ldots,N\} \text{ such that } a_i = k)}{N}.$$

Now we state and prove the famous Gauss-Kuzmin Theorem.

Theorem 4.42 (Gauss-Kuzmin Theorem)**.** *For almost every real number $x = [a_1, a_2, \ldots] \in (0,1)$, i.e. for all $x \in [0,1] \setminus \mathbb{Q}$, the positive natural number k appears in the continued fraction with frequency*

$$\lim_{N\to\infty} D_N(x,k) = \frac{1}{\log 2} \log\left(\frac{(k+1)^2}{k(k+2)} \right).$$

Proof. The digit k appears in the first N digits with frequency

$$\begin{aligned} D_N(x,k) &= \frac{\text{card}(i \in \{1,2,\ldots,N\} \text{ such that } a_i = k)}{N} \\ &= \frac{1}{N}\text{card}\left(i \in \{0,1,2,\ldots,N-1\} \text{ such that } T^i(x) \in \left(\frac{1}{k+1}, \frac{1}{k} \right) \right) \\ &= \sum_{i=0}^{N-1} \mathbf{1}(T^i(x))_{(\frac{1}{k+1}, \frac{1}{k})}. \end{aligned}$$

By Pointwise Ergodic Theorem (see Theorem 4.16), since T is ergodic with respect to the Gauss measure and $\mathbf{1}(x)_{\left(\frac{1}{k+1},\frac{1}{k}\right)} \in L^1_\mu$, we get

$$\lim_{N\to\infty}\sum_{i=0}^{N-1}\mathbf{1}(T^i(x))_{\left(\frac{1}{k+1},\frac{1}{k}\right)} = \frac{1}{\log 2}\int_{\frac{1}{(k+1)}}^{\frac{1}{k}}\frac{1}{1+x}\,\mathrm{d}x = \frac{1}{\log 2}\log\left(\frac{(k+1)^2}{k(k+2)}\right).$$

□

Theorem 4.43. *We can also derive the following results: for almost every real number $x = [a_1, a_2, \ldots] \in (0,1)$, i.e. for all $x \in [0,1]\setminus\mathbb{Q}$,*

1. $\lim_{n\to\infty}(a_1 a_2 \ldots a_n)^{\frac{1}{n}} = \prod_{k=1}^{\infty}\left(\frac{(k+1)^2}{k(k+2)}\right)^{\frac{\log k}{\log 2}}$.

2. $\lim_{n\to\infty}\frac{1}{n}(a_1 + a_2 + \cdots + a_n) = \infty$.

Proof. Let's prove the first result: define a function f on $(0,1)$ by $f(x) = \log(a)\mathbf{1}(x)_{\left(\frac{1}{a+1},\frac{1}{a}\right)}$. So for $x = [a_1, a_2, \ldots]$, $f(x) = \log(a_1)$. Then

$$\begin{aligned}\int_0^1 |f(x)|\mathrm{d}x &= \int_0^1 f(x)\mathrm{d}x\\ &= \sum_{a=1}^{\infty}\left(\frac{1}{a} - \frac{1}{a+1}\right)\log a\\ &\leqslant \sum_{a=1}^{\infty}\frac{1}{a^2}\log a < \infty \text{ since } \log(a) < \sqrt{a}.\end{aligned}$$

So $\int_0^1 f\,\mathrm{d}\mu < \infty$ also, since the density $\frac{\mathrm{d}\mu}{\mathrm{d}x} = \frac{1}{(1+x)\log 2}$ is bounded on $[0,1]$ and because the Gauss measure μ and the Lebesgue measure are equivalent on $[0,1]$ (see Definition 4.32). By the Pointwise Ergodic Theorem (Theorem 4.16) we therefore have, for almost every x,

$$\frac{1}{n}\sum_{j=1}^{n}\log a_j = \frac{1}{n}\sum_{j=0}^{n-1} f\left(T^j x\right) \longrightarrow \int_0^1 f(x)\mathrm{d}\mu.$$

Now,

$$\int_0^1 f\,\mathrm{d}\mu = \sum_{a=1}^{\infty}\frac{\log a}{\log 2}\int_{\frac{1}{1+a}}^{\frac{1}{a}}\frac{1}{1+x}\,\mathrm{d}x = \sum_{a=1}^{\infty}\frac{\log a}{\log 2}\log\left(\frac{(a+1)^2}{a(a+2)}\right).$$

Now take exponential on both sides, and we get for $n \to \infty$

$$\frac{1}{n}\sum_{j=1}^{n}\log a_j \to \sum_{k=1}^{\infty}\frac{\log k}{\log 2}\log\left(\frac{(k+1)^2}{k(k+2)}\right)$$

$$\exp\left(\frac{1}{n}\sum_{j=1}^{n}\log a_j\right) \to \exp\left(\sum_{k=1}^{\infty}\log\left(\left(\frac{(k+1)^2}{k(k+2)}\right)^{\frac{\log k}{\log 2}}\right)\right)$$

$$\exp\left(\sum_{j=1}^{n}\log a_j^{\frac{1}{n}}\right) \to \exp\left(\sum_{k=1}^{\infty}\log\left(\left(\frac{(k+1)^2}{k(k+2)}\right)^{\frac{\log k}{\log 2}}\right)\right)$$

$$(a_1 a_2 \dots a_n)^{\frac{1}{n}} \to \prod_{k=1}^{\infty}\left(\frac{(k+1)^2}{k(k+2)}\right)^{\frac{\log k}{\log 2}}.$$

So we proved (1). Now we have to prove (2). Let us define the function g such that for $x = [a_1, a_2, \dots, a_n], g(x) = a_1 = a\mathbf{1}(x)_{\left(\frac{1}{a+1},\frac{1}{a}\right)}$. So we get

$$\frac{1}{n}\left(a_1 + \dots + a_n\right) = \frac{1}{n}\sum_{j=0}^{n-1} g\left(T^j x\right) = \frac{1}{n}\sum_{i=0}^{n-1}\sum_{k=1}^{\infty} k\mathbf{1}(T^i(x))_{\left(\frac{1}{k+1},\frac{1}{k}\right)}.$$

Unfortunately, the Pointwise Ergodic Theorem cannot be applied to g, since

$$\begin{aligned}\int_0^1 |g(x)|\mathrm{d}x &= \int_0^1 g(x)\mathrm{d}x \\ &= \sum_{a=1}^{\infty}\left(\frac{1}{a} - \frac{1}{a+1}\right)a \\ &= \sum_{a=1}^{\infty}\frac{1}{a+1} = \infty,\end{aligned}$$

that is the harmonic series without the first term, which we know diverges.
For the same reason as before, $\int_0^1 g\,\mathrm{d}\mu = \infty$ also. However, for any fixed N the truncated function

$$g_N(x) = \begin{cases} g(x) & \text{if } g(x) \leqslant N \\ 0 & \text{if not} \end{cases}$$

is in L^1_μ since

$$\int_0^1 g_N\,\mathrm{d}\mu = \frac{1}{\log 2}\sum_{a=1}^{N} a\int_{1/(a+1)}^{1/a}\frac{1}{1+x}\,\mathrm{d}x = \sum_{a=1}^{N}\frac{a}{\log 2}\log\left(\frac{(a+1)^2}{a(a+2)}\right) < \infty$$

for all N in $1 \leq N < \infty$. Notice that $\int_0^1 g_N\,\mathrm{d}\mu \to \infty$ as $N \to \infty$. We can see it letting N tend to infinity in the previous integral or seeing the Lebesgue integral

$$\int_0^1 g_N \,\mathrm{d}x = \sum_{a=1}^{N} \frac{1}{a+1} = \infty,$$

and again since $\frac{\mathrm{d}\mu}{\mathrm{d}x} = \frac{1}{(1+x)\log 2}$ is bounded on $[0,1]$, also $\int_0^1 g_N \,\mathrm{d}\mu \to \infty$. By the Theorem 4.16,

$$\begin{aligned}\lim_{N\to\infty}\lim_{n\to\infty}\frac{1}{n}\sum_{j=0}^{n-1} g_N\left(T^j x\right) &= \lim_{N\to\infty}\int_0^1 g_N \,\mathrm{d}x \\ &= \int_0^1 \lim_{N\to\infty} g_N \,\mathrm{d}\mu = \int_0^1 g \,\mathrm{d}\mu = \infty.\end{aligned}$$

By Monotone Convergence Theorem, which we can apply since as N increases g_N also increases, we can bring the limit inside the integral. □

Chapter 5

Continued Fractions and Geodesics on $\mathbb{H}$

The idea of continued fractions emerged a really long time ago. It can be detected in the Greece of the second century A.D. and in India in the tenth century. In the modern era, Fermat and Lagrange studied them at the end of the eighteenth century [6]. In this chapter we analyze the relation between continued fractions and geodesics in $\mathbb{H}$ introducing the concept of cutting sequence. At the end we present one historical result that states that a geodesic on the Modular surface $\mathrm{PSL}(2,\mathbb{Z})\backslash\mathbb{H}$ has a dense trajectory. The main reference will be the paper by Caroline Series titled *Continued Fractions and Hyperbolic Geometry* [25].

5.1 Farey Tesselation, Continued Fractions and Cutting Sequences

We introduce the Farey Tesselation in order to construct the cutting sequence for a geodesic.

Definition 5.1 (Farey Neighbours)**.** Two rational numbers $\frac{p}{q}, \frac{r}{s} \in \mathbb{Q}$ are called *neighbours* if $|ps - rq| = 1$.

Definition 5.2 (Farey Sum)**.** The Farey sum of two neighbours $\frac{p}{q}, \frac{r}{s}$ is defined as $\frac{p+r}{q+s}$, and it's notation is $\frac{p}{q} \oplus_F \frac{r}{s}$.

Remark 5.3. If $\frac{p}{q} < \frac{r}{s}$ are neighbours, then so are $\frac{p}{q} < \frac{p}{q} \oplus_F \frac{r}{s}$ and $\frac{p}{q} \oplus_F \frac{r}{s} < \frac{r}{s}$. Indeed $|p(q+s) - (p+r)q| = 1$ verifies the first case.

Definition 5.4 (Farey Tesselation)**.** Let us see how we construct the Farey Tesselation:

1. Draw vertical lines starting from each integer on the real line. We denote these points as $\frac{n}{1}$ for $n \in \mathbb{Z}$. Note that for each $n \in \mathbb{Z}$, $\frac{n}{1}$ and $\frac{n+1}{1}$ are neighbours.
2. Join each pair $(\frac{n}{1}, \frac{n+1}{1})$ with a semicircle perpendicular to the real line.
3. Take the Farey sum $\frac{n}{1} \oplus_F \frac{n+1}{1} = \frac{2n+1}{2}$. Join the adjacent neighbours $\frac{n}{1}, \frac{2n+1}{2}$ and $\frac{2n+1}{2}, \frac{n+1}{1}$ with semicircles centred on $\mathbb{R}$.

4. Inductively, if $\frac{p}{q} < \frac{r}{s}$ are Farey neighbours joined by an arc, then join $\frac{p}{q}$ to $\frac{p+r}{q+s}$ and $\frac{p+r}{q+s}$ to $\frac{r}{s}$ with semicircles.
5. Continue in this way.

Inductively, if $\frac{p}{q}, \frac{r}{s}$ are joined by an arc of $\mathcal{F}$, then $\begin{bmatrix} p & r \\ q & s \end{bmatrix}$ has determinant ± 1.
The resulting picture should be the following (5.1) :

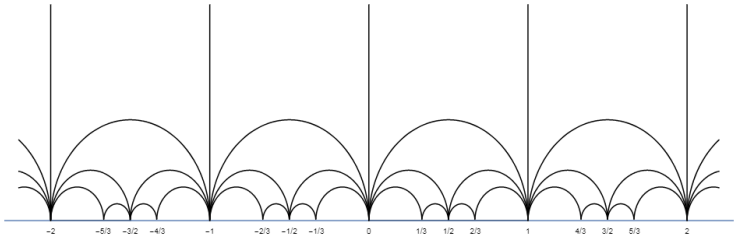

Figure 5.1: Farey tesselation of the hyperbolic plane [24].

Proposition 5.5. *The interior of ideal triangles (that means with all vertices in $\infty \cup \mathbb{R}$) in the Farey tessellation $\mathcal{F}$ cover the hyperbolic plane without overlaps. Moreover, if $g \in \mathrm{PSL}(2,\mathbb{Z})$, then $g(\Delta)$ is an ideal triangle in $\mathcal{F}$, where Δ is an ideal triangle in $\mathcal{F}$.*

Proof. Consider the basic triangle Δ with vertices at 0,1 and ∞. First note that every triangle in the tessellation is the image of Δ under some element in $\mathrm{PSL}(2,\mathbb{Z})$. By construction if $\frac{p}{q} > \frac{r}{s}$ are joined by an arc of $\mathcal{F}$, then $\det\left(\begin{bmatrix} p & r \\ q & s \end{bmatrix}\right) = 1$. Therefore $T = \begin{bmatrix} p & r \\ q & s \end{bmatrix} \in \mathrm{PSL}(2,\mathbb{Z})$, and $T(0) = \frac{r}{s}, T(\infty) = \frac{p}{q}$, and $T(1) = \frac{p}{q} \oplus_F \frac{r}{s}$. From Theorem 2.23, we know that geodesics are mapped to geodesics by Möbius transformations, hence every triangle in the tessellation is the image of Δ under some integral matrix of determinant one.

Now we want to show that the intersection of every two triangles of the tessellation is empty. From the image it is pretty clear, but let us go into the proof. Let $\mathcal{T}$ be the set of triangles in $\mathcal{F}$. If $E \in \mathcal{T}$, let $\mathring{E}$ denote its interior. Let $\mathring{E}_1$, $\mathring{E}_2$ the interiors of two $E_1, E_2 \in \mathcal{T}$. From above, $\exists g_i \in \mathrm{SL}(2,\mathbb{Z})$ such that $E_i = g_i(\Delta)$, thus it is enough to show that $\mathring{\Delta} \cap g\left(\mathring{\Delta}\right) = \emptyset$ for any $g \in \mathrm{PSL}(2,\mathbb{Z})$, since we want that $g_j\left(\mathring{\Delta}\right) \cap g_i\left(\mathring{\Delta}\right) = \emptyset$ for any $g_i, g_j \in \mathrm{PSL}(2,\mathbb{Z})$, but this is the same as writing $\mathring{\Delta} \cap g_j^{-1}\left(g_i\left(\mathring{\Delta}\right)\right) = \emptyset$ for any $g_i, g_j \in \mathrm{SL}(2,\mathbb{Z})$. Let $g = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ so that $\frac{a}{c} > \frac{b}{d}$. By translating and rotating Δ (if needed), we may assume that the side of $g(\Delta)$ joining $\frac{a}{c}$ to $\frac{b}{d}$ cuts the imaginary axis, and we can also assume that $\frac{a}{c} > 0 > \frac{b}{d}$. We claim that this is impossible for $g \in \mathrm{PSL}(2,\mathbb{Z})$. Without loss of generality, we assume $d > 0$, that implies $b < 0$. Then a, c have the same sign. If both are positive then $1 = ad - bc \geq 1 + 1 = 2$, and if both

are negative, then $1 = ad - bc \leq -1 + (-1) = -2$. Since we reached a contradiction, this completes the proof. □

Notation 5.6. We denote the imaginary axis as $\mathbb{I}$.

