# Dynamics on Hyperbolic Surfaces with Quantitative Estimates

written by:
Salome Althaus

supervised by:
Prof. Alexander Gorodnik

Master's Thesis

Institute of Mathematics
University of Zurich
July 2024

**Abstract**

This thesis explores the dynamical properties of the geodesic and the horocycle flow defined on the unit tangent bundle of a hyperbolic Riemannian surface. We begin with the study of the complex upper half plane, which serves as a model for a hyperbolic space. We introduce basic concepts of hyperbolic geometry, including the hyperbolic distance and locally length minimizing curves.
The identification of the unit tangent bundle of the upper half plane with the linear group $PSL_2(\mathbb{R})$ allows a new perspective on the flows in consideration. It turns out that they can be understood as groups actions on $PSL_2(\mathbb{R})$. The dynamical properties of the geodesic and the horocycle flow can then be discovered by analyzing the group structure of $PSL_2(\mathbb{R})$ and the study of representation theory.
To obtain a meaningful dynamical system, we must identify a measure-theoretically small action space for these flows. This leads to the study of the quotient space $\Gamma \setminus PSL_2(\mathbb{R})$ by a lattice $\Gamma \leq PSL_2(\mathbb{R})$, revealing a natural finite measure on this space.
After establishing the qualitative aspects of convergence of the geodesic and the horocycle flow in the first part of the thesis, we develop the theory necessary to estimate the rate of convergence of their *mixing* properties. This involves a substantial amount of representation theory. The final part of the thesis deals with the work of M. Ratner in [Rat87], which is about this rate of convergence.

## Acknowledgements

I would like to express my sincere gratitude to Prof. Alexander Gorodnik not only for the support and guidance throughout the writing process of my Master's thesis, but also for all of his great lectures I have attended in the last two years, which sparked my interest in the theory of dynamical systems. I would also like to thank the people I've met at the University of Zurich for going with me through the good and not so good times.

Finally, I want to thank my cat Kleopatra, my family and Kevin for their invaluable support in many aspects of my life.

# Contents

**Introduction** **4**

**1 The Geodesic Flow and the Horocycle Flow** **8**
1.1 The Hyperbolic Plane viewed as a Riemannian Manifold . . . . 8
1.2 A Group Action on the Hyperbolic Plane . . . . . 10
1.3 The Geodesic Flow on the Unit Tangent Bundle of the Hyperbolic Plane . . . . . 15
1.4 The Horocycle Flow and its Connection to the Geodesic Flow . . 21

**2 A Metric on the Unit Tangent Bundle of the Hyperbolic Plane** **25**
2.1 A closed linear Group . . . . . 25
2.2 The Lie Algebra of the Closed linear Group G . . . . . 27
2.3 A Left-Invariant Riemannian Metric on a closed linear Group . . 29
2.4 The stable Manifold for the Geodesic Flow . . . . . 34
2.5 Metric on Quotients by discrete Subgroups of closed linear Groups . . . . . 35

**3 Dynamics on Quotient Spaces** **38**
3.1 Fundamental Domains . . . . . 38
3.1.1 The Boundary of a Dirichlet Region . . . . . 41
3.1.2 The Fundamental Domains for $PSL_2(\mathbb{R})$ . . . . . 42
3.2 Finite Measure on the Quotient Space $X = \Gamma \setminus PSL_2(\mathbb{R})$ . . . . . 43
3.2.1 Lattices in $PSL_2(\mathbb{R})$ . . . . . 44
3.2.2 Haar Measure . . . . . 46
3.3 Dynamics on $\Gamma \setminus PSL_2(\mathbb{R})$ . . . . . 50

**4 Ergodic Theory and Unitary Representations** **53**
4.1 Ergodicity of measure-preserving Transformations . . . . . 53
4.2 Associated Unitary Operators and Unitary Representations of metrizable Groups . . . . . 54
4.3 Ergodicity of a Group Action . . . . . 58

**5 Ergodicity of Transformations on on the Quotient Space of $SL_2(\mathbb{R})$ by a Lattice** **61**
5.1 Preliminaries . . . 62
5.2 Ergodicity of the Geodesic Flow on $\Gamma \setminus SL_2(\mathbb{R})$ . . . 63
5.3 Ergodicity for measure-preserving Transformations on $\Gamma \setminus SL_2(\mathbb{R})$ by individual Elements in $SL_2(\mathbb{R})$ . . . 65
5.4 Ergodicity of the Horocycle Flow on $\Gamma \setminus SL_2(\mathbb{R})$ . . . 68

**6 Mixing Dynamical Systems** **69**
6.1 Preliminaries . . . 69
6.2 Mixing of $G$ on the Quotient Space by a Lattice $\Gamma \leq G$ . . . 70

**7 Representation Theory on $SL_2(\mathbb{R})$** **74**
7.1 Preliminaries . . . 74
7.1.1 An Algebra Homomorphism associated to G . . . 76
7.1.2 Dirac Sequences on G . . . 78
7.1.3 Equivalent Representations and Schur's Lemma . . . 80
7.1.4 Direct Sums of finite-dimensional Representations . . . 82
7.1.5 Direct Integral . . . 83
7.2 A Decomposition for $SL_2(\mathbb{R})$ . . . 84
7.2.1 Preliminaries . . . 85
7.2.2 Eigenvectors for the Rotation Group $K$ . . . 89
7.3 Smooth Maps and Lie Derivatives . . . 91
7.3.1 The smooth Structure of $G$ . . . 91
7.3.2 The Lie Derivative on the Lie Algebra of $SL_2(\mathbb{R})$ . . . 93

**8 Quantitative Estimates of Mixing for the Geodesic and the Horocycle Flows** **95**
8.1 An orthonormal Basis consisting of Eigenvectors of K . . . 96
8.2 The Set of $K^p$-Vectors . . . 98
8.3 Rate of Decay for Unitary Irreducible Representations of G . . . 104
8.4 Rate of Mixing for the Geodesic and the Horocycle Flows on $L^2(X, \mu)$ . . . 123

# Introduction

In this thesis, we aim to provide an introduction to homogeneous dynamics, specifically focusing on the dynamical properties of the geodesic and the horocycle flows.

The concept of the *geodesic flow* provides insights into how paths known as *geodesics* evolve on hyperbolic surfaces. Geodesics are essentially the shortest paths between points on these surfaces (with respect to a hyperbolic metric), and the geodesic flow and its "perpendicular" counterpart, the *horocycle flow*, help us understand how these paths evolve over time. It is worth noting that the term "horocycle" derives from the Greek words "horros", meaning "limit", and "kyklos", meaning "circle". This term is used to describe a specific type of curve in hyperbolic geometry, which we will explain in more detail. Importantly, the term "horocycle" has no connection to the word "horror".

The study of geodesic and horocycle flows belongs to the field of homogeneous dynamics. Homogeneous dynamics refers to the study of flows on homogeneous spaces. A *flow* is a group action of the real numbers on a set. More formally, we refer to the map $\phi : M \times \mathbb{R} \to M$ satisfying the conditions

$$\phi(x, 0) = x \quad \text{and} \quad \phi(\phi(x, t), s) = \phi(x, t + s), \quad x \in M$$

as a flow on the space $M$.
*Homogeneous spaces* are spaces on which a group acts transitively, meaning the group can move any point to another point, making every point in the space look the same under the group's action. We will consider homogeneous spaces of the form

$$X = \Gamma \setminus G = \{\Gamma h \mid h \in G\}$$

where $G$ is a closed matrix group and $\Gamma \leq G$ is a *lattice* of $G$, which is a subgroup, such that the space $X = \Gamma \setminus G$ admits a finite measure $m_X$. The action of $G$ on $X$ is given by the mapping

$$(g, x) \mapsto xg^{-1}$$

$g \in G$, $x \in X$. The choice of this map is not arbitrary; it is the only way to define a (left) *action* of $G$ on $X$.

If there is a one-parameter subgroup $< g >$ of $G$, then the family of maps

$$\{\phi_t(x) = x(g^{-1})^t \quad | \quad t \in \mathbb{R}\}$$

is called a *flow* on $X$.

Given the rather geometric description of the geodesic and the horocycle flow at the beginning of this introduction, the relation to homogeneous dynamics is unclear. We will explain it in the first chapter of this thesis. To summarize, we begin by introducing the upper half of the complex plane $\mathbb{H}$ equipped with the hyperbolic metric, which represents our hyperbolic space. The unit tangent bundle $\mathrm{T}^1\mathbb{H}$ serves us as a model for a space on which the action of the geodesic and the horocycle flow is defined. We proceed by showing that the closed matrix group $PSL_2(\mathbb{R}) = SL_2(\mathbb{R})/\{\pm I_2\}$ (where $I_2$ denotes the identity of $SL_2(\mathbb{R})$) acts by Möbius transformations on $\mathbb{H}$ and that the derivative of this action is simply transitive on the unit tangent bundle above. In this way, we get an identification

$$\mathrm{T}^1\mathbb{H} \cong PSL_2(\mathbb{R})$$

through a specified map. This leads to the reinterpretation of these flows as group actions of $PSL_2(\mathbb{R})$ (restricted to a subgroup) on $PSL_2(\mathbb{R})$. After defining them on the quotient space $\Gamma \setminus PSL_2(\mathbb{R})$ of $PSL_2(\mathbb{R})$ by a lattice $\Gamma$, Chapter 3, will set the start of ergodic theory, by revealing a natural finite measure on the space $\Gamma \setminus PSL_2(\mathbb{R})$, invariant for the action of $PSL_2(\mathbb{R})$.

## Number Theory and Homogeneous Dynamics

Over the last four decades, the theory of homogeneous dynamics has emerged as a valuable tool for analyzing various problems in number theory, including Diophantine approximation and lattice counting problems. In studying the dynamical properties of systems such as the geodesic and horocycle flow discussed in this thesis, the initial focus is often on establishing qualitative properties of certain types of asymptotic behaviors, such as *ergodicity* and *mixing*. However, the *rate* of convergence of these properties typically are not captured in classical ergodic theory. Since this rate of convergence is crucial for insights into number theoretic problems, it is of particular interest to connect these qualitative properties to quantitative estimates in ergodic theorems.
While the first part of this thesis is devoted to the analysis of classical (qualitative) ergodic properties of the geodesic and the horocycle flow, the second part focuses on understanding the rate of a specific type of convergence. More precisely, we aim to explain the work of Ratner in [Rat87], which addresses the decay rate of mixing of these flows.

We shall illustrate the use of homogeneous dynamics in number theory in a few examples.

A particularly important area in Number Theory is *Diophantine approximation*, which investigates how "well" a vector $x \in \mathbb{R}^d$, $d \geq 1$ can be approximated by vectors with rational coefficients. Given the density of rational numbers in $\mathbb{R}$, it is evident that any vector $x \in \mathbb{R}^d$ can be approximated arbitrarily close by a rational vector. Diophantine approximation aims to define relevant criteria to classify the quality of these approximations and to explore the properties of points and sets satisfying these criteria. As is explained in [EFS09], the quality of approximation is classified by notions like *well approximable* (WA), *Dirichlet improvable* (DI), or *badly approximable* (BA).
Historically, Diophantine approximation has been studied using the continued fraction expansion (CFE) of real numbers. The limitation of this technique is that CFE is only applicable to the real numbers. To extend the analysis to $\mathbb{R}^d$ for $d > 1$, one can use homogeneous dynamics. However, even in the one-dimensional case, homogeneous dynamics proves to be a valuable tool for addressing questions in Diophantine approximation. We will illustrate this with the *Dani correspondence*, introduced by S.G. Dani in [Dan85].

Intuitively, $x \in \mathbb{R}$ is well approximable, if it can be approximated by rational numbers with *small* denominators. The well-known Theorem of Dirichlet states that

$$\forall x \in \mathbb{R}, \forall Q \geq 1 : \exists p, q \in \mathbb{Q} : |x - \frac{p}{q}| < \frac{1}{q \cdot Q} \text{ and } q \leq Q.$$

This Theorem provides a bound on how closely any real number can be approximated by rational numbers with denominators within a given range. It yields the definition of a badly approximable number as follows.

**Definition 1.** A number $x \in \mathbb{R}$ is called *badly approximable* if there exists a constant $c(x) > 0$, depending on $x$, such that for all integers $p, q$ with $q > 0$ it holds that

$$|x - \frac{p}{q}| \geq \frac{c(x)}{q^2}.$$

The Dani correspondence classifies badly approximable numbers by a property of the geodesic flow on a folded space $\Gamma \setminus PSL_2(\mathbb{R})$ (one may think of the example where $\Gamma = PSL_2(\mathbb{Z})$ is the *modular group*). It states that the orbit of the geodesic flow at a given point in $\Gamma \setminus PSL_2(\mathbb{R})$ is bounded if and only if the inverse of its limit point as $t \to \infty$ when viewed as a flow on $\mathrm{T}^1\mathbb{H}$ (which is a real number) is badly approximable.
The connection between the geodesic flow and badly approximable numbers is introduced in [Dyn23]. For a more advanced treatment, the reader is refered to [GGN14].

The next example we provide is the *Gauss circle problem*. It refers to the question about the number of integer points $N(r)$ of $\mathbb{R}^n$ contained in a ball of certain radius $r$ centered at the origin in this space. It can be generalized as follows. Let $G$ denote a closed matrix group and let $\Gamma \leq G$ be a lattice of $G$. Since $G$ is a closed matrix group (in particular, $G$ is a Lie group), it is

metrizable. Thus, we may choose a metric on $G$ and denote by $B_t$, $t > 0$ the balls of radius $t$ with respect to this metric, around the identity in $G$. The analogue to the Gauss circle problem above is the estimate of the quantity of

$$N(t) = |\Gamma \cap B_t|.$$

Now, as is explained in [Gor18], a mixing action of $G$ on $\Gamma \setminus G$ serves as a tool to understand the asymptotic behaviour of $|\Gamma \cap B_t|$ as $t \to \infty$. Informally, if $G = PSL_2(\mathbb{R})$ and $\Gamma \leq G$ is a lattice, then the quantity $N(t)$ can be visualized as the number of "integer" (or more precisely: lattice) points contained in a hyperbolic circle in $\mathbb{H}$ with center at $i$ and with unit outward normal vectors on its boundary (this is a circle in the unit tangent bundle of $\mathbb{H}$). It is a hyperbolic analogue to the Gauss circle problem.
The analysis of $N(t)$ belongs to the field of *hyperbolic counting problems*, which, like its Euclidean counterpart, is a subject of significant interest in modern research.

For further reading, A. Gorodnik in [GN13] and G. Margulis in [Mar08] offer detailed expositions with various examples illustrating the applications of homogeneous dynamics in number theory.

# Chapter 1

# The Geodesic Flow and the Horocycle Flow

We begin this chapter by introducing the *hyperbolic plane*, following closely Chapter 9 in [EW10]. Equipped with a specific distance measurement called the *hyperbolic metric*, the hyperbolic plane serves as a concrete example of a Riemannian manifold. Through our study of the hyperbolic plane, we will lay the foundation for defining the space upon which the geodesic and the horocycle flow are defined.

## 1.1 The Hyperbolic Plane viewed as a Riemannian Manifold

Let $M$ be a smooth manifold. For a point $p \in M$, we write $\mathrm{T}_p(M)$ for the set containing all tangent vectors at $p$. Then $\mathrm{T}_p(M)$ is a vector space called the *tangent space* of $M$ at $p$. We may think of $\mathrm{T}_p(M)$ as a plane touching the manifold $M$ tangentially at the point $p$ while having no other intersections with $M$. It captures all possible directions in which a curve can pass through that point, making it the natural space for studying the differential of curves in the manifold.

**Definition 2.** A *Riemannian metric* on a smooth manifold $M$ is a collection of inner products $<,>_p$, $p \in M$ defined on the tangent space $\mathrm{T}_p(M)$.

A *Riemannian manifold* refers to a smooth manifold endowed with a Riemannian metric. A classical example for a Riemannian manifold is the upper half plane $\mathbb{H}$ defined as

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

together with the hyperbolic metric, which is a specific example of a Riemannian metric.

**Definition 3.** The *hyperbolic Riemannian metric* is defined as the collection of inner products

$$< v, w >_z = \frac{1}{y^2}(v \cdot w)$$

for $z = x + iy \in \mathbb{H}$ and $v, w \in \mathrm{T}_z\mathbb{H}$ where $(v \cdot w)$ denotes the usual inner product in $\mathbb{C}$ under the identification with $\mathbb{R}^2$ as a real vector space.

The hyperbolic metric on $\mathbb{H}$ is induced by the hyperbolic Riemannian metric through a specific interpretation of the tangent planes

$$\mathrm{T}_z\mathbb{H} = \{z\} \times \mathbb{C}$$

$z \in \mathbb{H}$. As indicated above, the structure of the $\mathrm{T}_z\mathbb{H}$ is a natural space for the derivative of a map $\phi : [0, 1] \to \mathbb{H}$ in the following way. We call $\phi$ as above *differentiable* at $t_0 \in [0, 1]$ if $\phi'(t_0)$ exists when viewed as a map into the ambient space $\mathbb{C}$. In that case, we define the derivative of $\phi$ at $t_0$ by

$$D\phi(t_0) = (\phi(t_0), \phi'(t_0)) \in \mathrm{T}_{\phi(t_0)}\mathbb{H}.$$

We may thus understand the space

$$\mathrm{T}\mathbb{H} = \mathbb{H} \times \mathbb{C} = \bigsqcup_{z \in \mathbb{H}} \mathrm{T}_z\mathbb{H}$$

as the space where the derivatives of curves as above are defined. We call $\mathrm{T}\mathbb{H}$ the *tangent bundle* of the manifold $\mathbb{H}$.
We let the *length* of a continuous and piecewise differentiable curve (also referred to as *path*) $\phi : [0, 1] \to \mathbb{H}$ be the value

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

where $\|D\phi(t)\|_{\phi(t)}$ denotes the length of the tangent vector $D\phi(t) \in \mathrm{T}_{\phi(t)}\mathbb{H}$ with respect to the norm derived from the inner product in the definition of the hyperbolic Riemannian metric at the point $\phi(t)$. We will refer to $\|D\phi(t)\|_{\phi(t)}$ as the *speed* of the path $\phi$ at time $t$.

The Riemannian hyperbolic metric induces the hyperbolic distance $d$ as follows.

**Definition 4.** The *hyperbolic distance* is defined as

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

$z_0, z_1 \in \mathbb{H}$, where the infimum is taken over all continuous piecewise differentiable paths $\phi$ with $\phi(0) = z_0$ and $\phi(1) = z_1$.

It is not clear a priori whether the infimum over paths connecting any two points $z_0, z_1 \in \mathbb{H}$ is attained. This problem will be adressed later.

We call the real line $\mathbb{R} \subseteq \mathbb{C}$ together with the point $\infty$ the *boundary* $\partial\mathbb{H}$ of $\mathbb{H}$. The scaling of the inner product in the Riemannian hyperbolic metric by the factor $\frac{1}{y^2}$ leads to an infinite distance between any point in $\mathbb{H}$ to a point $r$ in its boundary, where the distance is defined as an infimum over paths

$$\phi : [0,1] \to \mathbb{H} \cup \partial\mathbb{H}$$

with $\phi(t) \in \mathbb{H}$ for $t \in [0,1)$ and $\phi(1) = r$.

## 1.2 A Group Action on the Hyperbolic Plane

Continuing the discussion about the hyperbolic plane, we introduce a specific transformation induced by a group action. The latter is generally defined as follows.

**Definition 5.** Let $G$ be a group and let $X$ be a set. A left *group action* of $G$ on $X$ is a map

$$* : G \times X \to X$$
$$(g, x) \mapsto g * x$$

so that

1. $\forall g_1, g_2 \in G, \forall x \in X : g_1 * (g_2 * x) = (g_1 g_2) * x$.
2. $e * x = x, \ \forall x \in X$, where $e$ denotes the identity element of $G$.

**Remark 1.** Similarly, a right group action is a map $* : X \times G \to X$, $(x, g) \mapsto x * g$ satisfying the analogous axioms 1 and 2 above. We will, for the most part, be concerned with left group actions and therefore will call the latter just *group actions*.

We call the action of $G$ on $X$ *transitive* if

$$\forall x_0, x_1 \in X : \exists\ g \in G \text{ such that } g * x_0 = x_1$$

and we call it *simply transitive* if the above $g$ is unique.

**Remark 2.** We will omit the notation $(g, x) \mapsto g * x$ for the action of $g \in G$ on $x \in X$. Instead, we prefer to write $(g, x) \mapsto gx$ if the context makes it clear to which action this refers. In case it does not, we will write $(g, x) \mapsto g \cdot x$.

The *special linear group* of degree two over $\mathbb{R}$, denoted as $SL_2(\mathbb{R})$, consists of invertible $2 \times 2$-matrices with real coefficients, whose determinant is equal to one. It acts on $\mathbb{H}$ by the *Möbius transformation*

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \mapsto \frac{az + b}{cz + d} \tag{1.2.0.1}$$

To see this, we may check that the above map indeed defines an action. We have

$$cz + d = 0 \iff z = \frac{-d}{c} \in \mathbb{R} \text{ or } c = d = 0.$$

but $c = d = 0$ violates the condition that $\det(g) = ad - bc = 1$. This shows that the denominator in equation 1.2.0.1 is not equal to zero. Moreover, the imaginary part $\mathrm{Im}(gz)$ for any $g \in G$ and $z \in \mathbb{H}$ satisfies

$$\begin{aligned}\mathrm{Im}(gz) &= \mathrm{Im}\left(\frac{az+b}{cz+d}\right) = \mathrm{Im}\left(\frac{(az+b)(c\bar{z}+d)}{|cz+d|^2}\right) \\ &= \mathrm{Im}\left(\frac{ac|z|^2 + (ad-bc)z + bd}{|cz+d|^2}\right) = \frac{\mathrm{Im}(z)}{|cz+d|^2} > 0\end{aligned}$$

since $ad - bc = 1$ and $\mathrm{Im}(z) > 0$. This shows $gz \in \mathbb{H}$ for all $g \in SL_2(\mathbb{R})$, $z \in \mathbb{H}$. Additionally, it reveals the useful equation

$$\mathrm{Im}(gz) = \frac{\mathrm{Im}(z)}{|cz+d|^2}. \tag{1.2.0.2}$$

One may check by a straightforward computation that the associativity property

$$g \cdot (hz) = (gh) \cdot z, \quad g, h \in SL_2(\mathbb{R}), z \in \mathbb{H}$$

holds. We thus conclude that equation 1.2.0.1 indeed defines an action of $SL_2(\mathbb{R})$ on $\mathbb{H}$.

In general, Möbius transformations defined on the complex plane $\mathbb{C}$ refer to functions of the form

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

$z \in \mathbb{C}$, $a, b, c, d \in \mathbb{C}$, $ad - bc \neq 0$. They are viewed as functions with values in the extended complex plane $\mathbb{C} \cup \{\infty\}$.

**Remark 3.** An important property of Möbius transformations is that they preserve angles between curves. Said differently: Möbius transformations are *conformal* maps on $\mathbb{C} \cup \{\infty\}$ (see Proposition 4.5 in Chapter 4 of [Isa17]). They can be expressed as a composition of translations, dilations, rotations and inversions, all of which individually preserve angles. In particular, the image of lines and circles in the extended complex plane are again lines and circles, since within this space, these are precisely the geometrical objects of constant curvatures.

Notice that the matrix

$$-I_2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$$

acts trivially on $\mathbb{H}$. Thus equation 1.2.0.1 also describes an action of the *projective special linear group* which is defined as

$$PSL_2(\mathbb{R}) = SL_2(\mathbb{R})/\{\pm I_2\}$$

on $\mathbb{H}$.

In what follows, we shall denote by $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ the element

$$\pm g = \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PSL_2(\mathbb{R}).$$

The group action of $PSL_2(\mathbb{R})$ on $\mathbb{H}$ extends to an action on the tangent bundle $\mathrm{T}\mathbb{H}$ as follows. When viewed as a map from $\mathbb{H}$ to $\mathbb{H}$, $g$ is complex differentiable. Using that $\det(g) = 1$, a straightforward computation shows

$$g'(z) = \frac{1}{(cz+d)^2}.$$

We define the derivative map $Dg$ on the tangent bundle of $\mathbb{H}$ by

$$Dg : \mathrm{T}\mathbb{H} \to \mathrm{T}\mathbb{H}$$

$$Dg(z,v) = (g(z), g'(z)v) = \left( \frac{az+b}{cz+d}, \frac{v}{(cz+d)^2} \right)$$

$(z,v) \in \mathrm{T}\mathbb{H}$. For a fixed $z \in \mathbb{H}$, this is a linear map

$$(Dg)_z : \mathrm{T}_z\mathbb{H} \to \mathrm{T}_{g(z)}\mathbb{H}.$$

We claim that the mapping $(z,v) \mapsto Dg(z,v)$, $(z,v) \in \mathrm{T}\mathbb{H}$, $g \in PSL_2(\mathbb{R})$ defines a group action of the projective special linear group on the tangent bundle of the hyperbolic plane. Indeed, it is clear that the identity element of $PSL_2(\mathbb{R})$ fixes any vector $(z,v) \in \mathrm{T}\mathbb{H}$. To prove the claim, it remains to show property 1 of Definition 5, which follows from the chain rule for differentiation. We have

$$Dg(Dh(z,v)) = Dg(h(z), h'(z)v) = (g(h(z)), g'(h(z))h'(z)v)$$

$$= ((gh)(z), (gh)'(z)v) = D(gh)(z,v)$$

for all $g, h \in PSL_2(\mathbb{R})$ and $(z,v) \in \mathrm{T}\mathbb{H}$.

**Lemma 1.2.1.** *The action of $PSL_2(\mathbb{R})$ on $\mathbb{H}$ as in 1.2.0.1 has the following properties.*

1. *For any $z_0, z_1 \in \mathbb{H}$ and $g \in PSL_2(\mathbb{R})$, it holds that $d(gz_0, gz_1) = d(z_0, z_1)$ and so the action is isometric. Moreover, the derivative $Dg$ of the action of $g \in PSL_2(\mathbb{R})$ defined above preserves the Riemannian metric.*
2. *The action is transitive on $\mathbb{H}$.*
3. *The stabilizer of the point $i \in \mathbb{H}$, defined as*

$$Stab_{PSL_2(\mathbb{R})}(i) = \{g \in PSL_2(\mathbb{R}) : g(i) = i\}$$

*is the projective special orthogonal group*

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

*where the matrix group* $SO(2)$ *is the group of elements of the form*

$$\begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix}$$

$\theta \in \mathbb{R}$.

Observe that the second and the third property of the above lemma allows for a canonical identification

$$\mathbb{H} \cong PSL_2(\mathbb{R})/PSO(2)$$

in which the coset $gPSO(2)$ is identified with the point $g(i) \in \mathbb{H}$.

*Proof.* (1) We start by proving the second claim of the first point in Lemma 1.2.1 since the hyperbolic metric is defined using the hyperbolic Riemannian metric. Fix a point $z \in \mathbb{H}$ and consider two vectors $v, w \in \mathrm{T}_z\mathbb{H}$ under the action of $(Dg)_z$. We have that

$$\begin{aligned} < (Dg)_z v, (Dg)_z w >_{g(z)} &= \frac{1}{\mathrm{Im}(g(z))^2} \left( \frac{v}{(cz+d)^2}, \frac{w}{(cz+d)^2} \right) \\ &= \frac{|cz+d|^4}{\mathrm{Im}(z)^2} \left( \frac{v}{(cz+d)^2}, \frac{w}{(cz+d)^2} \right) \\ &= \frac{1}{\mathrm{Im}(z)^2}(v, w) = < v, w >_z \end{aligned} \tag{1.2.1.1}$$

where we used equation 1.2.0.2 and the property that multiplication of the vectors $v, w$ by $\frac{1}{(cz+d)^2} \in \mathbb{C}$ changes the Euclidean scalar product by a factor $|\frac{1}{(cz+d)^2}|^2$. Equation 1.2.1.1 then shows that the action of $(Dg)_z$ does not change the length of vectors in $\mathrm{T}_z\mathbb{H}$ for any point $z \in \mathbb{H}$. By definition of the function $L$, this implies for any continuous piecewise differentiable map $\phi : [0, 1] \to \mathbb{H}$ that $L(g\phi) = L(\phi)$ and hence

$$d(g(z_0), g(z_1)) = d(z_0, z_1)$$

$\forall z_0, z_1 \in \mathbb{H}$.

(2) The action of $PSL_2(\mathbb{R})$ on $\mathbb{H}$ is transitive if the point $i$ can be mapped to any other point $z = x + iy \in \mathbb{H}$. The matrix

$$\begin{pmatrix} \sqrt{y} & \frac{x}{\sqrt{y}} \\ 0 & \frac{1}{\sqrt{y}} \end{pmatrix}$$

is invertible with determinant equal to one and it maps $i$ to $z$.

(3) If $g(i) = i$, equation 1.2.0.2 implies $|ci + d|^2 = 1$. Since $c$ and $d$ are real, we then must have $c^2 + d^2 = 1$. Therefore, there exists some $\theta \in \mathbb{R}$ such that $c = \sin(\theta)$ and $d = \cos(\theta)$. Then

$$\frac{ai + b}{\sin(\theta)i + \cos(\theta)} = i \;\Leftrightarrow\; ai + b = i\cos(\theta) - \sin(\theta)$$

and so $a = \cos(\theta), b = -\sin(\theta)$.
The stabilizer of $i$ is thus the subgroup $SO(2)$, which, as a subset of $PSL_2(\mathbb{R})$, is precisely $PSO(2)$.

□

We established in Lemma 1.2.1 (i) that the action of $Dg$, $g \in PSL_2(\mathbb{R})$ on the tangent bundle $\mathrm{T}\mathbb{H}$ preserves the length of the tangent vectors in $\mathrm{T}_z\mathbb{H}$, $z \in \mathbb{H}$. Therefore, the restriction of $Dg$ on the *unit tangent bundle*

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

defines an action of $PSL_2(\mathbb{R})$ on $\mathrm{T}^1\mathbb{H}$. The next Lemma provides an important identification by classifying this action as simply transitive.

**Lemma 1.2.2.** *The action of $PSL_2(\mathbb{R})$ on $T^1\mathbb{H}$ is simply transitive. In this way, $T^1\mathbb{H}$ may be identified with $PSL_2(\mathbb{R})$.*

Notice that this resulting identification $\mathrm{T}^1\mathbb{H} \cong PSL_2(\mathbb{R})$ is not canonical. Instead, we have to explicitly fix a reference point $(z_0, v_0) \in \mathrm{T}^1\mathbb{H}$ and let it correspond to the identity element in $PSL_2(\mathbb{R})$. The identification is then a map given by $g \mapsto Dg(z_0, v_0)$ and the structure of group actions implies that under this map, the action of $PSL_2(\mathbb{R})$ on $\mathrm{T}^1\mathbb{H}$ corresponds to left multiplication of $PSL_2(\mathbb{R})$ on itself. It will turn out to be convenient to choose the reference vector $(z_0, v_0) = (i, i) \in \mathrm{T}^1\mathbb{H}$.

*Proof of Lemma 1.2.2.* Recall from Lemma 1.2.1 that the action on $\mathbb{H}$ is transitive. Thus, any base point $z$ of $(z, v) \in \mathrm{T}^1_z\mathbb{H}$ can be reached by some $g$, and it suffices to consider a unit tangent vector $v \in \mathrm{T}^1_i\mathbb{H}$ with base point $i$ and show that it can be mapped to any other vector in $\mathrm{T}^1_i\mathbb{H}$ by some unique $g$ in the stabilzer of $i$. The action of $g \in PSO(2)$ on $v \in \mathrm{T}_i\mathbb{H}$ is given by

$$(Dg)_i(v) = \frac{v}{(\sin(\theta)i + \cos(\theta))^2} = (\cos(2\theta) - i\sin(2\theta))v$$

Hence, by varying $\theta \in \mathbb{R}$, $(Dg)_i v$ could be any vector of unit length. Moreover, the condition $(Dg)_i v = v$ implies that $\theta \equiv 0 \pmod{\pi}$ and thus $g = \pm I_2$ which is the identity element of $PSO(2)$.

□

## 1.3 The Geodesic Flow on the Unit Tangent Bundle of the Hyperbolic Plane

We start by discussing the geodesic flow on the unit tangent bundle $\mathrm{T}^1\mathbb{H}$ that we introduced in the last section. The aim is to establish the necessary knowledge of basic hyperbolic geometry, including the notion of distances, geodesics and their parametrizations.
We have defined the hyperbolic distance $d(z_0, z_1)$ between two points $z_0, z_1 \in \mathbb{H}$ by an infimum over all continuous and piecewise differentiable curves connecting $z_0$ and $z_1$. As we mentioned, it is not clear whether such minimizing path exists. The first result of this section provides an answer to this question in the special case where both $z_0$ and $z_1$ lie on the imaginary axis contained in $\mathbb{H}$. Additionally, it indicates an important method for investigating problems concerning the hyperbolic geometry in $\mathbb{H}$: By using the distance-preserving property of the action by $PSL_2(\mathbb{R})$ on $\mathbb{H}$, metric questions about general curves in $\mathbb{H}$ will often be solved on the imaginary axis in $\mathbb{H}$ by a suitable transformation given by some element of $PSL_2(\mathbb{R})$.

**Lemma 1.3.1.** *Let $z_0 = y_0 i$, $z_1 = y_1 i$ with $0 < y_0 < y_1$. Then*

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

*and*

$$\phi(t) = y_0 \left(\frac{y_1}{y_0}\right)^t i$$

*for $t \in [0,1]$ defines a path in $\mathbb{H}$ from $z_0$ to $z_1$ with constant speed $\log(y_1) - \log(y_0)$, so that $L(\phi) = \log(y_1) - \log(y_0)$. Moreover, the curve $\phi$ is unique in the sense that if there exists another path $\psi : [0,1] \to \mathbb{H}$ going from $z_0$ to $z_1$ with $L(\psi) = d(z_0, z_1)$, then there is some increasing, piecewise differentiable map $f : [0,1] \to [0,1]$ with $\psi = \phi \circ f$.*

We remark that this Lemma does not only confirm the existence of a minimizing path for the points in the imaginary axis contained in $\mathbb{H}$, but tells us that it is unique up to reparametrization and lies itself in $\{yi \mid y > 0\}$. Due to this, the whole set $\{yi \mid y > 0\}$ will be called a *geodesic* in $\mathbb{H}$ and the minimizing path $\phi$ will be called a *geodesic path*.

*Proof.* The speed $\|D(\phi(t))\|_{\phi(t)}$ of the path $\phi$ at time $t$ equals

$$\begin{aligned}
\|D(\phi(t))\|_{\phi(t)} &= \|d/dt\ y_0 \exp(t \log(y_1/y_0))i\|_{\phi(t)} = \|y_0 \log(y_1/y_0)(y_1/y_0)^t i\|_{\phi(t)} \\
&= \sqrt{\frac{1}{\mathrm{Im}(\phi(t))}} \left| y_0 \log\left(\frac{y_1}{y_0}\right) \left(\frac{y_1}{y_0}\right)^t \right| \\
&= \left(\frac{y_0^{t-1}}{y_1^t}\right) \left(\frac{y_1^t}{y_0^t}\right) y_0 \log\left(\frac{y_1}{y_0}\right) \\
&= \log\left(\frac{y_1}{y_0}\right)
\end{aligned}$$

and therefore,

$$d(z_0, z_1) \leq L(\phi) = \log(y_1) - \log(y_0).$$

To see that $\phi$ is the smallest path connecting $z_0$ and $z_1$, consider any differentiable curve $\eta : [0,1] \to \mathbb{H}, \eta(t) := \eta_x(t) + i\eta_y(t)$ connecting $z_0$ and $z_1$. We have

$$\begin{aligned} L(\eta) &= \int_0^1 \frac{\|\eta'(t)\|_2}{\eta_y(t)} dt \geq \int_0^1 \frac{|\eta_y'(t)|}{\eta_y(t)} dt \\ &\geq \int_0^1 \frac{\eta_y'(t)}{\eta_y(t)} dt = \log(y_1) - \log(y_0) \end{aligned}$$

where we used the substitution $t = \eta_y(t)$ in the last equality. 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 $\eta_y'(t) > 0$ for all $t \in [0,1]$. Therefore, $\eta$ is a minimizing path connecting $z_0$ and $z_1$ if and only if it is increasing and it lies entirely on $\{yi \mid y > 0\}$, verifying the last statement of the Lemma. □

It will be convenient to parametrize geodesic paths, so that they have constant speed equal to one (and therefore have a domain whose length is equal to the length of the path). In view of Lemma 1.3.1, we then get a stronger notion of uniqueness for geodesics connecting points on the set $\{yi \mid y > 0\}$. More precisely, we obtain that for any $z_0 = iy_0$ and $z_1 = iy_1$ lying on the upper imaginary axis, there is a unique path

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

of unit speed with $\phi(0) = z_0$ and $\phi(d(z_0, z_1)) = z_1$. Since the unit interval gets scaled by the factor $d(z_0, z_1) = \log(y_1) - \log(y_0)$, the substitution $t = (\log(y_1) - \log(y_0))t$ and Lemma 1.3.1 shows that this path $\phi$ is then defined by

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

The isometry $g \in PSL_2(\mathbb{R})$ satisfies

$$L(g\phi) = L(\phi)$$

for any continuous and piecewise differentiable path $\phi : [a,b] \to \mathbb{H}$ as a consequence of Lemma 1.2.1 (1). From this we see that if $\phi$ is a geodesic path on the imaginary axis, then the curve $g\phi$, defined by the mapping $t \mapsto g\phi(t)$ is another geodesic path, connecting points in the left translated image of $\phi$ by $g$. We are now interested in the converse of this observation; whether all geodesics in $\mathbb{H}$ arise as an image of the imaginary axis under some $g \in PSL_2(\mathbb{R})$.
In order to answer this question, a closer look at the generators of the group $G = SL_2(\mathbb{R})$ is needed. Let

$$w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \in G$$

and let

$$U^- = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} : b \in \mathbb{R} \right\}$$

be the subgroup of $G$ consisting of upper unitriangular matrices. The following Lemma is proven similarly as in [htta].

**Lemma 1.3.2.** *The subgroup $U^-$ together with the element $w$ defined above generate $SL_2(\mathbb{R})$.*

*Proof.* We have

$$wU^-w^{-1} = \left\{ \begin{pmatrix} 1 & 0 \\ -b & 1 \end{pmatrix} : b \in \mathbb{R} \right\}$$

so the subgroup $U^-$ together with $w$ generate the subgroup of lower unitriangular matrices in $G$. We may generate the diagonal elements of $G$ by a suitable product of upper and lower unitriangular matrices as follows. We have

$$\begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix} \begin{pmatrix} 1 & \delta \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1+\beta\gamma & \delta(1+\beta\gamma)+\beta \\ \alpha(1+\beta\gamma)+\gamma & \alpha\delta(1+\beta\gamma)+\alpha\beta+\gamma\delta+1 \end{pmatrix} \tag{1.3.2.1}$$

First consider the case $1+\beta\gamma \neq 0$. Choose

$$\alpha = \frac{-\gamma}{1+\beta\gamma},\ \delta = \frac{-\beta}{1+\beta\gamma}$$

to get the anti-diagonal terms to be zero. The product above then becomes

$$\begin{pmatrix} 1+\beta\gamma & 0 \\ 0 & \frac{1}{1+\beta\gamma} \end{pmatrix}$$

as required.
Given an element

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$$

with $a \neq 0$, we have

$$\begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} \begin{pmatrix} x & 0 \\ 0 & 1/x \end{pmatrix} \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} x & \beta x \\ \alpha x & \alpha\beta x + 1/x \end{pmatrix}$$

$\alpha, \beta \in \mathbb{R}$ and $x \in \mathbb{R} \setminus \{0\}$, is equal to $g$ for a suitable choice of the parameters involved.
Otherwise, when $a = 0$ in the above matrix $g$, set $1+\beta\gamma = 0$. Equation 1.3.2.1 becomes

$$\begin{pmatrix} 0 & \beta \\ \gamma & 1+\alpha\beta+\gamma\delta \end{pmatrix}.$$

Then let $\delta = 0$ and we may choose the other parameters suitably, so that we get a matrix of the form

$$\begin{pmatrix} 0 & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R}).$$

□

Observe that the proof of Lemma 1.3.2 also shows that $SL_2(\mathbb{R})$ is generated by its upper and lower unitriangular matrices.

**Lemma 1.3.3.** *An element $g \in PSL_2(\mathbb{R})$, acting by Möbius transformation on $\mathbb{H}$, maps the geodesic $\{yi \mid y > 0\}$ either to another vertical line, or to the upper half of the circle*

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

*with center in the real axis (equivalently, meeting the real line at right angles).*

*Proof.* To prove these statements, we begin by noticing that elements $g \in PSL_2(\mathbb{R})$ of the form

$$\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$$

for some $b \in \mathbb{R}$ map the vertical line $\{yi \mid y > 0\}$ to the translated vertical line defined by $\{z \mid \mathrm{Re}(z) = b\}$. Moreover, the image of a upper half circle $C$ of the form $(x-c)^2 + y^2 = r^2$ for some $a \in \mathbb{R}$ under this element is the circle obtained by shifting the center of $C$ by a real constant $b$.
The mapping $z \mapsto -\frac{1}{z}$, corresponding to the element

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

and is given in polar coordinates by $(r, \phi) \mapsto (\frac{1}{r}, \pi - \phi)$, for $z = re^{i\phi}$. Both vertical lines and circles with real centers can be defined by equations of the form

$$\alpha r^2 + \beta r \cos(\phi) + \gamma = 0$$

which verifies the assertion of the Lemma for $w$. The proof is then complete by Lemma 1.3.2. □

For the proof of the next result, we borrowed some of the ideas of [BM00] in the proof of Lemma 1.6 in Chapter 2.

**Proposition 1.** For any two points $z_0, z_1 \in \mathbb{H}$ there is a unique path

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

of unit speed with $\phi(0) = z_0$ and $\phi(d(z_0, z_1)) = z_1$. Moreover, there is a unique element $g \in PSL_2(\mathbb{R})$ such that $\phi(t) = g(e^t i)$. Thus any geodesic is the image of the set $\{yi \mid y > 0\}$ under some $g$. It follows that the geodesics in $\mathbb{H}$ are the vertical lines and the half-circles with real center.

*Proof.* We first consider the case where $z_0, z_1 \in \mathbb{H}$ have the same imaginary part. Then the matrix

$$g = \begin{pmatrix} 1 & -\mathrm{Re}(z_0) \\ 0 & 1 \end{pmatrix}$$

maps the points $z_0$ and $z_1$ to the imaginary axis and the first claim in the Proposition follows from Lemma 1.3.1. Otherwise, we consider a circle $C$ going through $z_0$ and $z_1$ with center on the real axis. Denote by $t_1$ and $t_2$ the intersecting points of $C$ with $\mathbb{R}$. Consider the Möbius transformation given by

$$g = \begin{pmatrix} 1 & -t_2 \\ t_2 - t_1 & -t_1(t_2 - t_1) \end{pmatrix}$$

$$g: \ z \mapsto \frac{-1}{z - t_1} + \frac{1}{t_2 - t_1}, \ z \in \mathbb{C} \cup \{\infty\}$$

and notice that $g$ maps $t_1$ to $\infty$ and $t_2$ to 0. It follows from Remark 3 that $C$ gets mapped to a line which contains the point 0 and intersects the real axis at a right angle (because the center of $C$ is real). Hence, we see that $C$ gets mapped to the imaginary axis.
Set $g' = \frac{1}{t_2 - t_1} g$ so that $\det(g') = 1$ and $g' \in SL_2(\mathbb{R})$. Observe that $g'$ maps $t_1$ and $t_2$ to the same points as $g$ does. Thus the argument above shows that the image of $C$ under $g'$ is still the imaginary axis. The first claim then follows again by Lemma 1.3.1.
To finish the proof, we are left to show that not only is the geodesic path $\phi$ connecting $z_0, z_1 \in \mathbb{H}$ unique, but there is a unique element $g \in PSL_2(\mathbb{R})$ satisfying $\phi(t) = g(\mathrm{e}^t i)$. Let $v_0$ denote the unit tangent vector at $z_0$ which points in the direction of the geodesic path connecting $z_0$ and $z_1$, at the point $z_0$. By Lemma 1.2.2, there exists a unique $g \in PSL_2(\mathbb{R})$ such that $Dg(z_0, v_0) = (i, i)$. The geodesic connecting the points $z_0$ and $z_1$ is either an upper half circle $C$ with real center or a vertical line $l$. In any case, Lemma 1.3.3 implies that $g$ maps $C$ (respectively, $l$) to the imaginary axis which is determined by $(i, i)$. Hence, $gz_1 = y_1 i$ for some $y_1 > 0$. By Lemma 1.3.1, there is a unique geodesic path $\phi_0(t) = \mathrm{e}^t i$ of unit speed from $i$ to $y_1 i$, so that $\phi(t) = g^{-1}(\mathrm{e}^t i)$ is the unique geodesic path of unit speed connecting $z_0$ and $z_1$.

□

The geodesics in the hyperbolic plane satisfy all the fundamental principles that underlie Euclidean geometry, apart from the parallel axiom, showing that the latter is not a consequence of the other axioms. For instance, given any two distinct points in the hyperbolic plane, we've shown that there exists a unique geodesic going through them, and any two geodesics intersect in at most one point. However, given any point $z \in \mathbb{H}$ and a geodesic curve $l$ not containing $z$, there are uncountably many geodesics through $z$ that do not intersect $l$ and which are therefore "parallel" to $l$.
Notice that Proposition 1 shows (together with its proof) that a geodesic $l$ is determined by the array $(z, v) \in \mathrm{T}^1\mathbb{H}$ under the requirement that $l$ passes through $z$ in direction $v$. Moreover, for an arbitrary point $(z, v) \in \mathrm{T}^1\mathbb{H}$ there exists a unique $g \in PSL_2(\mathbb{R})$ with $Dg(z, v) = (i, i)$ and thus any geodesic in $\mathbb{H}$ can be mapped to the set $\{yi \mid y > 0\}$. Equivalently, all geodesics arise from an image of $\{yi \mid y > 0\}$ .

Given an array $(z, v) \in \mathrm{T}^1\mathbb{H}$, we now proceed by investigating the evolution of a point $p_t = p_t(z, v)$ following the unique parametrization of the geodesic determined by $(z, v)$ for time $t \in \mathbb{R}$ with starting point $p_0 = p_0(z, v) = (z, v)$.

**Definition 6.** The family of maps

$$p_t : \mathrm{T}^1\mathbb{H} \to \mathrm{T}^1\mathbb{H},\ t \in \mathbb{R}$$

$$(z, v) \mapsto p_t((z, v))$$

is called the *geodesic flow* on $\mathrm{T}^1\mathbb{H}$.

In the case of the geodesic $(i, i)$, this means that

$$p_t((i, i)) = (\mathrm{e}^t i, \mathrm{e}^t i).$$

Denote

$$a_t = \begin{pmatrix} \mathrm{e}^{-t/2} & 0 \\ 0 & \mathrm{e}^{t/2} \end{pmatrix}.$$

A straightforward computation shows that the derivative action of $a_t^{-1}$ on the array $(i, i)$ is equal to $(\mathrm{e}^t i, \mathrm{e}^t i)$. Hence

$$p_t((i, i)) = Da_t^{-1}(i, i). \tag{1.3.3.1}$$

This motivates the following consideration. The action of the geodesic flow is defined on the space $\mathrm{T}^1\mathbb{H}$, which we identified with $PSL_2(\mathbb{R})$ in Lemma 1.2.2. This identification sets up a bijection

$$\mathrm{T}^1\mathbb{H} \to PSL_2(\mathbb{R}),\ (z, v) \mapsto g_{z,v}$$

where $g_{z,v}$ is the unique element in $PSL_2(\mathbb{R})$ such that $Dg_{z,v}(i, i) = (z, v)$. Then for any $h \in PSL_2(\mathbb{R})$, $g_{z,v}$ is the unique element such that $D(hg_{z,v})(i, i)) = Dh(Dg_{z,v}(i, i)) = Dh(z, v)$. Hence

$$Dh(z, v) \mapsto hg_{z,v}. \tag{1.3.3.2}$$

This shows that the derivative action of $PSL_2(\mathbb{R})$ on $\mathrm{T}^1\mathbb{H}$ commutes with the left multiplication in this group. One may thus wonder whether the above bijection provides a correspondence between the geodesic flow on $\mathrm{T}^1\mathbb{H}$ and an action by a subgroup in $PSL_2(\mathbb{R})$ on $PSL_2(\mathbb{R})$ itself. Indeed, we obtain a correspondence as follows. The next Proposition is Proposition 3.2 in Chapter 2 of [BM00].

**Proposition 2.** The geodesic flow on $\mathrm{T}^1\mathbb{H}$ corresponds to the flow on the group $PSL_2(\mathbb{R})$ given by the right translations

$$g \mapsto ga_t^{-1}$$

$g \in PSL_2(\mathbb{R})$, $t \in \mathbb{R}$.

*Proof.* Given any point $(z, v) \in \mathrm{T}^1\mathbb{H}$, let $g_{z,v} \in PSL_2(\mathbb{R})$ be defined as above. We have seen that $Dg_{z,v}$ acts as an isometry on geodesics. Therefore, $Dg_{z,v}$ maps the unit speed parametrization of the geodesic defined by $(i, i)$ to the unit speed parametrization of the geodesic defined through $(z, v)$. It follows that the action of $Dg_{z,v}$ and the one of the geodesic flow $p_t$ on $(i, i)$ must commute. We have

$$p_t((z, v)) = p_t(Dg_{z,v}(i, i)) = Dg_{z,v}(p_t(i, i)) = D(g_{z,v}a_t^{-1})(i, i)$$

by equation 1.3.3.1. □

This describes the geodesic flow on $PSL_2(\mathbb{R})$ as right multiplication by the inverse of the matrix $a_t$, which we will denote as

$$R_{a_t}(g) = ga_t^{-1}$$

$g \in PSL_2(\mathbb{R})$. To summarize, we have the correspondence

$$(z, v) \mapsto p_t(z, v) \quad \longleftrightarrow \quad g_{z,v} \mapsto R_{a_t}(g_{z,v}) = g_{z,v}a_t^{-1}$$

$g_{z,v} \in PSL_2(\mathbb{R})$, $(z, v) \in \mathrm{T}^1\mathbb{H}$, $t \in \mathbb{R}$.

## 1.4 The Horocycle Flow and its Connection to the Geodesic Flow

We begin this section by some observations concerning the geodesic flow, which will motivate the definition of a certain transverse flow, called the *horocycle flow* defined on the unit tangent bundle $\mathrm{T}^1\mathbb{H}$ of the hyperbolic plane $\mathbb{H}$.

The geodesic curve going through the vector $(x + i, i)$, $x \in \mathbb{R}$ is a vertical line parallel to the geodesic $\{yi \mid y > 0\}$ . The respective geodesic flows therefore move along parallel to each other. We have

$$p_t((i, i)) = (\mathrm{e}^t i, \mathrm{e}^t i);$$

$$p_t((x + i, i)) = (x + \mathrm{e}^t i, \mathrm{e}^t i),\ t \in \mathbb{R}.$$

We claim that the hyperbolic distance between their base points does not exceed $\frac{|x|}{\mathrm{e}^t}$. Indeed, the horizontal path $\phi$ from $\mathrm{e}^t i$ to $x + \mathrm{e}^t i$ defined by

$$\phi(s) = \mathrm{e}^t i + sx,\ s \in [0, 1]$$

is of length

$$L(\phi) = \int_0^1 \|D\phi(s)\|_{\phi(s)} ds = \frac{1}{\mathrm{e}^t} \int_0^1 |x| ds = \frac{|x|}{\mathrm{e}^t}$$

proving the claim. We may assume that points $p_t((i, i))$ and $p_t((x + i, i))$ on the orbits of the geodesic flow approach each other as $t \to \infty$. However, as we

have not yet defined a metric on $\mathrm{T}^1\mathbb{H}$, there is no rigorous argument proving this statement about convergence. We will come back to this issue in the next chapter. For now, we may nonetheless continue this discussion using only the hyperbolic metric on $\mathbb{H}$. We claim that if the vector $(z, v) \in \mathrm{T}^1\mathbb{H}$ has the property that the distance between the base point of $p_t((z, v))$ and that of $p_t((i, i))$ goes to zero as $t \to \infty$, then $\mathrm{Im}(z) = 1$ and $v = i$.
First, we remark that $v$ must be purely imaginary, for otherwise the geodesic trajectory $p_t(z, v)$ describes a half circle, which intersects a point $r \in \mathbb{R}$ at infinity. Since the base point of $p_t(i, i)$ goes to the cusp $i\infty$, one can check that it would not be close to $r$. Write $v = \lambda i$, for some $\lambda > 0$. Then the condition

$$1 = \|v\|_z = \frac{\|v\|}{\mathrm{Im}(z)}$$

shows that $\mathrm{Im}(z) = \lambda$. Denote $z = x + i\lambda$. The coefficients $a, b, c, d$ of $g_{z,v}$ must satisfy

$$\left(\frac{ai + b}{ci + d},\ \frac{i}{(ci + d)^2}\right) = (x + i\lambda,\ i\lambda).$$

This implies

$$\frac{1}{\lambda} = (ci + d)^2 = -c^2 + 2icd + d^2$$

and so $cd = 0$. If $d = 0$, then $\frac{1}{\lambda} = -c^2$, which is impossible. Thus, $c = 0$ and $d = \lambda^{-1/2}$. It follows that

$$g_{z,v} = \begin{pmatrix} \lambda^{1/2} & x\lambda^{-1/2} \\ 0 & \lambda^{-1/2} \end{pmatrix} \quad \text{and} \quad p_t(z, v) = (\lambda(x + \mathrm{e}^t i), \lambda \mathrm{e}^t i).$$

We have

$$d(\lambda(x + \mathrm{e}^t i), \mathrm{e}^t i) = d(x + \mathrm{e}^t i, \lambda \mathrm{e}^t i) \geq |d(x + \mathrm{e}^t i, \mathrm{e}^t i) - d(\lambda \mathrm{e}^t i, \mathrm{e}^t i)|$$

where the inequality holds since the metric is invariant under the multiplication of $\begin{pmatrix} \lambda^{-1/2} & 0 \\ 0 & \lambda^{1/2} \end{pmatrix}$. If $\lambda \neq 1$, the distance $d(\lambda \mathrm{e}^t i, \mathrm{e}^t i)$ is equal to $|\log(\lambda)|$ by Lemma 1.3.1, and thus independent of $t$. Moreover, we have seen that $d(x + \mathrm{e}^t i, \mathrm{e}^t i) \to 0$ as $t \to \infty$. Hence, the distance between the base points of $p_t(z, v)$ and $p_t(i, i)$ is bounded from below for all $t \in \mathbb{R}$ and thus does not converge to zero. We conclude that $\lambda = 1$, proving our claim.

Combining the two statements above, we may presume that the set of points $(z, v)$ with the property that the distance between $p_t((i, i))$ and $p_t((z, v))$ converges to zero as $t \to \infty$ is precisely the horizontal line through $i$ with upward pointing vectors. Once we defined a metric $d$ on $\mathrm{T}^1\mathbb{H}$, coming from the left-invariant Riemannian metric on $PSL_2(\mathbb{R})$, we will call this set the *stable manifold* for the geodesic flow going through the vector $(i, i)$, and we denote it by

$W^{ss}(i,i)$. We will see that it is equal to the orbit of $(i,i)$ under the action of the subgroup

$$U^- = \left\{ \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, \ s \in \mathbb{R} \right\}. \tag{1.4.0.1}$$

In general, given a point $(z,v) \in \mathrm{T}^1\mathbb{H}$, the set

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

defines the stable manifold for the geodesic flow going through $(z,v)$. The above description of the set $W^{ss}(i,i)$ shows that it must be equal to the orbit of $(z,v)$ under the action of the subgroup $U^-$. This can be seen by the usual argument which we briefly shall repeat. Let $g \in PSL_2(\mathbb{R})$ be the unique element with $Dg(i,i) = (z,v)$. As $Dg$ is an isometry on $\mathrm{T}^1\mathbb{H}$, the image of $W^{ss}(i,i)$ under $Dg$ is precisely the set $W^{ss}(z,v)$. Then $W^{ss}(z,v)$ is the orbit of $(i,i)$ under $Dg$ applied to the action of the subgroup $U^-$. Said differently, $(z',v') \in W^{ss}(z,v)$ if and only if there exists an $s \in \mathbb{R}$ with $Dg(Du^-(s)(i,i)) = D(gu^-(s))(i,i) = (z',v')$, where

$$u^-(s) = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$$

The (stable) *horocycle flow* through $(z,v) \in \mathrm{T}^1\mathbb{H}$ can be understood by following the trajectory determined by the manifold $W^{ss}(z,v)$ for time $t \in \mathbb{R}$. More precisely, we define it as follows.

**Definition 7.** The *(stable) horocycle flow* $\{u^-(s), s \in \mathbb{R}\}$ is defined on the unit tangent bundle $\mathrm{T}^1\mathbb{H}$ by

$$(z,v) \mapsto u^-(s)((z,v)) = D\left( g_{z,v} \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix} \right)(i,i)$$

$s \in \mathbb{R}$. We get a corresponding description in $PSL_2(\mathbb{R})$ by

$$(z,v) \mapsto u^-(s)((z,v)) \quad \longleftrightarrow \quad g_{z,v} \mapsto R_{u^-(s)}(g_{z,v}) = g_{z,v}u^-(-s).$$

The *unstable manifold* for the geodesic flow going through $(z,v) \in \mathrm{T}^1\mathbb{H}$ is defined as the set

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

A similar argument as above shows that $W^{su}(z,v)$ is equal to the orbit of $(z,v)$ under the action of the subgroup

$$U^+ = \left\{ \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}, \ s \in \mathbb{R} \right\}.$$

We define the *unstable horocycle flow* to be the set of functions $\{u^+(s), s \in \mathbb{R}\}$ defined on $\mathrm{T}^1\mathbb{H}$ by the mapping

$$(z,v) \mapsto u^+(s)((z,v)) = D\left( g_{z,v} \begin{pmatrix} 1 & 0 \\ -s & 1 \end{pmatrix} \right)(i,i)$$

where $u^+(s) = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix} \in U^+$. The corresponding description as an action of the group $PSL_2(\mathbb{R})$ on itself is given analogously as above.

**Remark 4.** We defined both the geodesic and the horocycle flow in $PSL_2(\mathbb{R})$ as right multiplication by the inverse of the matrices $a_t$, $u^-(s)$ and $u^+(s)$ on $PSL_2(\mathbb{R})$ itself. The inversion is needed so that the right multiplication becomes a (left) action of $PSL_2(\mathbb{R})$ on itself, which we are going to consider later. If the matrices were not inverted, we would have

$$R_{g_1}(R_{g_2}h) = R_{g_1}(hg_2) = hg_2g_1 \neq R_{g_1g_2}(h)$$

$g_1, g_2, h \in PSL_2(\mathbb{R})$ and condition (1) in Definition 5 would not hold.