The previous proposition directly implies the following Corollary:

Corollary 5.7. *1. Every pair of neighbours is the endpoint of some side of $\mathcal{F}$.*

2. Every point $\frac{p}{q} \in \mathbb{Q}$ is a vertex of $\mathcal{F}$.

3. The Farey tessellation $\mathcal{F}$ is invariant under the action of $\mathrm{PSL}(2,\mathbb{Z})$.

Proof. Let's prove (1). Take a general pair of neighbours $\frac{p}{q}$, $\frac{r}{s}$. Then $\begin{bmatrix} p & r \\ q & s \end{bmatrix}(\mathbb{I})$ is a side of $\mathcal{F}$ by Proposition 5.5. Indeed $\frac{p}{q}$ and $\frac{r}{s}$ are the endpoints of the Farey side because they are solutions of the application of the integral Möbius map at ∞ and 0, which are the two endpoints of $\mathbb{I}$. For proving (2), consider a reduced fraction $\frac{p}{q} \in \mathbb{Q}$. Since the fraction is reduced, the greatest common divisor between p and q is 1. By Bézout's Identity, we can find $a, b \in \mathbb{Z}$ such that $ap - bq = 1$. Now considering that $\begin{bmatrix} p & b \\ q & a \end{bmatrix}(\infty) = \frac{p}{q}$, by Proposition 5.5 $\frac{p}{q}$ is a vertex of a Farey side. (3) comes directly from previous points and from Proposition 5.5. Indeed, by the proof of point 1 of this corollary, we can see that we can reach any Farey side applying an appropriate integral Möbius map to $\mathbb{I}$. As a consequence, we can reach any Farey triangle applying this map to the three sides of the Farey triangle with vertices $0, 1, \infty$. What is left to prove is actually that every $g \in \mathrm{PSL}_(2,\mathbb{Z})$ considered as a map that sends Farey trinagles to themselves is bijective. So take $g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{PSL}_(2,\mathbb{Z})$. Represent any Farey triangle by the three vertices in $\mathbb{Q} \cup \infty$. g is trivially a bijection of $\mathbb{R} \cup \infty$, so this directly means that g is also a bijection considered as a map that sends triples of elements in $\mathbb{Q} \cup \infty$ to themselves. □

Definition 5.8 (Continued Fractions)**.** For every real number $x \in \mathbb{R}$, we can express x through its *continued fraction form*, namely

$$x = a_0 + \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \dots}}},$$

where $a_0 \in \mathbb{N}$ and $a_i \in \mathbb{N}_{>0}$ for $i \geq 1$. The notation for a continued fraction is $x = [a_0; a_1, a_2, a_3, \dots]$ and $x = [a_0, a_1, a_2, a_3, \dots]$ for $x \in (0,1)$. Note that for $x \in (0,1) \setminus \mathbb{Q}$ and $\theta_n \in (0,1) \quad \forall n \geq 1$, $a_{i+1} = \lfloor \frac{1}{\theta_i} \rfloor$ (integer part) if $\frac{1}{\theta_i} = a_{i+1} + \theta_{i+1}$.

Theorem 5.9 (Existence and Uniqueness of Finite Continued Fractions [17])**.** *Any rational number can be expressed with a unique (actually two, but correlated) finite continued fractions, and conversely any finite continued fraction corresponds to a unique rational number.*

Proof. Suppose we start with a rational number, then Euclid's algorithm terminates in finitely many steps. This is because the successive reminders are strictly decreasing as they have to be less than the respective quotients. By construction, the successive quotients in Euclid's algorithm precisely give us a simple continued fraction expansion for the rational number we started with. Uniqueness of the Euclidian algorithm implies uniqueness of the continued fraction form, but we prove it by induction. The base case is for the integral part of the number, which is unique. Then we use the same argument if we think to the shifted $[a_1; a_2, \ldots]$ of $[a_0; a_1, \ldots]$. The integral part is unique here as well. We continue in this way to conclude uniqueness of the continued fraction form. Conversely, it is obvious from the definition of continued fraction that every finite continued fraction corresponds to a rational number. The reason why for a rational number there are actually two different finite continued fractions is because for $x \in \mathbb{Q}$, $x = [a_0; a_1, \ldots, a_n] = [a_0; a_1, \ldots, a_n - 1, 1]$. □

Theorem 5.10 (Existence and Uniqueness of Infinite Continued Fractions [17])**.** *Any irrational number is represented by a unique infinite continued fraction and any infinite continued fraction corresponds to a unique irrational number.*

Proof. Here the proof use notions that we introduce later in this thesis (see Definition 5.18) and can be found in A.J. Chinčin's work (Theorem 14, [33]). It also contains a similar version for rational numbers. □

Example 5.11 (Continued Fraction of $x = \frac{1+\sqrt{5}}{2}$)**.** Take $x = \frac{1+\sqrt{5}}{2}$. Since $1 < x < 2$ we have $\lfloor x \rfloor = 1$ and $x - 1 = \frac{-1+\sqrt{5}}{2}$.
Then $\frac{1}{(x-1)} = \frac{2}{(-1+\sqrt{5})} = \frac{2(1+\sqrt{5})}{4} = x$. Unwinding we get $x = 1 + \frac{1}{x}$ from which it follows that $x = [1; 1, 1, 1, \ldots]$.

Example 5.12 (Continued Fraction of $x = \sqrt{2}$)**.** Take $x = \sqrt{2}$.
Note that $(\sqrt{2}+1)(\sqrt{2}-1) = 1$, thus since $1 < x < 2$, we have $\lfloor x \rfloor = 1$ and from the equality above we get $(\sqrt{2}-1) = \frac{1}{1+\sqrt{2}}$. Thus recursively we get $x = 1 + \frac{1}{1+x}$, and therefore $x = [1; 2, 2, 2, \ldots]$.

Now we want to read continued fractions of real numbers from the Farey tesselation.
Let us take $x \in \mathbb{R}_{>0}$ and connect it to the imaginary axis by a semicircle arc γ (hyperbolic geodesic). Take the orientation of γ to go from the imaginary axis to x. This arc cuts a succession of tiles of $\mathcal{F}$. Each tile is an ideal triangle, so γ cuts exactly two sides of each tile in $\mathcal{F}$. The sides of an ideal triangle intersected by γ meet either to the left or to the right in a vertex (or at ∞). Label γ by L or R (in the exceptional case in which γ terminates in a vertex of a triangle, choose either label). The resulting sequence $L^{n_0}R^{n_1}L^{n_3}\ldots, n_i \in \mathbb{N}_{>0}$ is called the cutting sequence of x. If $x > 1$ the sequence begins with L, while if $0 < x < 1$ the sequence begins with R. Note that the cutting sequence is independent of the initial point of γ on $\mathbb{I}$. To visualize it, look at Figure 5.2.

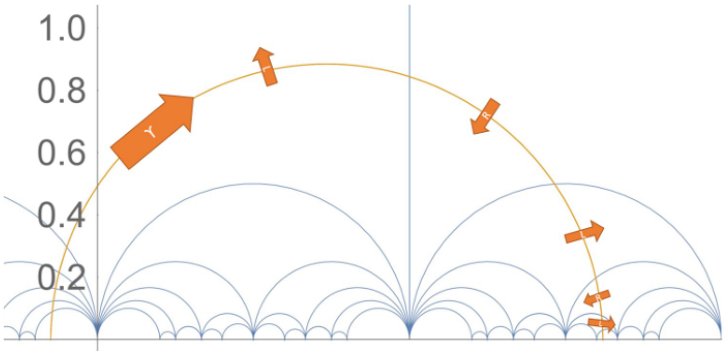

Figure 5.2: Cutting sequence of a geodesic ending up in $x = \frac{81}{50}$ [8].

Proposition 5.13. *Let $x > 1$ have cutting sequence $L^{n_0}R^{n_1}L^{n_3}\ldots, n_i \in \mathbb{N}_{>0}$. Then $x = [n_0; n_1, n_2, \ldots]$. Likewise if $0 < x < 1$ has cutting sequence $R^{n_1}L^{n_3}\ldots, n_i \in \mathbb{N}_{>0}$, then $x = [0; n_1, n_2, \ldots]$.*

Proof. Assume first that $x > 1$. The first fact that we note is that the first non-vertical geodesic cut by γ is after an integer n_0 if $n_0 \leq x < n_0 + 1$. Thus the cutting sequence begins with $L^{n_0}R$.

Let $P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. Note that $P^n = \begin{bmatrix} 1 & n \\ 0 & 1 \end{bmatrix}$, so it is simply a horizontal translation by n, while J is the Möbius inversion map $J(z) = \frac{-1}{z}$. Clearly P^n does not change the cutting sequence (considering also that the starting point moves n steps away from $\mathbb{I}$). We can conclude that γ cuts an ideal triangle T with the same symbol as $P(\gamma)$ cuts $P(T)$ (without forgetting about the shifting). What about J? It is evident that J preserves $\mathbb{I}$ and $\mathbb{R}$ so the direction is also preserved, it starts from $\mathbb{I}$ and it ends in the negative real numbers. So clearly the triangles cut by $J(\gamma)$ that were previously denoted by R in the right half of $\mathbb{H}$ cut by γ, are now denoted by L and vice versa, for example in the case of a triangle with a vertex at the origin (see Figure 5.3). What is interesting to see is that also in this case J preserves the cutting sequence (without shifting the starting point this time). There are several reasonings that can explain it: one of them is that we know that triangles are sent by J to other triangles (since geodesic are sent to other geodesics from Theorem 2.23), and we also know that if the vertex is "inside" the half-circle created by the geodesic we denote this triangle with L, otherwise with R. Now the map J sends all the vertices that are bigger than x inside the geodesic $J(\gamma)$, since in absolute value, after that J is applied, these vertices are smaller than $J(x)$, and therefore the triangle cut by $J(\gamma)$ is denoted with the same letter by the discussion done at the beginning. Let's mark the points where the segments of γ pass from L to R or vice-versa with $z_{-1}, z_0, z_1, \ldots$.

Consider the effect of JP^{-n_0} on γ. Note that $P^{-n_0}(z_0) \in \mathbb{I}$. Now the geodesic $P^{-n_0}(\gamma)$ starts from $P^{-n_0}(z_0)$ and ends at $P^{-n_0}(x) \in [0, 1)$, hence the cutting sequence of $P^{-n_0}(\gamma)$ begins with R. Now J maps $(x - n_0)$ to $\frac{-1}{(x-n_0)} < -1$. From the continued fraction expansion of x, $\lfloor \frac{1}{(x-n_0)} \rfloor = n_1$ and so $-(n_1 + 1) < \frac{-1}{(x-n_0)} \leq -n_1$. This means

that the cutting sequence of $JP^{-n_0}(\gamma)$, starting from the point $JP^{-n_0}(z_0)$ where it crosses the imaginary axis, starts with R^{n_1}, since going from right to left the vertex at infinity of vertical lines making a triangle is to the right of the geodesic.
Let's take another step to make it clearer. Apply now JP^{n_1}.
The geodesic $JP^{n_1}JP^{-n_0}(\gamma)$ meets the imaginary axis at $JP^{n_1}JP^{-n_0}(z_1)$ and ends in the point $\frac{-1}{n_1-\frac{1}{(x-n_0)}} > 1$. From the continued fraction expansion of x we have $\lfloor \frac{1}{(\frac{1}{(x-n_0)}-n_1)} \rfloor = n_2$. Indeed,

$$x = n_0 + \frac{1}{n_1 + \frac{1}{n_2+r}} \text{ with r} \in (0,1),$$

and so

$$\frac{1}{x-n_0} = n_1 + \frac{1}{n_2+r}.$$

Finally,

$$\frac{1}{(\frac{1}{(x-n_0)} - n_1)} = n_2 + r.$$

Therefore $JP^{n_1}JP^{-n_0}(\gamma)$ has a cutting sequence which starts with L^{n_2}. This sequence is the same as the sequence of γ read starting from z_1, so γ itself has sequence $L^{n_0}R^{n_1}L^{n_2}R$. Continuing the argument, we can conclude and prove our claim. Starting with x in $(0,1)$, we apply the same argument; the only difference is that the cutting sequence skips the first step. □

Figure 5.3: Examples of Right or Left triangles [23].

Proposition 5.14. $\mathrm{PSL}(2,\mathbb{Z})$ *is generated by* $P = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ *and* $J = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Proof. Proof of this can be found in the work done by Sarah Zerbes [34]. □

Corollary 5.15. *Cutting sequences described in Proposition 5.13 are preserved by* $\mathrm{PSL}(2,\mathbb{Z})$ *as long as we take keep track of the starting points.*

Proof. In the proof of Proposition 5.13 we proved that J preserves the cutting sequence of a geodesic γ, and we proved also that the cutting sequence of P^n is the same of the one of γ but shifted n steps (so we have to keep track of the starting point). By

Proposition 5.14 we know that J and P generates $\mathrm{PSL}(2,\mathbb{Z})$, therefore we can conclude that for a geodesic γ finishing at $x \in \mathbb{R}$ and intersecting $\mathbb{I}$, its cutting sequence is the same as the one of $g(\gamma)\ \forall g \in \mathrm{PSL}(2,\mathbb{Z})$, as long that we keep track of where it moves the starting point of the geodesic in origin in $\mathbb{I}$. □

Remark 5.16 (Negative numbers)**.** Throughout the section we will prove many theorems for positive real numbers only, because the proof for negative numbers is always very similar. Indeed, for $x = [a_0; a_1, a_2, a_3, \ldots]$, we observe that x and $J(x)$ have the same cutting sequence, so it is usually enough to replace a negative number y by $-1/y = [b_0; b_1, b_2, b_3, \ldots]$ with $b_0 \geq 0$.

5.2 Convergents of a number

Now we want to go deeply into continued fractions, and this will help us prove some interesting results about geodesics in $\mathbb{H}$.

Lemma 5.17. *Let γ, γ' be oriented geodesics in $\mathbb{H}$ (not necessarily cutting $\mathbb{I}$) with the same positive endpoint x. Then the cutting sequences of γ, γ' eventually coincide, i.e. from sufficently large indexes, the two cutting sequences coincide.*

Proof. Pick a side s of $\mathcal{F}$ which is cut by both γ and γ', and we know that there is some $A \in \mathrm{PSL}(2,\mathbb{Z})$ such that $A(s) = \mathbb{I}$.
Then we can conclude using the fact that the cutting sequence of a real number is independent from the starting point on the imaginary axes. □

Definition 5.18 (Convergents of x)**.** Let $x = [a_0; a_1, a_2, \ldots]$ and $\frac{p_n}{q_n} = [a_0; a_1, a_2, \ldots, a_n]$. The reduced fractions $\frac{p_n}{q_n}$ are called the convergents of x.

It is easy to see that each time we add an element to the continued fraction expansion we switch to be grater and less than our real number x, and because all rational numbers are the ends of sides of ideal triangles of the Farey tessellation, we claim that $\frac{p_n}{q_n}, \frac{p_{n+1}}{q_{n+1}}$ are the two ends of the side of $\mathcal{F}$ cut by γ at the moment the sequence of γ changes from L to R or vice versa. To verify this, we prove the following theorem and its corollary.

For the proof of Theorem 5.19, we follow a seminar-paper by Clemens Fruböse "Continued Fractions And Cutting Sequences" [8].