# Chapter 2

# A Metric on the Unit Tangent Bundle of the Hyperbolic Plane

Any *closed linear group* $G$ can be shown to be locally diffeomorphic to its *Lie algebra*, a real vector space, thereby endowing $G$ with a smooth manifold structure. Moreover, this Lie Algebra allows for an identification of the tangent bundle $\mathrm{T}G$ of $G$, and thus for the construction of a Riemannian metric on $G$. As indicated in the last section, it is necessary to define a metric on $\mathrm{T}^1\mathbb{H}$ in order to establish a notion of convergence within this space. We do this using the identification $\mathrm{T}^1\mathbb{H} \cong PSL_2(\mathbb{R})$. More precisely, we aim to define a left-invariant Riemannian metric on the linear group $PSL_2(\mathbb{R})$ which naturally serves as a metric on $\mathrm{T}^1\mathbb{H}$ through the distance preserving identification $g \mapsto Dg(i, i)$. Rather than offering a detailed introduction to the underlying theory leading to these results, our focus lies on presenting some of the key foundational concepts essential for our subsequent discussion. A more careful treatment concerning the Lie algebra of a closed linear group can be found in Chapter 9 of [EW10], which we follow in less detail in this chapter.

## 2.1 A closed linear Group

A linear group $G$ is called *closed* if it can be embedded into the group $GL_d(\mathbb{R})$ for some $d \geq 1$, such that the image is a closed set in $GL_d(\mathbb{R})$. Here, $GL_d(\mathbb{R})$ refers to the group of invertible $d \times d$ matrices with real coefficients, endowed with the topology inherited from the Euclidean topology of the vector space $\mathbb{R}^{d^2}$.

**Lemma 2.1.1.** *The matrix group $SL_2(\mathbb{R})$ is closed.*

*Proof.* This follows from the fact that the determinant map is continuous. Indeed, define

$$\det : SL_2(\mathbb{R}) \to \mathbb{R}, \quad g \mapsto \det(g).$$

Then the image of $SL_2(\mathbb{R})$ is the closed set $\{1\} \subset \mathbb{R}$. □

Importantly, the linear group $PSL_2(\mathbb{R})$ turns out to be a closed linear group. To see this, we follow the presentation in [EW10] on pages 284-285. Consider the map

$$\phi : Sl_2(\mathbb{R}) \to Gl(Mat_{22}(\mathbb{R})) \cong Gl_4(\mathbb{R})$$
$$(\phi(g))m = gmg^{-1}$$

where $Gl(V)$ denotes the group of invertible and linear maps $g : V \to V$ on a vector space $V$. It is clear that $\phi$ is a group homomorphism. Moreover, both $\phi(I)$ and $\phi(-I)$ is the identity element of $Gl(Mat_{22}(\mathbb{R})$. Therefore, $\phi$ descends to

$$\hat{\phi} : PSL_2(\mathbb{R}) \to Gl(Mat_{22}(\mathbb{R}))$$

We need to show that the limit point of any converging sequence $(h_n) \subseteq \hat{\phi}(PSL_2(\mathbb{R}))$ is contained in $\hat{\phi}(PSL_2(\mathbb{R}))$. To do so, it is enough to show that $\hat{\phi}$ is an injective and *proper* map, where proper means that the pre-image of any compact set $K \subseteq Gl(Mat_{22}(\mathbb{R}))$ under $\hat{\phi}$ is again compact as a subset in $PSL_2(\mathbb{R})$.
Indeed, suppose $\hat{\phi}$ is injective and proper. Let $(h_n)$ denote any sequence in the image of $PSL_2(\mathbb{R})$ under $\hat{\phi}$, converging to some element $h \in Gl(Mat_{22})$. As $\hat{\phi}$ is injective, the pre-image $\hat{\phi}^{-1}(h_n)$, $n \in \mathbb{N}$ corresponds to exactly one element $g_n \in PSL_2(\mathbb{R})$. Moreover, the converging sequence $(h_n)$ is bounded and thus contained in a compact set $K' \subseteq Gl(Mat_{22})$. Since we assume that $\hat{\phi}$ is proper, this shows that the sequence $(g_n)$ of the elements $g_n = \hat{\phi}^{-1}(h_n)$, $n \in \mathbb{N}$ is contained in a compact set $K \subseteq PSL_2(\mathbb{R})$. Hence, there exists a subsequence, which we denote as $(g_{n_k})$ of $(g_n)$, converging to some element $g \in PSL_2(\mathbb{R})$. We then have $h = \hat{\phi}(g)$ by continuity of the map $\hat{\phi}$.
Said differently, if $\hat{\phi}$ is proper, then the limit point of any converging sequence in the image of $PSL_2(\mathbb{R})$ under $\hat{\phi}$ is contained in this image. This means that the group $PSL_2(\mathbb{R})$ is closed.

We first show that $\hat{\phi}$ is injective, or equivalently, that $\hat{\phi}(g)$, $g \in PSL_2(\mathbb{R})$ is the identity element of $Gl(Mat_{22}(\mathbb{R}))$ if and only if $g$ is the identity in $PSL_2(\mathbb{R})$. Consider the basis vector

$$m = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

of $Gl(Mat_{22}(\mathbb{R}))$. If the image

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

is $m$, then $a = \pm 1$ and $c = 0$. Furthermore, if this image is bounded, then both $a$ and $c$ must be bounded. Setting

$$m = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

a similar computation on the condition $gmg^{-1} = m$ shows that $d = \pm 1$ and $b = 0$. This proves that $\hat{\phi}$ is injective. Moreover, we see that if the image is bounded, then $d$ and $b$ are bounded. Thus, any bounded image of $\hat{\phi}$ has a bounded pre-image in $PSL_2(\mathbb{R})$, and so $\hat{\phi}$ is indeed a proper map. This and the above show that the image of $PSL_2(\mathbb{R})$ under $\hat{\phi}$ is closed in $Gl(Mat_{22}(\mathbb{R}))$.

## 2.2 The Lie Algebra of the Closed linear Group G

We begin by analyzing the local structure of a closed linear group $G$. As mentioned above, $G$ is locally diffeomorphic to a real vector space, called the *Lie algebra* $\mathfrak{g}$ of $G$. This diffeomorphism will be given by the exponential map (respectively, the logarithm map) defined on a space of square matrices with real coefficients of appropriate dimension.

The *exponential map* exp: $Mat_{dd}(\mathbb{R}) \to Mat_{dd}(\mathbb{R})$ is defined by

$$\exp : x \mapsto e^x = \sum_{n=0}^{\infty} \frac{1}{n!} x^n$$

The series converges for all $x \in Mat_{dd}(\mathbb{R})$, so the exponential map is well-defined. To see this, notice that the (Frobenius) norm $\|x\|$ of any $x \in Mat_{dd}(\mathbb{C})$ satisfies

$$\|x^n\| \leq \|x\|^n.$$

Hence, if $\|x\| \leq R$, $R \in \mathbb{R}_+$, we have that $\|x\|^n \leq R^n$ and therefore the above series converges absolutely. In fact, it converges uniformly on the set $\{x \in Mat_{dd}(\mathbb{R}) \mid \|x\| \leq R\}$ so that the exponential map is continuous.
The *logarithm map* is defined on the set $\{x \in Mat_{dd}(\mathbb{R}) : \|x - I\| < 1\}$ (where $I$ denotes the identity in $GL_d(\mathbb{R})$) by the series

$$\log x = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} (x - I)^n$$

which converges for all $x \in Mat_{dd}(\mathbb{R})$ for the same reason as above.
An important property of these maps is the following.

**Lemma 2.2.1.** *The exponential map and the logarithm map are, locally, the inverses of each other. More precisely, we have*

1. $\forall x \in Mat_{dd}(\mathbb{R})$ *with* $\|x - I\| < 1 : \exp(\log x)) = x$.

2. $\forall x \in Mat_{dd}(\mathbb{R})$ *with* $\|x\| < \log(2) : \|\exp(x) - I\| < 1$ *and* $\log(\exp(x)) = x$.

*Proof.* We give a sketch for the proof. First, it is immediately checked that if $d = (\mathrm{diag}(x_i))_{1 \leq i \leq d}$ is a diagonal matrix, the exponential map applied to $d$ is equal to $(\mathrm{diag}(\exp(x_i)))_{1 \leq i \leq d}$, where by $\exp(x_i)$, $1 \leq i \leq d$, we mean the usual exponential defined on the real numbers. Similarly, we have $\log(d) = (\mathrm{diag}(\log(x_i)))_{1 \leq i \leq d}$ when $\|d - I\| < 1$.
Moreover, given a matrix $a \in GL_d(\mathbb{R})$, it holds that $(axa^{-1})^n = ax^n a^{-1}$ for all $n \geq 1$. This implies

$$\exp(axa^{-1}) = a \exp(x) a^{-1}.$$

Also, we have $(axa^{-1} - I) = a(x - I)a^{-1}$ and so $(axa^{-1} - I)^n = a(x - I)^n a^{-1}$. Then, for $x$ and $axa^{-1}$ close enough to the identity, it holds that $\log(axa^{-1}) = a \log(x) a^{-1}$.
The assertions of the Lemma are now easily proven for diagonalizable matrices. In the general case, one considers the jordan decomposition of the matrix and approximates its jordan blocks by diagonalizable matrices. One gets to the desired result by a continuity argument on the exponential and the logarithm map and the observations from above.

□

The following Proposition states essentially the same as Proposition 9.5 in Chapter 9 of [EW10], to which we refer the reader to for a proof.

**Proposition 3.** Let $G \subseteq GL_d(\mathbb{R})$ be a closed linear group. Then there is a neighbourhood $B$ of the identity $I \in G$ with the property that $\log B \subseteq Mat_{dd}(\mathbb{R})$ is a neighbourhood of 0 inside a linear subspace $\mathfrak{g} \subseteq Mat_{dd}(\mathbb{R})$ which is characterized by the following equivalent properties.

1. $\mathfrak{g}$ is the subspace consisting of all derivatives $\phi'(t_0)$ of paths $\phi : [a, b] \to G \subseteq GL_d(\mathbb{R})$ at the points $t_0 \in [a, b]$ with $\phi(t_0) = I$.

2. $\mathfrak{g} = \{x \in Mat_{dd}(\mathbb{R}) : \exp tx \in G \text{ for all } t \in \mathbb{R} \}$.

The Proposition above indicates that, locally, $G$ looks like a real vector space. The exponential map and the logarithm map provide a local coordinate system within a neighbourhood of identity in $G$, identifying this neighbourhood with an Euclidean space given by the linear subspace $\mathfrak{g} \subseteq Mat_{dd}(\mathbb{R})$. Moreover, it states that the tangent space at identity of $G$ is equal to $\mathfrak{g}$, which we shall henceforth call the *Lie Algebra* of $G$.

**Lemma 2.2.2.** *The Lie Algebra* $\mathfrak{g}$ *of* $SL_2(\mathbb{R})$ *is given by the set of real* $2 \times 2$ *matrices with trace equal to zero.*

*Proof.* Let $\phi : [0, 1] \to SL_2(\mathbb{R})$ be any continuous and differentiable path

$$t \mapsto \begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix}$$

with $\phi(0) = I$. As $\phi(t) \in SL_2(\mathbb{R})$ for all $t$, it holds that $1 \equiv \det \phi(t) = a(t)d(t) - b(t)c(t)$, implying that

$$0 = d/dt[\det \phi(t)]_{t=0} = a'(0)d(0) + d'(0)a(0) - b'(0)c(0) - b(0)c'(0).$$

Moreover, $\phi(0) = I$ and so $b(0) = c(0) = 0$. This shows $a'(0)+d'(0) = 0$. Hence,

$$\mathfrak{g} = \left\{ \begin{pmatrix} a & b \\ c & -a \end{pmatrix} = \phi'(0),\ \phi : [0,1] \to SL_2(\mathbb{R}) \text{ differentiable with } \phi(0) = I \right\}$$

□

## 2.3 A Left-Invariant Riemannian Metric on a closed linear Group

Proposition 3 describes the tangent space at the identity of $G$. Conveniently, using the group structure of $G$ allows for a definition of the tangent space at any point $g \in G$. We will do this by using the left-translation of curves passing through the identity. In particular, this allows for a definition of the tangent bundle T$G$. To make this discussion precise, we follow the pages $291 - 293$ of Chapter 9 in [EW10].

**Definition 8.** Let $G$ denote a closed linear group. We assign to any point $t_0 \in [0,1]$ and any path $\phi : [0,1] \to G$ which is differentiable at $t_0$ the corresponding tangent vector

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

In this way, the *tangent bundle* T$G$ of $G$ is defined to be $G$ x $\mathfrak{g}$. We define the *tangent space* $\mathrm{T}_g G$ at $g \in G$ by

$$\mathrm{T}_g G = \{g\} \ \times \mathfrak{g}.$$

**Remark 5.** As usual, we call a path

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

*differentiable* at $t_0 \in [0,1]$, if the derivative $\phi'(t_0)$ exists when viewed as a function with values in the Euclidean space $\mathbb{R}^{d^2}$.

It may not be directly clear why the above Definition of the tangent bundle T$G$ makes sense. For $D\phi(t_0) \in \mathrm{T}G$ to hold, one needs to verify that

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

To see this, define the curve

$$\eta(t) = \phi(t_0)^{-1}\phi(t)$$

which has values in $G$ and is such that $\eta(t_0) = I$. Hence, by Proposition 3 we have

$$\eta'(t_0) = \phi(t_0)^{-1}\phi'(t_0) \in \mathfrak{g}$$

as required.

**Proposition 4.** Let $G$ be a closed linear group and consider a continuous curve $\phi : [0,1] \to G$ which is differentiable at $t_0 \in [0,1]$. Then the curves $(g\phi)(t) = g\phi(t)$ and $(\phi g^{-1})(t) = \phi(t)g^{-1}$, $g \in G$, are also differentiable at $t_0$, and if $D\phi(t_0)$ = $(\phi(t_0)$,$v)$ then

$$D(g\phi)(t_0) = (g\phi(t_0), v)$$

and

$$D(\phi g^{-1})(t_0) = (\phi(t_0)g^{-1}, gvg^{-1})$$

*Proof.* We have

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

by definition of the derivative $D(g\phi)$ at $t_0$. This proves the first claim because the $\mathfrak{g}$-component is unchanged by $g$. Similarly,

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

□

The left translation by some element $g \in G$, denoted as $L_g$, is the map

$$L_g : G \to G, \quad h \mapsto gh.$$

The chain rule applied to $D(g\phi) = D(L_g\phi)$ yields

$$D(L_g\phi)(t_0) = D(L_g)_{\phi(t_0)} D\phi(t_0).$$

Using the right hand side of this equality, Proposition 4 essentially says that the map $D(L_g)$ at a point $h \in G$ is defined as

$$(DL_g)_h : T_hG \to T_{gh}G, \quad (h, v) \mapsto (gh, v).$$

Said differently, $(DL_g)_h$ moves the base point $h$ but acts as the identity on the tangent vector $v$.
We define the *Adjoint map* $Ad_g$ at $g \in G$ by

$$Ad_g : \mathfrak{g} \to \mathfrak{g}, \quad v \mapsto gvg^{-1}$$

Since the exponential map satisfies

$$\exp(gxg^{-1}) = g \ \exp(x) \ g^{-1}$$

for all $g \in Gl_d(\mathbb{R})$ and $x \in Mat_{dd}(\mathbb{R})$, it follows from the second characterization of $\mathfrak{g}$ in Proposition 3 that $Ad_g$ is well-defined.
From these observations we similarly discover that the right translation

$$R_g : G \to G, \quad h \mapsto hg^{-1}$$

whose derivative at a point $h \in G$, by Proposition 4, is the map

$$D(R_g)_h : T_hG \to T_{hg^{-1}}G, \quad (h, v) \mapsto (hg^{-1}, gvg^{-1})$$

translates the base point on the right and acts on the fiber by the map $Ad_g$.

Now choose any inner product $<,>$ on $\mathfrak{g}$ (for instance, by identifying $\mathfrak{g}$ with $\mathbb{R}^{d^2}$ where $d = \dim(\mathfrak{g})$, and choosing the usual inner product on $\mathbb{R}^{d^2}$). Define a Riemannian metric on the tangent space $T_gG$, $g \in G$, by

$$< (g, u), (g, v) > = < u, v >_g := < u, v >, \quad u, v \in T_gG. \tag{2.3.0.1}$$

Equation 2.3.0.1 means that for any two vectors in the same tangent space $T_gG$, we define their inner product through the initially chosen inner product $<,>$ on the Lie Algebra $\mathfrak{g}$. Observe that this construction is similar to the hyperbolic Riemannian metric in Definition 3 - apart from the correcting factor $\frac{1}{y^2}$ used in the latter. The reason for this difference is due to the chosen construction of the tangent bundles T$\mathbb{H}$ and $TG$. We defined T$\mathbb{H}$ as $\mathbb{H} \times \mathbb{C}$ with the understanding that when we take derivatives of paths in $\mathbb{H}$, we take derivatives of paths in $\mathbb{C}$. We later showed that the factor $\frac{1}{y^2}$ leads to a Riemannian metric invariant under the action of $SL_2(\mathbb{R})$. In the case of $G$, we defined $TG$ as the product of $G$ with the tangent space at the identity, namely the Lie Algebra $\mathfrak{g}$. Then we used left translation to define the derivatives of curves passing through other points in $G$ in Proposition 4. We have seen that a consequence of this definition is that the derivative $DL_g$ of the left translation $L_g$ acts as the identity on the fiber of points in $TG$. This has the effect of making the Riemannian metric defined in equation 2.3.0.1 left-invariant, as we desire. The main idea for the construction of a metric $d$ on $G$ will be similar to the one chosen for the hyperbolic plane: The distance of two points in $G$ shall be the infimum over the length of the paths $\phi$ "connecting" them in a specific way, where we define the length of such paths as

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

just as in the case for the paths in $\mathbb{H}$. The main difficulty in the definition of such a metric lies in the fact that $G$ may not be path-connected. Fortunately, the next Proposition (which is similar to Corollary 9.7 in Chapter 9 of [EW10]) provides the final result to make the above discussion rigorous.

Recall the following notions of connectedness in a topological space.

**Definition 9.** Let $X$ denote a topological space.

1. $X$ is *connected* if there exist no open, non-empty disjoint subsets $U, V \subset X$ such that $X = U \sqcup V$.

2. $X$ is *path connected* if for all $x, y \in X$, there exists a continuous curve $\alpha : [a, b] \to X$ with $\alpha(a) = x$ and $\alpha(b) = y$.

3. $X$ is *locally path connected* if for all $x \in X$, there exists an open and path connected neighbourhood $U \subset X$ of $x$.

**Definition 10.** The connected component of a point in a topological space $X$ is the largest connected subset of $X$ containing it.

While in general, a path-connected set is connected, the converse direction does not necessarily hold.

**Proposition 5.** The connected component $G^\circ$ of the identity in $G$ is a path-connected, open and normal subgroup of $G$.

*Proof.* Set

$$H = \bigcup_{n=1}^{\infty} (\exp(\mathfrak{g}))^n$$

where

$$(\exp(\mathfrak{g}))^n = \{\exp(v_1)\ \exp(v_2)\ ...\exp(v_n)\ |v_1, ..., v_n \in \mathfrak{g}\}$$

$n \geq 1$. Observe that $H$ is connected; it is a union of connected sets $(\exp(\mathfrak{g}))^n$, $n \geq 1$ such that the intersection of these sets is non-empty, as they all contain the identity element of $G$. In particular, this shows that $H \subseteq G^\circ$. Furthermore, observe that $H$ is path-connected. To see this, it suffices to consider the identity $I$ and an arbitrary

$$g = \exp(v_1)...\exp(v_n) \in H$$

since translation by a group element is a homeomorphism. The path defined by

$$\phi(t) = \exp(tv_1)...\exp(tv_n)$$

for $t \in [0, 1]$ is smooth, and has $\phi(0) = I$ and $\phi(1) = g$.
By Proposition 3, $\exp(\mathfrak{g})$ is an open neighbourhood of $I$ in $G$. Hence, for every $g \in (\exp \mathfrak{g})^n$, the set $g(\exp \mathfrak{g}) \subseteq (\exp \mathfrak{g})^{n+1}$ is an open neighbourhood of $g$ inside $G$. Thus $H$ is open in $G$. If we can show that $H$ is closed, then it follows that $G^\circ \subseteq H$. Indeed, for if $H$ is closed but $H \subsetneq G^\circ$, then

$$G^\circ = H \sqcup (G^\circ \setminus H)$$

would be a disjoint union of non-empty open sets (with respect to the subset topology), contradicting the connectedness of $G^\circ$.
We first claim that $H \leq G$ is a subgroup of $G$. It is clear that $H$ is closed under multiplication. One may check that $\exp(v)^{-1} = \exp(-v)$ for $v \in \mathfrak{g}$ and therefore, $\exp(v)^{-1} \in \mathfrak{g}$ whenever $v \in \mathfrak{g}$, since the Lie Algebra $\mathfrak{g}$ is a vector space. Hence, $H$ is closed under taking inverses and the claim is verified. Now, we see with the above that

$$G \setminus H = \bigsqcup_{g \in G, g \neq I} gH$$

is open, which implies that $H$ is closed. Thus $G^\circ \subseteq H$.

It remains to check that $H \leq G$ is a normal subgroup. For $g \in G$ the conjugation map $h \mapsto ghg^{-1}$ sends $G$ to $G$, and for some neighbourhood $B'$ of $I$ we also have

$$gB'g^{-1} \subseteq B$$

where $B$ is chosen as in Proposition 3 so that $\log(B)$ is a neighbourhood of 0 inside $\mathfrak{g}$. This implies that

$$g(\log(B'))g^{-1} = \log(gB'g^{-1}) \subseteq \mathfrak{g}$$

and so $g\mathfrak{g}g^{-1} \subseteq \mathfrak{g}$ by linearity. Taking the exponential shows

$$g\exp(\mathfrak{g})g^{-1} \subseteq \exp(\mathfrak{g})$$

and therefore $gHg^{-1} \subseteq H$.

□

**Corollary 1.** For any closed linear group $G$, the Riemannian metric defined by 2.3.0.1 is a left-invariant metric on the connected component $G^\circ$ of $G$. In other words, if we define the length of a piecewise smooth curve $\phi : [0,1] \to G$ by

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

and use this to define a metric

$$d_G(g_0, g_1) = \inf_\phi L(\phi)$$

where the infimum is taken over all such curves with $\phi(0) = g_0$, $\phi(1) = g_1$, then

$$d_G(hg_0, hg_1) = d_G(g_0, g_1)$$

$\forall h, g_0, g_1 \in G^\circ$.

Since the proof of Proposition 5 shows that $G^\circ$ is path-connected, the first statement in Corollary 1 is an immediate consequence of this Proposition and Proposition 4. The rest follows analogously as in the case of the metric on the hyperbolic plane $\mathbb{H}$. The only difference lies in the fact that we need to restrict the metric to the path-connected part of $G$, that is to $G^\circ$, in order for it to make sense. We derive a metric $d_G$ for the whole group $G$ as follows. It is enough to require that the arguments of the function $d_G$ lie in the same coset of $G^\circ$ by initially defining $d_G(g_0, g_1) = \infty$ if $g_0G^\circ \neq g_1G^\circ$ and then defining a metric $\overline{d_G}$ on $G$ by

$$\overline{d_G}(g_0, g_1) = \begin{cases} \frac{d_G(g_0,g_1)}{1+d_G(g_0,g_1)} & \text{if } g_0G^\circ = g_1G^\circ \\ 1 & \text{otherwise.} \end{cases}$$

## 2.4 The stable Manifold for the Geodesic Flow

We are now in position to make the discussion about the stable manifold $W^{ss}(z_0, v_0)$ for the geodesic flow going through $(z_0, v_0) \in \mathrm{T}^1\mathbb{H}$ meaningful. Throughout this section, we fix a left-invariant metric $d$ on the group $PSL_2(\mathbb{R})$, with respect to which the notion of convergence is defined.

Recall that we denoted by a *stable manifold* for the geodesic flow $p_t$ on $\mathrm{T}^1\mathbb{H}$ through a vector $(z_0, v_0)$ the set of vectors $(z, v) \in \mathrm{T}^1\mathbb{H}$ satisfying

$$d(p_t((z, v)), p_t((z_0, v_0)) \to 0$$

as $t \to \infty$, where $d$ denotes the metric on $\mathrm{T}^1\mathbb{H}$, inherited from the left-invariant Riemannian metric on $PSL_2(\mathbb{R})$ through the identification $\mathrm{T}^1\mathbb{H} \cong PSL_2(\mathbb{R})$ . We hypothesized that the stable manifold for the geodesic flow on the geodesic determined by the vector $(z_0, v_0) \in \mathrm{T}^1\mathbb{H}$ is precisely the orbit of this vector under the horocycle flow. Using the correspondence of the geodesic flow on $\mathrm{T}^1\mathbb{H}$ with the action of $PSL_2(\mathbb{R})$ on itself, we can reformulate this assumption accordingly:

Let $g_0 \in PSL_2(\mathbb{R})$ be the unique element such that $Dg_0(i, i) = (z_0, v_0)$. We claim that a point $g_1 \in PSL_2(\mathbb{R})$ is in the orbit $g_0U^-$ if and only if

$$d(R_{a_t}(g_0), R_{a_t}(g_1)) \to 0 \tag{2.4.0.1}$$

as $t \to \infty$.
The first step in verifying the assumption is to note that by the left-invariance of $d$, we have

$$d(R_{a_t}(g_0), R_{a_t}(g_1)) = d(g_0a_t^{-1}, g_1a_t^{-1}) = d(I_2, a_tg_0^{-1}g_1a_t^{-1}).$$

Write

$$g_0^{-1}g_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Then

$$a_t \begin{pmatrix} a & b \\ c & d \end{pmatrix} a_t^{-1} = \begin{pmatrix} a & b\ \mathrm{e}^{-t} \\ c\ \mathrm{e}^{t} & d \end{pmatrix}$$

and hence the convergence in 2.4.0.1 holds if and only if

$$g_0^{-1}g_1 = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$$

for some $b \in \mathbb{R}$, or equivalently, $g_1 \in g_0U^-$.

This explains the local dynamics of the geodesic flow in the sense that it identifies its stable directions. However, since the group $PSL_2(\mathbb{R})$ is large both in topological and measure-theoretic sense, the global dynamics of the geodesic flow on $\mathrm{T}^1\mathbb{H}$ is not very interesting. In fact, the orbit of any element $g \in PSL_2(\mathbb{R})$

eventually leaves any compact set, implying that the dynamics exhibit no recurrence. As will become evident, in order to establish a meaningful setting for the ergodic theory of the geodesic flow, we need to look at its action on quotients of $PSL_2(\mathbb{R})$ by discrete subgroups $\Gamma$, while ensuring that the space $X = \Gamma \setminus PSL_2(\mathbb{R})$ is, in a certain sense, small.

## 2.5 Metric on Quotients by discrete Subgroups of closed linear Groups

As we have seen above, any closed linear group $G$ admits a left-invariant metric $d_G$. The goal of this section is to show that this metric induces a left-invariant metric $d_X$ for the quotient space $X = \Gamma \setminus G$, where $\Gamma \leq G$ is a discrete subgroup of $G$.

Recall that a subset $D$ of a metric space is *discrete* if every point $x \in D$ has a neighbourhood intersecting $D$ only in the point $x$. A discrete subgroup $\Gamma \subseteq G$ is, in a sense, *uniformly discrete*: The distances between any points in $\Gamma$ are uniformly bounded from below. To see this, suppose for a contradiction that there is a sequence of a pair of points $\gamma_n, \gamma_n' \in \Gamma$ with $d_G(\gamma_n, \gamma_n') \to 0$ as $n \to \infty$. Then by the left invariance of $d_G$, we also have that $d_G(\gamma_n^{-1}\gamma_n', I) \to 0$ which gives points $\gamma_n^{-1}\gamma_n'$ in $\Gamma$ arbitrarily close to the identity, contradicting that $\Gamma$ is discrete. Hence, there exists some $\eta > 0$ such that $d_G(\gamma_n, \gamma_n') > \eta$ for all $\gamma_n, \gamma_n' \in \Gamma$.

We define a metric on the space $X = \Gamma \setminus G$ of right cosets $\Gamma g$ of $\Gamma$ in $G$ by

$$d_X(\Gamma g_1, \Gamma g_2) = \inf_{\gamma_1, \gamma_2 \in \Gamma} d_G(\gamma_1 g_1, \gamma_2 g_2) = \inf_{\gamma \in \Gamma} d_G(g_1, \gamma g_2) \tag{2.5.0.1}$$

Notice that the second equality in 2.5.0.1 follows from the left-invariance of $d_G$. The triangle inequality for $d_X$ is shown by concatenation of smallest paths: If $p_1$ and $p_2$ denote paths going from $\Gamma g_1$ to $\Gamma g_2$ and from $\Gamma g_2$ to $\Gamma g_3$ respectively, then the concatenation $p$ of these paths is a path going from $\Gamma g_1$ to $\Gamma g_3$ of length equal to the sum of the lengths of $p_1$ and $p_2$. As there may be other paths connecting $\Gamma g_1$ and $\Gamma g_3$, the infimum over all these paths will be smaller or equal to the infimum of the above concatenation.
The discussion following Proposition 4 shows that for any differentiable path $\phi : [0, 1] \to G$ we have

$$L(R_g \phi) \leq \|Ad_g\| \; L(\phi)$$

$g \in G$, and where $\|Ad_g\|$ denotes the operator norm of the adjoint representation at $g$ with respect to the chosen inner product on $\mathfrak{g}$. From this we conclude that the map $R_g : G \to G$ is Lipschitz continuous with Lipschitz constant given by $\|Ad_g\|$. Then if $\Gamma$ is uniformly discrete, the same holds for the right cosets of $\Gamma$ in $G$. Hence

$$\inf_{\gamma \in \Gamma} d_G(g_1, \gamma g_2) > 0$$

unless $g_1 = \gamma g_2$. In view of the second equality in 2.5.0.1 we conclude that $d_X$ indeed defines a metric on $X$. We will henceforth always use the topology induced by this metric. Another consequence of $d_X$ is that $X$ and $G$ are locally isometric in the following sense.

**Proposition 6.** Let $\Gamma$ be a discrete subgroup of the closed linear group $G$. For any compact subset $K \subseteq X = \Gamma \setminus G$, there exists some $r = r(K) > 0$, called the *injectivity radius* on $K$, with the property that for any $x_0 \in K$ the map

$$B_r^G = \{g \in G \mid d_G(g, I) < r\} \to B_r^X(x_0) = \{y \in X \mid d_X(x_0, y) < r\},$$

$$g \mapsto x_0 g$$

is an isometry between $B_r^G$ and $B_r^X(x_0)$. If $K = \{x_0\}$ where $x_0 = \Gamma h$ for some $h \in G$, then

$$r = \frac{1}{4} \inf_{\gamma \in \Gamma \setminus \{I\}} d_G(h^{-1}\gamma h, I) \tag{2.5.0.2}$$

has this property.

*Proof.* We begin by proving the assertion for $K = \{x_0\}$. Let $r$ be as in equation 2.5.0.2 and notice that $r > 0$, since $h^{-1}\Gamma h$ is a (uniformly) discrete subgroup of $G$. For $g_1, g_2 \in B_r^G$,

$$d_X(\Gamma h g_1, \Gamma h g_2) = \inf_{\gamma \in \Gamma} d_G(h g_1, \gamma h g_2) = \inf_{\gamma \in \Gamma} d_G(g_1, h^{-1}\gamma h g_2)$$

by left-invariance of $d_G$. We need to show that $d_X(\Gamma h g_1, \Gamma h g_2) = d_G(g_1, g_2)$, so we want the infimum above to be attained for $\gamma = I$. Suppose $\gamma \in \Gamma$ is such that

$$d_G(g_1, h^{-1}\gamma h g_2) \leq d_G(g_1, g_2) < 2r.$$

Since $g_1, g_2 \in B_r^G$, we have

$$d_G(h^{-1}\gamma h g_2, I) \leq d_G(h^{-1}\gamma h g_2, g_1) + d_G(g_1, I) < 3r.$$

Similarly

$$\begin{aligned} d_G(h^{-1}\gamma h, I) = d_G(I, h^{-1}\gamma^{-1}h) &\leq d_G(I, g_2) + d_G(g_2, h^{-1}\gamma^{-1}h) \\ &\leq r + d_G(h^{-1}\gamma h g_2, I) \\ &< 4r. \end{aligned}$$

Hence, by the choice of $r$,

$$d_G(h^{-1}\gamma h, I) < \inf_{\gamma \in \Gamma \setminus \{I\}} d_G(h^{-1}\gamma h, I)$$

implying $\gamma = I$.
For a general compact set $K \subseteq X$, pick $x_0 \in K$ and let $r$ be as in 2.5.0.2. A computation analogous to the above shows that for any $y_0 \in B_{r/2}^X(x_0)$, $g_1, g_2 \in B_{r/2}^G$ we still have

$$d_X(y_0 g_1, y_0 g_2) = d_G(g_1, g_2).$$

By compactness of $K$, it has a finite covering of open balls $B^X_{r(x_0)/2}(x_0)$, $x_0 \in K$ and where $r(x_0)$ defined as in 2.5.0.2 is the uniform injectivity radius for the respective ball. By choosing $r = r(K)$ from the Proposition to be the minimum of the associated injectivity radii for this covering, the desired result follows. □

We define the canonical quotient map

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

$$g \mapsto \Gamma g.$$

By definition of the metric $d_X$, it satisfies

$$d_X(\pi(g_1), \pi(g_2)) \leq d_G(g_1, g_2)$$

$g_1, g_2 \in G$ and therefore is continuous. In particular, Proposition 6 states that $\pi_X$ is a local isometry. This observation motivates a different way to understand the quotient space $X = \Gamma \setminus G$: We may consider a subset $F' \subset G$ such that $\pi_X$ restricted to $F'$ is a bijection. In fact, the analysis of such sets will be the goal of the next chapter. They will serve as a bridge between the topological properties of $X$ we considered here, and the understanding of $X$ as a natural measure space.

# Chapter 3

# Dynamics on Quotient Spaces

The geodesic flow acting on the group $G = PSL_2(\mathbb{R})$ exhibits little interesting behaviour regarding the convergence of orbits. The action needs to take place on a quotient space $X = \Gamma \setminus G$ considered in the preceding section, such that $X$ is finite with respect to a specific natural measure. Our objective is to develop the necessary theory towards this setting, which involves understanding the structure required for the discrete subgroup $\Gamma \leq G$, as well as the construction of the measure mentioned above. We follow Chapter 2 in [BM00] and Chapter 9 in [EW10].

## 3.1 Fundamental Domains

Given an action of a group $\Gamma$ on a locally compact space $X$, the *orbit* of $x \in X$ denotes the set

$$\mathcal{O}(x) = \{g \cdot x,\ g \in \Gamma\}.$$

The sets of orbits $\mathcal{O}(x)$, $x \in X$ form a partition of $X$ under the equivalence relation

$$x \sim y \iff \exists h \in \Gamma : h \cdot x = y$$

$x, y \in X$. They can be viewed as a quotient space of $X$, induced by the group action. Roughly speaking, the concept of *fundamental domains* serves as a geometrical description of this quotient space. A measurable subset $F \subseteq X$ is called a fundamental domain if it contains one point of each orbit $\mathcal{O}(x)$, $x \in X$. More precisely, we define such $F$ as follows.

**Definition 11.** Let $(X, \mu)$ be a locally compact measure space on which a group $\Gamma$ acts by homeomorphisms. A (borel) measurable subset $F$ of $X$ is a *fundamental domain* for the action of $\Gamma$ if
(i) $\bigcup_{\gamma \in \Gamma} \gamma \bar{F} = X,$ where $\bar{F}$ is the closure of $F,$ and
(ii) $\gamma_1 F \cap \gamma_2 F = \emptyset,$ for all $\gamma_1, \gamma_2 \in \Gamma, \gamma_1 \neq \gamma_2$.

For instance, a fundamental domain $F \subseteq \mathbb{R}^n$ for the group $\mathbb{Z}^n$ acting as a translation is given by

$$F = \{(x_1, ..., x_n) \in \mathbb{R}^n |\ 0 < x_i < 1,\ i = 1, ...n\}$$

**Remark 6.** We are going to use a slightly weaker definition of a fundamental domain, where the union above holds up to measure zero and the intersection is a null set. Whenever the distinction is relevant, we will refer to a fundamental domain as above as a *strict* fundamental domain.

We are interested in fundamental domains for the action of discrete subgroups of $PSL_2(\mathbb{R})$ on the hyperbolic plane $\mathbb{H}$. These domains will lay the basis for the construction of a meaningful dynamical system of the form as mentioned in the introduction of this chapter.

We begin with an important characterization of discrete subgroups of $PSL_2(\mathbb{R})$.

**Definition 12.** The action of a group $\Gamma$ on a locally compact space $X$ is said to be *properly discontinuous* if for any compact $K \subseteq X$, the set

$$\{\gamma \in \Gamma |\ \gamma K \cap K \neq \emptyset\}$$

is finite.

**Lemma 3.1.1.** *An infinite subgroup $\Gamma \subseteq PSL_2(\mathbb{R})$ is discrete if and only if its action on $\mathbb{H}$ is properly discontinuous.*

*Proof.* Assuming that $\Gamma$ is not discrete, we may choose a sequence of elements $(\gamma_n) \subseteq \Gamma$ with $\gamma_n \neq I$ for all $n \geq 1$ and $\gamma_n \to I$ as $n \to \infty$. Then, for any point $z_0 \in \mathbb{H}$, it holds that $\gamma_n z_0 \to z_0$. Thus, for an open ball $B$ centered around $z_0$, we see that that the intersection $\gamma_n \bar{B} \cap \bar{B}$, where $\bar{B}$ denotes the closure of the open ball $B$, is non-empty for infinitely many $n \geq 1$.
Conversely, assume that $\Gamma$ is discrete. We claim that it suffices to prove the following:
Let $K$ be any compact subset of $\mathbb{H}$. Then the set

$$E = \{g \in PSL_2(\mathbb{R}) \mid gK \cap K \neq \emptyset\}$$

is compact.
Indeed, suppose $E$ is compact. Choose $\eta > 0$ such that

$$d(\gamma_1, \gamma_2) > \eta$$

for all $\gamma_1, \gamma_2 \in \Gamma$. Let $\bigcup_{g \in E} B_\eta(g)$ be an open cover of $E$ given by open balls $B_\eta(g)$ of radius $\eta$ centered around $g \in E$. By compactness we may choose $g_1, ..., g_N \in E$, for some $N \in \mathbb{N}$, such that

$$E = \bigcup_{i=1}^{N} B_\eta(g_i).$$

By the choice of the radius $\eta$, it follows that the set in Definition 12 is finite.

Let $(g_n) \subseteq E$ be a sequence converging to $g \in PSL_2(\mathbb{R})$ as $n \to \infty$. Consider any $z \in K$. We have that $g_n z \in K$ for all $n \geq 1$ and as $K$ is closed (because it is a compact set in $\mathbb{H}$) and the action of $PSL_2(\mathbb{R})$ on $\mathbb{H}$ is continuous, we have

$$g_n z \to gz \in K$$

and $g \in E$. This shows that $E$ is closed.
We are left to show that $E$ is a bounded set in $PSL_2(\mathbb{R})$ when viewed as a subset of $\mathbb{R}^4$.
As $K$ is compact, there exist some constants $R, \epsilon > 0$ such that

$$|z| \ \leq R, \ \mathrm{Im}(z) \geq \epsilon, \ \forall z \in K.$$

Let

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in E.$$

Since $gz \in K$ whenever $z \in K$, we have

$$\left| \frac{az+b}{cz+d} \right| \leq R$$

and

$$\mathrm{Im}\left( \frac{az+b}{cz+d} \right) = \frac{\mathrm{Im}(z)}{|cz+d|^2} \geq \epsilon$$

by equation 1.2.0.2.
Thus

$$|cz+d|^2 \leq \frac{1}{\epsilon}\mathrm{Im}(z) \leq \frac{R}{\epsilon}$$

and

$$|az+b|^2 \leq R^2|cz+d|^2 \leq \frac{R^3}{\epsilon}$$

Since $z$ belongs to the compact set $K$, the set of all $(a,b,c,d) \in \mathbb{R}^4$ satisfying these inequalities is obviously bounded.

□

There is an explicit method for the construction of a fundamental domain for an infinite discrete subgroup $\Gamma \leq PSL_2(\mathbb{R})$ acting on $\mathbb{H}$, using *Dirichlet regions*, which we now introduce.
Choose a point $p \in \mathbb{H}$ which is not fixed by any $\gamma \in \Gamma \setminus \{I\}$. Observe that such a point exists, since $\Gamma$ is countable and any nontrivial element $g \in SL_2(\mathbb{R})$ acting as a Möbius transformation in $\mathbb{C}$ has at most two fixed points $z, \bar{z} \in \mathbb{C}$ (and hence at most one in $\mathbb{H}$). Indeed, the fixed points of an element $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in$

$PSL_2(\mathbb{R})$, acting as a Möbius transformation on $\mathbb{C}$ are obtained through the quadratic equation

$$z = \frac{az+b}{cz+d}.$$

Since the coefficients of this equation are real, the roots are complex conjugate to each other.

For each $\gamma \in \Gamma \setminus \{I\}$, define the sets

$$L_\gamma = L_\gamma(p) := \{z \in \mathbb{H} \mid d(z,p) = d(z,\gamma p)\} \tag{3.1.1.1}$$

and

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

where $p \in \mathbb{H}$ is as above and where $d$ denotes the hyperbolic metric on $\mathbb{H}$.

**Definition 13.** The set

$$D = D(p) = \bigcap_{\gamma \in \Gamma \setminus \{I\}} H_\gamma(p)$$

is called a *Dirichlet region* for $\Gamma$.

### 3.1.1 The Boundary of a Dirichlet Region

We shall improve our understanding of the sets $L_\gamma$ and $H_\gamma$ used in the definition of a Dirichlet region. We follow the explanations in [Lab22].
First, we claim that the sets $L_\gamma(p)$ are geodesic curves if $\gamma \in \Gamma$ is non-trivial. To see this, let $\gamma \in \Gamma \setminus \{I\}$ and first consider the case where $p = r + i$ and $\gamma p = -r + i$ , for some $r \in \mathbb{R}$. Then the set of points in $\mathbb{H}$ equidistant to $p$ and $\gamma p$ is precisely the imaginary axis, which is a geodesic. For general points $p$ and $\gamma p$, there exists (as can be seen in the proof of Proposition 1) an isometry $g \in PSL_2(\mathbb{R})$ such that $gp = r + i$ and $g\gamma p = -r + i$. Then $g^{-1}$ applied to the imaginary axis, whose image is $L_\gamma(p)$, is another geodesic, containing the points equidistant to $p$ and $\gamma p$.
Furthermore, the sets $H_\gamma(p)$, $\gamma \in \Gamma \setminus \{I\}$ are hyperbolically convex regions, with the understanding that the geodesic segment $[z, w]$ for any pair of points $z, w \in H_\gamma(p)$ is again contained in $H_\gamma(p)$. This is proven by a similar argument as above. Let $\gamma \in \Gamma \setminus \{I\}$. We may apply $g \in PSL_2(\mathbb{R})$ to obtain points $gp = r+i$ and $g\gamma p = -r+i$. Then the geodesic segments $[z, w]$ for $z, w \in H_\gamma(p)$ are half circles whose arcs from any $x \in [z, w]$ to $p$ are shorter than those of the circles connecting $x$ with $p\gamma$. The above arcs remain the same after applying the isometry $g^{-1}$ to the segments $[x, p]$ and $[x, \gamma p]$.
The closure $\overline{H_\gamma(p)}$, $\gamma \in \Gamma \setminus \{I\}$ is given by

$$\overline{H_\gamma}(p) = \{z \in \mathbb{H} \mid d(z,p) \leq d(z,\gamma p)\}$$

Indeed, in the case $p = r + i$ and $\gamma p = -r + i$, $r \in \mathbb{R}_+$, we have for any $\epsilon > 0$ and $z_0 \in L_\gamma(p)$, that there exists an element $x \in [z_0, p]$ such that $d(z_0, x) < \epsilon$.

This can be seen by traversing the geodesic segment $[z_0, p]$ starting from $z_0$ by a distance smaller than $\epsilon$. As $x$ lies on the right of the imaginary axis, it is immediate that the length of the path from $x$ to $p$ is smaller than the one to $\gamma p$. Therefore, the intersection of an open ball centered at $z_0$ of radius $\epsilon$ with $H_\gamma(p)$ is nonempty. This shows

$$\{z \in \mathbb{H} \mid d(z,p) \leq d(z,\gamma p)\} \subseteq \overline{H_\gamma}(p).$$

The reverse inclusion is clear as $\{z \in \mathbb{H} \mid d(z,p) \leq d(z,\gamma p)\}$ is a closed set. Hence

$$\overline{H_\gamma}(p) = H_\gamma(p) \cup L_\gamma(p).$$

For general points $p$ and $\gamma p \in \mathbb{H}$ we apply as usual some isometry $g \in PSL_2(\mathbb{R})$ to obtain points of the form as above. We have

$$\overline{D(p)} \subseteq \bigcap_{\gamma \in \Gamma \setminus \{I\}} \overline{H_\gamma}(p) = \bigcap_{\gamma \in \Gamma \setminus \{I\}} (H_\gamma(p) \cup L_\gamma(p)).$$

The boundary of the Dirichlet region is thus given by

$$\partial D(p) = \overline{D(p)} \setminus D(p)^\circ \subseteq \overline{D(p)} \setminus D(p) \subseteq \bigcup_{\gamma \in \Gamma \setminus \{I\}} L_\gamma(p).$$

Said differently, the boundary of any Dirichlet region is given by geodesic curves.

### 3.1.2 The Fundamental Domains for $PSL_2(\mathbb{R})$

The crucial property of Dirichlet regions is the following.

**Theorem 1.** Any Dirichlet region $D$ for an infinite discrete subgroup $\Gamma \leq PSL_2(\mathbb{R})$ is an open and convex fundamental domain for the action of $\Gamma$ on $\mathbb{H}$.

*Proof.* We know that $D = D(p)$, $p \in \mathbb{H}$ is the intersection of convex sets, and so it is convex. To show that $D$ is open, suppose for a contradiction that it is not. Then $\mathbb{H} \setminus D$ is not closed and thus there exist $z \in D$ and sequences $(z_n) \subset \mathbb{H}$, $(\gamma_n) \subseteq \Gamma \setminus \{I\}$ such that

$$z_n \to z \quad \text{and} \quad d(z_n, p) \geq d(z_n, \gamma_n p).$$

This implies that the sequence $(\gamma_n p)$ is bounded. Thus $\{\gamma_n p,\ n \geq 1\}$ is contained in some compact subset of $\mathbb{H}$. Since the action of $\Gamma$ on $\mathbb{H}$ is properly discontinuous, we conclude that the set $\{\gamma_n, n \geq 1\}$ is finite. Hence, upon passing to the subsequence $\gamma_{n_k} \to \gamma$ for some $\gamma \in \Gamma \setminus \{I\}$ if necessary, we get

$$d(z,p) = \lim_{k\to\infty} d(z_{n_k}, p) \geq \lim_{k\to\infty} d(z_{n_k}, \gamma_{n_k} p) = d(z, \gamma p)$$

But this contradicts that $z \in D$, so we conclude that $D$ is open.

The action of $\Gamma$ on $\mathbb{H}$ being discrete, each $\Gamma$-orbit $\mathcal{O}$ is discrete, and hence, with help of the Axiom of Choice, we may pick a point $p_{\mathcal{O}} \in \mathcal{O}$ such that

$$\forall z \in \mathcal{O} : \ d(p_{\mathcal{O}}, p) \leq d(z, p)$$

Set $F = \{p_{\mathcal{O}} \mid \mathcal{O} \subseteq \mathbb{H}, \ \mathcal{O} \text{ is a } \Gamma - \text{orbit}\}$ and notice that $F$, containing exactly one point of each orbit under $\Gamma$, is by definition a fundamental domain for the action of $\Gamma$ on $\mathbb{H}$.
We want to show that the sets $D$ and $F$ are equal up to measure zero.
Observe that $p_{\Gamma z} = z$ for all $z \in D$. Indeed, $p_{\Gamma z}$ is of the form $p_{\Gamma z} = \gamma_0 z$ for some $\gamma_0 \in \Gamma$ and satisfies $d(\gamma_0 z, p) \leq d(\gamma z, p)$ for all $\gamma \in \Gamma$. In particular

$$d(\gamma_0 z, p) \leq d(z, p) \Leftrightarrow d(z, \gamma_0^{-1} p) \leq d(z, p).$$

But then the condition $z \in D$ requires $\gamma_0$ to be trivial. This proves the assertion and that $D \subseteq F$.
For the converse inclusion, pick $p_{\mathcal{O}} \in F$. Suppose the geodesic segment $[p, p_{\mathcal{O}})$ meets some point in $L_\gamma(p)$. Since $p \notin L_\gamma(p)$, we have

$$(p, p_{\mathcal{O}}) \cap L_\gamma(p) \neq \emptyset$$

and by the convexity of $H_\gamma(p)$, it follows that $d(p, p_{\mathcal{O}}) > d(\gamma p, p_{\mathcal{O}}) = d(p_{\mathcal{O}}, \gamma p)$ contradicting the condition imposed on $p_{\mathcal{O}}$.
Hence, $[p, p_{\mathcal{O}}) \subseteq D$ which shows that $p_{\mathcal{O}} \in \overline{D}$ and thus $F \subseteq \overline{D}$. Now, we know that the boundary of $D$ consists of countably many geodesic curves. These curves can be shown to be of measure zero with respect to the *hyperbolic area form*, which we will consider later in this chapter. This completes the proof.
□

## 3.2 Finite Measure on the Quotient Space $X = \Gamma \setminus PSL_2(\mathbb{R})$

The goal of this section is to define a finite measure on the quotient space $X = \Gamma \setminus PSL_2(\mathbb{R})$ by an infinite discrete subgroup $\Gamma \leq PSL_2(\mathbb{R})$, invariant for the left action of $PSL_2(\mathbb{R})$ on $X$. This leads us to the notion of the *Haar measure*, which is a natural measure defined on a metric, $\sigma$-compact, locally compact group $G$. It turns out that the Haar measure on $PSL_2(\mathbb{R})$ induces a finite measure on $X$ if a fundamental domain for $X$ has finite measure with respect to the Haar measure on $PSL_2(\mathbb{R})$. Here and for the rest of this section, we mean by a fundamental domain for the quotient space $X$ a fundamental domain $F$ for the action of $\Gamma$ on $PSL_2(\mathbb{R})$. Thus, $F \subset PSL_2(\mathbb{R})$ is a measurable set such that

$$|F \cap \Gamma g| = 1, \quad \forall g \in PSL_2(\mathbb{R}).$$

### 3.2.1 Lattices in $PSL_2(\mathbb{R})$

**Definition 14.**

1. The *hyperbolic area form* $A$ on the hyperbolic plane $\mathbb{H}$ is defined as $dA = \frac{1}{y^2}dxdy$, $z = x + iy \in \mathbb{H}$.

2. The *hyperbolic volume form* $m$ on the unit tangent bundle $\mathrm{T}^1\mathbb{H}$ is defined as $dm = \frac{1}{y^2}dxdyd\theta$, where $\theta$ denotes a unit tangent vector at $z = x + iy$.

**Definition 15.** A discrete subgroup $\Gamma \leq PSL_2(\mathbb{R})$ is called a *lattice* in $PSL_2(\mathbb{R})$ if a fundamental domain for the quotient space $X = \Gamma \backslash PSL_2(\mathbb{R})$ has finite measure with respect to the volume form $m$.

The Theorem of Siegel provides a geometric description of a Dirichlet region for any lattice $\Gamma$ in $PSL_2(\mathbb{R})$.

**Theorem 2** (Theorem of Siegel)**.** A Dirichlet region for a lattice $\Gamma$ in $PSL_2(\mathbb{R})$ is a convex hyperbolic polygon.

The Theorem of Siegel states that the Dirichlet regions for lattices in $PSL_2(\mathbb{R})$ only have finitely many sides, or, by the previous discussion: The boundary of any Dirichlet region is made of finitely many geodesic curves. The proof of this Theorem can be found in Chapter 2 of [BM00]. We shall illustrate this result in the case where the discrete subgroup of $PSL_2(\mathbb{R})$ is given by the *modular group*

$$PSL_2(\mathbb{Z}) = SL_2(\mathbb{Z})/\{\pm I_2\}$$

which is, as will turn out, a lattice. We follow the computations in Example 2.7 of [BM00].

**Proposition 7.** The set $E = \{z \in \mathbb{H} \mid |z| > 1, |\mathrm{Re}(z)| < \frac{1}{2}\}$ is a fundamental domain for the action of $PSL_2(\mathbb{Z})$ on $\mathbb{H}$.

*Proof.* In view of the Theorem 1, it is enough to show that there exists a point $p \in \mathbb{H}$ such that $E$ is the Dirichlet region $D = D(p)$ corresponding to $p$.
Consider $p = 2i$. This point is not fixed by any non-trivial element of $PSL_2(\mathbb{Z})$. Indeed, suppose

$$\frac{a(2i) + b}{c(2i) + d} = 2i$$

for some $a, b, c, d \in \mathbb{Z}$ with $ad - bc = 1$. Then

$$2 = \mathrm{Im}\left(\frac{a(2i) + b}{c(2i) + d}\right) = \frac{2}{|c(2i) + d|^2}$$

where the second equality follows from equation 1.2.0.2. This implies $c = 0$ and $d \in \{-1, 1\}$ and so $a = d$ and $b = 0$. Hence, $\gamma p = p$ holds only for $\gamma = I_2$.
Write

$$\gamma_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

These elements describe the mappings $\gamma_1 : z \mapsto z+1$ and $\gamma_2 : z \mapsto -1/z$. From this we see that the sides of $E$ are precisely the geodesic segments given by $L_{\gamma_1}, L_{\gamma_1^{-1}}, L_{\gamma_2}$ (defined as in 3.1.1.1). Thus, if a point $z \in \mathbb{H}$ satisfies $d(z,p) < d(z,\gamma p)$ for all $\gamma \in PSL_2(\mathbb{Z})$ then it must be contained in $E$. Hence, we have $D \subseteq E$.
We argue by contradiction to obtain the reverse inclusion. So suppose $D$ is strictly contained in $E$. This means that there is a $z \in E$, $\gamma \in PSL_2(\mathbb{Z}) \setminus \{I_2\}$ such that

$$d(z,p) \geq d(z,\gamma p) \iff d(z,p) \geq d(\gamma^{-1}z, p)$$

by the left-invariance of the metric $d$. Observe that $\gamma^{-1}z \in E$, for otherwise $\gamma \in \{\gamma_1, \gamma_1^{-1}, \gamma_2\}$ and $z$ would lie on a side of $E$ which is not contained in $E$. Now, write

$$\gamma^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in PSL_2(\mathbb{Z}).$$

If $c = 0$, then $a, d = \pm 1$, $b \neq 0$ and

$$|\mathrm{Re}(\gamma^{-1}z)| = |\mathrm{Re}(\pm z + b)| = |\mathrm{Re}(z) + b| \geq |b| - \frac{1}{2} \geq \frac{1}{2}.$$

But this contradicts $\gamma^{-1}z \in E$ and thus $c \neq 0$. Moreover,

$$|cz+d|^2 = c^2|z|^2 + 2cd\mathrm{Re}(z) + d^2 > c^2 + d^2 - |cd| = (|c| - |d|)^2 + |cd|$$

where the inequality holds because of $|z| > 1$ and $|\mathrm{Re}(z)| < 1/2$. Observe that $(|c| - |d|)^2 + |cd| = 0$ only if $c = d = 0$, which is impossible. Thus the integer $(|c| - |d|)^2 + |cd|$ is greater or equal to one, implying by the above inequality that $|cz + d|^2 > 1$. Using once more equation 1.2.0.2 we see

$$\mathrm{Im}(\gamma^{-1}z) = \frac{\mathrm{Im}(z)}{|cz+d|^2} < \mathrm{Im}(z).$$

Observe that by the choice of $z$, we can apply the same argument as above replacing $z$ by $\gamma^{-1}z$ and $\gamma^{-1}$ by $\gamma$, leading to the contradiction

$$\mathrm{Im}(z) < \mathrm{Im}(\gamma^{-1}z).$$

We conclude that $D = E$.

□

To see that the modular group is indeed a lattice in $PSL_2(\mathbb{R})$, we want to show that the volume of the fundamental domain $E$ is finite with respect to the volume form $m$. Notice that for any $z \in E$, it holds that $\mathrm{Im}(z) > \sqrt{3}/2$ and hence

$$\begin{aligned} m(E) = \int_E dA &\leq \int_{\sqrt{3}/2}^{\infty} \int_{-1/2}^{1/2} \frac{dxdy}{y^2} \\ &= \int_{\sqrt{3}/2}^{\infty} \frac{1}{y^2} dy = \frac{2}{\sqrt{3}/2} < \infty. \end{aligned}$$

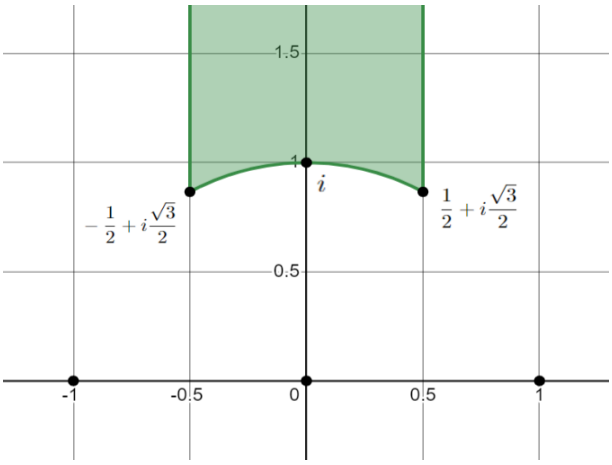


Figure 3.1: A fundamental domain for $PSL_2(\mathbb{Z})$.

We obtain a fundamental domain $F$ for $X = PSL_2(\mathbb{Z}) \setminus PSL_2(\mathbb{R})$ as follows. Define

$$F = \{g \in PSL_2(\mathbb{R}) \mid g(i) \in E\} \tag{3.2.0.1}$$

and notice that

$$\begin{aligned} |F \cap PSL_2(\mathbb{Z})\ h| &= |\{g \in PSL_2(\mathbb{R}) \mid g(i) \in E,\ g = \gamma h \text{ for some } \gamma \in PSL_2(\mathbb{Z})\}| \\ &= |\{\gamma \in PSL_2(\mathbb{Z}) \mid \gamma h(i) \in E\}| \\ &= 1 \end{aligned}$$

for all $h \in PSL_2(\mathbb{R})$.