Theorem 5.19. *Let $x = [a_0; a_1, a_2, \ldots]$ and $\frac{p_n}{q_n} = [a_0; a_1, a_2, \ldots, a_n]$, so in particular $p_0 = a_0, q_0 = 1$. Then:*

1. $\begin{bmatrix} p_n & p_{n+1} \\ q_n & q_{n+1} \end{bmatrix}$ *has determinant ± 1 for all $n \geq 0$.*
2. $p_n = a_n p_{n-1} + p_{n-2}$ *and* $q_n = a_n q_{n-1} + q_{n-2}$*, where $n \geq 2$.*
3. $\frac{p_{2n}}{q_{2n}} \leq \frac{p_{2n+2}}{q_{2n+2}} \leq x \leq \frac{p_{2n+1}}{q_{2n+1}} \leq \frac{p_{2n-1}}{q_{2n-1}}$ *for all $n \geq 0$, with equality on one or other side if and only if x is rational and the sequence terminates. Exceptionally we suppose $q_{-1} = 0, p_{-1} = 1$.*

Proof. Choose $x > 0$. We define a matrix $M := JP^{-a_{2n}}JP^{+a_{2n-1}}\dots JP^{+a_1}JP^{-a_0}$. Since J and P have determinant 1, M also has determinant 1. We analyze the way M acts on the continued fraction of x of order $2n$:

$$\begin{aligned} M\left([a_0; a_1, a_2, \dots, a_{2n}]\right) &= JP^{-a_{2n}}JP^{+a_{2n-1}}\dots JP^{+a_1}J\left([0; a_1, a_2, \dots, a_{2n}]\right) \\ &= JP^{-a_{2n}}JP^{+a_{2n-1}}\dots JP^{+a_1}\left([-a_1; -a_2, \dots, -a_{2n}]\right) \\ &= \dots = JP^{-a_{2n}}\left(a_{2n}\right). \end{aligned}$$

By definition of the continued fraction form, we note that $x = [a_0; a_1, a_2, \dots, a_n, \alpha] = [a_0; a_1, a_2, \dots, a_n + \frac{1}{\alpha}]$, $\alpha \in \mathbb{N}$. Now we consider the effect of M on $[a_0; a_1, a_2, \dots, a_{2n}, \alpha]$, i.e. an expansion of x up to order $2n+1$:

$$M\left([a_0; a_1, a_2, \dots, a_{2n}, \alpha]\right) = JP^{-a_{2n}}\left(a_{2n} + \frac{1}{\alpha}\right) = J\left(0 + \frac{1}{\alpha}\right) = J\left(\frac{1}{\alpha}\right) = -\alpha.$$

If $\alpha = 0$, then

$$M\left([a_0; a_1, a_2, \dots, a_{2n}, \alpha]\right) = JP^{-a_{2n}}\left(a_{2n} + \frac{1}{0}\right) = J\left(\frac{1}{0}\right) = 0.$$

Accordingly, if $\alpha = \infty$, then

$$M\left([a_0; a_1, a_2, \dots, a_{2n}, \alpha]\right) = JP^{-a_{2n}}\left(\left[a_{2n} + \frac{1}{\infty}\right]\right) = J\left(\frac{1}{\infty}\right) = -\infty.$$

Since a matrix M^{-1} exists because M has determinant 1, we write M^{-1} with matrix representation

$$M^{-1} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

From above we found that $\frac{a}{c} = \frac{a\cdot(-\infty)+b}{c\cdot(-\infty)+d}$ and $\frac{b}{d} = \frac{a\cdot 0+b}{c\cdot 0+d}$, therefore

$$M^{-1}(0) = [a_0; a_1, a_2, \dots, a_{2n-1}] = \frac{p_{2n-1}}{q_{2n-1}}, \quad M^{-1}(-\infty) = [a_0; a_1, a_2, \dots, a_{2n}] = \frac{p_{2n}}{q_{2n}},$$

where $M^{-1}(0) = [a_0; a_1, a_2, \dots, a_{2n-1}]$ since $\frac{1}{a_{2n}+\frac{1}{0}} = \frac{1}{a_{2n}+\infty} = \frac{1}{\infty} = 0$. The same argument is used for $M^{-1}(-\infty) = [a_0; a_1, a_2, \dots, a_{2n}]$. Note that we are abusing the definition of continued fraction (Definition 5.8), since all a_i, for $i \geq 1$, should be natural numers grather or equal than 1. It follows that

$$M^{-1} = \begin{bmatrix} -p_{2n} & p_{2n-1} \\ -q_{2n} & q_{2n-1} \end{bmatrix} \text{ or } \begin{bmatrix} p_{2n} & -p_{2n-1} \\ q_{2n} & -q_{2n-1} \end{bmatrix}. \tag{5.1}$$

The signs in the matrix follow from the following discussion:

$$\begin{aligned} M^{-1} &= \left(JP^{-a_{2n}}JP^{+a_{2n-1}}\dots JP^{+a_1}JP^{-a_0}\right)^{-1} \\ &= P^{+a_0}J^{-1}P^{-a_1}J^{-1}\dots P^{-a_{2n-1}}J^{-1}P^{+a_{2n}}J^{-1} \end{aligned}$$

with

$$J^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \text{ and } P^{-a} = \begin{bmatrix} 1 & -a \\ 0 & 1 \end{bmatrix}.$$

We can decompose M^{-1} into blocks of the form $P^{-\alpha_i}J^{-1}P^{+\alpha_j}J^{-1}$ for $\alpha_j, \alpha_i > 0$ for all $i, j > 0$ and a single block $P^{+a_0}J^{-1}$. The entries of M^{-1} are given by (5.1). Moreover, computing the matrix representation of the block $P^{-\alpha_i}J^{-1}P^{+\alpha_j}J^{-1}$, we obtain

$$\begin{bmatrix} -\alpha_j\alpha_i & \alpha_i \\ \alpha_j & -1 \end{bmatrix}.$$

We can recognize the following pattern: a composition of an even number of these blocks yields a matrix of the form

$$\begin{bmatrix} A_i & -A_j \\ -A_k & A_l \end{bmatrix}, \quad A_m > 0.$$

Accordingly, a composition of an odd number of these blocks yields

$$\begin{bmatrix} -A_i & A_j \\ A_k & -A_l \end{bmatrix}, \quad A_m > 0.$$

In order to construct the full matrix M^{-1}, we consider the remaining block

$$P^{+a_0}J^{-1} = \begin{bmatrix} -a_0 & 1 \\ -1 & 0 \end{bmatrix}.$$

A composition of these blocks results in a matrix with either first negative column or second negative column. Then either leads to

$$M^{-1} = \begin{bmatrix} -p_{2n} & p_{2n-1} \\ -q_{2n} & q_{2n-1} \end{bmatrix} \text{ or } \begin{bmatrix} p_{2n} & -p_{2n-1} \\ q_{2n} & -q_{2n-1} \end{bmatrix}.$$

Remark 5.20. From the determinant of M^{-1}, we can deduce the following result:

$$\begin{aligned} & p_{2n-1}q_{2n} - p_{2n}q_{2n-1} = 1 \\ & \Rightarrow p_{2n-1}q_{2n} - p_{2n}q_{2n-1} > 0 \\ & \Rightarrow p_{2n-1}q_{2n} > p_{2n}q_{2n-1} \\ & \Rightarrow \frac{p_{2n}}{q_{2n}} < \frac{p_{2n-1}}{q_{2n-1}}. \end{aligned}$$

We also define N as $N = JP^{+a_{2n+1}}JP^{-a_{2n}} \dots JP^{+a_1}JP^{-a_0}$.
N acts on $[a_0; a_1, a_2, \dots, a_{2n+1}]$ (expansion of x to order $2n+1$) in the following way:

$$\begin{aligned} N\left([a_0; a_1, a_2, \dots, a_{2n+1}]\right) &= JP^{+a_{2n+1}}JP^{-a_{2n}} \dots JP^{+a_1}J\left([0; a_1, a_2, \dots, a_{2n+1}]\right) \\ &= JP^{+a_{2n+1}}JP^{-a_{2n}} \dots JP^{+a_1}\left([-a_1; -a_2, \dots, -a_{2n+1}]\right) \\ &= \dots = JP^{a_{2n+1}}\left(-a_{2n+1}\right). \end{aligned}$$

We also remark that:

$$N\left([a_0; a_1, a_2, \ldots, a_{2n+1}, \beta]\right) = JP^{+a_{2n+1}}\left(-a_{2n+1} - \frac{1}{\beta}\right) = J\left(0 - \frac{1}{\beta}\right) = J\left(-\frac{1}{\beta}\right) = \beta.$$

It follows that $N^{-1}(0) = [a_0; a_1, a_2, \ldots, a_{2n+1-1}] = \frac{p_{2n}}{q_{2n}}$ and
$N^{-1}(\infty) = [a_0; a_1, a_2, \ldots, a_{2n+1}] = \frac{p_{2n+1}}{q_{2n+1}}$. In matrix form we get

$$N^{-1} = \begin{bmatrix} p_{2n+1} & p_{2n} \\ q_{2n+1} & q_{2n} \end{bmatrix} \text{ or } \begin{bmatrix} -p_{2n+1} & -p_{2n} \\ -q_{2n+1} & -q_{2n} \end{bmatrix} \tag{5.2}$$

with $\det\left(N^{-1}\right) = 1$. The two possibilities of N^{-1} follow from a similar argument to the one we did for M^{-1}. For $N^{-1} = P^{+a_0}J^{-1}P^{-a_1}J^{-1}\ldots P^{-a_{2n}}J^{-1}P^{+a_{2n+1}}J^{-1}$ with

$$J^{-1} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \text{ and } P^{-a} = \begin{bmatrix} 1 & -a \\ 0 & 1 \end{bmatrix},$$

we can decompose N^{-1} into blocks of the form $P^{+\alpha_i}J^{-1}P^{-\alpha_j}J^{-1}$. The entries of N^{-1} are given by (5.2). Moreover, computing the matrix representation of the block $P^{+\alpha_i}J^{-1}P^{-\alpha_j}J^{-1}$, we obtain

$$\begin{bmatrix} -\alpha_j\alpha_i & -\alpha_i \\ -\alpha_j & -1 \end{bmatrix}.$$

A composition of an even number of these blocks yields a matrix of the form

$$\begin{bmatrix} A_i & A_j \\ A_k & A_l \end{bmatrix}, \quad A_m > 0.$$

Accordingly, a composition of an odd number of these blocks yields

$$\begin{bmatrix} -A_i & -A_j \\ -A_k & -A_l \end{bmatrix}, \quad A_m > 0.$$

Then either

$$N^{-1} = \begin{bmatrix} -p_{2n+1} & -p_{2n} \\ -q_{2n+1} & -q_{2n} \end{bmatrix} \text{ or } \begin{bmatrix} p_{2n+1} & p_{2n} \\ q_{2n+1} & q_{2n}. \end{bmatrix}$$

Remark 5.21. From the determinant of N^{-1}, we can deduce the following result:

$$\begin{aligned} & p_{2n+1}q_{2n} - p_{2n}q_{2n+1} = 1 \\ & \Rightarrow p_{2n}q_{2n+1} < p_{2n+1}q_{2n} \\ & \Rightarrow \frac{p_{2n}}{q_{2n}} < \frac{p_{2n+1}}{q_{2n+1}} \text{ as before in Remark 5.20.} \end{aligned}$$

We conclude,

$$\det \begin{bmatrix} p_n & p_{n+1} \\ q_n & q_{n+1} \end{bmatrix} = \pm 1 \quad \forall n \geq 0.$$

Now we want to prove point 2.
If we use $\frac{p_{2n+1}}{q_{2n+1}} = [a_0; a_1, a_2, \ldots, a_{2n}, a_{2n+1}]$, from (5.1) we obtain $M^{-1}\left(-a_{2n+1}\right) =$

$\frac{p_{2n+1}}{q_{2n+1}}$. On the other hand, by directly applying the action (5.1) to $(-a_{2n+1})$, we obtain the following expression:

$$M^{-1}(-a_{2n+1}) = \frac{p_{2n+1}}{q_{2n+1}} = \frac{\pm p_{2n-1} \mp (-a_{2n+1})\, p_{2n}}{\pm q_{2n-1} \mp (-a_{2n+1})\, q_{2n}}.$$

As a result, we conclude:

$$\frac{p_{2n+1}}{q_{2n+1}} = \frac{p_{2n-1} + a_{2n+1}p_{2n}}{q_{2n-1} + a_{2n+1}q_{2n}}.$$

Considering the numerator and denominator separately, we get the result:

$$p_{2n+1} = p_{2n-1} + a_{2n+1}p_{2n},\ q_{2n+1} = q_{2n-1} + a_{2n+1}q_{2n}.$$

We now proceed in the same way as above with the matrix N : $N^{-1}(a_{2n+2}) = \frac{p_{2n+2}}{q_{2n+2}}$.
By plugging in a_{2n+2} to the mapping which is represented by

$$N^{-1} = \pm \begin{bmatrix} p_{2n+1} & p_{2n} \\ q_{2n+1} & q_{2n} \end{bmatrix}$$

we get

$$N^{-1}(a_{2n+2}) = \frac{p_{2n+2}}{q_{2n+2}} = \frac{p_{2n} + a_{2n+2}p_{2n+1}}{q_{2n} + a_{2n+2}q_{2n+1}}.$$

Combining our results about M and N, we conclude our final result:

$$p_{n+1} = a_{n+1}p_n + p_{n-1} \text{ and } q_{n+1} = a_{n+1}q_n + q_{n-1}$$

for $n \geq 1$.

Now we prove point 3. If we define

$$(a_n)_{n\in\mathbb{N}} := \frac{p_{2n}}{q_{2n}} \text{ and } (b_n)_{n\in\mathbb{N}} := \frac{p_{2n+1}}{q_{2n+1}}$$

we want to show $(a_n)_{n\in\mathbb{N}}$ is increasing, $(b_n)_{n\in\mathbb{N}}$ is decreasing. If $x \in \mathbb{Q}$, the continued fraction terminates. In this case equality holds. So choose $x \notin \mathbb{Q}$ in order to have an infinte continued fraction. From Remarks 5.20 and 5.21, we know that $\frac{p_0}{q_0} < \frac{p_1}{q_1}$ and $\frac{p_2}{q_2} < \frac{p_1}{q_1}$. We claim $\frac{p_0}{q_0} < \frac{p_2}{q_2}$. Here is the reason:

$$\frac{p_0}{q_0} = a_0 < a_0 + \frac{1}{a_1 + \frac{1}{a_2}} = \frac{p_2}{q_2}$$
$$\Rightarrow \frac{p_0}{q_0} < \frac{p_2}{q_2} < \frac{p_1}{q_1}.$$

Again as above, we remark that $\frac{p_2}{q_2} < \frac{p_1}{q_1}$ and $\frac{p_2}{q_2} < \frac{p_3}{q_3}$. Since $\frac{p_2}{q_2}$ and $\frac{p_3}{q_3}$ are neighbours, $\frac{p_3}{q_3} < \frac{p_1}{q_1}$, because the Farey-tessellation has no intersections (see Figure 5.4). One

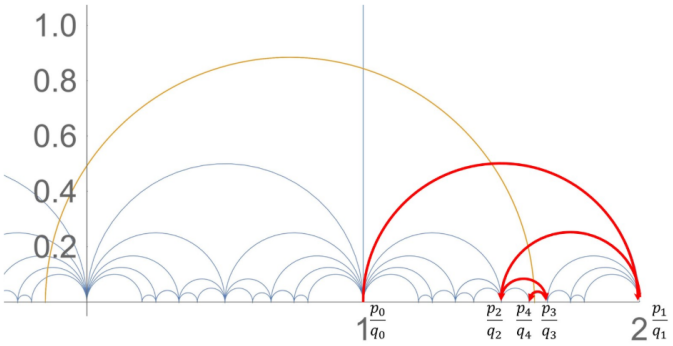

Figure 5.4: Alternating property of $\left(\frac{p_n}{q_n}\right)_{n\in\mathbb{N}}$ on the example $x = [1; 1, 1, 1, 1, 1, 2]$ [8].

begins to see the alternating property of the sequence $\left(\frac{p_n}{q_n}\right)_{n\in\mathbb{N}}$. By induction of that argument we finally reach the result

$$\frac{p_{2n}}{q_{2n}} \leq \frac{p_{2n+2}}{q_{2n+2}} \leq x \leq \frac{p_{2n+1}}{q_{2n+1}} \leq \frac{p_{2n-1}}{q_{2n-1}} \quad \forall n \geq 1,$$

which is equivalent to saying that $(a_n)_{n\in\mathbb{N}} := \frac{p_{2n}}{q_{2n}}$ is increasing and that $(b_n)_{n\in\mathbb{N}} := \frac{p_{2n+1}}{q_{2n+1}}$ is decreasing. We can extend that for $n \geq 0$ if we assume $q_{-1} = 0, p_{-1} = 1$. Indeed, $\frac{p_{-1}}{q_{-1}} = \frac{1}{0} = \infty$. □

Corollary 5.22. *With the notation of Theorem 5.19,* $\frac{p_n}{q_n} \to x$ *as* $n \to \infty$. *Moreover* $\left|x - \frac{p_n}{q_n}\right| \leq \frac{1}{q_n q_{n+1}}$.