### 3.2.2 Haar Measure

Here and for the rest of this section, let $G$ denote a metric, $\sigma$-compact, locally compact group. The (left-invariant) *Haar measure* $m_G$ on $G$ is characterized (up to scalar multiples) by the following properties:

1. $m_G(K) < \infty$ for any compact set $K \subseteq G$.
2. $m_G(\mathcal{O}) > 0$ for any open set $\mathcal{O} \subseteq G$.
3. $m_G(gB) = m_G(B)$ for any $g \in G$ and any measurable $B \subseteq G$.

The Haar measure is a natural measure on $G$ in the following sense.

**Theorem 3** (Haar)**.** Let $G$ be a metrizable, $\sigma$-compact and locally compact group. Then there exists a left-invariant Haar measure satisfying properties $(1) - (3)$ above.

A sketch for the proof of this Theorem is given in Chapter 8 of [EW10]. An important property of the Haar measure in $G$ is its uniqueness up to scalar multiple (see p. 246 in [EW10]). The other Haar measures of $G$ are obtained through the push-forward measure $(R_g)_*(m_G)$ given by the right translation $R_g$, $g \in G$ as is shown in the following. We have, for any measurable set $B \subseteq G$ and any $h \in G$,

$$(R_g)_* m_G(hB) = m_G(hBg) = m_G(Bg) = (R_g)_* m_G(B).$$

This shows that $(R_g)_*(m_G)$ is another (left-invariant) Haar measure on $G$.

The *modular* function

$$\mathrm{mod} \ : \ G \to \mathbb{R}_{>0}$$

assigns to each $g \in G$ the scaling factor $\mathrm{mod}(g)$ that maps $m_G$ to $(R_g)_* m_G$. More precisely,

$$(R_g)_* m_G = \mathrm{mod}(g) \ m_G$$

for $g \in G$. Notice that

$$\begin{aligned} \mathrm{mod}(gh) m_G &= (R_{gh})_*(m_G) = (R_g \circ R_h)_*(m_G) \\ &= (R_g)_*(\mathrm{mod}(h) m_G) \\ &= \mathrm{mod}(g)\mathrm{mod}(h) m_G \end{aligned}$$

for all $g, h \in G$. Hence, the modular function is a group homomorphism.

**Definition 16.** A group $G$ equipped with a left Haar measure $m_G$ is called *unimodular* if $\mathrm{mod}(G) = \{1\}$.

In a unimodular group, the left Haar measure is therefore also a right Haar measure. The following Proposition states an important result that allows for a natural construction of finite measures on the quotient space $X = \Gamma \setminus G$, by a *lattice* $\Gamma \leq G$. In this setting, we mean by a lattice a discrete subgroup $\Gamma \leq G$ such that there exists a fundamental domain for $\Gamma \setminus G$ which is of finite measure with respect to the (left-invariant) Haar measure on $G$. We will denote this fundamental domain by $F$. The proof of the Proposition significantly relies on the observation that if $G$ has a lattice, then $G$ is unimodular.

**Proposition 8.** Let $\Gamma \leq G$ be a lattice of $G$. Then any fundamental domain has the same measure as $F$, $G$ is unimodular, and the Haar measure $m_G$ induces a finite measure $m_X$ on $X$ by

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

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

*Proof.* We first show that any two measurable sets $B, B' \subseteq G$ with the property that the restrictions $\pi|_B$ and $\pi|_{B'}$ of the quotient map

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

$$g \mapsto \Gamma g$$

are injective and are such that $\pi(B) = \pi(B')$, satisfy

$$m_G(B) = m_G(B').$$

This proves the first part of the Proposition, since, as is immediate from the definition, a measurable subset $F \subseteq G$ is a fundamental domain for $X$ if and only if $\pi|_F$ is a bijection.
Since we assume that $\pi(B) = \pi(B')$, there exists, for any $g \in B$, by injectivity of $\pi|_B$, a unique $b' \in B'$ with $\Gamma g = \Gamma b'$. This shows $g \in \gamma B'$ for a unique $\gamma \in \Gamma$. We have

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

and similarly

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

These two decompositions can be derived from each other in the following sense. An element $B \cap \gamma B'$ of the first partition corresponds to a unique element of the second, as for $\gamma' := \gamma^{-1}$, we have

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

Therefore

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

where we used the left-invariance of the Haar measure in the second equation.
Next, we claim that the right-translation $Fg$ of any fundamental domain $F$ by $g \in G$ is again a fundamental domain. To see this, first pick any $h \in G$. Since $F$ is a fundamental domain, there exists a unique $\gamma \in \Gamma$ with $hg^{-1} \in \gamma F$, implying that $h \in \gamma F g$ for some unique $\gamma \in \Gamma$.
From the above, two injective sets with the same image under $\pi$ have the same measure, we see that

$$m_G(F) = m_G(Fg) = \mathrm{mod}(g) m_G(F).$$

Moreover, $0 < m_G(F) < \infty$ by the assumption $m_G(F) < \infty$ and the the discreteness of $\Gamma$. Therefore, the above implies $\mathrm{mod}(g) \in \{1\}$ and that $G$ is unimodular.

We now define a measure $m_X$ on $X$ as follows. For any measurable set $B \subseteq X$, let

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

and note that this definition is independent of the choice of the fundamental domain $F$ by the first part of the proof. To see that $m_X$ is invariant under the action of $R_g$, $g \in G$ , let $C = \pi^{-1}(B) \cap F$ and remark that for $F' = Fg$

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

and so $m_X(Bg) = m_G(Cg)$. We have $m_G(C) = m_G(Cg)$ by the unimodularity of $m_G$. Thus

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

which shows $m_X(B) = m_X(R_g^{-1}(B))$ as desired.

□

**Remark 7.** Although the space $X = \Gamma \setminus G$ is not a group in general, we will refer to the invariant measure $m_X$ as the *Haar measure* of $X$.

The next Lemma states that the volume element $m$ defined in 14 is invariant for the action of $PSL_2(\mathbb{R})$ on $\mathrm{T}^1\mathbb{H}$. In view of the identification $PSL_2(\mathbb{R}) \cong \mathrm{T}^1\mathbb{H}$, this means that there exists a measure on $PSL_2(\mathbb{R})$ induced by $m$, which is invariant under the left action of $PSL_2(\mathbb{R})$ on itself. Thus, this measure is in fact a Haar measure on $PSL_2(\mathbb{R})$. Together with Proposition 8, this will provide the desired finite measure on $X = \Gamma \setminus PSL_2(\mathbb{R})$, invariant under the action of $PSL_2(\mathbb{R})$.

**Lemma 3.2.1.** *The hyperbolic area form $dA$ on $\mathbb{H}$ and the hyperbolic volume form $dm$ on $T^1\mathbb{H}$ are both invariant under the respective action of $PSL_2(\mathbb{R})$.*

*Proof.* We begin by showing that the area form $dA$ above is invariant under the action of $PSL_2(\mathbb{R})$. Equivalently, we want to show that for any continuous function $f : \mathbb{H} \to \mathbb{R}$ with compact support, it holds that

$$\int_{\mathbb{H}} f \circ g dA = \int_{\mathbb{H}} f dA$$

$g \in PSL_2(\mathbb{R})$. As the complex derivative of $z \mapsto g(z) = \frac{az+b}{cz+d}$ is equal to $\frac{1}{(cz+d)^2}$, the determinant of the Jacobian of $g$, when viewed as a map from $\mathbb{C} \cong \mathbb{R}^2$ to $\mathbb{C} \cong \mathbb{R}^2$, equals $\frac{1}{|cz+d|^4}$ (this follows immediately from the Cauchy-Riemann equations). Therefore, the substitution $z' = g(z)$ in the above integral yields

$$\int_{\mathbb{H}} f(g(z)) \frac{1}{y(z)^2} dxdy = \int_{\mathbb{H}} f(g(z)) \frac{\frac{1}{|cz+d|^4}}{\frac{y(z)^2}{|cz+d|^4}} dxdy$$

$$= \int_{\mathbb{H}} f(z') \frac{1}{y(z')^2} dx'dy' = \int_{\mathbb{H}} f dA$$

where $z := x + iy$ and $z' = x(z') + y(z')i$ and by noticing that $y(z') = \frac{y(z)}{|cz+d|^2}$ by equation 1.2.0.2.
The proof for the volume form $dm$ on $\mathrm{T}^1\mathbb{H}$ is done analogously by first computing the Jacobian of $Dg$ and then performing a substitution in the integral.

□

**Corollary 2.** For any lattice $\Gamma \leq PSL_2(\mathbb{R})$, the hyperbolic measure defined by the volume form $dm = \frac{1}{y^2} dxdyd\theta$ induces a finite $PSL_2(\mathbb{R})$-invariant measure $m_X$ on $X = \Gamma \setminus PSL_2(\mathbb{R})$: For the canonical quotient map

$$\pi : PSL_2(\mathbb{R}) \to X$$

$$g \mapsto \Gamma g,\ g \in PSL_2(\mathbb{R})$$

and a fundamental domain $F$ for $X$, the $PSL_2(\mathbb{R})$-invariant measure $m_X$ on $X$ is defined by

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

for measurable sets $B \subseteq X$.

*Proof.* Since the group $PSL_2(\mathbb{Z})$ is a lattice of $PSL_2(\mathbb{R})$, we know from Proposition 8 that $PSL_2(\mathbb{R})$ is unimodular. Moreover, it follows from Lemma 3.2.1 that under the identification $PSL_2(\mathbb{R}) \cong \mathrm{T}^1\mathbb{H}$, the Haar measure on $PSL_2(\mathbb{R})$ is given (up to a constant multiple) by the volume form $m$ on $\mathrm{T}^1\mathbb{H}$. Therefore, the statement of this Corollary is now just a consequence of Proposition 8. □

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

In the following, let $\Gamma \leq PSL_2(\mathbb{R})$ denote a lattice and set $X = \Gamma \setminus PSL_2(\mathbb{R})$. We endow $X$ with the metric defined by equation 2.5.0.1. Let

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

The geodesic flow on $X$ is still defined by an action of $PSL_2(\mathbb{R})$ on $X$, given by

$$R_{a_t} : x \mapsto x a_t^{-1}$$

$x \in X$, $t \in \mathbb{R}$. For this name to be appropriate, one needs to show that $\Gamma$ can be used to define a hyperbolic surface $\Gamma \setminus \mathbb{H}$ and that the unit tangent bundle of this surface can be identified with $X$. For more details on the former see Chapter 3 in [BM00] and [FK12].

**Proposition 9.** The identification $\mathrm{T}^1\mathbb{H} \cong PSL_2(\mathbb{R})$ induces an identification

$$\mathrm{T}^1(\Gamma \setminus \mathbb{H}) \cong \Gamma \setminus PSL_2(\mathbb{R})$$

of the unit tangent bundle of $\Gamma \setminus \mathbb{H}$ with the orbit space $\Gamma \setminus PSL_2(\mathbb{R})$. The geodesic flow on $\mathrm{T}^1(\Gamma \setminus \mathbb{H})$ corresponds to the flow

$$\Gamma \setminus PSL_2(\mathbb{R}) \to \Gamma \setminus PSL_2(\mathbb{R})$$

$$\Gamma g \mapsto \Gamma g\ a_t^{-1}$$

$t \in \mathbb{R}$.

*Proof.* The identification $\mathrm{T}^1\mathbb{H} \cong PSL_2(\mathbb{R})$, given by the map $\varphi$ (sending $(z, v)$ to the unique $g \in PSL_2(\mathbb{R})$ with $Dg(i, i) = (z, v)$) shows in equation 1.3.3.2 that the action of $PSL_2(\mathbb{R})$ on $\mathrm{T}^1\mathbb{H}$ commutes with the left multiplication in this group. The result is thus a consequence of Proposition 2. □

**Remark 8.** Given a lattice $\Gamma \leq PSL_2(\mathbb{R})$, the above Proposition tells us that the unit tangent bundle of the hyperbolic surface $\Gamma \setminus \mathbb{H}$ takes the form $\Gamma \setminus PSL_2(\mathbb{R})$. It turns out that $\mathrm{T}^1(\Gamma \setminus \mathbb{H})$ can also be represented by $\Gamma' \setminus SL_2(\mathbb{R})$, for the symmetric lattice subgroup $\Gamma' \leq SL_2(\mathbb{R})$, $\Gamma' = \pi^{-1}(\Gamma)$ and this is what we will do in the discussion of dynamics of the geodesic and horocycle flows on hyperbolic surfaces, starting from Chapter 5. The reason why this is true is that $\Gamma' \setminus SL_2(\mathbb{R})$ is isomorphic to $\pi(\Gamma') \setminus PSL_2(\mathbb{R})$, and that any symmetric lattice subgroup $\Gamma' \leq SL_2(\mathbb{R})$ corresponds to a unique lattice $\Gamma \leq PSL_2(\mathbb{R})$.

Heuristically speaking, the geodesic flow for time $t$ on $X$ can be understood by following an arrow in $\mathrm{T}^1(\Gamma \setminus \mathbb{H})$ for the given amount of time until it intersects the boundary of a fundamental domain $F$. In order to understand how this action looks like, a specific fundamental $F$ domain for the action of a lattice $\Gamma \leq PSL_2(\mathbb{R})$ on $PSL_2(\mathbb{R})$ needs to be established, as only the identifications of the boundaries in $F$ tell us how the geodesic flow crossing such a boundary when viewed as a flow on $PSL_2(\mathbb{R})$ looks like in $\Gamma \setminus PSL_2(\mathbb{R})$: Since as soon as the flow reaches a boundary of $F$ pointing outwards, there is exactly one element of $\Gamma$ that maps the flow to an array pointing inwards the fundamental domain $F$.
We may look at the example $\Gamma = PSL_2(\mathbb{Z})$ and the fundamental domain $F$ established in equation 3.2.0.1.
Suppose $x = \Gamma g \in X = PSL_2(\mathbb{Z}) \setminus PSL_2(\mathbb{R})$ is such that $g \in F$ and represent $x$ by the arrow $(g(i), v_x) \in \mathrm{T}^1(\Gamma \setminus \mathbb{H})$. Then for all $t \in \mathbb{R}$, the geodesic flow $R_{a_t}(g)$ lies on the geodesic at time $t$ defined by this arrow, as long as it does not hit the boundary of $E$. As soon as it does (see Figure 3.2), depending on which boundary, a unique element of $\Gamma$ maps $R_{a_t}(g)$ to the arrow pointing inwards to $E$. Then the geodesic flow follows the geodesic determined by this new arrow until it hits (if it ever does) the next boundary of $E$.

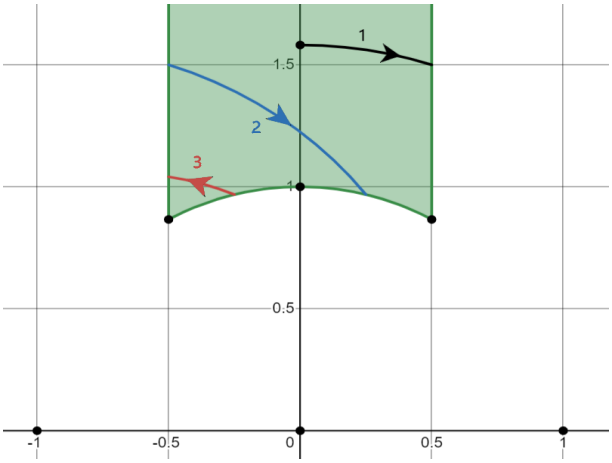


Figure 3.2: The trajectory of a geodesic flow hitting the boundaries of a fundamental domain.

# Chapter 4

# Ergodic Theory and Unitary Representations

In this chapter, we give a brief introduction to fundamental concepts from ergodic theory and unitary representations, laying the basis for the theory of dynamical systems on the quotient space of the group $SL_2(\mathbb{R})$ by a lattice.

## 4.1 Ergodicity of measure-preserving Transformations

Consider a probability space $(X, \mathcal{B}_X, \mu)$. Let $T : X \to X$ be a measure-preserving transformation, that is, $T$ is measurable and such that

$$\mu(T^{-1}(B)) = \mu(B)$$

for all measurable sets $B \in \mathcal{B}$, where $T^{-1}(B)$ denotes the pre-image of $B$ under $T$. The system $(X, \mathcal{B}_X, \mu, T)$ is called a *measure-preserving system*. We call a set $B \subseteq X$ *invariant* under $T$, if $T^{-1}(B) \subseteq B$.
Intuitively, the map $T$ is called *ergodic* (for $\mu$) if the space $X$ cannot be divided into two subsets of positive measure each of which is invariant under $T$.

**Definition 17.** A measure-preserving transformation $T : X \to X$ of a probability space $(X, \mathcal{B}, \mu)$ is *ergodic* if for any $B \in \mathcal{B}$,

$$T^{-1}(B) \subseteq B \;\Rightarrow\; \mu(B) = 0 \text{ or } \mu(B) = 1 \tag{4.1.0.1}$$

Notice that ergodicity is a term concerning both the transformation $T$ and the probability measure $\mu$. In case where the emphasis is on the transformation $T$, we will call $\mu$ an *ergodic measure*, meaning that $T$ is ergodic with respect to the measure $\mu$.
The next Proposition is a classical result in Ergodic theory and it is a partial result of Proposition 2.14 in [EW10] and we refer the reader to this source for a proof.

**Proposition 10.** Let $T$ be a measure-preserving transformation of the probability space $(X, \mathcal{B}, \mu)$. Then the following are equivalent.

1. T is ergodic.
2. For any measurable function $f : X \to \mathbb{C}$, $f \circ T = f$ almost everywhere implies that $f$ is equal to a constant almost everywhere.

The Proposition above is particularly important for the subsequent discussion. It allows for a reinterpretation of ergodicty as a property of a specific unitary operator, which is introduced in the next section. We will see that it generalizes to the case where the dynamical system is given by a group action, instead of the transformation $T$.

## 4.2 Associated Unitary Operators and Unitary Representations of metrizable Groups

Given a linear and continuous operator between Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$

$$U : \mathcal{H}_1 \to \mathcal{H}_2$$

the *adjoint operator* $U^* : \mathcal{H}_2 \to \mathcal{H}_1$ is defined by the relation

$$\langle Uv, w\rangle_{\mathcal{H}_2} = \langle v, U^* w\rangle_{\mathcal{H}_1}$$

for all $v \in \mathcal{H}_1$, $w \in \mathcal{H}_2$. As usual, continuity or any topological statements concerning $\mathcal{H}_1$ and $\mathcal{H}_2$ are meant with regards to the norm induced by the respective inner product. The operator $U$ is called an *isometry* if $\|Uv\|_{\mathcal{H}_2} = \|v\|_{\mathcal{H}_1}$ for all $v \in \mathcal{H}_1$.

**Definition 18.** The linear operator $U : \mathcal{H}_1 \to \mathcal{H}_2$ is called *unitary* if $U$ is invertible and $U^{-1} = U^*$ or equivalently, if $U$ is invertible and

$$\langle Uv_1, Uv_2\rangle_{\mathcal{H}_2} = \langle v_1, v_2\rangle_{\mathcal{H}_1}$$

for all $v_1, v_2 \in \mathcal{H}_1$.

**Remark 9.** A unitary operator as above is an isometry.

There is a a close connection between the theory of unitary operators and ergodicity of measure-preserving systems. Any measure-preserving system $(X, \mathcal{B}, \mu, T)$ gives rise to a specific unitary operator $U$ as follows. Recall that the space

$$L^2_\mu(X) = \left\{ f : X \to \mathbb{C} \text{ measurable } \; : \sqrt{\int_X |f(x)|^2 d\mu(x)} \; < \infty \right\}$$

is a Hilbert space with inner product given by

$$\langle f_1, f_2\rangle_{L^2_\mu} = \int_X f_1(x) \; \overline{f_2(x)} d\mu(x)$$

for $f_1, f_2 \in L^2_\mu(X)$. A natural operator on $L^2_\mu(X)$ associated to the transformation $T$ is the so-called *Koopman operator* $U_T$ for $T$, given by the mapping

$$U_T : f \mapsto f \circ T$$

for $f \in L^2_\mu(X)$. The Koopman operator is unitary. Indeed, we have

$$\langle U_T f_1, U_T f_2 \rangle_{L^2_\mu} = \int_X (f_1 \circ T)(x)\overline{(f_2 \circ T)(x)}\, d\mu(x)$$

$$= \int f_1(x)\overline{f_2(x)}\, d\mu(x) = < f_1, f_2 >_{L^2_\mu}$$

where we used the substitution $x \mapsto T(x)$ which cancels out the change to the pushforward measure $(T * \mu)(x) = \mu(T^{-1}(x))$, as $\mu$ is $T$-invariant.
The next result is Lemma 2.18 in [EW10].

**Lemma 4.2.1.** *A measure-preserving transformation $T$ is ergodic if and only if 1 is an eigenvalue of the operator $U_T$ on $L^2_\mu(X)$ with multiplicity one.*

*Proof.* This follows from Proposition 10 (v) which states that the transformation $T$ is ergodic if and only if all $\mu$-measurable functions $f : X \to \mathbb{C}$ satisfying $U_T(f) = f$ almost everywhere (a.e.) are equal to a constant a.e.
Indeed, first assume that $T$ is ergodic. The function $f \equiv 1$ is an eigenfunction of $U_T$ with eigenvalue 1. If $g$ is another eigenfunction to the eigenvalue 1, then by assumption, $g$ is constant a.e. and so in particular $g = cf$ a.e. for some constant $c$.
Conversely, assume that 1 is an eigenvalue of $U_T$ with multiplicity one. Then all functions, and so in particular all measurable functions fixed by $U_T$ are a constant multiple of 1, which again, by Proposition 10 (v) implies that $T$ is ergodic. □

**Remark 10.** Lemma 4.2.1 shows that ergodicity of $T$ is a property of its associated unitary operator. In general, any characteristic of a measure-preserving transformation that can be discovered by studying the associated Koopman operator on the Hilbert space $L^2_\mu(X)$ is called a *spectral* or a *unitary* property.

So far, we have only considered a dynamical system involving a single measure-preserving transformation $T$ on $X$. The aim is to extend this notion to the setting where the transformation $T$ is replaced by a continuous and measure-preserving group action of a closed linear group $G$ on a locally compact metric space $X$ equipped with a finite measure $\mu$. The continuity of the action is meant not only by continuity of the individual maps $g \in G$: $x \mapsto g(x)$, $x \in X$, but in the following stronger sense.

**Definition 19.** The group $G$ as above acts continuously on $X$ if the map $G \times X \to X$ defined by

$$(g, x) \mapsto g(x)$$

is continuous with respect to the product topology induced by the metrics of $G$ and $X$ and the metric on $X$.

Let

$$U_g : L^2_\mu(X) \to L^2_\mu(X)$$

denote the Koopman operator defined by

$$U_g(f) = f \circ R_{g^{-1}} = x \mapsto f(xg)$$

The family of operators $\{U_g, g \in G\}$ will be crucial for the analysis of the dynamical system $(X, \mathcal{B}_X, \mu, G)$. There are some preliminary steps to be made. We follow the proof of Lemma 8.7 in [EW10] for the next result. We will write $g^{-1} \cdot x = xg$ to indicate the action of $G$ on $X$ given by the inverse of right multiplication.

**Lemma 4.2.2.** *Let $G$ be a $\sigma$-compact, locally compact metrizable group acting continuously on a $\sigma$-compact, locally compact metric space $X$, equipped with a finite measure $\mu$. Assume that $\mu$ preserves the action of $G$. Then for any fixed function $f \in L^2_\mu(X)$,*

$$g \mapsto U_g f$$

*is a continuous map from $G$ to $L^2_\mu(X)$ with respect to the norm $\| \cdot \|_{L^2_\mu}$.*

*Proof.* Denote by $C_c(X)$ the space of continuous functions on $X$ with compact support and fix $f \in L^2_\mu(X)$. Recall from functional analysis that $C_c(X)$ is dense in $L^2_\mu(X)$, so given $\epsilon > 0$, there exists $h \in C_c(X)$ with

$$\|f - h\|_{L^2_\mu} < \epsilon. \tag{4.2.2.1}$$

As $G$ is locally compact, there exists a compact neighbourhood $L_0$ of the identity $I \in G$. Then the set $L = L_0 \cap L_0^{-1}$ is a *symmetric* compact neighbourhood of $I$. That is, $L^{-1} = L$, where $L^{-1} = \{l^{-1} \mid l \in L\}$. Now consider

$$K = L \cdot \mathrm{Supp}(h) = \{v \cdot x \mid v \in L, x \in \mathrm{Supp}(h)\}$$

which is also compact by the continuity of the $G$-action ($K$ is the image of the continuous group action restricted to the compact set $L \times \mathrm{Supp}(h)$). Since $h$ is uniformly continuous when restricted to the compact set $K$, there exists some $\delta > 0$ such that for all $x, y$ with $d_X(x, y) < \delta$, we have

$$|h(x) - h(y)| < \epsilon/(\mu(K))^{1/2}$$

Similarly, the continuous group action restricted to the compact set $L \times K$ is uniformly continuous. Therefore, we can choose a small enough symmetric neighbourhood $L' \subseteq L$ such that $g \in L'$ implies $d_X(I \cdot x, g \cdot x) < \delta$ for all $x \in K$ and the identity $I \in G$. Combining these two facts, we get $g \in L'$ implies $|h(x) - h(g^{-1} \cdot x)| < \epsilon/(\mu(K))^{1/2}$ and so

$$\int_K |h(x) - h(g^{-1} \cdot x)|^2 d\mu < \epsilon^2$$

If $x \notin K$, we have $h(x) = 0$ (as the support of $h$ is contained in $K$) and $h(g^{-1} \cdot x) = 0$ for $g \in L'$. Therefore, $|h(x) - h(g^{-1} \cdot x)| = 0$ and we get by integrating over all of $X$ that

$$\int_X |h(x) - h(g^{-1} \cdot x)|^2 d\mu < \epsilon^2$$

for $g \in L'$. As $U_g$ is an isometry, 4.2.2.1 implies $\|U_g f - U_g h\|_{L^2_\mu} < \epsilon$ and so for $g \in L'$,

$$\|f - U_g f\|_{L^2_\mu} < 3\epsilon$$

by the triangle inequality. In general, $g_0 \in G$ and $g \in g_0 L'$ implies that

$$\|U_{g_0} h - U_g h\|_{L^2_\mu} = \|h - U_{g_0^{-1} g} h\|_{L^2_\mu} < \epsilon$$

which proves the Lemma. □

**Definition 20.** Let $V$ be a vector space and denote by $Gl(V)$ the general linear group of invertible linear transformations on $V$. A *representation* of $G$ on $V$ is a group homomorphism

$$\pi : G \to Gl(V)$$

**Remark 11.** A representation of $G$ as above is sometimes, depending on the context, denoted by the tuple $(\pi, V)$.

**Definition 21.** A representation $(\pi, V)$ of a linear group $G$ onto the vector space $V$ is called *trivial* if for all $g \in G$, the map $\pi(g)$ is the identity element of $Gl(V)$.

**Definition 22.** A representation $U$ of $G$ on a Hilbert space $\mathcal{H}$ is called a *strongly continuous unitary* representation if $\forall g \in G : U(g) =: U_g$ is unitary and if the mappings

$$G \to \mathcal{H},\ g \mapsto U_g(\xi)$$

are continuous for all $\xi \in \mathcal{H}$ (equivalently, $U$ is continuous in the strong operator topology).

**Remark 12.** In the rest of this thesis, all unitary representations of $G$ in consideration will be strongly continuous. Therefore, to make notation easier, we will refer to a strongly continuous unitary representation as just a unitary representation.

Under the assumption that the finite measure $\mu$ leaves the $G$-action invariant, Lemma 4.2.2 shows that the representation $U : G \to Gl(L^2_\mu(X))$ sending $g \in G$ to the Koopman operator $U_g$ is in fact a unitary representation of $G$ onto the Hilbert space $L^2_\mu(X)$. This representation is called the *regular representation* of $G$ on $L^2_\mu(X)$.

## 4.3 Ergodicity of a Group Action

We proceed by characterizing an *ergodic group action* of $G = SL_2(\mathbb{R})$ on the measure-preserving system $(X, \mathcal{B}_X, m_X)$, where $X = \Gamma \setminus G$ is the quotient of $G$ by a lattice $\Gamma$. This amounts in investigating whether a family of functions given by the action of $G$ on $X$ is ergodic. We follow Chapter 1 in [BM00].

**Definition 23.** A measurable subset $A \subseteq X$ is called *invariant* under (the action of) a subgroup $H \leq G$ of $G$, if

$$h^{-1} \cdot A \subseteq A$$

for all $h \in H$.

**Definition 24.** The action of a subgroup $H \leq G$ of $G$ on $X$ is called *ergodic* if there are no non-trivial invariant subsets of $X$. That is, for any measurable subset $A$ of $X$, invariant under the action of $H$, we have $m_X(A) \in \{0, 1\}$.

**Definition 25.** Let $H \leq G$ denote a closed subgroup of $G$. A measurable function $f : X \to \mathbb{C}$ is called *essentially $H$-invariant* if, for any $h \in H$, one has $f(h \cdot x) = f(x)$ for $m_X$-almost all $x \in X$. The function $f : X \to \mathbb{C}$ is called *$H$-invariant* if $f(h \cdot x) = f(x)$ for all $h \in H$ and all $x \in X$.

The equivalence between a single ergodic transformation $T$ and the absence of measurable and almost everywhere (a.e.) invariant functions that are not constant a.e. (see 10) cannot be imposed into the setting of $H$-ergodic actions directly. It is clear that if any measurable and essentially $H$-invariant function $f : X \to \mathbb{C}$ is constant almost everywhere, then the action of $H$ is ergodic (just consider a measurable, $H$-invariant subset $A \subseteq X$ and set $f = \mathbb{1}_A$). However, the converse implication may not hold. Suppose $H$ acts ergodically on $X$ and consider a measurable function $f$ on $X$ that is essentially $H$-invariant. Then in particular, it is invariant almost everywhere under the action of any fixed $h \in H$, say on the set $X \setminus X_h$. The problem now is that the union of nullsets $\bigcup_{h \in H} X_h$ on which $f$ is not invariant may not be itself a nullset, as it may not be countable. In particular, even if $f$ can be shown to be constant on $X \setminus \bigcup_{h \in H} X_h$, it may not be constant almost everywhere. In case if $f$ is $H$-invariant, we do not have this issue and, as is shown in the proof of the next Theorem, the implication holds in both directions. Fortunately, as one may check in Lemma 1.2 of Chapter 1 in [BM00]; an essentially $H$-invariant function on $X$ agrees almost everywhere with an $H$-invariant function. We formulate this statement below.

**Lemma 4.3.1.** *Let $f : X \to \mathbb{C}$ be a measurable essentially $H$-invariant function on $X$. Then there exists a measurable $H$-invariant function $\tilde{f} : X \to \mathbb{C}$ such that $f = \tilde{f}$ almost everywhere on $X$.*

**Theorem 4.** Let $H \leq G$ be a closed subgroup of $G$, acting on the finite measure space $(X, \mathcal{B}_X, m_X)$. Then the following are equivalent:

(1) The action of $H$ is ergodic;

(2) If $f : X \to \mathbb{C}$ is measurable and essentially $H$-invariant, then $f$ is constant almost everywhere.

*Proof.* We may assume without loss of generality that $m_X(X) = 1$. Let $A \subseteq X$ be a $H$-invariant subset. Then $\mathbb{1}_A$ is $H$-invariant and constant almost everywhere. Thus $m_X(A) \in \{0, 1\}$. Hence, (2) implies (1).
To prove the other direction, assume that the action of $H$ on $X$ is ergodic. Let $f$ be a measurable and essentially $H$-invariant function on $X$. Using the result of Lemma 4.3.1, we may assume that $f$ is $H$-invariant. Consider the set

$$X(k, n) = \left\{ x \in X : \frac{k}{2^n} \leq \mathrm{Re}(f(x)) < \frac{k+1}{2^n} \right\}$$

$k \in \mathbb{Z}$, $n \in \mathbb{N}$. Then $X(k, n)$ is measurable and $H$-invariant, and so by assumption

$$m_X(X(k, n)) \in \{0, 1\}, \quad \forall k \in \mathbb{Z}, \forall n \in \mathbb{N}.$$

We may partition the space $X$ as follows. For a fixed $n \in \mathbb{N}$ we have

$$X = \bigsqcup_{k \in \mathbb{Z}} X(k, n).$$

Next, associate to each $n$ a number $k_n \in \mathbb{Z}$ such that $m_X(X(k_n, n)) = 1$. Since the partition is refining for increasing $n$, we have

$$X(k_n, n) \cap X(k_{n+1}, n+1) = X(k_{n+1}, n+1).$$

Set $Y = \bigcap_{n=1}^{\infty} X(k_n, n)$. Then $m_X(Y) = 1$ and $\mathrm{Re}(f)$ is constant on $Y$.
We may analogously construct a set $Y'$, $m_X(Y') = 1$ on which $\mathrm{Im}(f)$ is constant. Then $f$ is constant on $Y \cap Y'$, and as $m_X(Y \cap Y') = 1$ the proof is complete. □

**Remark 13.** From now on, we will refer to essentially invariant functions as invariant functions, because they differ only on sets of measure zero, making the distinction irrelevant in upcoming computations.

The next result is a strengthening of Theorem 4, which will be useful later. Define the space

$$L^2_0(X) = \left\{ f \in L^2_\mu(X) \mid \int_X f(x) d\mu(x) = 0 \right\}$$

consisting of square-integrable functions with zero mean.

**Theorem 5.** Denote by $U^0$ the unitary representation of a closed subgroup $H \leq G$ obtained by restricting its regular representation to $L^2_0(X)$. The action of $H$ is ergodic if and only if there is no non-zero subspace on which $U^0$ is trivial.

*Proof.* If the action of $H$ is ergodic, then by Theorem 4 the only trivial representation of $H$ is the one restricted to the constant functions in $L^2_\mu(X)$, which are, except for the zero-function, not contained in $L^2_0(X)$. Conversely, if the representation $U^0$ is trivial on a non-zero subspace of $L^2_0(X) \subseteq L^2_\mu(X)$, then the eigenvalue 1 of $U$ is not simple (since the constant functions are always eigencvectors of $U$ for the eigenvalue 1) and by Theorem 4, the action of $H$ is not ergodic. □

**Remark 14.** We may thus characterize ergodicity of the group action $H \leq G$ by the absence of invariant functions in $L^2_\mu(X)$ with zero mean.

# Chapter 5

# Ergodicity of Transformations on on the Quotient Space of $SL_2(\mathbb{R})$ by a Lattice

Throughout this chapter, let $\Gamma \leq SL_2(\mathbb{R})$ denote a lattice in $SL_2(\mathbb{R})$. If not otherwise specified, we denote $G = SL_2(\mathbb{R})$, $X = \Gamma \setminus SL_2(\mathbb{R})$ and we let $\mathcal{B}_X$ denote the Borel $\sigma$-algebra of $X$.
We analyze the measure-preserving dynamical systems on the space $(X, \mathcal{B}_X)$ endowed with the Haar measure $m_X$. More precisely, we investigate ergodic properties of transformations on that space, given by one-parameter subgroups or single elements of $G$. We may assume without loss of generality that the measure $m_X$ is normalized, meaning that $m_X(X) = 1$. We follow Section 3 of Chapter 11 in [EW10].

The main ingredient in the proofs of ergodic properties of transformations on the dynamical system $(X, \mathcal{B}_X, m_X)$ is a translation of the ergodic property into the language of representation theory, using results obtained on the group structure of $G$ and different versions of the *Mautner phenomenon*.
The Mautner phenomenon, first stated by F. Mautner in [Mau57] in 1957, is a general assertion about the fixed points of unitary representation of a Lie group $G$ on a Hilbert space $\mathcal{H}$ restricted to a one parameter subgroup $< a(t) >$ of $G$. It states that if a vector $v \in \mathcal{H}$ is fixed by the representation restricted to $< a(t) >$, then $v$ must also be fixed by a (much) larger subgroup $H$ of $G$. As is explained in [Moo80], the "size" of $H$ (in comparison to the size of $< a(t) >$) depends on how *non-commutative* the Lie group $G$ is. Thus, naturally, the Mautner phenomenon can be formulated in different ways, of which some are more general than the others.

We will first use a very specific form of this phenomenon, namely Lemma 5.2.2, which applies to the regular representation of $G = SL_2(\mathbb{R})$ on the Hilbert space $L^2_{m_X}(X)$, restricted to the one parameter subgroup of $G$ given by its positive diagonal elements. It turns out that this is already enough to prove ergodicity of the geodesic flow on $(X, \mathcal{B}_X, m_X)$.
The generalization of this Lemma, which is stated in Proposition 11, will provide more insights into the dynamical system $(X, \mathcal{B}_X, m_X, R_g)$, $g \in G$, given by the right translation of an element $g \in G$. More explicitly, it will enable us to answer the question about whether $(X, \mathcal{B}_X, m_X, R_g)$ is ergodic. The last result of this section is a proof for ergodicity of the horocycle flows on $X$, which will then be a direct consequence of the results obtained before.

Let $A$ and $U^-$ denote the one-parameter subgroups of $G$ given by

$$A = \left\{ a(t) = \begin{pmatrix} \mathrm{e}^{-t/2} & 0 \\ 0 & \mathrm{e}^{t/2} \end{pmatrix} \ : \ t \in \mathbb{R} \right\}, \ U^- = \left\{ u^-(s) = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \ : \ s \in \mathbb{R} \right\}$$

Recall that the geodesic and the horocycle flows on $X$ are respectively defined by

$$R_{a_t}(x) = x a_t^{-1} = x \begin{pmatrix} \mathrm{e}^{t/2} & 0 \\ 0 & \mathrm{e}^{-t/2} \end{pmatrix}$$

$$R_{u^-(s)}(x) = x u^-(-s) = x \begin{pmatrix} 1 & -s \\ 0 & 1 \end{pmatrix}$$

$x \in X$ and $s, t \in \mathbb{R}$.
The geodesic and the horocycle flows can thus be understood as the actions of the one-parameter subgroups $A \leq G$, respectively $U^- \leq G$ on $X$.

## 5.1 Preliminaries

**Definition 26.** The elments $g_1, g_2 \in G$ of a group $G$ are said to be *conjugate* if there exists some $h \in G$ with $g_1 = h g_2 h^{-1}$.

**Lemma 5.1.1.** *Any $g \in SL_2(\mathbb{R})$ is conjugate to some element of $\pm A, \pm U^-$ or $SO(2)$.*

*Proof.* Given an element $g \in G$, there are three cases to consider:

*Case 1*: $g$ is diagonalizable over $\mathbb{R}$.
Such element $g$ is called *hyperbolic*. It is conjugate to

$$\pm D = \begin{pmatrix} d & 0 \\ 0 & 1/d \end{pmatrix}$$

for some $d \in \mathbb{R}$ (remark that $1 = \det(hDh^{-1}) = \det(D)$ and so $D$ must be if this form). Then set $t \in \mathbb{R}$ so that $|d| = \mathrm{e}^{-t/2}$.

*Case 2*: $g$ is diagonalizable over $\mathbb{C} \setminus \mathbb{R}$.
Such element $g$ is called *elliptic*. Its eigenvalues $\lambda_1$ and $\lambda_2$ are complex conjugate to each other and they are the entries of the conjugate diagonal matrix $D$, so again because the determinant of $g$ is equal to one, $\lambda_1 = \mathrm{e}^{i\theta}$, $\lambda_2 = \mathrm{e}^{-i\theta}$ for some $\theta \in \mathbb{R}$. Thus, $D \in SO(2)$ and $g$ is conjugate to an element of $SO(2)$.

*Case 3*: $g$ is not diagonalizable.
Such element $g$ is called *parabolic*. The Jordan normal form of $g$ is

$$\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$$

for some $s \in \mathbb{R}$ which is an element of $U^-$. □

**Lemma 5.1.2.** *Let $g_1, g_2, h \in G$ such that $g_2 = hg_1h^{-1}$. Then*

$$R_{g_2} = R_h R_{g_1} R_{h^{-1}}$$

*viewed as maps on $X$. Moreover, the measure preserving systems $(X, \mathcal{B}_X, m_X, R_{g_1})$ and $(X, \mathcal{B}_X, m_X, R_{g_2})$ are conjugate (i.e. measurably isomorphic).*

*Proof.* The first statement immediately follows from the definition of the right translation $R_g$ by some $g \in G$. In particular, since $\Gamma$ is a lattice, the action of $R_h$ on $X$ is invariant under the finite measure $m_X$ by Proposition 8, proving the second part of the Lemma. □

Lemma 5.1.2 shows that measure-theoretic qualities of the action by elements in $G$ are preserved under conjugation. In particular, ergodicity of the action by some $g \in G$ implies ergodicity of the action by any element in the conjugacy class of $g$. Now, we know that the group $SO(2)$ is compact, and so the orbit of a point $x \in X$ under right translation by elements $g \in SO(2)$ is compact and therefore exhibits little interesting dynamics. Lemma 5.1.2 tells us that in fact, all elements of the larger group $G$ conjugate to such $g$ (i.e. all elliptic elements of $G$) won't be of interest when investigating ergodic transformations of $X$. Therefore, it remains to study ergodicity of elements conjugate to the geodesic or the horocycle flow on $X$.

## 5.2 Ergodicity of the Geodesic Flow on $\Gamma \backslash SL_2(\mathbb{R})$

Denote the set of lower unitriangular matrices in $G$ by $U^+$. Recall from Lemma 1.3.2 that $U^-$ and $U^+$ generate $G$. The next result is Lemma 3.5 in [BM00].

**Lemma 5.2.1.** *Let*

$$g = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \in A.$$

*If $a < 1$, the convergence*

$$\lim_{n \to \infty} g^n h g^{-n} = I$$

*holds for $h \in U^-$. Otherwise, if $a > 1$, then it holds for $h \in U^+$.*

*Proof.* We prove this by computing the relevant matrices. We have

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & a^2 x \\ 0 & 1 \end{pmatrix}$$

and so

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}^{n} \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}^{-n} = \begin{pmatrix} 1 & a^{2n} x \\ 0 & 1 \end{pmatrix}$$

which tends to the identity if $a < 1$. If instead $a > 1$, then a similar computation on the transposed of $h$ above (which is then in $U^+$) gives

$$g^n h g^{-n} = \begin{pmatrix} 1 & 0 \\ a^{-2n} x & 1 \end{pmatrix}$$

proving the lemma. □

A specific version of the Mautner Phenomenon is stated below and is similar to Lemma 3.6 in [BM00].

**Lemma 5.2.2** (Mautner's Lemma)**.** *Let $U$ denote the regular representation of $G$ on the Hilbert space $L^2_{m_X}(X)$. Assume that $g, h \in G$ satisfy*

$$\lim_{n\to\infty} g^n h g^{-n} = I.$$

*Then if $f \in L^2_{m_X}(X)$ is fixed by $U_g$, it is also fixed by $U_h$.*

**Remark 15.** As the $L^2_{m_X}$-norm on $X$ identifies elements that are equal up to measure zero, the fixed vectors $f \in L^2_{m_X}$ of $U_g$, $g \in G$ are precisely the functions in $L^2_{m_X}$ that are invariant under the transformation $R_g$ almost everywhere.

*Proof of Lemma 5.2.2.* It holds that

$$\|U_h f - f\|_{L^2_{m_X}(X)} = \|U_h U_{g^{-n}} f - U_{g^{-n}} f\|_{L^2_{m_X}(X)}$$

as the substitution $f \to U_{g^{-n}} f$ does not change the norm of $(U_h f - f)$ by right-invariance of the measure $\mu$. By unitarity of $U_g$, the above is equal to

$$\|U_{g^n} U_h U_{g^{-n}} f - f\|_{L^2_{m_X}(X)}.$$

Since by assumption $\lim_{n\to\infty} g^n h g^{-n} = I$, we see that

$$\|U_h f - f\|_{L^2_{m_X}(X)} = \lim_{n\to\infty} \|U_{g^n} U_h U_{g^{-n}} f - f\|_{L^2_{m_X}(X)} = 0$$

by the strong continuity of the operator $U$. This shows $U_h f = f$ in $L^2_{m_X}(X)$. □

**Corollary 3.** The geodesic flow on $X = \Gamma \setminus G$ is ergodic with respect to the Haar measure $m_X$ on $X$.

*Proof.* The geodesic flow is represented by the one-parameter subgroup

$$A = \left\{ a(t) = \begin{pmatrix} \mathrm{e}^{-t/2} & 0 \\ 0 & \mathrm{e}^{t/2} \end{pmatrix} \ : \ t \in \mathbb{R} \right\}.$$

Suppose a function $f \in L^2_{m_X}(X)$ is $A$-invariant. Then $U_a f = f$ for all $a \in A$. Given any $h \in U^+ \cup U^-$, Lemma 5.2.1 tells us that there exists an element $g_a \in A$ so that $\lim_{n\to\infty} g_a^n h g_a^{-n} = I$. Hence, by the previous Lemma, we have $U_h f = f$ in $L^2_{m_X}(X)$. Equivalently, $f$ is invariant a.e. under $R_h$. This holds for all $h \in U^+ \cup U^-$. But since $U^+ \cup U^-$ generates $G$, it follows that $f$ is essentially $G$-invariant and so it is equal to a $G$-invariant function $f'$ almost everywhere. As the group action is transitive, $f'$ must be constant. Since we have characterized ergodicity by the absence of invariant functions in $L^2_0(X) \subseteq L^2_{m_X}(X)$ in Theorem 5, it follows that the geodesic flow is ergodic. □

## 5.3 Ergodicity for measure-preserving Transformations on $\Gamma \backslash SL_2(\mathbb{R})$ by individual Elements in $SL_2(\mathbb{R})$

The ergodicity of the geodesic flow proven above may be considered as a rather weak property, since only the functions $f \in L^2(X)$ invariant under the whole action of the subgroup $A \leq G$ are demanded to be constant. But in fact, a much stronger result holds as well.

**Theorem 6.** If $g \in G$ is not conjugate to an element of $SO(2)$, then $R_g$ acts ergodically on $(X, \mathcal{B}_X, m_X)$.

The proof of this Theorem requires a similar but stronger result than Mautner's Lemma 5.2.2 which was discussed in the last section. The main component is Proposition 11, which we state below.

In a metric group $G$ equipped with a left-invariant metric $d_G$, let $B^G_\delta(I)$, $\delta > 0$ denote the open ball around the identity $I \in G$ of radius $\delta$. Write

$$B^G_\delta(h) = hB^G_\delta(I)$$

$h \in G$ for the open ball of radius $\delta$ around $h$.

**Proposition 11.** Consider a unitary representation $(\pi, \mathcal{H}_\pi)$ of a metric group $G$ on a Hilbert space $\mathcal{H}_\pi$. Suppose that a subgroup $L \leq G$ fixes a vector $v_0 \in \mathcal{H}$. Then any $h \in G$ satisfying

$$B^G_\delta(h) \cap LB^G_\delta L \neq \emptyset$$

for all $\delta > 0$ also fixes $v_0$.

We can apply this Proposition to $G = SL_2(\mathbb{R})$ and the regular representation of $G$ to get a strengthening of Lemma 5.2.2 as follows. Define the subgroup $L = g^{\mathbb{Z}}$ for some

$$g = g_a := \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}$$

$a \neq 1$. We have for all $h \in U^+ \cup U^-$, $g^n h g^{-n} \to I$ either for $n \to \infty$ or $n \to -\infty$. In particular, $B_\delta^G(h) \cap LB_\delta^G L \neq \emptyset$ for all $\delta > 0$. Therefore, a consequence of this Proposition is:

**Corollary 4.** For every non-trivial $g$ contained in the subgroup $A \leq G = SL_2(\mathbb{R})$, the measure-preserving system $(X, \mathcal{B}_X, m_X, R_g)$ is ergodic.

*Proof.* Let $U$ denote the regular representation of $G$ on $L^2_{m_X}(X)$ and let $g \in A$. If $f \in L^2_{m_X}(X)$ is fixed by $U_g$, then it follows from Proposition 11 applied to the subgroup $L = g^{\mathbb{Z}}$ that $f$ is also fixed by $U_h$, for any $h \in U^+ \cup U^-$. As $U^+ \cup U^-$ generates $G$, $f$ is essentially $G$-invariant and so, as we argued in the proof of Corollary 3, it must be equal to a constant almost everywhere. The proof is now complete by Theorem 5. □

**Definition 27.** The functions

$$\lambda_{\xi,\eta} : G \to \mathbb{C}, \ g \mapsto < \pi(g)\xi, \eta >$$

where $\xi, \eta \in \mathcal{H}_\pi$ are called the *matrix coefficients* of the $G$-representation $(\pi, \mathcal{H}_\pi)$.

*Proof of Proposition 11.* After scaling the vector $v_0$ if necessary, we may assume that $\|v_0\| = 1$. Consider the matrix coefficient $\lambda = \lambda_{v_0,v_0} : h \mapsto < \pi(h)v_0, v_0 >$. Recall that we require $(\pi, \mathcal{H}_\pi)$ to be strongly continuous. Hence, by continuity of the inner product on $\mathcal{H}_\pi$, we see that $\lambda$ defines a continuous map on $G$. Furthermore, it is invariant under multiplication by elements in the group $L$. Indeed, given $g_1, g_2 \in L$ and $h \in G$ we have

$$\lambda(g_1 h g_2) = < \pi(g_1)\pi(h)\pi(g_2)v_0, v_0 > = < \pi(h)v_0, \pi(g_1^{-1})v_0 > = \ \lambda(h) \tag{5.3.0.1}$$

since $\pi$ is a unitary group homomorphism and $\pi(L)v_0 = v_0$.
Suppose now that $\lambda(h) = 1$. As $\pi$ is an isometry, we have $\|\pi(h)v_0\| = 1$. Thus

$$1 = < \pi(h)v_0, v_0 > \leq \ \|\pi(h)v_0\|\|v_0\| = 1$$

by the Cauchy-Schwartz inequality which is an equality only if the vectors in the inner product $< v, w >$ are linearly dependent. Thus $\pi(h)v_0 = cv_0$ for some constant $c \neq 0$, implying $\pi(h)(v_0) = v_0$. This shows

$$\lambda(h) = 1 \ \Rightarrow \ \pi(h)(v_0) = v_0.$$

Next, let $(g_n)$ denote a sequence converging to the identity $I \in G$. If $h \in G$ with $B_\delta^G(h) \cap LB_\delta^G L \neq \emptyset$ for all $\delta > 0$, then there exist sequences $(l_n)$ and $(l'_n)$ in $L$ with

$$l_n g_n l'_n \to h, \quad \text{as } n \to \infty.$$

By equation 5.3.0.1 we then see that

$$\lambda(l_n g_n l_n') = \lambda(g_n) \to \lambda(I) = \|v_0\|^2 = 1$$

for $n \to \infty$ by continuity of $\lambda$. On the other hand, again by continuity of $\lambda$

$$\lambda(l_n g_n l_n') \to \lambda(h) \text{ as } n \to \infty.$$

Thus $\lambda(h) = 1$ which, by the above, implies that $v_0$ is fixed by $\pi(h)$. □

*Proof of Theorem 6.* We know from Lemma 5.1.1 that an element $g \in SL_2(\mathbb{R})$ that is not conjugate to one in $SO(2)$ is conjugate to an element of either the group $\pm A$ or the group $\pm U^-$. Lemma 5.1.2 states that the map $R_g$ is ergodic, if and only if the right translation by the conjugated element of $g$ is ergodic. Therefore, it is enough to consider the maps $R_{g_a}$ for some nontrivial $g_a \in A$ and $R_{g_{u_s}}$ for a nontrivial $u_s \in U^-$. We have already established ergodicity of the former in Corollary 4, so all that remains is the assertion for the right translation by an element of $U^-$.
Set

$$g = u_s := \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}, \ s \neq 0$$

and suppose that $f \in L^2_{m_X}(X)$ is invariant under $R_g$. Let $U$ denote the regular representation of $SL_2(\mathbb{R})$ on $L^2_{m_X}(X)$. We proceed by applying Proposition 11 as follows. Let $m, n \in \mathbb{Z}$ and set

$$\begin{aligned} h_\epsilon = g^n u_\epsilon^+ g^m &= \begin{pmatrix} 1 & ns \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \epsilon & 1 \end{pmatrix} \begin{pmatrix} 1 & ms \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 + ns\epsilon & (1 + ns\epsilon)ms + ns \\ \epsilon & 1 + ms\epsilon \end{pmatrix} \end{aligned} \tag{5.3.0.2}$$

Given $\epsilon > 0$ small enough, let $n(\epsilon) = n \in \mathbb{Z}$ such that

$$|(1 + ns\epsilon) - 2| \ < |s|\epsilon < 1.$$

This is possible because $s \in \mathbb{R}$ is fixed and so the above amounts to solve

$$|nc(\epsilon) - 1| < |c(\epsilon)| \iff \left| n - \frac{1}{c(\epsilon)} \right| < 1$$

for $n$ where we denote the constant $c(\epsilon) = s\epsilon > 0$. Similarly choose $m(\epsilon) = m \in \mathbb{Z}$ with

$$\left| ms + \frac{ns}{1 + ns\epsilon} \right| < 1.$$

The upper right entry of the matrix in equation 5.3.0.2 then satisfies

$$|(1 + ns\epsilon)ms + ns| \ < |1 + ns\epsilon| < \ 3$$

by the triangle inequality. Passing to a subsequence if necessary, the matrix $h_\epsilon$ converges to

$$h = \begin{pmatrix} 2 & r \\ 0 & 1/2 \end{pmatrix} \tag{5.3.0.3}$$

as $\epsilon \to 0$, for some $r$ with $|r| \leq 3$. Indeed, this holds by the choices of $n(\epsilon)$, $m(\epsilon)$ and the fact that $G$ is closed. Thus, given $\delta > 0$, there exists $0 < \epsilon < \delta$ so that $h_\epsilon \in B_\delta^G(h)$ and $u_\epsilon^+ \in B_\delta^G$. Then

$$h_\epsilon \in B_\delta^G(h) \cap g^{\mathbb{Z}} B_\delta^G g^{\mathbb{Z}}$$

and therefore

$$B_\delta^G(h) \cap g^{\mathbb{Z}} B_\delta^G g^{\mathbb{Z}} \neq \emptyset.$$

It follows from Proposition 11, that $f$ is invariant almost everywhere under $R_h$. To finish the proof, observe that $h$ is conjugate to the element

$$a := \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \in A.$$

Thus the Theorem follows from the ergodicity of the map $R_a$ and Lemma 5.1.2. □

## 5.4 Ergodicity of the Horocycle Flow on $\Gamma\backslash SL_2(\mathbb{R})$

The proof of Theorem 6 in fact reveals the ergodicity of the horocycle flow on $(X, \mathcal{B}_X, m_X)$ as follows.

**Corollary 5.** The horocycle flow on $(X, \mathcal{B}_X, m_X)$ is ergodic.

*Proof.* Suppose $f \in L^2_{m_X}(X)$ is $U^-$-invariant and consider

$$u(s_1)u_\epsilon^+ u(s_2) = \begin{pmatrix} 1 + s_1\epsilon & s_2(1 + s_1\epsilon) + s_1 \\ \epsilon & s_2\epsilon + 1 \end{pmatrix}$$

$s_1, s_2 \in \mathbb{R}$, $\epsilon > 0$. The application of Proposition 11 is similar as in the proof of Theorem 6, but more straightforward, since we can let the top right-hand entry of the above matrix vanish. Set $s_2 = s_2' := -\frac{s_1}{1+s_1\epsilon}$. We have

$$u(s_1)u_\epsilon^+ u(s_2') = \begin{pmatrix} 1 + s_1\epsilon & 0 \\ \epsilon & \frac{1}{1+s_1\epsilon} \end{pmatrix}.$$

Now, pick any $a > 1$ and consider

$$g_a = \begin{pmatrix} a & 0 \\ 0 & 1/a \end{pmatrix}.$$

Let $s_1 = \frac{a-1}{\epsilon}$ and $0 < \epsilon < \delta$. Then $g_a \in U^- B_\delta^G U^-$. It follows from Proposition 11 that $f$ invariant under $R_{g_a}$ and thus equal to a constant almost everywhere, by ergodicity of the transformation $R_{g_a}$. □

# Chapter 6

# Mixing Dynamical Systems

In this chapter, we introduce a dynamical property stronger than ergodicity, discussed in the previous two chapters. Specifically, we introduce the notion of a *mixing* dynamical system.
As in Chapter 5, if not otherwise specified, we denote $G = SL_2(\mathbb{R})$ and let $\Gamma \leq G$ be any lattice in $G$. We consider the space $(X, \mathcal{B}_X, m_X)$, where $X = \Gamma \setminus G$ is the quotient space of $G$ by a lattice, and where $m_X$ is the Haar measure on $X$, normalized such that $m_X(X) = 1$.
The main result is Theorem 7, which states that the action of $G$ on $(X, \mathcal{B}_X, m_X)$ is mixing. We follow Section 4 of Chapter 11 in [EW10].

## 6.1 Preliminaries

Let us first consider a general measure-preserving system $(X, \mathcal{B}_X, m_X, G)$ given by the action of a locally compact group $G$ on a probability space $(X, \mathcal{B}_X, m_X)$.

**Definition 28.** A sequence $\lambda = (\lambda_n)$ in $G$ is said to converge to infinity if for any compact subset $K \subseteq G$, the number of $n$ such that $\lambda_n \in K$ is finite.

**Definition 29.** The measure-preserving action of a non-compact group $G$ on $X$ is *mixing* if for all $f, g \in L^2_{m_X}(X)$ it holds that

$$\int_X f(g_n \cdot x) g(x) dm_X \to \int_X f(x) dm_X \int_X g(x) dm_X$$

as $n \to \infty$, for any sequence $(g_n) \subseteq G$ with $g_n$ converging to infinity as $n \to \infty$.

**Lemma 6.1.1.** *Suppose the action of $G$ on $X$ is mixing. Then the action of any non-compact subgroup $H \leq G$ on $X$ is mixing.*

*Proof.* This follows immediately from Definition 29. □

## 6.2 Mixing of $G$ on the Quotient Space by a Lattice $\Gamma \leq G$

The goal of this section is to prove the following result.