Proof. Let us show that $\left(\frac{p_n}{q_n}\right)_{n\in\mathbb{N}}$ is a Cauchy-sequence and converges to x :

$$\left|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}\right| = \left|\frac{p_n q_{n+1} - p_{n+1} q_n}{q_n q_{n+1}}\right| = \left|\frac{\pm 1}{q_n q_{n+1}}\right| = \frac{1}{q_n q_{n+1}}.$$

Note that $(q_n)_{n\in\mathbb{N}}$ is an increasing sequence, because $a_i \geq 1 \quad \forall i > 0$. This can be easily checked by transforming the continued fraction form into a unique fraction. So, for every $\epsilon > 0$, we can choose $N(\epsilon) \in \mathbb{N}_{>0}$, such that

$$\left|\frac{p_n}{q_n} - \frac{p_{n+1}}{q_{n+1}}\right| \leq \left|\frac{\pm 1}{q_{N(\epsilon)} q_{N(\epsilon)+1}}\right| < \epsilon.$$

Thus $\left(\frac{p_n}{q_n}\right)_{n\in\mathbb{N}}$ is a Cauchy-sequence. This immediately leads to the fact that $\left(\frac{p_n}{q_n}\right)_{n\in\mathbb{N}}$ converges, since $\mathbb{R}$ is a complete metric space. According to the construction of the sequence, it is obvious that it converges to x for the Squeeze Theorem. Therefore we can also conclude that since $\frac{p_n}{q_n} \leq x \leq \frac{p_{n+1}}{q_{n+1}}$ for n even (or the opposite for n odd), we get $\left|x - \frac{p_n}{q_n}\right| \leq \left|\frac{p_{n+1}}{q_{n+1}} - \frac{p_n}{q_n}\right| \leq \frac{1}{q_n q_{n+1}}$. □

Corollary 5.23. *Let $s_0, s_1, s_2, \ldots$ be the sides of $\mathcal{F}$ which mark the changes in the cutting sequence of γ from L to R and vice versa, starting with s_0 being the vertical line from a_0 to ∞. Then for $n \geq 0$, the endpoints of s_{2n} are the points $\frac{p_{2n}}{q_{2n}} < \frac{p_{2n-1}}{q_{2n-1}}$ while the endpoints of s_{2n+1} are the points $\frac{p_{2n}}{q_{2n}} < \frac{p_{2n+1}}{q_{2n+1}}$.*

Proof. We already proved in Theorem 5.19 that the even sequence of $\frac{p_n}{q_n}$ is increasing and that the odd one is decreasing. So it is enough to prove that for a general n, two consecutive terms of the sequence $\frac{p_n}{q_n}$ are indeed the endpoints of the geodesics intersected when the cutting sequence changes from R to L or vice-versa. We remind also that from point 1 of Theorem 5.19, we know that consecutive terms of the sequence $\frac{p_n}{q_n}$ are endpoints of a side of a Farey's tail. Let us take $x = [a_0; a_1, a_2, \ldots]$ and consider as usual the geodesic connecting the imaginary axis to x. Then its cutting sequence is $L^{a_0}R^{a_1}L^{a_2}\ldots$. We know s_0 has endpoints a_0 and ∞. We define $p_{-1} = 1$, and $q_{-1} = 0$ so that $\frac{p_{-1}}{q_{-1}} = \infty$ and $\frac{p_0}{q_0} = a_0$ by construction. We know that the left endpoint of s_1 must therefore still be $a_0 = \frac{p_0}{q_0}$. If $a_1 = 1$, then x is in the second half of the interval $[a_0, a_0 + 1]$, so the right endpoint is $a_0 + 1$. Another way to arrive to this conclusion is the following: since $a_1 = 1$, there is only one segment of the geodesic that is denoted with R, so in order to have the sides of the ideal triangle meet to the right, the vertex will be $a_0 + 1$. Let's analyze when $a_1 \geq 2$: if $a_1 = 2$, then there are two segments denoted with R, and so the right endpoint of a next Farey side with left endpoint a_0 is $a_0 \oplus a_0 + 1 = \frac{2a_0+1}{2}$. Recall that the way to find endpoints of smaller radius sides of the Farey tesselation is to compute the Farey sum between two integers, and then inductively between Farey fractions. If $a_1 = 3$, then the endpoint is $a_0 \oplus a_0 \oplus a_0 + 1 = \frac{3a_0+1}{3}$. The right endpoint of s_1 is therefore

$$\frac{a_1a_0 + 1}{a_1} = \frac{a_1p_0 + p_{-1}}{a_1q_0 + q_{-1}} = \frac{p_1}{q_1}$$

by Theorem 5.19. We now have a_2 many L's in the cutting sequence. Therefore we know that the right endpoint of s_2 is $\frac{p_1}{q_1}$. Similarly to the last step, the left hand endpoint moves through a_2 steps from $\frac{p_0}{q_0}$ towards $\frac{p_1}{q_1}$. If $a_2 = 1$, then the left endpoint is $\frac{p_1}{q_1} \oplus \frac{p_0}{q_0} = \frac{p_1+p_0}{q_1+q_0}$. If $a_2 = 2$, then this left endpoint is $\frac{p_1}{q_1} \oplus \frac{p_1}{q_1} \oplus \frac{p_0}{q_0} = \frac{2p_1+p_0}{2q_1+q_0}$. Thus the left endpoint of s_2 is $\frac{a_2p_1+p_0}{a_2q_1+q_0} = \frac{p_2}{q_2}$ by the above theorem. We can continue in this way to show that given a step n, the endpoints are $\frac{p_{n-1}}{q_{n-1}}$ and $\frac{p_n}{q_n}$, thus completing the proof. □

Definition 5.24 (Tail of a number)**.** Two real numbers $x = [a_0; a_1, \ldots], y = [b_0; b_1, \ldots]$ are said to have *the same tails* if there exist $k, l \in \mathbb{N}$ such that $a_{k+r} = b_{l+r}$ for all $r \geq 1$. They have the same tails mod 2 if $k + l$ is even.

Proposition 5.25. *$x = [a_0; a_1, \ldots], y = [b_0; b_1, \ldots]$ have the same tails mod 2 if and only if there exists $g \in SL(2, \mathbb{Z})$ such that $g(x) = y$.*

Proof. Firstly assume that x and y have the same tails mod 2, i.e. $k + l$ is even, which implies that either both k and l are even or that both are odd. Let us analyze the case where both are even, therefore $k = 2n, l = 2m$. Then

$$JP^{-a_{2n}} \dots JP^{a_1} JP^{-a_0}(x) = -[a_{2n}; a_{2n+1}, \dots] = JP^{-b_{2m}} \dots JP^{b_1} JP^{-b_0}(y)$$

$$\Rightarrow (JP^{-b_{2m}} \dots JP^{b_1} JP^{-b_0})^{-1} JP^{-a_{2n}} \dots JP^{a_1} JP^{-a_0}(x) = y.$$

So there exists $g(x) = y$ for some $g \in SL(2, \mathbb{Z})$. If both are odd, therefore $k = 2n-1, l = 2m-1$, then

$$JP^{a_{2n-1}} \dots JP^{a_1} JP^{-a_0}(x) = [a_{2n}; a_{2n+1}, \dots] = JP^{b_{2m-1}} \dots JP^{b_1} JP^{-b_0}(y)$$

$$\Rightarrow (JP^{b_{2m-1}} \dots JP^{b_1} JP^{-b_0})^{-1} JP^{a_{2n-1}} \dots JP^{a_1} JP^{-a_0}(x) = y.$$

Note that $g = (JP^{b_{2m-1}} \dots JP^{b_1} JP^{-b_0})^{-1} JP^{a_{2n-1}} \dots JP^{a_1} JP^{-a_0}$ (or the g of the case of k and l even) is an integral matrix of determinant 1 since it is the composition of integral matrices and of inverses of integral matrices of determinant one.

Now assume that $g(x) = y$ for some $g \in \mathrm{SL}(2, \mathbb{Z})$ and assume $x, y > 0$. Choose a $w < 0$ and let γ_x, γ_y be the oriented geodesics with endpoints (w, x) and (w, y) and call as z_x, z_y the points where γ_x, γ_y cut $\mathbb{I}$. Thus the continued fraction expansions of x, y can be read off from the cutting sequences of γ_x, γ_y, starting from z_x, z_y. Now $g(\gamma_x)$ ends in the point $g(x) = y$, therefore using Lemma 5.17, the cutting sequences of $g(\gamma_x)$ and γ_y eventually coincide. But choosing appropriate starting points under g for the cutting sequence of $g(\gamma_x)$, cutting sequences of γ and $g(\gamma)$ are the same by Corollary 5.15. So the tails of the cutting sequences of γ_x, γ_y must agree. Moreover $k + l$ must be even because the alternating L, R symbols in the cutting sequences must match. □

5.3 Dense Trajectories of Geodesics on $\Gamma\backslash\mathbb{H}$

Now we are interested in analyzing geodesics and cutting sequences on the Modular surface, which is the quotient space of $\mathbb{H}$ by the modular group $\Gamma = \mathrm{PSL}(2, \mathbb{Z})$, obtained gluing the sides of the fundamental domain in the Figure 5.5. The aim here is to prove that geodesics on the modular surfaces have a dense trajectory. We will provide a sketch of the proof.

Firstly, we have to prove that the what we have in Figure 5.5 is actually a fundamental domain.

Proposition 5.26. *The set in Figure 5.6 $\Delta = \{z \in \mathbb{H}$ such that $\Re(z) \in [0, 1]$ and $|z| \geq 1$ and $|z - 1| \geq 1\}$ is a fundamental domain.*

Proof. From Proposition 5.5 and Corollary 5.7 we already proved all the necessary properties listed in Definition 3.14. The map r in the image that sends Δ to other fundamental domains in the same Farey triangle is $r = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix}$. Moreover, recalling E in Example 3.21, we can see that Area(E) = Area (Δ), since we simply move the

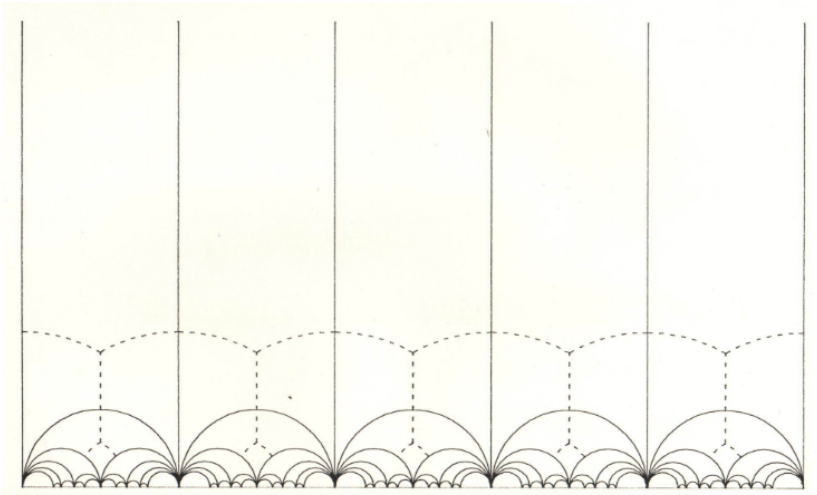

Figure 5.5: Fundamental domain for the action of $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$ on $\mathbb{H}$ [25].

left-half of E to the other side of the right-half of E. The area is the same since the high y is the same and the surface-size covered is the same. □

Definition 5.27 (Projection Map)**.** As introduced before, π is the usual projection map $\pi : \mathbb{H} \to \Gamma\backslash\mathbb{H}$.

Remark 5.28. Geodesics on $\Gamma\backslash\mathbb{H}$ are exactly the images under π of geodesics in $\mathbb{H}$. Orientation is also preserved.

Let us understand a bit more about our Modular surface.

Proposition 5.29. *Viewing $M = \Gamma\backslash\mathbb{H}$ in a topologically way, M is the thrice - punctured sphere with singular points at $\pi(i), \pi(\frac{1}{2}(1+i\sqrt{3}))$ and $\pi(\infty)$.*

A picture of our surface (Figure 5.6) can be found in Dal'Bo's book "Geodesic and Horocyclic Trajectories" on page 81 [5].

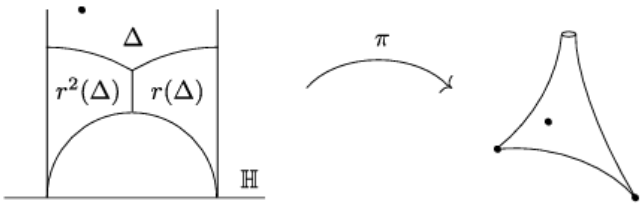

Figure 5.6: The Modular Surface [5].

Proposition 5.30. *The geodesic $\pi(\gamma)$ terminates if and only if $\pi(\gamma)$ begins or ends in the cusp $\pi(\infty)$.*

Proof. The statement comes from the fact that cutting sequences of geodesics terminate if and only if the geodesics terminate in $\mathbb{Q}$, which are the endpoints of the sides of Farey's tails, and each side of a tail in the Farey tesselation is equivalent in the modular surface to the imaginary axis. □

Remark 5.31. By the definition of quotient space, all sides in our Freay tesselation are projected through π to the vertical line connecting i and ∞, which we call S, so every geodesic passing through the imaginary axis in $\mathbb{H}$ is projected onto a geodesic on M that cuts S infinitely many times.

Now pick a point where a geodesic $\overline{\gamma}$ on M cuts S. Then we can lift $\overline{\gamma}$ to a unique γ in $\mathbb{H}$ that cuts the imaginary axis between i and ∞. After applying J and a certain amount of times P, we can transport γ to a geodesic with endpoints $\gamma_{-\infty} \in (-1, 0]$ and $\gamma_{\infty} \in [1, \infty)$. Let $\frac{-1}{\gamma_{-\infty}} = [b_0; b_1, b_2, ..]$ and $\gamma_{\infty} = [a_0; a_1, a_2, ..]$.
We represent the geodesic with endpoints $(\gamma_{-\infty}, \gamma_{\infty})$ by the doubly infinite sequence $\ldots, b_1, b_0, \mid a_0, a_1, \ldots$, where the bar indicates the position where γ cuts $\mathbb{I}$.

Now we state an important proposition to reach density of trajectories of geodesics on the Modular surface.

Proposition 5.32. *γ' is another lift with its endpoints $(\gamma'_{-\infty}, \gamma'_{\infty})$ in the same intervals $(-1, 0]$, $[1, \infty)$ if and only if its representative sequence $\ldots, b'_1, b'_0, \mid a'_0, a'_1, \ldots$ is a shift of $\ldots, b_1, b_0, \mid a_0, a_1, \ldots$.*

Remark 5.33. Remember that if geodesics γ, δ have close endpoints $(\eta, x), (\alpha, \beta)$, then γ and δ are close for a long section of their trajectories. Equally, (η, x) is close to (α, β) if the corresponding cutting sequences agree over a long block surrounding the zero bar.

Now choose a doubly infinite sequence in which every possible finite block of positive integers occurs infinitely often. This sequence represents infinitely many distinct geodesics in $\mathbb{H}$ and they are all images of one another under $\mathrm{PSL}(2, \mathbb{Z})$, hence by Corollary 5.15 project to a unique geodesic on the Modular surface whose trajectory is dense.

5.4 Periodic Continued Fractions

In this section we will concentrate on (eventually) periodic continued fraction. There are indeed a few theorems that connect periodicity of continued fractions to solutions of integral coefficients quadratic equations. Firstly we will give an algebraic proof of these results borrowing ideas from Sierpinski's book "Elementary Theory of Numbers" [26], Hatcher's book "Topology of numbers" [11] and from "An Introduction to the Theory of Numbers" written by Hardy and Wright [10]. That said, we will use as main source Bruce Ikenaga's works "Periodic Continued Fractions" [13] and "Purely Periodic Continued Fractions" [14]. Then we will prove these results with the discussion about geodesics and cutting sequences introduced in this chapter again looking at the work of Caroline Series "The Modular Surface and Continued Fractions" [24], but picking some ideas from section 2.4 of "Geodesic and Horocyclic Trajectories"[5]. The

goal is to astonish the reader proposing two opposite approaches to the same problems.

Definition 5.34 (Eventually Periodic Continued Fraction)**.** A real number $x = [a_0; a_1, a_2, ..]$ is said to be *eventually periodic* if $x = [a_0; a_1, a_2, ..] = [a_0; a_1, a_2, \ldots, \overline{a_k, a_{k+1}, \ldots, a_l}]$, i.e it starts with a preperiod and then repeats another period an infinite number of times.

Definition 5.35 (Quadratic Irrational)**.** A quadratic irrational α is an irrational real root in a quadratic equation

$$Ax^2 + Bx + C = 0$$

where $A, B, C \in \mathbb{Z}$. We let α' denote the other root of the above quadratic equation. The other root is called the (algebraic) conjugate of α.

Proposition 5.36. *A number is a quadratic irrational if and only if it can be written in the form $\frac{p+\sqrt{q}}{r}$, where $p, q, r \in \mathbb{Z}, r \neq 0$ and q is positive and not a perfect square.*

Proof. Suppose x is a quadratic irrational. Then x is a root of

$$ax^2 + bx + c = 0, \quad a, b, c \in \mathbb{Z}, a \neq 0.$$

By the quadratic formula,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

If $b^2 - 4ac = 0$, then $x = -\frac{b}{2a}$, which is a rational number, contrary to assumption.
If $b^2 - 4ac < 0$, then x is complex, and again we have a contradiction. Hence, $b^2 - 4ac > 0$.
Finally, if $b^2 - 4ac$ is a perfect square, then $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$ is rational. Hence, $b^2 - 4ac$ is not a perfect square.

For the converse, suppose $x = \frac{p+\sqrt{q}}{r}$, where $p, q, r \in \mathbb{Z}, r \neq 0$, and q is positive and not a perfect square. Then

$$\begin{aligned} rx - p &= \sqrt{q} \\ (rx - p)^2 &= q \\ r^2x^2 - 2rpx + \left(p^2 - q\right) &= 0. \end{aligned}$$

This is a quadratic equation with integer coefficients, and $r^2 \neq 0$ since $r \neq 0$. Therefore, x is a quadratic irrational. □

Lemma 5.37. *If x is a quadratic irrational and a_0 is an integer, then $a_0 + \frac{1}{x}$ is a quadratic irrational.*

Proof. Write $x = \frac{a+\sqrt{b}}{c}$, where $a, b, c \in \mathbb{Z}, c \neq 0$, and b is positive and not a perfect square. Then

$$a_0 + \frac{1}{x} = a_0 + \frac{1}{\frac{a+\sqrt{b}}{c}} = a_0 + \frac{c}{a+\sqrt{b}} = \frac{\left(a_0a^2 + ac - a_0b\right) - c\sqrt{b}}{a^2 - b}.$$

The last expression is a quadratic irrational. Note that $a^2 - b \neq 0$, because b is not a perfect square.

□

The following lemma is a direct inductive consequence of Lemma 5.37.

Lemma 5.38. *If x is a quadratic irrational, then*

$$a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n + \cfrac{1}{x}}}}$$

is also a quadratic irrational.

Lemma 5.39. *Let $a_0, a_1, \ldots, a_n \in \mathbb{Z}$. Let*

$$y = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n + \cfrac{1}{x}}}}.$$

Then y can be written as $\frac{ax+b}{cx+d}$, where $a, b, c, d \in \mathbb{Z}$.

Proof. The proof is again by induction.
As basecase we take $n = 0$, where $a_0 + \frac{1}{x} = \frac{a_0x+1}{x}$. Assume that this is true for $n - 1$. Then, there exist $a, b, c, d \in \mathbb{Z}$ such that

$$\frac{ax+b}{cx+d} = a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n + \cfrac{1}{x}}}.$$

Therefore we conclude that

$$y = a_0 + \frac{1}{\frac{ax+b}{cx+d}} = \frac{(a_0a + c)\,x + (a_0b + d)}{ax + b}.$$

□

Theorem 5.40 (Lagrange's Theorem)**.** *Let $x \in \mathbb{R} \setminus \mathbb{Q}$. The continued fraction of x is eventually periodic if and only if x is a quadratic irrational.*

Proof. The proof follows the one by Professor Ikenaga [13]. Let x be eventually periodic, i.e $x = [a_0; a_1, a_2, \ldots, \overline{a_k, a_{k+1}, \ldots, a_l}]$. Let's denote with y the periodic part of x, so $y = [\overline{a_k; a_{k+1}, a_{k+2} \ldots, a_l}]$, which means

$$y = a_k + \cfrac{1}{a_{k+1} + \cfrac{1}{a_{k+2} + \ddots + \cfrac{1}{a_l + \cfrac{1}{y}}}}.$$

Then,

$$\begin{aligned}
y &= [\overline{a_k; a_{k+1}, \ldots, a_l}] \\
y - a_k &= [0; \overline{a_{k+1}, \ldots, a_l, a_k}] \\
\frac{1}{y - a_k} &= [\overline{a_{k+1}; \ldots, a_l, a_k}] \\
\frac{1}{y - a_k} - a_{k+1} = \frac{1 - (y - a_k)a_{k+1}}{1 - a_k y} &= [0; \overline{a_{k+2}, \ldots, a_l, a_k, a_{k+1}}] \\
\frac{1 - a_k y}{1 - (y - a_k)a_{k+1}} &= [\overline{a_{k+2}; \ldots, a_l, a_k, a_{k+1}}]. \\
&\ldots
\end{aligned}$$

This is exaclty the same idea of Lemma 5.39, i.e there exist first degree functions g, f such that

$$\frac{g(y)}{f(y)} = [\overline{a_k; a_{k+1}, \ldots, a_l}],$$

$$\frac{g(y)}{f(y)} = y,$$

with g, f two functions of degree 1 with integral coefficients since $a_i > 0 \quad \forall i > 0$. Therefore $yf(y) = g(y)$ is a quadratic equation with integral coefficients and y is an irrational root of this equation. Remember, that y is irrational since it has an infinite continued fraction form.
Now, since

$$x = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \ddots + \cfrac{1}{a_n + \cfrac{1}{y}}}},$$

by Lemma 5.38 x is also a quadratic irrational. The other direction of the proof is not complicated but very dense in terms of computations. The reader can read it in Bruce Ikenaga's work [13]. □

Definition 5.41 (Purely Periodic Continued Fraction)**.** A real number $x = [a_0; a_1, a_2, \ldots]$ is said to be *purely periodic* if $x = [\overline{a_0; a_1, \ldots, a_n}]$, i.e it starts with a period and then repeats it an infinite number of times.

Theorem 5.42 (Galois's Theorem)**.** *Let x be a quadratic irrational and x' its conjugate. The continued fraction of x is purely periodic if and only if $x > 1$ and $x' \in (-1, 0)$.*

Proof. The proof follows the one by Bruce Ikenaga [14], but the first person to give this proof was the famous Évariste Galois in 1828.
Suppose $x > 1$ and $-1 < \bar{x} < 0$. Define x_n as a number such that $\lfloor x_n \rfloor = a_n$. Therefore we have

$$\begin{aligned} x_{n+1} &= \frac{1}{x_n - a_n} \\ \frac{1}{x_{n+1}} &= x_n - a_n \\ \frac{1}{\overline{x_{n+1}}} &= \overline{x_n} - a_n. \end{aligned}$$

Note that $a_n \geq 1$ for $n \geq 1$. But $x_0 = x > 1$, so $a_0 \geq 1$. Thus, $a_n \geq 1$ for $n \geq 0$. Now we claim that for all $n \geq 0$,

$$-1 < \overline{x_n} < 0.$$

We prove the claim using induction.
As basecase take $\bar{x} = \overline{x_0}$, so by assumption $-1 < \overline{x_0} < 0$. Assume that $-1 < \overline{x_n} < 0$. Then $a_n \geq 1$ gives $-a_n \leq -1$, so adding $\overline{x_n} < 0$ gives

$$\begin{aligned} \overline{x_n} - a_n &< -1 \\ \frac{1}{\overline{x_{n+1}}} &< -1 \\ \overline{x_{n+1}} &> -1. \end{aligned}$$

Since the middle inequality shows $\overline{x_{n+1}} < 0$, we get

$$-1 < \overline{x_{n+1}} < 0.$$

This proves the claim by induction. Now

$$\begin{aligned} \overline{x_n} - a_n &= \frac{1}{\overline{x_{n+1}}} \\ \overline{x_n} &= \frac{1}{\overline{x_{n+1}}} + a_n. \end{aligned}$$

Using the claim, we have

$$\begin{aligned} -1 < \frac{1}{\overline{x_{n+1}}} + a_n &< 0 \\ 1 > \left(-\frac{1}{\overline{x_{n+1}}}\right) - a_n &> 0. \end{aligned}$$

This inequality says that $-\frac{1}{\overline{x_{n+1}}} > a_n$, and also that a_n is an integer which differs from $-\frac{1}{\overline{x_{n+1}}}$ by less than 1 . It follows that

$$a_n = \lfloor -\frac{1}{\overline{x_{n+1}}} \rfloor.$$

Since x is a quadratic irrational, its continued fraction is periodic. Thus, there are indices i and j such that

$$x_i = x_j \quad \text{where} \quad 0 < i < j.$$

Hence,

$$\begin{aligned}
\overline{x_i} &= \overline{x_j} \\
\left\lfloor -\frac{1}{\overline{x_i}} \right\rfloor &= \left\lfloor -\frac{1}{\overline{x_j}} \right\rfloor \\
a_{i-1} &= a_{j-1} \\
a_{i-1} + \frac{1}{x_i} &= a_{j-1} + \frac{1}{x_j} \\
x_{i-1} &= x_{j-1}.
\end{aligned}$$

Thus, $x_i = x_j$ implies $x_{i-1} = x_{j-1}$. Continuing to reduce indices in this way, since we assumed $i < j$, we obtain

$$x_0 = x_{j-i}.$$

Therefore,

$$x = x_0 = \left[\overline{a_0; a_1, \ldots, a_{j-i-1}}\right].$$

Hence, x is purely periodic.

Let's prove the other direction now.
Suppose x is a quadratic irrational that is purely periodic, so $x = [\overline{a_0; \ldots, a_{n-1}}]$, where $a_0, \ldots, a_{n-1} \in \mathbb{N}_{>0}$ because $a_i \geq 1 \quad \forall i \geq 1$, but since a_0 comes again later due to periodicity, we also have $a_0 \geq 1$. We have

$$x_n = x = \left[\overline{a_0; \ldots, a_{n-1}}\right].$$

Hence,

$$x = \frac{x_n p_{n-1} + p_{n-2}}{x_n q_{n-1} + q_{n-2}} = \frac{x p_{n-1} + p_{n-2}}{x q_{n-1} + q_{n-2}} \text{ by Theorem 5.19.}$$

So

$$\begin{aligned}
q_{n-1}x^2 + q_{n-2}x &= p_{n-1}x + p_{n-2} \\
q_{n-1}x^2 + (q_{n-2} - p_{n-1})\, x - p_{n-2} &= 0.
\end{aligned}$$

The quadratic equation $\psi(x) = q_{n-1}x^2 + (q_{n-2} - p_{n-1})\, x - p_{n-2}$ has x and $\bar{x}$ as solutions. We already know $x > 1$. It remains to show that this equation has a solution in $(-1, 0)$, and since that root can't be x, it must be $\bar{x}$. Let's consider the function $\psi(t) = q_{n-1}t^2 + (q_{n-2} - p_{n-1})\, t - p_{n-2}$. Let's take $t = 0$: then $\psi(0) = -p_{n-2} < 0$. Next, take $t = -1$,

$$\begin{aligned}
\psi(-1) &= q_{n-1} + (p_{n-1} - q_{n-2}) - p_{n-2} \\
&= (a_{n-1}q_{n-2} + q_{n-3}) + (a_{n-1}p_{n-2} + p_{n-3}) - p_{n-2} - q_{n-2} \\
&= (a_{n-1} - 1)\, p_{n-2} + (a_{n-1} - 1)\, q_{n-2} + p_{n-3} + q_{n-3} \\
&= (a_{n-1} - 1)\, (p_{n-2} + q_{n-2}) + (p_{n-3} + q_{n-3}) \\
&\geq p_{n-3} + q_{n-3} \\
&> 0.
\end{aligned}$$

Then continuity of ψ and the fact that $\psi(-1) > 0$ and $\psi(0) < 0$ imply that there is a root between -1 and 0 by the Intermediate Value Theorem. As noted above, that root must be $\bar{x}$. □

Now that we have presented nice and straightforward proofs for these two interesting results, we want to prove them using the theory about cutting sequences that we developed in this chapter. Again, the reference for this is Caroline Series' work "The Modular Surface and Continued Fraction" [24]. Before we do that, let us prove another useful theorem also taken from Ikanega's paper [14].

Theorem 5.43. *Consider the continued fractions*

$$x = [a_0; a_1, \dots a_{n-1}, a_n] \text{ and } y = [a_n; a_{n-1}, \dots a_1, a_0]\,.$$

Note that y is always constructed following the order $0, \dots, n$, so it is constructed from right to left. Let p_k and q_k denote the numerator and denominator of the k^{th} convergent of x. Let p'_k and q'_k denote the numerator and denominator of the k^{th} convergent of y. Then:

1. $$\frac{p_n}{p_{n-1}} = \frac{p'_n}{q'_n} \quad \text{and} \quad \frac{q_n}{q_{n-1}} = \frac{p'_{n-1}}{q'_{n-1}} \quad \text{for} \quad n \geq 1;$$

2. $$p_n = p'_n, \quad p_{n-1} = q'_n, \quad q_n = p'_{n-1}, \quad q_{n-1} = q'_{n-1}.$$

Remark 5.44. By reversing the roles of x and y, it also follows that

$$\frac{p'_n}{p'_{n-1}} = \frac{p_n}{q_n} \quad \text{and} \quad \frac{q'_n}{q'_{n-1}} = \frac{p_{n-1}}{q_{n-1}}.$$

Proof. We want to work by induction on n.
As basecase, consider $n = 1$. So by Theorem 5.19

$$\frac{p_1}{p_0} = \frac{a_0 a_1 + 1}{a_0} = \frac{p'_1}{q'_1},$$
$$\frac{q_1}{q_0} = a_1 = \frac{p'_0}{q'_0}.$$

The result holds for $n = 1$. Assume as induction-assumption that the result holds for n. We need to prove the result for $n + 1$. We know by Theorem 5.19 that

$$\frac{p_{n+1}}{p_n} = \frac{a_{n+1}p_n + p_{n-1}}{p_n} = a_{n+1} + \frac{p_{n-1}}{p_n} = a_{n+1} + \frac{1}{\left(\frac{p_n}{p_{n-1}}\right)}.$$

We apply the induction hypothesis to $[a_0; a_1, \dots a_n]$ and $[a_n; a_{n-1}, \dots a_0]$. Note that p_{n-1} and p_n are the same for $[a_0; a_1, \dots a_n]$ and $[a_0; a_1, \dots a_n, a_{n+1}]$. However, the p's and q's for $[a_{n+1}; a_n, \dots a_1, a_0]$ and for $[a_n; \dots a_1, a_0]$ are different, indeed for example $p'_n = a_0 p'_{n-1} + p'_{n-2}$ for $[a_n; \dots a_1, a_0]$, but for $[a_{n+1}; a_n, \dots a_1, a_0]$, $p'_n = a_1 p'_{n-1} + p'_{n-2}$,

so we will use double-primed p's and q's for $[a_n; \dots a_1, a_0]$. With this understanding, the induction hypothesis gives

$$\frac{p_n}{p_{n-1}} = \frac{p_n''}{q_n''} = [a_n; \dots a_1, a_0]\,.$$

Then we get

$$\begin{aligned}\frac{p_{n+1}}{p_n} &= a_{n+1} + \frac{1}{\left(\frac{p_n}{p_{n-1}}\right)} \\ &= a_{n+1} + \frac{1}{[a_n; \dots a_1, a_0]} \\ &= [a_{n+1}; a_n, \dots a_1, a_0] \\ &= \frac{p'_{n+1}}{q'_{n+1}}.\end{aligned}$$

Similarly,

$$\frac{q_{n+1}}{q_n} = \frac{a_{n+1}q_n + q_{n-1}}{q_n} = a_{n+1} + \frac{q_{n-1}}{q_n} = a_{n+1} + \frac{1}{\left(\frac{q_n}{q_{n-1}}\right)}.$$

By induction,

$$\frac{q_n}{q_{n-1}} = [a_n; a_{n-1}, \dots a_1]\,.$$

So

$$\frac{q_{n+1}}{q_n} = a_{n+1} + \frac{1}{[a_n; a_{n-1}, \dots a_1]} = [a_{n+1}; a_n, \dots a_1] = \frac{p'_n}{q'_n}.$$

Thus the result holds for all $n \geq 1$ by induction. Let us now prove the second point. Recall that from Theorem 5.19

$$p_n q_{n-1} - p_{n-1} q_n = \pm 1.$$

It follows that $\frac{p_n}{p_{n-1}}$ and $\frac{q_n}{q_{n-1}}$ are in lowest terms by Bézout's Identity. But $\frac{p'_n}{q'_n}$ and $\frac{p'_{n-1}}{q'_{n-1}}$ are convergents of a continued fraction, so they're in lowest terms as well. $\frac{p_n}{p_{n-1}} = \frac{p'_n}{q'_n}$ is an equality between fractions in lowest terms, so we can conclude $p_n = p'_n$ and $p_{n-1} = q'_n$. Likewise, $\frac{q_n}{q_{n-1}} = \frac{p'_{n-1}}{q'_{n-1}}$ is an equality between fractions in lowest terms, so $q_n = p'_{n-1}$ and $q_{n-1} = q'_{n-1}$. □

This result relates a finite continued fraction $[a_0; a_1, \dots a_n]$ and its "reverse" $[a_n; a_{n-1}, \dots a_0]$. The next result (also due to Galois) considers the relationship between the purely periodic continued fractions $[\overline{a_0; a_1, \dots a_n}]$ and $[\overline{a_n; a_{n-1}, \dots a_0}]$.