**Theorem 7.** The action of $G$ on $(X, \mathcal{B}_X, m_X)$ is mixing.

**Corollary 6.** The actions of the geodesic and the horocycle flows on the space $(X, \mathcal{B}_X, m_X)$ are mixing.

*Proof.* This is an immediate consequence of Theorem 7 and of Lemma 6.1.1. □

The proof of Theorem 7 uses a consequence of a result concerning unitary representations of locally compact groups, similar to the statement in Proposition 11.
Let $G$ be a locally compact group and let $\alpha = (a_n) \subset G$ be a sequence in $G$. Define the set

$$S(\alpha) = \{g \in G | I \in \overline{\{a_n^{-1} g a_n \mid n \in \mathbb{N}\}}\}$$

Remark that in case

$$a_n = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}^n \in A,$$

Lemma 5.2.1 tells us that either the subgroup $U^+$ or $U^-$ is contained in $S(\alpha)$. In order to prove Theorem 7 we need to improve the understanding of the set $S(\alpha)$ for more general sequences $\alpha$ in $G$.

**Proposition 12.** Let $G$ be a locally compact group and let $(\pi, \mathcal{H}_\pi)$ be a unitary representation of $G$ on a Hilbert space $\mathcal{H}_\pi$. Let $\alpha = (a_n)$ be a sequence in $G$ with the property that for some $v \in \mathcal{H}_\pi$, $(\pi(a_n)v)$ weak$^*$ converges to a vector $v_0 \in \mathcal{H}_\pi$. Then $v_0$ is fixed by all elements $g$ in the closure of the subgroup generated by the set $S(\alpha)$.

*Proof.* Let $g \in S(\alpha)$ and let $(a_{n_k})$ be a subsequence of $(a_n)$ such that

$$\lim_{k \to \infty} a_{n_k}^{-1} \; g \; a_{n_k} = I$$

By definition of weak$^*$-convergence in a Hilbert space, we get for any $w \in \mathcal{H}_\pi$, that

$$< \pi(g)v_0, w > = < v_0, \pi(g)^{-1} w > = \lim_{k \to \infty} < \pi(a_{n_k})v, \pi(g)^{-1} w >$$

$$= \lim_{k \to \infty} < \pi(g a_{n_k})v, w > .$$

In particular, this holds for $g = I$. So we have

$$< v_0, w > = \lim_{k \to \infty} < \pi(a_{n_k})v, w > .$$

It then follows from the Cauchy-Schwartz inequality that

$$
\begin{aligned}
&| < \pi(g)v_0, w > - < v_0, w > | \\
&= \lim_{k\to\infty} | < \pi(a_{n_k}^{-1})\pi(g a_{n_k})v, \pi(a_{n_k})^{-1} w > - < v, \pi(a_{n_k})^{-1} w > | \\
&\leq \lim_{k\to\infty} \|\pi(a_{n_k}^{-1} g a_{n_k})v - v\| \|w\| = 0
\end{aligned}
$$

by the strong continuity of $\pi$. As this holds for all $w \in \mathcal{H}_\pi$, this proves the statement for $g \in S(\alpha)$. For an element $g$ in the closure of the subgroup generated by $S(\alpha)$, say $g = \lim_{n\to\infty} g_n$ with $g_n = g_n^1 ... g_n^k \in S(\alpha), k \geq 1$, we have $\pi(g)v_0 = \lim_{n\to\infty} \pi(g_n)v_0$ by strong continuity of $\pi$. We have

$$
\begin{aligned}
\|\pi(g_n)v_0 - v_0\| &= \|\pi(g_n^1)...\pi(g_n^k)v_0 - v_0\| \\
&= \|\pi(g_n^1)...\pi(g_n^{k-1})v_0 - v_0\| = ... = \|\pi(g_n^1)v_0 - v_0\| = 0
\end{aligned}
$$

and hence

$$
\|\pi(g)v_0 - v_0\| = \lim_{n\to\infty} \|\pi(g_n)v_0 - v_0\| = 0.
$$

□

In order to make use of this result, we need to ensure that $S(\alpha)$ is non-trivial, possibly depending on the sequence $(\alpha) \subseteq G$ .

**Definition 30.** An element $g \in GL_{dd}(\mathbb{R})$ is called *unipotent*, if $(g - I)^d = 0$, where $I \in GL_{dd}(\mathbb{R})$ denotes the identity.

**Lemma 6.2.1.** *Let $\alpha = (g_n)$ be a sequence in $SL_2(\mathbb{R})$ converging to infinity. Then the set $S(\alpha)$ contains a non-trivial unipotent element.*

*Proof.* Recall from the beginning of Chapter 2 that the map $\phi : SL_2(\mathbb{R}) \to Gl(Mat_{22}(\mathbb{R}))$ sending $g$ to

$$
\phi(g) : v \mapsto g\ v\ g^{-1}
$$

is proper, meaning that the pre-image of $\phi$ of any compact set $K \subseteq SL_2(\mathbb{R})$ is compact (since $\phi$ is proper if and only if the map $\hat{\phi}$ considered in said chapter is). The set $\{g_n, n \geq 1\}$ is contained in the pre-image of $\{\phi(g_n), n \geq 1\}$ under $\phi$, so, in particular, if $\{g_n, n \geq 1\}$ leaves any compact set, then $\{\phi(g_n), n \geq 1\}$ is not compact and thus, $\lim_{n\to\infty} \|\phi(g_n)\| = \infty$. Indeed, for if $\lim_{j\to\infty} \|\phi(g_{n_j})\| = k$ for some $k > 0$, then there would be infinitely many $j$ with $\|\phi(g_{n_j})\| < k + 1$ and the pre-image of these points would be compact, and so infinitely many $g_{n_j}$ would be contained in a compact set, contradicting the assumption made in the Lemma.
Moreover, observe that $\phi$ is trivial for matrices $v \in \mathbb{R}I$. Since the Lie Algebra $\mathfrak{g}$ of $SL_2(\mathbb{R})$ is the set of matrices with trace equal to zero (see Lemma 2.2.2), we can decompose $Mat_{22}(\mathbb{R})$ as

$$
Mat_{22}(\mathbb{R}) = \mathbb{R}I \oplus \mathfrak{g}.
$$

Then for each $n \geq 1$, the norm of the adjoint operator at $g_n$

$$Ad_{g_n} : \mathfrak{g} \to \mathfrak{g},\ v \mapsto g_n\ v\ g_n^{-1}$$

goes to infinity. Hence, we may choose a sequence $(v_n) \subseteq \mathfrak{g}$ with $\|v_n\| \to 0$ and $\|Ad_{g_n}(v_n)\| = c > 0$ for some constant $c > 0$. We may choose $c$ small enough so that exp is injective on the ball of radius $2c$ around $0 \in \mathfrak{g}$. Then, after passing to a subsequence if necessary, we get $\exp Ad_{g_n}(v_n) \to u \neq I$ (since $\|Ad_{g_n}(v_n)\| = c > 0$) while on the other hand, $h_n := \exp v_n \to I$. As $h_n$ approaches the identity, its eigenvalues must be close to one for large $n$. Thus the same must hold for $g_n\ h_n\ g_n^{-1}$, since the eigenvalues of a matrix are invariant under conjugation. Therefore, the limit of $g_n\ h_n\ g_n^{-1}$ has one as its only eigenvalue and the matrix $u$ is unipotent and non-trivial.

□

*Proof of Theorem 7.* Let $X$ be as in the statement of the Theorem and consider a sequence $(a_n) \subseteq G$ converging to infinity. Let $U$ denote the regular representation of $G$. Fix any $f \in L^2_{m_X}(X)$. We want to show that $U_{a_n}(f)$ converges in the weak$^*$-topology to the constant function equal to $\int_X f(x)dm_X(x)$. Said differently, we want for all $h \in L^2_{m_X}(X)$, that the convergence

$$\begin{aligned}\langle U_{a_n}(f), \overline{h}\rangle \to \left\langle \int_X f(x)dm_X(x), \overline{h}\right\rangle &= \int_X \int_X f(x)dm_X(x)\overline{h(x)}dm_X(x) \\ &= \int_X f(x)dm_X(x) \int_X \overline{h(x)}dm_X(x) \\ &= < f, 1 > < \overline{h}, 1 >\end{aligned}$$

holds as $n \to \infty$. Notice that by Definition 29, this proves mixing for the action of $G$ on $X$.
As the operators $U_{a_n}, n \geq 1$ are unitary, we have that

$$\|U_{a_n}(f)\|_2 = \|f\|_2.$$

Now, the Banach-Alaoglu Theorem states that the balls

$$B_r = \{f \in L^2_{m_X}(X) : \|f\| \leq r\}$$

$r \geq 0$, are compact in the weak$^*$-topology, where we use the identification of the dual space $L^2_{m_X}(X)^* \cong L^2_{m_X}(X)$ coming from the Riez Representation Theorem. Therefore, as the sequence $U_{a_n}(f)$ is contained in a weak$^*$-compact set, it has a subsequence, say, $U_{a_{n_k}}(f)$ weak$^*$-converging to some $f_0 \in L^2_{m_X}(X)$.
Consider the sequence $\alpha = (a_{n_k})$. By Lemma 6.2.1, there is a non-trivial unipotent element $u$ of $G$ contained in the set $S(\alpha)$. We know from Proposition 12 that for all $a \in S(\alpha)$, $U_a$ fixes the vector $f_0$. So in particular, it holds that $U_u$ fixes $f_0$. But since $u$ is conjugate to an element of the subgroup $U^- \leq G$ by Lemma 5.1.1, and since the action of such element is ergodic, $f_0$ must be constant almost everywhere, implying that $f_0 = \int_X f(x)dm_X(x)$.

□

We end this section with a version of the *Howe-Moore Theorem*, which we will use in the last chapter of this thesis. The Theorem, originally stated in [HM79] by R. Howe and C. Moore, addresses the vanishing of matrix coefficients for unitary representations of $SL_2(\mathbb{R})$. Initially formulated for this group, the Theorem has since been extended to apply to more general semi-simple Lie groups.
The version we prove below is an immediate consequence of Proposition 12. The more general version applying to unitary representations of $SL_2(\mathbb{R})$ is Theorem 11.25 in Chapter 11 of [EW10]. For a careful explanation of the proof stating the Theorem more generally for semi-simple Lie group, we refer the reader to Chapter 3 in [BM00].

**Theorem 8** (Howe-Moore)**.** Let $(T, H(T))$ denote a unitary representation of $SL_2(\mathbb{R})$ on the Hilbert space $H(T)$. Assume that $T$ leaves no non-zero vector in $H(T)$ invariant. Then for all $v, w \in H(T)$, the matrix coefficients $< T_{a_n} v, w >$ for $a_n \in A$

$$a_n = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}^n$$

$a \in \mathbb{R}^+$, vanish at infinity:

$$\lim_{n \to \infty} \ < T_{a_n} v, w > = 0.$$

*Proof.* Pick any $v \in H(T)$ and suppose without loss of generality that the diagonal element $a$ of the matrix $a_n$ in the statement of the Theorem is greater than one (the proof for $a < 1$ is essentially the same, as will be evident). We have $\|T_{a_n} v\| = \|v\|$, as $T$ is unitary. Thus, as in the previous proof, it follows from the Banach-Alaoglu Theorem that there is a subsequence $(T_{a_{n_k}} v)$ converging in the weak$^*$ topology to some vector which we shall denote as $v_0$. This means that

$$< T_{a_{n_k}} v, w > \to < v_0, w >$$

as $k \to \infty$, for all $w \in H(T)$. Proposition 12 tells us that all elements contained in the closure of the subgroup generated by $S(\alpha)$, $\alpha = (a_{n_k})$ fix the vector $v_0$ and Lemma 5.2.1 tells us that this subgroup contains $U^+$. It follows from a computation similar to the one in the proof of Corollary 5 that there exists a diagonal element $d$ of $SL_2(\mathbb{R})$ which is contained in the set $U^+ B_\delta^G U^+$. Then by Proposition 11, we have that $d$ fixes $v_0$. We may thus apply Proposition 11 to the subgroup $L = d^{\mathbb{Z}}$. Since any $u \in U^- \cup U^+$ is contained in the set $d^{\mathbb{Z}} B_\delta^G d^{\mathbb{Z}}$ by Lemma 5.2.1, it follows that $v_0$ is fixed by all of $SL_2(\mathbb{R})$ (since $U^- \cup U^+$ generates $SL_2(\mathbb{R})$). But we assume that the representation $T$ has no non-zero invariant vectors, and so $v_0 = 0$. This completes the proof. $\square$

# Chapter 7

# Representation Theory on $SL_2(\mathbb{R})$

Given the typical outcomes of ergodic theory, it is common that the applications of geodesic and horocycle flows to other domains lean towards *qualitative* analyzes, which assure certain asymptotic behavior, typically without providing a precise rate of convergence. As mentioned in the introduction of this thesis, there are instances, particularly in certain fields like Number Theory, where this is not sufficient. In the final chapter of this thesis, we aim to find a rate of convergence for the mixing property of the geodesic and the horocycle flow on the usual measure space $(X, \mathcal{B}_X, m_X)$, following closely the work of Ratner in [Rat87]. Here, we introduce some elementary results needed to follow along the arguments of said paper. We follow Chapters 1, 2 and 6 in [Lan85].

## 7.1 Preliminaries

Throughout this section, let $G$ denote a locally compact Hausdorff group and let $H$ be a Banach space. We will always assume that the Haar measure $\mu$ on $G$ is unimodular, thus invariant for left- and right-multiplication in $G$.

An important observation about unimodular measures on $G$ is the following. The unimodular measure of $G$ is invariant under taking the inverse. More precisely, we have for all $f \in C_c(G)$, $g \in G$

$$\int_G f(g^{-1}) d\mu(g) = \int_G f(g) d\mu(g).$$

This result will be useful in upcoming computations. The proof needs a preliminary Lemma.

**Lemma 7.1.1.** *Consider a unimodular closed matrix group $G$. Let*

$$\tau : G \to G, \quad \sigma : G \to G$$

*denote an antiautomorphism and an automorphism on $G$ of degree two, respectively. This means that the maps $\tau$ and $\sigma$ satisfy*

$$\tau(xy) = \tau(y)\tau(x), \quad \sigma(xy) = \sigma(x)\sigma(y)$$

*and*

$$\tau^2(x) = x, \quad \sigma^2(x) = x$$

*for all $x, y \in G$.*
*Then the Haar measure on $G$ is invariant under transformations by $\tau$ and $\sigma$.*

*Proof.* We shall prove the assertion for the antiautomorphism $\tau$, as the one for $\sigma$ is done similarly.

We have

$$\int_G f(\tau(y)x)dy = \int_G f(\tau(\tau(x)y))dy = \int_G (f \circ \tau)(\tau(x)y)dy = \int_G (f \circ \tau)(y)dy$$

$$= \int_G f(\tau(y))dy$$

For all $f \in C_c(G)$, $x \in G$. This shows that the measure $(\tau)_* dy$ is right-invariant. By uniqueness of the Haar measure, there exists a positive number $\Delta(\tau)$ with

$$\int_G f(\tau(y))dy = \Delta(\tau) \int_G f(y)dy.$$

It holds that

$$\begin{aligned} \int_G f(\tau^2(y))dy &= \int_G f(y)dy \\ &= \Delta(\tau) \int_G f(\tau((y))dy = \Delta(\tau)\Delta(\tau) \int_G f(y)dy \\ &= \Delta^2(\tau) \int_G f(y)dy \end{aligned}$$

so $\Delta^2(\tau) = 1$ and hence $\Delta(\tau) = 1$, since $\Delta(\tau)$ is positive. This implies that the Haar measure is invariant under the transformation $\tau$. □

**Corollary 7.** The Haar measure of a unimodular group $G$ is invariant under taking inverses in $G$.

*Proof.* This follows from the observation that

$$\tau : G \to G, \quad x \mapsto x^{-1}$$

$x \in G$ is an antiautomorphism of the type as in Lemma 7.1.1. □

We denote by

$$\pi : G \to Gl(H)$$

the representation of $G$ on $H$, assumed to be continuous in the strong topology.

**Definition 31.** A representation $\pi : G \to Gl(H)$ is called *bounded*, if there exists a constant $C > 0$ such that $\|\pi(g)\| \leq C$ for all $g \in G$, where $\|\cdot\|$ denotes the operator norm. It is called *locally bounded* if for any compact subset $K \subseteq G$, the set $\pi(K)$ is bounded in $Gl(H)$, in the sense that $\sup_{k\in K}\{\ \|\pi(k)\|\ \} \leq C$ for some constant $C > 0$.

**Remark 16.** A unitary representation of $G$ on a Hilbert space $H$ is always bounded by 1.

For the convenience of the reader, we state the Banach-Steinhaus Theorem, which concerns the uniform boundedness of a family of linear operators on a Banach space. This Theorem is a fundamental result in Functional Analysis that ensures, under certain conditions, a uniform bound on the operator norms. The proof can be reviewed on page 2 in [Lan85].

**Theorem 9** (Banach-Steinhaus)**.** Let $E$ be a Banach space and let $\{T_i\}_{i\in I}$ be a collection of bounded operators on $E$. If for all $v \in E$, it holds that

$$\sup_{i\in I}\{\ \|T_i v\|_E\ \} < \infty$$

then $\{T_i\}_{i\in I}$ is bounded as a subset of the algebra $End(E)$ of endomorphisms on $E$.

**Lemma 7.1.2.** *A representation $\pi : G \to Gl(H)$ is locally bounded.*

*Proof.* Let $K$ be any compact subset of $G$. Since the representation $\pi$ is strongly continuous, the set $\pi(K)v$, $v \in H$ is compact in $H$, and so it is bounded. By the Banach-Steinhaus Theorem, it follows that $\pi(K)$ is bounded in $Gl(H)$. □

### 7.1.1 An Algebra Homomorphism associated to G

We define $C_c(G)$ as the set of continuous functions on $G$ with compact support. The *convolution operator* $*$ on $C_c(G)$ is given by

$$\psi * \phi(x) = \int_G \psi(xy^{-1})\phi(y)dy$$

where $dy$ is the Haar measure on $G$. This operator induces a product on $C_c(G)$, thereby endowing the space $(C_c(G), *, +)$ with an algebraic structure.
A representation $(\pi, H)$ of $G$ as in the previous section induces an algebra homomorphism $\pi^1$ as follows. Denote by $End(H)$ the algebra of endomorphisms on $H$. Then the map

$$\pi^1 : C_c(G) \to End(H)$$

defined by

$$\pi^1(\psi)v = \int_G \psi(x)\pi(x)vdx$$

$v \in H$, is an algebra homomoprhism on $C_c(G)$, in the sense that

$$\pi^1(\psi * \phi) = \pi^1(\psi) \circ \pi^1(\phi) =: \pi^1(\psi)\pi^1(\phi) \tag{7.1.2.1}$$

and for all $g \in G$,

$$\pi(g)\pi^1(\psi) = \pi^1(L_{g^{-1}}\psi) \tag{7.1.2.2}$$

where $L_{g^{-1}}\psi(x) = \psi(g^{-1}x)$. Indeed, the last equation follows from the left invariance of the Haar measure. The integral in equation 7.1.2.1 is defined since the map $x \mapsto \psi(x)\pi(x)v$ going from $G$ to $H$ is continuous with compact support. Moreover, $\pi^1$ is a homomorphism for the convolution product. To see this, remark that

$$\pi^1(\psi * \phi)v = \int_G (\psi * \phi)(x)\pi(x)vdx = \int_G \int_G \psi(xy^{-1})\phi(y)\pi(x)vdydx$$

$v \in H$. By Fubini's Theorem, the order of integration can be reversed and substituting $x \mapsto xy$ yields

$$= \int_G \int_G \psi(x)\phi(y)\pi(x)\pi(y)vdxdy = \int_G \int_G \psi(x)\pi(x)\phi(y)\pi(y)vdydx$$

$$= \pi^1(\psi)\pi^1(\phi)v.$$

Since $\psi \in C_c(G)$ has compact support and as $\pi$ is locally bounded by Lemma 7.1.2,

$$\|\pi^1(\psi)\| = \sup_{v\in H, \|v\|=1} \|\pi^1(\psi)v\| = \sup_{v\in H, \|v\|=1} \left\| \int_G \psi(x)\pi(x)vdx \right\|$$

$$\leq \|\psi\|_\infty \sup_{x\in Supp(\psi)} \|\pi(x)\| \ \mu(Supp(\psi)) < C_{\psi,\pi}$$

where $C_{\psi,\pi} > 0$ is a constant depending on $\psi$ and $\pi$ and where we used that the Haar measure of the compact set $Supp(\psi)$ is finite. Hence, the operator $\pi^1(\psi)$ is bounded and we conclude that $\pi^1(\psi) \in End(H)$.

Assume now that $\pi$ is unitary and $H$ is a Hilbert space. Then the condition on the arguments on $\pi^1$ can be relaxed. Namely, instead of taking functions in $C_c(G)$, the equations 7.1.2.1 and 7.1.2.2 remain true for functions in $L^1(G)$, since $\pi$ is bounded by 1.
Denote by $\pi^1(\psi)^*$ the adjoint operator of $\pi^1(\psi)$, $\psi \in C_c(G)$, and define $\psi^*(x) = \overline{\psi(x^{-1})}$. We have $\pi^*(x) = \pi(x)^{-1} = \pi(x^{-1})$, $x \in G$ and

$$\begin{aligned} < \pi^1(\psi)v, w > &= < \int_G \psi(x)\pi(x)vdx, w > = \int_G \psi(x) < \pi(x)v, w > dx \\ &= \int_G \psi(x) < v, \pi(x)^{-1}w > dx = < v, \int_G \overline{\psi(x)}\pi(x^{-1})wdx > \\ &= < v, \int_G \overline{\psi(x^{-1})}\pi(x)wdx > \end{aligned}$$

$v, w \in H$ and where $< , >$ denotes the inner product in $H$. Here we used that the Haar measure on $G$ is invariant under taking inverses. We see

$$\pi^1(\psi)^* = \pi^1(\psi^*). \tag{7.1.2.3}$$

Moreover,

$$\|\pi^1(\psi)\| \leq \sup_{g\in G} \|\pi(g)\|\|\psi\|_1 \leq \|\psi\|_1.$$

Therefore, the map $\psi \to \pi^1(\psi)$ is continuous and $\pi^1$ is a continuous linear Banach-Algebra homomorphism from $L^1(X)$ to $End(H)$.

### 7.1.2 Dirac Sequences on G

The *Dirac sequence* on $G$ denotes a sequence $(\varphi_n)$ in $C_c(G)$ satisfying the following three properties.

(1) $\varphi_n \geq 0$ for all $n$.

(2) $\int_G \varphi_n(x)dx = 1$ for all $n$.

(3) For any neighbourhood $V$of the identity $e \in G$, the support of $\varphi_n$ is contained in $V$ for all sufficiently large $n$.

The third property indicates that the area of $\varphi_n$ concentrates around the origin for large $n$. It will be convenient at one point to use a slightly weaker condition than the third above:

> $(3')$: Given a neighbourhood $V$ of $e$ in $G$, and its complement $Z$, and $\epsilon > 0$, we have
>
> $$\int_Z \varphi_n(x)dx < \epsilon$$
>
> for all $n$ sufficiently large.

Said differently, instead of assuming that the supports of the functions $\varphi_n$ shrink to the identity, we assume the corresponding $L^1$ condition.
We can similarly consider Dirac sequences $(\varphi'_n)$ at other points $g \in G$, in the sense that $\varphi'_n$ satisfies (1), (2) above and the area under $\varphi'_n$ concentrates around $g$.

**Lemma 7.1.3.** *Let $(\varphi_n)$ be a Dirac sequence. Then for every $v$ in the Banach space $H$, the sequence $(\pi^1(\varphi_n)v)$ converges to $v$.*

*Proof.* Denote by $S_n$ the support of $\varphi_n$. We have that

$$\begin{aligned}\int_G \varphi_n(x)\pi(x)vdx - v &= \int_G (\varphi_n(x)\pi(x) - \varphi_n(x))vdx \\ &= \int_{S_n} \varphi_n(x)(\pi(x)v - v)dx \\ &\leq \sup_{x\in S_n} \|\pi(x)v - v\|\end{aligned}$$

where in the first equation, we used $\int_G \varphi_n(x)dx = 1$. As the representation $\pi$ is strongly continuous and locally bounded, and because the support of $\varphi_n$ gets arbitrarily close to the identity for large $n$, the above integral converges to 0 as $n \to \infty$. □

The crucial property of Dirac sequences on $G$ is that they provide information about the values of the mappings $\pi(g), g \in G$, when the values for $\pi^1(\varphi_n)$ are known. To see this, observe that for a given $g \in G$ and a Dirac sequence $(\varphi_n)$, the sequence $(L_{g^{-1}}\varphi_n)$ is a Dirac sequence at $g$, in the sense that the support of $L_{g^{-1}}\varphi_n$ is concentrated around $g$ (as $Supp(L_{g^{-1}}\varphi_n) = gSupp(\varphi_n)$). By equation 7.1.2.2 and Lemma 7.1.3, we get the convergence

$$\pi^1(L_{g^{-1}}\varphi_n)v = \pi(g)\pi^1(\varphi_n)v \to \pi(g)v$$

for $n \to \infty$. The vector $\pi(g)v$ can therefore be computed using a suitable sequence of functions in $C_c(G)$ applied to the operator $\pi^1$.

**Definition 32.** A subspace $W$ of $H$ is called *$G$-invariant* if $\pi(g)W \subseteq W$ for all $g \in G$. Similarly, $W$ is invariant for $C_c(G)$ if $\pi^1(f)W \subseteq W$ for every $f \in C_c(G)$. In this case, $W$ is called *$C_c(G)$-invariant*.

We will always assume that a subspace $W$ of $H$ is closed.

**Lemma 7.1.4.** *A subspace $W$ of $H$ is $G$-invariant if and only if it is $C_c(G)$-invariant.*

*Proof.* Assume that $W$ is $G$-invariant. Let $w \in W$ and $f \in C_c(G)$. Then

$$\pi^1(f)w = \int_G f(x)\pi(x)w dx = \int_G f(x)w'(x)dx \in \overline{W} = W$$

where $w'(x) = \pi(x)w \in W$. Conversely, if $W$ is $C_c(G)$-invariant, then by Lemma 7.1.3, we have that $\pi(g)w = \lim_{n\to\infty} \pi^1(L_{g^{-1}}\varphi_n)w \in \overline{W} = W$ for all $g \in G$, $w \in W$, since $\pi$ is locally bounded. □

**Lemma 7.1.5.** *Let $\mathcal{A}$ be a dense subspace of $L^1(G)$ and assume that the representation $\pi$ is bounded. Then a subspace $W$ of $H$ is $G$-invariant if and only if it is $\mathcal{A}$-invariant.*

*Proof.* If $W$ is $G$-invariant, then by Lemma 7.1.4 it is also $C_c(G)$-invariant. Pick any function $f \in L^1(G)$ and let $(g_n) \subset C_c(G)$ be a sequence $L^1$-converging to $f$. Since we assume that $\pi$ is bounded, it follows that $\pi^1(g_n)$ converges in the strong topology to $\pi^1(f)$. Then for all $w \in W$

$$\pi^1(f)w = \lim_{n\to\infty} \pi^1(g_n)w \in W$$

as $W$ is closed. This shows that $W$ is $L^1(G)$ invariant and so in particular, it is invariant for $\mathcal{A}$.
Conversely, if $W$ is $\mathcal{A}$ invariant, then given any $f \in L^1(G)$ and a sequence $(h_n) \subset \mathcal{A}$ converging in $L^1$ to $f$, the same argument as above shows that $W$ is invariant under $f$. Since this holds for any $f \in L^1(G)$, we have that $W$ is $L^1(G)$-invariant and so it is $C_c(G)$-invariant. The desired result follows from Lemma 7.1.4.

□

**Lemma 7.1.6.** *Let $f \in L^1(G)$ and let $K \subseteq G$ be a compact set on which $f$ is continuous. Let $(\varphi_n)$ be a Dirac sequence. Then $\varphi_n * f$ converges to $f$ uniformly on $K$.*

*Proof.* Let $x \in K$. Observe that $f(x) = \int_G \varphi_n(y) f(x) dy$ by the second property of a Dirac sequence. Moreover,

$$\varphi_n * f(x) = \int_G \varphi_n(xy^{-1}) f(y) dy = \int_G \varphi_n(y) f(y^{-1}x) dy$$

because the Haar measure $dy$ is unimodular. Therefore,

$$\varphi_n * f(x) - f(x) = \int_G (f(y^{-1}x) - f(x)) \varphi_n(y) dy.$$

By the uniform continuity of $f$ on $K$, given any $\epsilon > 0$ there exists a neighbourhood $U$ of the identity $e$ in $G$ such that $y \in U$ implies for all $x \in K$ that

$$|f(y^{-1}x) - f(x)| < \epsilon.$$

Since the support of $\varphi_n$ is contained in $U$ for $n$ large enough, we get

$$\sup_{x \in K} |\varphi_n * f(x) - f(x)| < \epsilon$$

which proves the Lemma.

□

**Corollary 8.** If $f \in C_c(G)$, then $\varphi_n * f$ converges uniformly to $f$ on the compact set $Supp(\varphi_n) Supp(f)$ and is 0 outside of this set. In particular, $(\varphi_n * f)$ converges in $L^1$ to $f$.

*Proof.* Notice that $\varphi_n(xy^{-1}) f(y)$ in the integral of the convolution $\varphi_n * f$ is non-zero only if $xy^{-1} \in Supp(\varphi_n)$ and $y \in Supp(f)$. Hence, the support of $\varphi_n * f$ is contained in the compact set $K = Supp(\varphi_n) Supp(f)$ and we can thus restrict $\varphi_n * f$ to $K$. The first statement of the Corollary then follows from Lemma 7.1.6 and the $L^1$-convergence is obtained by the uniform convergence of $\varphi_n * f$ to $f$ on $K$. □

### 7.1.3 Equivalent Representations and Schur's Lemma

**Definition 33.** Let $H_1$, $H_2$ be Banach spaces and consider two representations

$$\pi_1 : G \to Gl(H_1), \quad \pi_2 : G \to Gl(H_2)$$

of $G$ on $H_1$ and $H_2$ respectively. A linear map $\alpha : H_1 \to H_2$ is called an *intertwining map* if

$$\alpha \circ \pi_1(g) = \pi_2(g) \circ \alpha$$

for all $g \in G$.