Theorem 5.45. *Let $x = [\overline{a_0; a_1, \dots a_{n-1}, a_n}]$ be a purely periodic quadratic irrational. Then $-\frac{1}{\bar{x}}$ is purely periodic, and*

$$-\frac{1}{\bar{x}} = [\overline{a_n; a_{n-1}, \dots a_1, a_0}]\,.$$

Proof. Let $\frac{p_k}{q_k}$ be the k^{th} convergent of x. Then

$$x = [\overline{a_0; a_1, \ldots a_{n-1}, a_n}] = [a_0; a_1, \ldots a_{n-1}, a_n, x].$$

The convergents algorithm (see Theorem 5.19) gives

$$\begin{aligned} x &= \frac{p_n x + p_{n-1}}{q_n x + q_{n-1}} \\ q_n x^2 + q_{n-1} x &= p_n x + p_{n-1} \\ q_n x^2 + (q_{n-1} - p_n)\, x - p_{n-1} &= 0. \end{aligned}$$

Let

$$y = [\overline{a_n; a_{n-1}, \ldots a_1, a_0}] = [a_n; a_{n-1}, \ldots a_1, a_0, y].$$

Let $\frac{p'_k}{q'_k}$ denote the k^{th} convergent of y. The convergents algorithm gives

$$y = \frac{p'_n y + p'_{n-1}}{q'_n y + q'_{n-1}}.$$

By Theorem 5.45, $p'_n = p_n, p'_{n-1} = q_n, q'_n = p_{n-1}$, and $q'_{n-1} = q_{n-1}$. So

$$\begin{aligned} y &= \frac{p_n y + q_n}{p_{n-1} y + q_{n-1}} \\ p_{n-1} y^2 + q_{n-1} y &= p_n y + q_n \\ p_{n-1} y^2 + (q_{n-1} - p_n)\, y - q_n &= 0 \\ p_{n-1} + (q_{n-1} - p_n) \left(\frac{1}{y}\right) - q_n \left(\frac{1}{y}\right)^2 &= 0 \\ q_n \left(-\frac{1}{y}\right)^2 + (q_{n-1} - p_n) \left(-\frac{1}{y}\right) - p_{n-1} &= 0. \end{aligned}$$

Thus, x and $-\frac{1}{y}$ are roots of the quadratic $q_n t^2 + (q_{n-1} - p_n)\, t - p_{n-1} = 0$, so they must be conjugates:

$$\begin{aligned} \bar{x} &= -\frac{1}{y}, \\ y &= -\frac{1}{\bar{x}}, \text{ so we get} \\ -\frac{1}{\bar{x}} &= [\overline{a_n; a_{n-1}, \ldots a_1, a_0}]. \end{aligned}$$

□

Now we prove the two main results about periodic and purely periodic continued fractions from an hyperbolic point of view, following Caroline Series's work [24] again.

Theorem 5.46. *Let x be a quadratic irrational and x' its conjugate. The continued fraction of x is purely periodic if and only if $x > 1$ and $x' \in (-1, 0)$.*

Proof. Suppose that $x = [\overline{a_1; \ldots, a_r}]$. Let γ be the geodesic with endpoints $x, x' \in [-1, 0)$, $\frac{-1}{x'} = [\overline{a_r; \ldots, a_1}]$ by Theorem 5.45. Then γ has a periodic cutting sequence with period r: in other words γ is fixed by a $g \in \mathrm{PSL}(2, \mathbb{Z}), g \neq I$, because following the process of construction of the cutting sequence (see Proposition 5.13), applying J and P is the same as shifting the natural numbers in the cutting sequence. Thus x, x' are fixed by $g(z) = \frac{az+b}{cz+d}$, i.e $x = \frac{ax+b}{cx+d}$ and $x' = \frac{ax'+b}{cx'+d}$. Hence they are conjugate roots of a quadratic equation with coefficients in $\mathbb{Z}$.

Conversely, let x and x' be solutions of a quadratic equation such that $x > 1$ and $x' \in (-1, 0)$, satisfying the equation $ax^2 + bx + c = 0$, where a, b, c are integral and coprime. Using the inequalities $x > 1, -1 < x' < 0$, we can deduce that $|a|, |b|, |c|$ are bounded in terms of $D = b^2 - 4ac$. Indeed we have $x = \frac{-b+\sqrt{D}}{2a} > 1$ and $x' = \frac{-b-\sqrt{D}}{2a} \in (-1, 0)$. Now,

$$x - x' = \frac{2\sqrt{D}}{2a} > 0 \Rightarrow a > 0.$$

Note that this also implies that $b \geq 0$ since $x' \in (-1, 0)$ and because b is an integer. Moreover,

$$x = \frac{-b + \sqrt{D}}{2a} > 1 \Rightarrow \sqrt{D} > b,$$

$$x' - x = -\frac{2\sqrt{D}}{2a} < -1 \Rightarrow \frac{2\sqrt{D}}{2a} > 1 \Rightarrow \sqrt{D} > a.$$

Finally, since we have two real solutions, $D > 0$. Therefore,

$$D = b^2 - 4ac > 0 \Rightarrow \frac{b^2}{4a} > c \Rightarrow \frac{D}{4a} > c.$$

Thus there are only a finite number of quadratic solutions $x > 1$ and $x' \in (-1, 0)$ with the same discriminant D, since $a, b \in \mathbb{N}$. Now as in Proposition 5.13 for $x = [a_0; a_1, \ldots]$, $JP^{-a_0}(z) = -[a_1; a_2, \ldots]$ and $JP^{a_1}(-[a_1; a_2, \ldots]) = [a_2; a_3, \ldots]$. Consider $JP^{a_1}JP^{-a_0}$ acting on the geodesic joining x' to x. Since $JP^{a_1}JP^{-a_0} \in \mathrm{PSL}(2, \mathbb{Z})$, it follows that $JP^{a_1}JP^{-a_0}(x'), JP^{a_1}JP^{-a_0}(x) = [a_3; a_4, \ldots]$ is another pair of quadratic solutions such that $JP^{a_1}JP^{-a_0}(x) > 1$ and $JP^{a_1}JP^{-a_0}(x') \in (-1, 0)$ with the same discriminant D since the irrational part of x and x' is not changed applying an integral matrix. The same holds for $JP^{a_r}JP^{-a_{r-1}} \ldots JP^{a_1}JP^{-a_0}(x'), JP^{a_r}JP^{-a_{r-1}} \ldots JP^{a_1}JP^{-a_0}(x)$ for any $r \in \mathbb{N}_{>0}$, and since we have a finite number of solutions $x > 1$ and $x' \in (-1, 0)$ with discriminant D and since by Corollary 5.7 applying J and P corresponds at most a shift in the cutting sequence of a number, this sequence repeats so that the endpoint continued fraction expansions are periodic. □

Theorem 5.47. *Let $x \in \mathbb{R} \setminus \mathbb{Q}$. The continued fraction of x is eventually periodic if and only if x is a quadratic irrational.*

Proof. Suppose that the expansion of x has periodic tail. As in the proof of Proposition 5.25 we may find $g \in \mathrm{PSL}(2, \mathbb{Z})$ such that $g(x)$ is purely periodic, $g(x) = [\overline{a_1; \ldots, a_r}]$. Applying Theorem 5.46, $g(x)$ is quadratic, hence so is x simply adjusting the quadratic equation.

Conversely, suppose that x is quadratic and let x' be the conjugate root. Let γ be the geodesic joining x to $\overline{x}$. Pick a lift $g(\gamma)$ of $\pi(\gamma)$ with $g(\gamma)$ that has one endpoint bigger than 1 and the other in $(-1, 0)$. By Theorem 5.46, $g(x)$ has a periodic expansion and hence, the tail of x is periodic since applying a $g \in \mathrm{PSL}(2, \mathbb{Z})$ the continued fraction expansion remains invariant or at most shifted (see Corollary 5.15). □

Chapter 6

Diopanthine Approximation

Diopanthine Approximation is about studying the approximation of real numbers by rational numbers. Besides what was said in the introduction, a crucial study was made in the eighteenth century, since mathematicians were interested into numerical resolution of equations. Dirichlet, Kronecker and Minkowski are some examples of mathematicians who proved important results in this field [6].

One branch of the theory of Diophantine approximations consists in constructing a correspondence between some algebraic properties of an irrational number and those of the sequence of its continued fraction expansion, for example as we did for (purely) periodic continued fractions at the end of Chapter 5.

Another branch of Diophantine approximations focuses on the speed of convergence of the sequence of rational numbers associated with a continued fraction expansion.

Our final aim in this last section is to prove three main results about approximation of irrational numbers from a hyperbolic point of view using settings of Chapter 5 about geodesics and continued fractions. In particular we often project on the Modular surface $M = \mathrm{PSL}(2,\mathbb{Z})\backslash\mathbb{H}$ through the usual projection π, connecting excursions into $\pi(H_t^+)$ of a geodesic ray $\pi([z,x))$, where x is irrational and H_t^+ is a horocylce disk, and an approximation of x by a sequence of rational numbers. These theorems were originally proved in a more algebraic way, but our purpose here is not to gain simplicity but to illustrate the fact that the mathematical world is not compartmentalized. In this chapter we follow chapter 7.3 of Dal'Bo's book *Geodesic and Horocyclic Trajectories* [5], which treats all we do in more general settings, in particular for lattices different from $\mathrm{PSL}(2,\mathbb{Z})$. For us that we always work on the Modular surface, $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$.

Notation 6.1. From now on we will denote geodesics and their orientations by writing their endpoints, e.g. (∞x) denotes the vertical line starting from ∞ and ending at x. Moreover, for a Möbius transformation $\gamma(z) = \frac{az+b}{cz+d}$, if we write $c(\gamma)$, we mean the coefficient c of the Möbius map γ.

Proposition 6.2 (Definition of Hyperbolic Isometries)**.** *Let g in $\mathrm{SL}(2,\mathbb{R}) \setminus \{\mathrm{e}\}$. Then, we call g hyperbolic if $|\mathrm{Tr}(g)| > 2$, then g fixes exactly two points of $\overline{\mathbb{H}}$, both of which are in $\mathbb{R}\cup\infty$, so g preserves the whole geodesic with these two endpoints.*

Definition 6.3 (Horocycles)**.** We define horocycles as horizontal lines in $\mathbb{H}$ (which can be interpretated as circles tangent to ∞) or circles in $\mathbb{H}$ tangent to $\mathbb{R}$. In symbols: $H_t = \{z \in \mathbb{H} \mid \Im(z) = t\}$ is an horizontal horocylce and $H_t^+ = \{z \in \mathbb{H} \mid \Im(z) \geq t\}$ is a horizontal horodisk.

Definition 6.4. For a point $z \in \mathbb{H}$ we associate to each $x \in \mathbb{H}$ the set $E([z,x)) \subset \mathbb{R}_{>0}$, that contains all $t > 0$ for which there exists an unbounded sequence $(s_n)_{n\geqslant 1}$ satisfying

$$\pi\left(r\left(s_n\right)\right) \in \pi\left(H_t\right).$$

where $(r\left(s\right))_{s\geq 0}$ is the arclength parametrization of the ray $[z,x)$ of the unique geodesic passing through z and x.

Definition 6.5 (Height function of a point)**.** Let $x \in \mathbb{R} \setminus \mathbb{Q}$. The upper bound of the set $E([z,x))$ is called the height of the ray $\pi([z,x))$ and is written as $h(x)$.

Definition 6.6 (Geometrically Badly Approximated Number)**.** A number $x \in \mathbb{R} \setminus \mathbb{Q}$ is said to be badly geometrically approximated if $h(x)$ is finite.

Definition 6.7 (Badly Approximated Number)**.** A badly approximated number is an x for which there is a positive constant c such that for all rational $\frac{p}{q}$ we have

$$\left|x - \frac{p}{q}\right| > \frac{c}{q^2}.$$

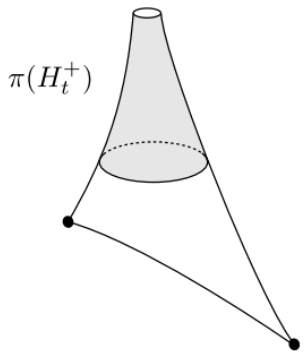

Figure 6.1: Representation of Proposition 6.8, where we have to think about a path that theoretically doesn't enter in the gray area for some $t > 0$ [5].

Proposition 6.8. *$x \in \mathbb{R}\setminus\mathbb{Q}$ is geometrically badly approximated if and only if the ray $\pi([z,x))$ is bounded.*

Proof. Proof comes directly from the definition of height. It can be found in the chapter 7.2 of Dal'Bo's book [5]. □

Lemma 6.9. *Let $g(z) = \frac{(az+b)}{(cz+d)}$ be a Möbius transformation, where a, b, c, d are real numbers satisfying $c \neq 0$ and $ad - bc = 1$. For $t > 0$, consider $H_t^+ = \{z \in \mathbb{H} \mid \Im(z) \geqslant t\}$. The horodisk $g(H^+)$ is the Euclidean disk tangent to the real axis at the point $\frac{a}{c}$, with Euclidean diameter $\frac{1}{tc^2}$.*

Proof. The proof follows Exercise 1.10 in [25]. Consider $H_t^+ = \{z \in \mathbb{H} \mid \Im(z) \geqslant t\}$ and $g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{PSL}_2(\mathbb{R})$. A basic result in Complex Analysis that is an extension to $\mathbb{C}$ of what we discovered for $\mathbb{H}$ is that Möbius transformation maps generalized circles (circles and straight lines that go to ∞) to generalized circles and preserve angles (see Proposition 2.22), so since H_t^+ is a generalized disk in $\mathbb{H}$ with boundary composed by a horizontal straight line that go to ∞, $g(H_t^+)$ is an Euclidian disk tangent to $g(\infty) = \frac{a}{c}$ in $\mathbb{H}$ since we cannot reach ∞ with g from H_t and since in Chapter 2 we proved that Möbius transformation maps $\mathbb{H}$ to itself. So considering r the Euclidian ray of this circle, $g^{-1}(\frac{a}{c} + 2ir)$ must lie on the boundary of H_t^+ so $\Im(g^{-1}(\frac{a}{c} + 2ir)) = t$. Since $g^{-1} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ we get

$$g^{-1}(\frac{a}{c} + 2ir) = \frac{d(\frac{a}{c} + 2ir) - b}{-c(\frac{a}{c} + 2ir) + a} = \frac{1 + 2ircd}{-2irc^2} \text{ and } \Im(g^{-1}(\frac{a}{c} + 2ir) = \frac{1}{2rc^2}$$

□

Now we state three lemmas that will be central in the proofs of the next theorems.

Lemma 6.10. *Let $x \in \mathbb{R}$ and let $(r(s))_{s\in\mathbb{R}}$ be an arclength parametrization of the veritical oriented geodesic from ∞ to x. $\Gamma = \mathrm{PSL}(2, \mathbb{R})$ as always. The following are equivalent:*

1. *there exists a sequence $(s_n)_{n\geqslant 1}$ of positive real numbers such that $\lim_{n\to+\infty} s_n = +\infty$ and $\pi(r(s_n))$ belongs to the horocycle $\pi(H_t)$;*
2. *there exists a sequence $(\gamma_n)_{n\geqslant 1}$ in $\Gamma \setminus \Gamma_\infty$ (where Γ_∞ denotes the subgroup of Γ fixing ∞) such that*

$$|x - \gamma_n(\infty)| \leqslant \frac{1}{(2tc^2(\gamma_n))} \quad \text{and} \quad \lim_{n\to+\infty} c(\gamma_n) = +\infty.$$

where we remind that $\gamma_n(\infty) = \frac{a(\gamma_n)}{c(\gamma_n)}$.

Proof. Let's prove firstly (1) ⇒ (2).
$\pi(r(s_n))$ belongs to $\pi(H_t)$ means that there exists a sequence $(\gamma_n)_{n\geq 1} \in \Gamma$ such that

$$r(s_n) \in \gamma_n(H_t) \quad \forall n \geq 1.$$

So the geodesic (∞x) intersects $\gamma_n(H_t)$, hence $|x - \gamma_n(\infty)| \leqslant \frac{1}{(2tc^2(\gamma_n))}$ since the radius of the horocycle $\gamma_n(H_t)$ is $\frac{1}{(2tc^2(\gamma_n))}$ by Lemma 6.9. Since s_n goes to $+\infty$ and $\gamma_n(H_t)$ does not meet H_t, then the sequence $(c(\gamma_n))_{n\geqslant 1}$ is not bounded.

Now let's prove (2) $\Rightarrow$ (1).
By hypothesis, the geodesic (∞x) intersects each circle $\gamma_n(H_t)$. If we suppose by contradiction that there exists s_n denoting the largest real number s such that $r(s) \in \gamma_n(H_t)$, then $r(s_n) = x + ibe^{-s_n}$ is the parametrization of the vertical geodesic $(x\infty)$ with $b > 0$. Let r_n denote the Euclidean ray of $\gamma_n(H_t)$. The center of this Euclidean circle is the point $\gamma_n(\infty) + ir_n$, thus the following equation holds:

$$(x - \gamma_n(\infty))^2 + \left(be^{-s_n} - r_n\right)^2 = r_n^2$$

The sequences $(x - \gamma_n(\infty))_{n\geqslant 1}$ and $(r_n)_{n\geqslant 1}$ converge to 0 . It follows that $\lim_{n\to+\infty} s_n = +\infty$. □

Remark 6.11. Let $x \in \mathbb{R}$, we have (by Fig. 6.2) and by Lemma 6.9

$$(\infty x) \cap \gamma(H_t) \neq \varnothing \Longrightarrow |x - a/c(\gamma)| \leqslant 1/\left(2tc^2(\gamma)\right) \tag{6.1}$$

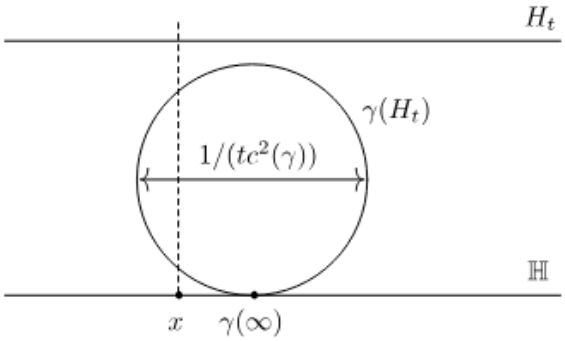

Figure 6.2: Illustration of inequality 6.1 [5].

Lemma 6.12. *For $\mathcal{T}_1(z) = z+1, \mathcal{T}_{-1}(z) = \frac{z}{(z+1)}$, we have $\mathcal{T}_1^n(z) = z+n$ and $\mathcal{T}_{-1}^n(z) = \frac{1}{n+\frac{1}{z}}$.*

Proof. $\mathcal{T}_1^n(z) = z + n$ is inductively trivial. Let's prove the other one by induction. For $n = 1$ it is true by definition. Assume that this holds for n. Now as induction proof, we prove that it holds for $n + 1$.

$$\mathcal{T}_{-1}^{n+1}(z) = \mathcal{T}_{-1}\left(\frac{1}{n+\frac{1}{z}}\right) = \frac{\frac{1}{n+\frac{1}{z}}}{\frac{1}{n+\frac{1}{z}}+1} = \mathcal{T}_{-1}^n(z) = \frac{1}{n+1+\frac{1}{z}}.$$

□

Lemma 6.13. *For $x = [0; n_1, n_2, \ldots]$, there exists a sequence $(g_i)_{i\geqslant 2}$ in Γ such that*

$$g_i^{-1}(\infty x) \cap H_{\frac{n_i+\frac{1}{(n_{i+1}+1)}}{2}} \neq \varnothing \quad \text{and} \quad \lim_{i\to+\infty} c\left(g_i^{-1}\right) = +\infty$$

Proof. Define $\mathcal{T}_1(z) = z+1, \mathcal{T}_{-1}(z) = \frac{z}{z+1}$ and $J(z) = \frac{-1}{z}$. $\mathcal{T}_1$ was previously called P, but to simplify the notation we denote now the usual function $P(z) = z+1$ with $\mathcal{T}_1(z)$. If i is even, define

$$g_i^{-1} = J \circ \mathcal{T}_1^{n_{i-1}} \circ \mathcal{T}_{-1}^{n_{i-2}} \circ \cdots \circ \mathcal{T}_1^{n_1} \circ J.$$

Thus we can see by Lemma 6.12 and that $g_i^{-1}(\infty)$ is the rational $-[0; n_{i-1}, n_{i-2}, \ldots, n_1]$ (we did this multiple times in Chapter 5), and the continued fraction expansion of $g_i^{-1}(x)$ is $([n_i; n_{i+1}, \ldots, n_k])_{k\geqslant 1}$.
If i is odd, define

$$g_i^{-1} = \mathcal{T}_{-1}^{n_{i-1}} \circ \mathcal{T}_1^{n_{i-2}} \circ \cdots \circ \mathcal{T}_1^{n_1} \circ J.$$

Thus $g_i^{-1}(\infty)$ is the rational $[0; n_{i-1}, n_{i-2}, \ldots, n_1]$ and $g_i^{-1}(x)$ has the sequence $-[n_i; n_{i+1}, \ldots, n_k]_{k\geqslant 1}$ by Lemma 6.12 as its continued fraction expansion. In these two cases one has:

$$\left|g_i^{-1}(\infty) - g_i^{-1}(x)\right| \geqslant n_i + \frac{1}{(n_{i+1}+1)}$$

which proves the first part of the lemma since the radius of the geodesic passing through $g_i^{-1}(\infty)$ and $g_i^{-1}(x)$ intersects the horizontal line of the horocycle $H_{\frac{n_i + \frac{1}{(n_{i+1}+1)}}{2}}$.

The second part of this result arises from the fact that the sequences $(c(g_{2i}))_{i\geqslant 1}$ and $(c(g_{2i+1}))_{i\geqslant 1}$ are sequences of positive integers which are strictly increasing by definition of g_i and by Lemma 6.12. □

Now let's state and prove the first theorem of the three about approximation of irrational numbers with rational numbers that we present in this chapter.

Theorem 6.14. *For all $x \in \mathbb{R}\setminus\mathbb{Q}$, there is a sequence $\left(\frac{p_n}{q_n}\right)_{n\geqslant 1}$ of rational numbers satisfying*

$$\left|x - \frac{p_n}{q_n}\right| \leqslant \frac{1}{(2q_n^2)} \quad \textit{and} \quad \lim_{n\to+\infty} |q_n| = +\infty$$

Proof. Let $x \in \mathbb{R}\setminus\mathbb{Q}$. Accordingly to Lemma 6.10, we will show that there exists a sequence $(s_n)_{n\geqslant 1}$ of positive real numbers such that

$$\pi(r(s_n)) \in \pi(H_1) \quad \text{and} \quad \lim_{n\to+\infty} s_n = +\infty.$$

We argue by contradiction. Suppose that there exists $z \in (\infty x)$ such that $\pi([z,x))$ does not intersect $\pi(H_1)$. Lift this situation up to $\mathbb{H}$ and consider $r(z) = \frac{(z-1)}{z}$. The transformations r and $r^2(z) = \frac{-1}{(z-1)}$ map H_1 to circles of Euclidean diameter 1 by Lemma 6.9, tangent to the real line at points $r(\infty) = \frac{1}{1} = 1$ and $r^2(\infty) = \frac{-1}{\infty} = 0$, and tangent to each other at the point $\frac{1}{2} + \frac{i}{2}$ (Fig. 6.3).

As showed in Proposition 5.26, the set Δ in Fig 5.5 defined by

$$\Delta = \{z \in \mathbb{H} \quad | \quad 0 \leqslant \operatorname{Re} z \leqslant 1, |z| \geqslant 1 \text{ and } |z-1| \geqslant 1\}$$

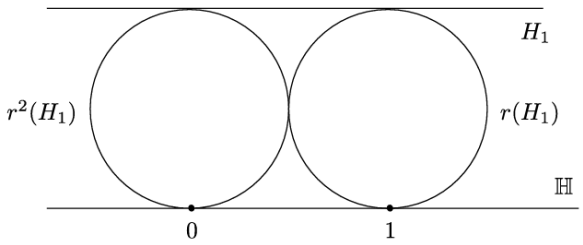

Figure 6.3: $r^2(H_1)$ and $r(H_1)$ [5].

is a fundamental domain of Γ. It follows from the definition of fundamental domain that there exists $\gamma \in \Gamma$ such that $\gamma(z)$ belongs to Δ (see Definition 3.14). The point x does not belong to the Γ-orbit of the point ∞ since it is irrational and so $\gamma(x)$ is also irrational for all γ in Γ, and the ray $[\gamma(z), \gamma(x))$, which is a circular arc, does not intersect H_1 since it does not also the projection in the Modular surface M . Hence this ray is in $\mathbb{H} \setminus H_1^+$. For the same reasons, it is also in $\mathbb{H} \setminus r\left(H_1^+\right)$ and $\mathbb{H} \setminus r^2\left(H_1^+\right)$. Since $\gamma(z)$ belongs to Δ, this ray is in a compact set (which is a subset of the area between H_1, $r(H_1)$ and $r(H_2)$ since $\gamma(z) \in \Delta \setminus H_1^+$) and cannot arrive to $\gamma(x)$ without intersecting $r(H_1)$ or $r^2(H_1)$ or both. So we finally reach a contradiction.

To achieve the proof, it suffices to apply Equation 6.1 and Lemma 6.10 as already said. □

Let's define now for each irrational number x the quantity

$$\nu(x) = \inf\left\{c > 0 \mid \exists \quad 1 \leq \left(\frac{p_n}{q_n}\right)_n \in \mathbb{Q}, \left|x - \frac{p_n}{q_n}\right| \leqslant \frac{c}{q_n^2} \text{ and } \lim_{n\to+\infty} |q_n| = +\infty\right\}.$$

By Theorem 6.14, this quantity is less or equal than $\frac{1}{2}$ for all x. The following theorem is more precise.

Theorem 6.15 (Hurwitz's Theorem)**.** *For every irrational number x, one has*

$$\nu(x) \leqslant \frac{1}{\sqrt{5}}.$$

Also $\nu(x) = \frac{1}{\sqrt{5}}$ if and only if there exist a, b, c, d in $\mathbb{Z}$ such that $ac - bd = 1$ and $x = \frac{a\mathcal{N}+b}{c\mathcal{N}+d}$, where $\mathcal{N} = \frac{1+\sqrt{5}}{2}$ is the golden ratio .

Proof. By Lemma 6.10, the quantities $\nu(x)$ and the height $h(x)$ of x are related by the following relation:

$$\nu(x) = \frac{1}{2h(x)}.$$

Indeed $h(x) = t$ in the Lemma 6.10, by Definition 6.5. Our theorem can therefore be restated in terms of the height of geodesic rays in the form

$$\inf_{x\in\mathbb{R}\setminus\mathbb{Q}} h(x) = \frac{\sqrt{5}}{2}.$$

Let us prove this statement. Consider the case where x is the fixed point of a hyperbolic isometry γ in Γ. Choose z on the geodesic preserved by this isometry, with the condition that $\pi([z,x)) = \pi\left((\gamma^-\gamma^+)\right)$, where $\pi\left((\gamma^-\gamma^+)\right)$ is a compact geodesic on M. Thus $h(x)$ is the largest $t > 0$ such that

$$\pi\left(\left(\gamma^-\gamma^+\right)\right) \cap \pi\left(H_t\right) \neq \varnothing.$$

In other words, lifting the situation on $\mathbb{H}$ one has the expression (Fig. 6.4)

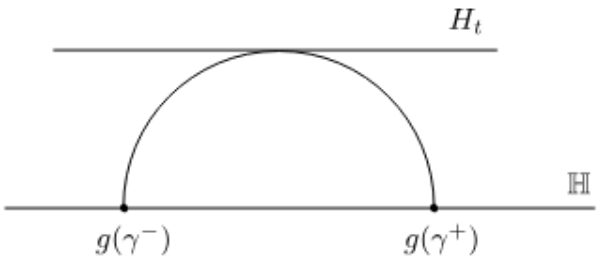

Figure 6.4: Representation of Equation 6.2 [5].

$$h(x) = \max_{g\in\Gamma} \frac{\left|g\left(\gamma^-\right) - g\left(\gamma^+\right)\right|}{2}. \tag{6.2}$$

We write $\gamma(z) = \frac{(az+b)}{(cz+d)}$ with $ad - bc = 1$. Since γ is hyperbolic, we have $c \neq 0$, otherwise we cannot have two fixed points in $\mathbb{R} \cup \infty$. Remember from linear algebra that the trace of a matrix is invariant under conjugation, i.e. for a matrix $g \in \Gamma$, $\mathrm{Tr}(g\gamma g^{-1}) = \mathrm{Tr}(\gamma) = (a + d)$. Then we also know that γ fixes γ^- and γ^+, and therefore $g\gamma g^{-1}$ fixes $g(\gamma^-)$ and $g(\gamma^+)$. Moreover, the two fixed points of γ are $\gamma^\pm = \frac{a-d\pm\sqrt{(a+d)^2-4}}{2c}$. So the two fixed points for $g\gamma g^{-1}$ are $g(\gamma^\pm) = \frac{a(g\gamma g^{-1})-d(g\gamma g^{-1})\pm\sqrt{(a+d)^2-4}}{2c(g\gamma g^{-1})}$. So we obtain

$$\left|g\left(\gamma^-\right) - g\left(\gamma^+\right)\right| = \frac{\sqrt{(a+d)^2-4}}{c\left(g\gamma g^{-1}\right)}. \tag{6.3}$$

Let us consider the case where x is the golden ratio $\mathcal{N} = \frac{(1+\sqrt{5})}{2}$. This point is fixed by the hyperbolic isometry $\gamma = \mathcal{T}_1\mathcal{T}_{-1}$ in Γ, where $\mathcal{T}_1(z) = z + 1, \mathcal{T}_{-1}(z) = \frac{z}{z+1}$, which can also be written in the form

$$\gamma(z) = \frac{(2z+1)}{z+1}.$$

Indeed, we know that the golden ratio is one of the two conjugate solutions of

$$z^2 - z - 1 = 0.$$

The following equality can be deduced from equation 6.3 replacing $a = 2, b = 1, c = 1, d = 1$:

$$h(\mathcal{N}) = \frac{\sqrt{5}}{2 \min_{g\in\Gamma} c\left(g\gamma g^{-1}\right)}.$$

For all $g \in \Gamma$ the isometry $g\gamma g^{-1}$ is hyperbolic (since γ preserves two real points), and we have $c\left(g\gamma g^{-1}\right) \neq 0$. Furthermore, this quantity is an integer thus $c\left(g\gamma g^{-1}\right) \geqslant 1$. In the particular case where $\gamma = \mathcal{T}_1\mathcal{T}_{-1}$, one has $c(\gamma) = 1$. It follows that

$$h(\mathcal{N}) = \frac{\sqrt{5}}{2}.$$

We prove now that for all $x \in \mathbb{R} \setminus \mathbb{Q}$,

$$h(x) \geqslant \frac{\sqrt{5}}{2}.$$

Let $x \in \mathbb{R} \setminus \mathbb{Q}$. The definition of the height $h(x)$ implies

$$h(\gamma x) = h(x)$$

for all $\gamma \in \Gamma$. Thus without loss of generality one can assume that x is in $[0, 1]$. We denote by $([0; n_1, \ldots, n_k])_{k\geqslant 1}$ the continued fraction expansion of x. We will use the isometries $\mathcal{T}_1(z) = z + 1, \mathcal{T}_{-1}(z) = \frac{z}{(z+1)}$ and $J(z) = \frac{-1}{z}$. These three isometries are interrelated by the equality $J\mathcal{T}_{-1}J = \mathcal{T}_1^{-1}$. Let $k \geqslant 2$ and set $\gamma_k = \mathcal{T}_{-1}^{n_1} \cdots \mathcal{T}_{(-1)^k}^{n_k}$. From Lemma 6.12 the following facts hold:

1. if k is even, then $\gamma_k(0) = [0; n_1, \ldots, n_k]$ and $\gamma_k(\infty) = [0; n_1, \ldots, n_{k-1}]$;
2. if k is odd, then $\gamma_k(0) = [0; n_1, \ldots, n_{k-1}]$ and $\gamma_k(\infty) = [0; n_1, \ldots, n_k]$;
3. $\lim_{k\to+\infty} [0; n_1, \ldots, n_k] = x$.

Consider an arclength parametrization $(r(s))_{s\in\mathbb{R}}$ of the oriented geodesic (∞x). If the sequence $(n_i)_{i\geqslant 1}$ contains infinitely many terms $\geqslant 3$, by Lemmas 6.10 and 6.13, there exists an unbounded sequence $(s_n)_{n\geqslant 1} \in \mathbb{R}_{>0}$ such that

$$\pi\left(r\left(s_n\right)\right) \in \pi\left(H_{\frac{3}{2}}\right)$$

at least, picking in Lemma 6.13 $n_i = 3$ and $n_{i+1} = \infty$ and then apply Lemma 6.10. Thus $h(x) \geqslant \frac{3}{2}$ which implies $h(x) > \frac{\sqrt{5}}{2}$. Otherwise, after replacing x with $\gamma_i^{-1}(x)$ with $i \geqslant 1$, two cases arise:

1. either the sequence $(n_i)_{i\geqslant 1}$ contains infinitely many terms equal to 2 and $1 \leqslant n_i \leqslant 2$ for all $i \geqslant 1$,
2. or every term of the sequence $(n_i)_{i\geqslant 1}$ is equal to 1.

In the first case, by Lemmas 6.10 and 6.13 as before, there exists an unbounded sequence $(s_n)_{n\geqslant 1}$ in $\mathbb{R}_{>0}$ such that

$$\pi\left(r\left(s_n\right)\right) \in \pi\left(H_{\frac{7}{6}}\right)$$

at least, picking in Lemma 6.13 $n_i = 2$ and $n_{i+1} = 2$ and then apply Lemma 6.10. Thus $h(x) \geqslant \frac{7}{6}$, and in particular, $h(x) > \frac{\sqrt{5}}{2}$. In the second case, x is related to the golden ratio. More precisely, one has $\mathcal{T}_1(x) = \mathcal{N}$, thus $h(x) = h(\mathcal{N})$ and, by the first part of the proof, $h(x) = \frac{\sqrt{5}}{2}$. Thus we have proved our claim. □

The badly approximated real numbers x for which $\nu(x)$ is strictly positive are of special interest. For example it is the case for $\sqrt{2}$.

The following theorem connects this property to a property of the continued fraction form $[n_0; n_1, n_2, \dots]$ of x.

Theorem 6.16. *Let x be an irrational number. The following are equivalent:*

1. *the sequence $(n_i)_{i\geqslant 0}$ is bounded;*
2. *the real number x is badly approximated.*

Proof. Let x be an irrational number. We know that $\nu(x)$ and $h(x)$ are related by the following equation:

$$\nu(x) = \frac{1}{2h(x)}.$$

Thus x is badly approximately if and only if x is geometrically badly approximated since geometrically approximated means that $h(x)$ is finite (see Definitions 6.7 and 6.6). Recall that, by Proposition 6.8, this property is equivalent to the fact that the ray $\pi([z, x))$ is bounded. We prove firstly (2) ⇒ (1) by contrapposition (so not (1) ⇒ not (2)). Suppose that the sequence $(n_i)_{i\geqslant 1}$ is not bounded, which it is equivalent to saying that for all $t > 0$, there exists a subsequence $(n_{i_k})_{k\geqslant 1}$ such that $n_{i_k} \geqslant 2t \quad \forall k \geq 1$. Let $(r(s))_{s\in\mathbb{R}}$ be the usual arclength parametrization of the oriented geodesic (∞x). By Lemmas 6.10 and 6.13, in particular applying Lemma 6.13 with $n_i = 2t$ and $n_{i+1} = \infty$ and then applying Lemma 6.10, there exists a sequence $(s_n)_{n\geqslant 1}$ in $\mathbb{R}_{>0}$ converging to $+\infty$ such that

$$\pi\left(r\left(s_n\right)\right) \in \pi\left(H_t\right).$$

One can conclude by Definition 6.5 that $h(x)$ is greater than t for all $t > 0$ and hence that x is not geometrically badly approximated.

We prove the other direction also by contrapposition, so not (2) ⇒ not (1). Suppose that $\pi([i, x))$ is not bounded. For all integers $k \geqslant 1$ the surface M without the cusp $\pi\left(H_k^+\right)$ is bounded. Hence there exists $K \geqslant 1$ such that for all $k \geqslant K$ (Fig. 6.5)

$$\pi([i, x)) \cap \pi\left(H_k\right) \neq \varnothing.$$

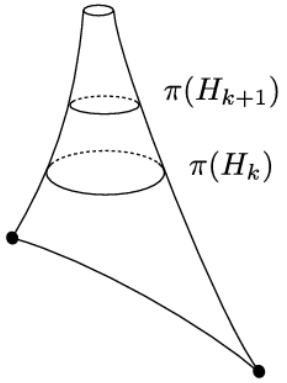

Figure 6.5: Representation of $\pi(H_k)$ in the surface M [5].

Lifiting the situation on $\mathbb{H}$, this property translates to the existence of a sequence $(g_k)_{k\geqslant K}$ in Γ satisfying

$$g_k^{-1}([i,x)) \cap H_k \neq \varnothing.$$

Since the coefficients of the Möbius transformation g_k are integers, we have $\Im(g_k^{-1}(i)) \leqslant 1$ by Equation 2.2. Therefore, the ray $g_k^{-1}([i,x))$ crosses H_k at two points since $g_k^{-1}(x) \in \mathbb{R}$, which implies the inequality

$$\left|\Re\left(g_k^{-1}(i)\right) - g_k^{-1}(x)\right| > k,$$

because $g_k^{-1}(i)$ comes before the maximum of the semicircle geodesic, so the difference between the two real parts is bigger than the ray of the semicircle, that is bigger than k. It follows that this ray intersects at least k vertical geodesics of the form

$$\left((m_k+1)\,\infty\right), \left((m_k+2)\,\infty\right), \ldots, \left((m_k+k)\,\infty\right).$$

Hence the ray $[i,x)$ intersects k consecutive Farey vertical lines

$$g_k\mathcal{T}_1^{m_k+1}(0\infty), \quad \ldots, \quad g_k\mathcal{T}_1^{m_k+k}(0\infty) \text{ with } m_k \in \left(\Re\left(g_k^{-1}(i)\right), g_k^{-1}(x)\right).$$

Returning to the geometric construction of the continued fraction expansion discussed in Chapter 5, one obtains that for all $k \geqslant K$, there exists $n_i \geqslant k$, since intersecting vertical lines means that we have $k \geq K$ "L" because cutting sequences are preserved by $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$ (see Corollary 5.7). □

For this last theorem, an algebraic proof is provided in Einsiedler and Wards' book at page 87 [7] using convergents of continued fractions. The reader can compare the two proofs as we did in Chapter 5 to once again be impressed by the huge difference approaches between them. So we have concluded this last chapter, proving interesting results about approximation of irrational numbers by rational numbers through an hyperbolic point of view.

Chapter 7

Conclusion

In the pursuit of a comprehensive understanding of flows on the Modular surface and their connection to continued fractions, this master thesis highlights the main concepts and proofs. However the topics can be hugely extended and faced in more detail, as always when we go deeply in mathematics. Regarding continued fractions, for example, other fascinating properties can be proved. An example could be a beautiful (almost) palindromic periodic continued fraction form for square roots of non perfect square natural numbers [25]. Moreover, in chapter 3 of Einsiedler's book there are other results about continued fractions or convergents of continued fractions proved by exploiting again the ergodicity of the Gauss map [7]. Furthermore, the whole approach to building cutting sequences is done by Dal'Bo in his book in chapter 2.4 [5] in a more detailed way, maybe a bit more dense in sense of notation, but we warmly advise to read it and compare it with our version taken from Caroline Series' work [25]. Moreover, note that in this thesis we faced problems on the Modular surface, but more advanced considerations can be made on bigger dimensional space like $\mathrm{SL}(n,\mathbb{Z})\backslash\mathrm{SL}(n,\mathbb{R})$. Furthermore, S. Patterson's thesis [21] is one of the founding texts which generalize this hyperbolic approach to Fuchsian groups. Dal'Bo also faced several of the problems that we faced, and some that we didn't, taking a generic Fuchsian group Γ instead of $\Gamma = \mathrm{PSL}(2,\mathbb{R})$.

Generally and morally thinking, I believe this thesis is a really good example of the extension and the power of mathematics.

Bibliography

[1] James W. Anderson. *Hyperbolic Geometry*. Springer-Verlag, 2005. ISBN 1852339349. doi: 10.1007/1-84628-220-9. URL http://dx.doi.org/10.1007/1-84628-220-9.

[2] Titu Andreescu, Dorin Andrica, and Ion Cucurezeanu. *An Introduction to Diophantine Equations: A Problem-Based Approach*. Birkhäuser Boston, 2010. ISBN 9780817645496. doi: 10.1007/978-0-8176-4549-6. URL http://dx.doi.org/10.1007/978-0-8176-4549-6.

[3] Roberto Bonola. *Non-Euclidean geometry; a critical and historical study of its development*. LaSalle, Ill. : Open Court Publishing Co., 1912. ISBN 9781015662803. URL https://download.tuxfamily.org/openmathdep/geometry_advanced/Non-Euclidean_Geometry-Bonola.pdf.

[4] Sami Comment. History of Hyperbolic Geometry . URL https://web.colby.edu/thegeometricviewpoint/2016/12/08/history-of-hyperbolic-geometry/.

[5] Françoise Dal'Bo. *Geodesic and Horocyclic Trajectories*. Springer London, 2011. ISBN 9780857290731. doi: 10.1007/978-0-85729-073-1. URL http://dx.doi.org/10.1007/978-0-85729-073-1.

[6] Pierre Dugac and Jean Alexandre Dieudonné. *Abrégé d'histoire des mathématiques 1700-1900*. Hermann, Paris, nouveau tirage edition, 1992. ISBN 2705660240.

[7] Manfred Einsiedler and Thomas Ward. *Ergodic Theory: with a view towards Number Theory*. Springer London, 2011. ISBN 9780857290212. doi: 10.1007/978-0-85729-021-2. URL http://dx.doi.org/10.1007/978-0-85729-021-2.

[8] Clemens Fruböse. Continued Fractions and Cutting Sequences. URL https://www.mathi.uni-heidelberg.de/~alessandrini/Arith_Reports/3-Continued%20fractions.pdf. Supervised by Beatrice Pozzetti and Daniele Alessandrini.

[9] Brian C. Hall. *Lie Groups, Lie Algebras, and Representations: An Elementary Introduction*. Springer International Publishing, 2015. ISBN 9783319134673. doi: 10.1007/978-3-319-13467-3. URL http://dx.doi.org/10.1007/978-3-319-13467-3.

[10] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Oxford, fourth edition, 1975.

[11] A. Hatcher. *Topology of Numbers*. Miscellaneous Book Series. American Mathematical Society, 2022. ISBN 9781470456115. URL https://books.google.ch/books?id=GV-vEAAAQBAJ.

[12] Eberhard Hopf. Ergodic theory and the geodesic flow on surfaces of constant negative curvature. *Bulletin of the American Mathematical Society*, 77:863–877, 1971. URL https://api.semanticscholar.org/CorpusID:30906722.

[13] Bruce Ikenaga. Periodic Continued Fractions, 2019. URL https://sites.millersville.edu/bikenaga/number-theory/periodic-continued-fractions/periodic-continued-fractions.pdf.

[14] Bruce Ikenaga. Purely Periodic Continued Fractions, 2019. URL https://sites.millersville.edu/bikenaga/number-theory/purely-periodic-continued-fractions/purely-periodic-continued-fractions.html.

[15] M. C. Irwin. *Smooth dynamical systems*, volume 94 of *Pure Appl. Math., Academic Press*. Academic Press, New York, NY, 1980.

[16] Wilhelm Klingenberg. *Lectures on Closed Geodesics*. Springer Berlin Heidelberg, 1978. ISBN 9783642618819. doi: 10.1007/978-3-642-61881-9. URL http://dx.doi.org/10.1007/978-3-642-61881-9.

[17] Gautam Gopal Krishnan. Continued Fractions, 2016. URL https://pi.math.cornell.edu/~gautam/ContinuedFractions.pdf.

[18] Niels Lauritzen. Continued Fractions and Factoring. URL https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=a510fb0836d8741a5ac0f9cb1d2cbbc190637185.

[19] Hartmut Monien and Vincent Delecroix. Polygons and triangles in hyperbolic geometry. URL https://doc.sagemath.org/html/en/reference/plotting/sage/plot/hyperbolic_polygon.html. Accessed: 2024-06-27.

[20] Gabriel P. Paternain. *Geodesic Flows*. Birkhäuser Boston, 1999. ISBN 9781461216001. doi: 10.1007/978-1-4612-1600-1. URL http://dx.doi.org/10.1007/978-1-4612-1600-1.

[21] S. J. Patterson. Diophantine approximation in fuchsian groups. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 282 (1309):527–563, 1976. ISSN 00804614. URL http://www.jstor.org/stable/74711.

[22] Ya B Pesin. Geodesic flows with hyperbolic behaviour of the trajectories and objects connected with them. *Russian Mathematical Surveys*, 36(4):1–59, August 1981. ISSN 1468-4829. doi: 10.1070/rm1981v036n04abeh002636. URL http://dx.doi.org/10.1070/RM1981v036n04ABEH002636.

[23] Simon Seignourel. The Farey Tessellation , 2023. URL https://math.uchicago.edu/~may/REU2023/REUPapers/Seignourel.pdf.

[24] Caroline Series. The Modular Surface and Continued Fractions, 1985. URL https://reu.dimacs.rutgers.edu/~ks1613/documents/series.pdf.

[25] Caroline Series. Continued Fractions and Hyperbolic Geometry, July 2015. URL https://homepages.warwick.ac.uk/~masbb/HypGeomandCntdFractions-2.pdf.

[26] W. Sierpinski. *Elementary Theory of Numbers: Second English Edition (edited by A. Schinzel)*. ISSN. Elsevier Science, 1988. ISBN 9780080960197. URL https://books.google.ch/books?id=ktCZ2MvgN3MC.

[27] Boris Solomyak. Lecture 5: Hyperbolic plane, geodesic and horocyclic flows, November 2020. URL https://u.math.biu.ac.il/~solomyb/TEACH/21/GrAct/Lec5b.pdf.

[28] Ian Stewart and David Tall. *Complex Analysis*. Cambridge University Press, August 2018. ISBN 9781108436793. doi: 10.1017/9781108505468. URL http://dx.doi.org/10.1017/9781108505468.

[29] John Stillwell. *Mathematics and Its History*. Springer New York, 2010. ISBN 9781441960535. doi: 10.1007/978-1-4419-6053-5. URL http://dx.doi.org/10.1007/978-1-4419-6053-5.

[30] Noam Szyfer. Elements of Geometry course, by Corinna Ulcigrai, June 2023.

[31] Wikipedia. Gauss–Kuzmin–Wirsing operator, . URL https://en.wikipedia.org/wiki/Gauss%E2%80%93Kuzmin%E2%80%93Wirsing_operator#/media/File:Gauss_function.svg. Accessed: 2024-06-27.

[32] Wikipedia. Fundamental Domain, . URL https://en.wikipedia.org/wiki/Fundamental_domain#/media/File:ModularGroup-FundamentalDomain.svg. Accessed: 2024-06-27.

[33] E. M. Wright. Continued fractions. by a. ya. khinchin. pp. xi, 95. 37s. 6d. (pap 14s.). 1964. (university of chicago press). *The Mathematical Gazette*, 49(369):334–334, 1965. doi: 10.1017/S0025557200053328.

[34] Sarah Zerbes. Modular Forms , 2021. URL https://metaphor.ethz.ch/x/2022/fs/401-4118-22L/notes/modular%20forms_1-26-27.pdf.