**Definition 34.** We call $\pi_1$ and $\pi_2$ as above *equivalent* if there exists a bijective intertwining map $\alpha : H_1 \to H_2$.
If $H_1$ and $H_2$ are Hilbert spaces, $\pi_1$, $\pi_2$ and $\alpha$ are unitary, then $\pi_1$ and $\pi_2$ are called *unitary equivalent* .

**Definition 35.** A representation $\pi : G \to Gl(H)$ on a Banach space $H$ is called *irreducible* if for all $G$-invariant subspaces $W \subseteq H$ we have either $W = \{0\}$ or $W = H$.

The following Lemma is known under the name of "Schur's Lemma". First stated by the German mathematician I. Schur in 1905, this Lemma is a fundamental result in the representation theory of groups, but it can also be applied more generally to the study of representations of algebras or even sets of linear operators on a Hilbert space.
Although we will not make use of this specific version of Lemma 7.1.7 concerning group representations, we have chosen to state and prove it here, to help build a sense of intuition for the statement. The proof of the more general Theorem, which we will use subsequently and which is stated afterwards in Theorem 10, relies specifically on spectral theory of hermitian operators on the Hilbert space $H$. This topic is not the primary focus of this thesis and will not be introduced here. For a detailed proof of this Theorem, we refer the interested reader to the appendix on page 367 of [Lan85].
The proof of the Lemma below is from the unpublished lecture notes of Prof. A. Gorodnik for the course "Analysis on Matrix Groups", held in the autumn semester 2023 at the University of Zurich.

**Lemma 7.1.7** (Schur's Lemma)**.** *Consider two irreducible representations $\pi_1$ and $\pi_2$ of $G$ on the Banach spaces $H_1$ and $H_2$ respectively.*

1. *Suppose $\alpha : H_1 \to H_2$ is an intertwining map for $\pi_1$ and $\pi_2$. Then $\alpha = 0$ or $\alpha$ is a bijection.*
2. *Suuppose $\alpha' : H_1 \to H_1$ is an intertwining map for $\pi_1$. Then $\alpha = \lambda I_{H_1}, \lambda \in \mathbb{C}$ where $I_{H_1}$ denotes the identity element of $Gl(H_1)$.*

*Proof.* (1) Let $U = \{v \in H_1 : \ \alpha(v) = 0\}$. We have

$$\alpha(\pi_1(g)v) = \pi_2(g)(\alpha(v)) = 0$$

$v \in U$, $g \in G$. Hence, $\pi_1(g)v \in U$ and $U$ is $\pi_1$-invariant. Since $\pi_1$ is irreducible, this shows $U = 0$ or $U = H_1$. Thus, $\alpha$ is injective or $\alpha = 0$.
Let $W = \alpha(H_1)$. Then

$$\pi_2(g)W = \alpha(\pi_1(g)H_1) = W$$

for all $g \in G$. Since $\pi_2$ is irreducible, this shows $W = 0$ or $W = H_2$. Thus, $\alpha = 0$ or $\alpha$ is surjective.
(2) Since $\alpha'$ is linear, there exists $\lambda \in \mathbb{C}$ with

$$H' = \{v \in H_1 : \ \alpha'(v) = \lambda v\} \neq \{0\}.$$

We have for $v \in H'$ and all $g \in G$

$$\alpha'(\pi(g)v) = \pi_1(g)\alpha'(v) = \pi_1(g)(\lambda v) = \lambda\pi_1(g)(v).$$

Hence, $H'$ is invariant for $\pi_1$. Since $\pi_1$ is irreducible, $H' = H_1$. □

The statement of Schur's Lemma can be generalized to a set $S$ of operators on a Hilbert space $H$ as follows.

**Theorem 10** (Schur's Lemma)**.** Let $S$ be a set of operators on the Hilbert space $H$, leaving no closed subspace invariant except $\{0\}$ and $H$ itself. Let $A$ be a hermitian operator such that $A \circ B = B \circ A$ for all $B \in S$. Then $A = cI_H$ for some real number $c$, where $I_H$ is the identity operator of $H$.

**Corollary 9.** Let $S$ be a set of operators on the Hilbert space $H$, leaving no closed subspace invariant except $\{0\}$ and $H$ itself. Let $A$ be an operator such that $A \circ A^* = A^* \circ A$, $A \circ T = T \circ A$, and $A^* \circ T = T \circ A^*$ for all $T \in S$. Then $A = cI_H$ for some complex number $c$, where $I_H$ is the identity operator of $H$.

*Proof.* Write

$$A = \frac{A + A^*}{2} + i\frac{A - A^*}{2i}.$$

This gives a decomposition of $A$ into a sum $B + iC$ of commuting hermitian operators $B = (A + A^*)/2$ and $C = (A - A^*)/(2i)$. We can apply Theorem 10 to each one of $B$ and $C$ to prove the Corollary. □

### 7.1.4 Direct Sums of finite-dimensional Representations

**Definition 36.** A finite-dimensional representation $(\pi, H)$ of $G$ is called *completely reducible* if $H$ is the orthogonal sum of irreducible and $G$-invariant subspaces. Said differently,

$$H = \bigoplus_{i=1}^{l} H_i$$

where $H_i$, $1 \leq i \leq l$, is irreducible and invariant for $G$.

**Lemma 7.1.8.** *Let $H$ be a finite-dimensional Hilbert space and let $\pi : G \to Gl(H)$ be a unitary representation of $H$. Then $\pi$ is completely reducible.*

*Proof.* We assume that $\pi$ is not irreducible, for otherwise the decomposition in Definition 36 exists in the obvious sense. So let $W \subseteq H$ be a $G$-invariant subspace and denote by $W^\perp$ its orthogonal complement. As $H$ is assumed to be a Hilbert space, we have

$$H = W \oplus W^\perp.$$

Let $v \in W^\perp$. Then for all $w \in W$ and $g \in G$,

$$< w, \pi(g)v > = < \pi(g^{-1})w, \pi(g^{-1})\pi(g)v > = < \pi(g^{-1})w, v > = \ 0$$

since $\pi(g^{-1})w \in W$ as $W$ is $G$-invariant. Hence, $\pi(g)v \in W^{\perp}$ and so $W^{\perp}$ is $G$-invariant.
If $W$ (or $W^{\perp}$) are not irreducible, the same argument as above decomposes $W$ (or $W^{\perp}$) into a direct sum of invariant subspaces, of lower dimension. As they have finite dimension, this procedure ends in finitely many steps, giving the desired decomposition into irreducible and invariant subspaces.

□

**Definition 37.** We define the *direct sum* of a countable family $\{(\pi_i, H_i), i \in I\}$ of representations of $G$, denoted by $\bigoplus_{i\in I} \pi_i$, as the representation $\pi$ on $H = \bigoplus_{i\in I} H_i$, given by

$$\pi(g)\left(\sum_{i\in I} v_i\right) = \sum_{i\in I} \pi_i(g)v_i, \quad v_i \in H_i, i \in I.$$

Lemma 7.1.8 states that in the context of finite-dimensional Hilbert spaces, any unitary representation can be decomposed into a direct sum of irreducible representations by restricting it to the irreducible and invariant subspaces. It turns out that this result can be generalized using the concept of *direct integrals*, which is due to von Neumann in [Neu49].
A preliminary step towards this generalization involves the analysis of whether it is possible to define a "direct sum" for a potentially uncountable family of representations. We won't provide a detailed introduction to this theory, as it lies beyond the primary focus of this thesis. The aim is to make sense of the work of Ratner in [Rat87] which we consider in the next chapter. For a thorough treatment, we direct the reader to Chapter 2 of [Mac76].

### 7.1.5 Direct Integral

We follow the section *Harmonic Analysis* in [Rat87]. Consider a standard measure space $(Y, \nu)$ (i.e. a Borel space which is measurably isomorphic to $[0, 1] \subset \mathbb{R}$ endowed with the Lebesgue measure and the Borel $\sigma$-algebra) and to each $y \in Y$ associate a complex separable Hilbert space $H(y)$. Let

$$\Psi(y) = \{\psi_n(y),\ n \in \mathbb{Z}\ \}$$

denote an orthonormal basis in $H(y)$ and consider a function

$$f : Y \to \prod_{y\in Y} H(y)$$

satisfying $f(y) \in H(y)$, $y \in Y$.

**Definition 38.** A function $f$ on $Y$ with $f(y) \in H(y)$ is called $\Psi$-*measurable* if

$$y \mapsto < f(y), \psi_n(y) >$$

is a measurable function of $y$ for all $n \in \mathbb{Z}$.

Let $H$ be the the set of all $\Psi$-measurable functions $f$ satisfying

$$\int_Y \|f(y)\|^2_{H(y)} d\nu(y) < \infty.$$

We say that two functions in $H$ are equivalent if they differ on a set of $\nu$-measure zero. Let

$$< f, h > = \int_Y < f(y), h(y) >_{H(y)} \, d\nu(y), \quad f, h \in H.$$

Then the space $H$ becomes a separable Hilbert space under the above inner product. With this definition, we call $H$ the *direct integral* of the spaces $H(y)$ and we write

$$H = \int_Y H(y) d\nu(y).$$

This construction inspires the following particular question. Suppose that to each $y \in Y$ there exists a non-trivial irreducible unitary representation $(\pi_y, H(y))$ of $G = SL_2(\mathbb{R})$. Does the above construction then induce a unitary representation on $H$?
Indeed, it does as follows. Assume that the orthonormal basis $\Psi(y)$ of $H(y)$ defined above additionally satisfies that the maps given by

$$(g, y) \mapsto < \pi_y(g)\psi_n(y), \psi_m(y) >_{H(y)}, \quad n, m \in \mathbb{Z}$$

are measurable for $g \in G$, $y \in Y$. We define a representation $(\pi, H)$ of $G$ by

$$(\pi(g)f)(y) = (\pi_y(g))(f(y))$$

$g \in G$, $y \in Y$, $f \in H$ and we may write $\pi$ as the *direct integral*

$$\pi = \int_Y \pi_y d\nu(y).$$

It follows immediately from the definition of the inner product in $H$ that the representation $(\pi, H)$ is unitary. The crucial property which leads to a generalization of Lemma 7.1.8 is the following fact, which is stated as *Fact 1.1* in [Rat87].
*Fact 1* : Let $T$ be a unitary representation of $G$ on the Hilbert space $H(T)$, having no non-zero invariant vectors in $H(T)$. Then $T$ is unitary equivalent to a direct integral $\pi$ as above.

## 7.2 A Decomposition for $SL_2(\mathbb{R})$

We now come back to the group $G = SL_2(\mathbb{R})$. We denote by $(\pi, H)$ a representation of $G$ on the Banach space $H$. Recall that the compact subgroup $K = SO(2)$ of $G$ consists of matrices

$$r(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad \theta \in \mathbb{R}.$$

### 7.2.1 Preliminaries

For $n, m \in \mathbb{Z}$, let $S_{n,m}$ denote the set of functions $f \in C_c(G)$ satisfying

$$f(r(\theta)yr(\theta')) = \mathrm{e}^{-in\theta} f(y) \mathrm{e}^{-im\theta'}$$

for all $y \in G$ and all $\theta, \theta' \in \mathbb{R}$. We denote by $\mathcal{A} = \sum S_{n,m}$ the set of algebraic sums of functions in $S_{n,m}, n, m \in \mathbb{Z}$.

**Lemma 7.2.1.** *The set of functions $\mathcal{A}$ is $L^1$-dense in $C_c(G)$. In fact, given $\epsilon > 0$ and $f \in C_c(G)$, there exists a function $g$ of the form $\sum S_{n,m}$ such that the support of $g$ is contained in $KSupp(f)K$ and such that $\|f - g\|_\infty < \epsilon$.*

*Proof.* For each $y \in G$ and $f \in C_c(G)$ we define

$$f^y(\theta, \theta') = f(r(\theta)yr(\theta')).$$

Then $f^y$ is a continuous function depending on two variables, integrable over $\mathbb{R}$. We may consider the Fourier expansion of $f^y$, given by

$$f^y(\theta, \theta') = \sum_{n,m \in \mathbb{Z}} f_{n,m}(y) \mathrm{e}^{in\theta} \mathrm{e}^{im\theta'}$$

and the $(n, m)$ Fourier coefficient $f_{n,m}$ of $f^y$

$$f_{n,m}(y) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(r(\theta)yr(\theta')) \mathrm{e}^{in\theta} \mathrm{e}^{im\theta'} d\theta d\theta'.$$

If $f_{n,m}(y) \neq 0$ then $f(r(\theta)yr(\theta')) \neq 0$. This shows that the support of $f_{n,m}$ is contained in $KSupp(f)K$.
It is a standard fact from Fourier Analysis that the sums

$$C_M(\theta) = \frac{1}{M} \sum_{N=0}^{M-1} \sum_{|n| \leq N} \mathrm{e}^{in\theta}$$

$M = 1, 2, ...$ form a Dirac sequence on $\mathbb{R}$.
Similarly, the product

$$C_M(\theta) C_M(\theta') = \frac{1}{M^2} \sum_{N=0}^{M-1} \sum_{K=0}^{M-1} \sum_{|n| \leq N} \sum_{|k| \leq K} \mathrm{e}^{in\theta + ik\theta'}$$

forms a Dirac sequence in two variables, say $(C_M(\theta, \theta'))$, $M = 1, 2, ....$ We have

$$f^y * C_M(0, 0) = \frac{1}{M^2} \sum_{N=0}^{M-1} \sum_{K=0}^{M-1} \sum_{|n| \leq N} \sum_{|k| \leq K} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(r(\theta)yr(\theta')) \mathrm{e}^{-in\theta} \mathrm{e}^{-ik\theta'}.$$

This shows that $f^y * C_M(0,0)$ is the sum of terms $c_{n,m} f_{n,m}(y)$ with appropriate constant coefficients $c_{n,m}$ arising from the sum in $C_M(\theta, \theta')$. We claim that $f^y * C_M(0,0)$ converges to $f^y(0,0) = f(y)$ uniformly on $y$. Since

$$1 = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} C_M(-\theta, -\theta') d\theta d\theta'$$

we have

$$f^y * C_M(0,0) - f^y(0,0)$$
$$= \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} [f^y(\theta, \theta') C_M(-\theta, -\theta') - f^y(0,0) C_M(-\theta, -\theta')] d\theta d\theta'.$$

We proceed by estimtaing the difference in the above equation. By property $(3')$ of the Dirac sequence $(C_M(\theta, \theta'))$, there exists a neighbourhood $U$ of $(0,0)$ such that

$$\int \int_{U^c} C_M(\theta, \theta') d\theta d\theta' < \epsilon \quad \text{and} \quad \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} C_M(\theta, \theta') d\theta d\theta' = 1$$

where $U^c$ denotes the complement of $U$. Therefore, the integral of the difference above is estimated by integrals over $U$ and $U^c$, i.e. by

$$\sup_{(\theta,\theta') \in U} |f^y(\theta, \theta') - f^y(0,0)| + 2\|f\|_\infty \int_{U^c} C_M(\theta, \theta') d\theta d\theta'.$$

Since $f$ has compact support, we have

$$|f(r(\theta) y r(\theta')) - f(y)| < \epsilon$$

for $U$ sufficiently small, uniformly in $y$, $(\theta, \theta') \in U$. This is because then $r(\theta)$ and $r(\theta')$ are close to the identity, by the choice of $U$ being a neighbourhood of $(0,0)$ in $[-\pi, \pi] \times [-\pi, \pi]$. This proves what we wanted.

□

Recall that for a function $f$ on $G$, we defined $f^*(x) = \overline{f(x^{-1})}$.

**Lemma 7.2.2.** *It holds that*

1. $S_{n,m} * S_{l,q} = 0$, *if* $m \neq l$.
2. $S_{n,m}^* = S_{m,n}$.
3. $S_{n,m} * S_{m,q} \subseteq S_{n,q}$.

*Proof.* Let $f \in S_{n,m}$, $g \in S_{l,q}$. Then

$$f * g(x) = \int_G f(xy^{-1}) g(y) dy = \int_G f(xy^{-1} r(-\theta))\ g(r(\theta) y) dy$$

$$= \int_G f(xy^{-1}) \mathrm{e}^{im\theta} \mathrm{e}^{-il\theta} g(y) dy$$

where we substituted $y \to r(\theta)y$ in the second equation. Hence, $f * g(x) = \mathrm{e}^{im\theta}\mathrm{e}^{-il\theta} f * g(x)$ for all $x \in G$ which is only possible if $m = l$ or if $f * g = 0$, proving the first assertion of the Lemma.

For $f \in S_{n,m}$,

$$f(r(\theta)xr(\theta'))^* = \overline{f(r(-\theta')x^{-1}r(-\theta))} = \overline{\mathrm{e}^{in\theta'} f(x^{-1})\mathrm{e}^{im\theta}} = \mathrm{e}^{-im\theta} f^*(x)\mathrm{e}^{-in\theta'}$$

and so $f^* \in S_{m,n}$.

The third assertion of the Lemma is proven similarly.

□

We deduce from Lemma 7.2.2 that $S_{n,n}$ is an algebra under convolution (i.e. it is closed under taking convolutions). In fact, it exhibits the following crucial algebraic property.

**Lemma 7.2.3.** *The algebra $S_{n,n}$ is commutative.*

We need two preliminary results to prove this Lemma.

**Lemma 7.2.4** (Polar Decomposition of $G$)**.** *Let $S \subset G$ denote the set of symmetric elements in $G$. Then $G$ decomposes as*

$$G \cong SK.$$

*Proof.* Pick $x \in G$ and set $y = xx^t$, where $x^t$ denotes the transposed of $x$. Then $y$ is symmetric and positive definite. Hence, there is a basis of $\mathbb{R}^2$ consisting of eigenvectors of $y$ such that

$$yv_i, = \lambda_i v_i, \quad \lambda_i > 0, \ i = 1, 2.$$

Set $s^2 = y$. The eigenvalues of the matrix $s$ for $v_1$,$v_2$ are $\sqrt{\lambda_i}$, $i = 1, 2$, respectively. Define $k = s^{-1}x$. Then $x = sk$. We are left to check that $k$ is an element of $K$. We have

$$k^t k = x^t (s^{-1})^t s^{-1} x = x^t s^{-1} s^{-1} x = x^t y^{-1} x = x^t (x^t)^{-1} x^{-1} x = I.$$

This proves our claim.

□

**Lemma 7.2.5.** *Let $S \subset G$ as above and let $\sigma$ denote and automorphism on $G$, defined by*

$$\sigma(x) = x^\sigma = \gamma x \gamma, \quad x \in G$$

*where*

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

1. *We have $k^\sigma = k^{-1}$, $k \in K$.*

*2. For all $s \in S$ there is a $k_1 \in K$ such that*

$$s^\sigma = k_1 s k_1^{-1}.$$

*Proof.* These assertions are proven by straightforward computations on the relevant matrices. In (2), one may recall from the Spectral Theorem that real symmetric matrices are diagonalizable by orthogonal matrices. Moreover, we have $\gamma = \gamma^{-1}$ and that $\gamma$ is symmetric. Hence, $s^\sigma$ is the conjugation of $s$ by $\gamma$ and also symmetric. Let $O(2)$ denote the set of orthogonal $2 \times 2$-matrices. Write $s = qdq^{-1}$, $q \in O(2)$, where $d$ is a diagonal element of $G$.
Suppose first that $\det(q) = -1$. Then we have $\gamma q \in SO(2)$ and

$$s^\sigma = \gamma q d q^{-1} \gamma = \gamma q d (\gamma q)^{-1} = k_1 s k_1^{-1}$$

where $k_1 = \gamma q \in SO(2)$.
If $q$ above is in $SO(2)$, then we use $q^\sigma = q^{-1}$ from the first part of the Lemma and the fact that $\gamma$ commutes with diagonal elements to obtain

$$s^\sigma = \gamma q d (\gamma q)^{-1} = \gamma (\gamma q \gamma)^{-1} d (\gamma q)^{-1} = \gamma\gamma q^{-1} \gamma d \gamma q \gamma\gamma = q^{-1} d q.$$

□

*Proof of Lemma 7.2.3.* Let $f, g \in S_{n,n}$. Define $f^*(x) = f(x^t)$. We first claim that

$$(g * f)^* = f^* * g^*. \tag{7.2.5.1}$$

We have

$$(g * f)^*(x) = (g * f)(x^t) = \int_G g(x^t y^{-1}) f(y) dy$$

and

$$\begin{aligned}(f^* * g^*)(x) &= \int_G f^*(xy^{-1}) g^*(y) dy = \int_G f((y^{-1})^t x^t) g(y^t) dy \\ &= \int_G f(y^{-1} x^t) g(y) dy = \int_G f(y^{-1}) g(x^t y) dy \\ &= \int_G f(y) g(x^t y^{-1}) dy\end{aligned}$$

where in the last three equations we used successively the substitutions $y \mapsto y^t$, $y \mapsto x^t y$ and $y \mapsto y^{-1}$, proving the claim. Next, let $\sigma$ denote the automorphism of Lemma 7.2.5. Define $f'(x) = f(x^\sigma)$. Then a similar computation as above shows that

$$(g * f)' = g' * f'. \tag{7.2.5.2}$$

Moreover, we have $f^* = f'$. To see this, we may pick $x \in G$ and write $x = sk$, for some $s \in S$, $k = r(\theta) \in K$, $\theta \in \mathbb{R}$, using Lemma 7.2.4. Then

$$\begin{aligned}f(x^t) &= f(k^t s^t) = f(k^{-1} s) = \mathrm{e}^{in\theta} f(s) \\ f(x^\sigma) &= f(s^\sigma k^{-1}) = \mathrm{e}^{in\theta} f(k_1 s k_1^{-1}) = \mathrm{e}^{in\theta} f(s)\end{aligned}$$

where $k_1$ is such that $s^\sigma = k_1 s k_1^{-1}$. This shows $f^* = f'$ and thus

$$f^* * g^* = g^* * f^* \tag{7.2.5.3}$$

via equations 7.2.5.1 and 7.2.5.2. Notice that

$$f^*(r(\theta)xr(\theta')) = f(r(-\theta')x^t r(-\theta)) = \mathrm{e}^{in\theta'} f(x^t)\mathrm{e}^{in\theta} = \mathrm{e}^{in\theta'} f^*(x)\mathrm{e}^{in\theta}$$

$\theta, \theta' \in \mathbb{R}$, $x \in G$. This shows $f^* \in S_{-n,-n}$. Hence, the operator $*$ maps the set $S_{n,n}$ bijectively to $S_{-n,-n}$. This, together with equation 7.2.5.3 implies

$$f * g = g * f$$

as desired. □

### 7.2.2 Eigenvectors for the Rotation Group $K$

We denote by $H_n \subseteq H$, $n \in \mathbb{Z}$ the set of vectors satisfying

$$\pi(r(\theta))v = \mathrm{e}^{in\theta} v$$

for all $\theta \in \mathbb{R}$. Then $H_n$ is a subspace of $H$. Moreover, $H_n$ is clearly closed, for if $(v_k)$ is a sequence in $H_n$ converging to some $v \in H$, then

$$\pi(r(\theta))v = \lim_{k\to\infty} \pi(r(\theta))v_k = \lim_{k\to\infty} \mathrm{e}^{in\theta} v_k = \mathrm{e}^{in\theta} v$$

and $v \in H_n$.

**Lemma 7.2.6.** *If $H$ is a Hilbert space and if the representation $\pi : G \to Gl(H)$ is unitary on $K$, then $H_n$ is orthogonal to $H_m$ for all $n \neq m \in \mathbb{Z}$.*

*Proof.* Let $v \in H_n, w \in H_m$. Since $\pi$ is unitary on $K$, it holds that $\pi(r(\theta))^* = \pi(r(\theta))^{-1} = \pi(r(\theta)^{-1}) = \pi(r(-\theta))$. Therefore,

$$\begin{aligned} < \pi(r(\theta))v, w > &= \mathrm{e}^{in\theta} < v, w > \\ &=< v, \pi(r(-\theta))w > \\ &= \mathrm{e}^{im\theta} < v, w > \end{aligned}$$

for all $\theta \in \mathbb{R}$. This shows $< v, w > = 0$ if $n \neq m$.

□

**Lemma 7.2.7.** *We have that*

1. $\pi^1(S_{n,m})H \subseteq H_n$.
2. $\pi^1(S_{n,m})H_q = \{0\}$ *if* $m \neq q$.

*Proof.* Suppose $m \neq q$. As the Haar measure on $G = SL_2(\mathbb{R})$ is unimodular, we can use the invariance of translations $y \to yr(\theta)$ in the integral $\int_G f(y)\pi(y)vdy$ as follows. For $f \in S_{n,m}$ and $v \in H_q$,

$$\int_G f(y)\pi(y)vdy = \int_G f(yr(\theta))\pi(yr(\theta))vdy = \int_G f(yr(\theta))\pi(y)\pi(r(\theta))vdy$$

$$= \mathrm{e}^{-im\theta}\mathrm{e}^{iq\theta}\int_G f(y)\pi(y)vdy$$

implying that $\int_G f(y)\pi(y)vdy = 0$ unless $m = q$. This proves the second statement of the Lemma.
Now let $v \in H$ and $f \in S_{n,m}$. For $k = r(\theta)$, we have that,

$$\pi(k)\pi^1(f)v = \pi(k)\int_G f(y)\pi(y)vdy = \int_G f(y)\pi(ky)vdy$$

$$= \int_G f(k^{-1}y)\pi(y)vdy = \mathrm{e}^{in\theta}\pi^1(f)v$$

and so $\pi^1(f)v \in H_n$. □

**Theorem 11.** Let $\pi$ be an irreducible representation of $G$ on a Banach space $H$. Then the sum $\sum_n H_n$ is dense in $H$. If $H$ is a Hilbert space and $\pi$ is unitary on $K$, this sum is an orthogonal decomposition of $H$.

*Proof.* Let $E$ be the (closed) subspace generated by the $H_n$. By Lemma 7.2.7, $E$ is invariant for the algebra $\mathcal{A}$. As the latter is dense in $C_c(G)$, it is dense in $L^1(G)$ (since $C_c(G)$ is dense in $L^1(G)$), and it follows from Lemma 7.1.5 that $E$ is $C_c(G)$-invariant, or equivalently, by Lemma 7.1.4, it is $G$-invariant. Since $\pi$ is irreducible, it follows that $E = H$. If $\pi$ is unitary on $K$, we know from Lemma 7.2.6 that the $H_n$ are pairwise orthogonal. □

**Lemma 7.2.8.** *Let $H$ be a Hilbert space and suppose the representation $\pi$ is irreducible and unitary on $K$. Then the space $H_q$ is irreducible for $S_{q,q}$, and if $H_q \neq \{0\}$, then $\pi^1(S_{q,q})H_q \neq \{0\}$.*

*Proof.* Recall that $H_q$ is irreducible for $S_{q,q}$ if there is no nontrivial proper subspace $W$ of $H_q$ with $\pi^1(f)W \subseteq W$ for all $f \in S_{q,q}$.
Suppose for a contradiction that such $W$ does exist. Let $w \in W$ and let $f$ be any finite sum of functions $f_{n,m} \in S_{n,m}$, say $f = f_{n_1,m_1} + ... + f_{n_l,m_l}$. By Lemma 7.2.7, we have

$$(\pi^1(f)w)_q = (\pi^1(f_{n_1,m_1})w)_q + ... + (\pi^1(f_{n_l,m_l})w)_q = \pi^1(f_{q,q})w \in W$$

if there is an $i$ with $f_{n_i,m_i} = f_{q,q}$, $1 \leq i \leq l$, in the definition of $f$ (otherwise, the above equals $0 \in W$). Thus for every $f \in \mathcal{A}$, the $q$-component of the vector $\pi^1(f)w$ lies in $W$. The algebra $\mathcal{A}$ is $L^1$-dense in $C_c(G)$ by Lemma 7.2.1. For

any given $\phi \in C_c(G)$, there thus exists a sequence $(f_n) \subset \mathcal{A}$ converging in $L^1$ to $\phi$. We have

$$(\pi^1(\phi)w)_q = \lim_{n\to\infty} (\pi^1(f_n)w)_q \in \overline{W} = W.$$

Now, for each $q \in \mathbb{Z}$, write $W_q \subseteq H_q$ for the largest $S_{q,q}$-invariant subspace and set $H' = \bigoplus_{q\in\mathbb{Z}} W_q$. From the above, we know that $H'$ is $C_c(G)$-invariant. But by irreducibility of $\pi$, it follows from Theorem 11 that $H' = H$ and thus $W_q = H_q$ for all $q \in \mathbb{Z}$.

□

**Theorem 12.** Let $\pi$ be an irreducible and unitary representation of $G$ on a Hilbert space $H$. Let $H_n$ be the subspace of vectors $v$ such that

$$\pi(r(\theta))v = \mathrm{e}^{in\theta}v.$$

Then $\dim(H_n) \in \{0, 1\}$.

*Proof.* Recall that we defined $f^*(x) = \overline{f(x^{-1})}$, $f \in C_c(G)$. Since we assume that $\pi$ is unitary, we have by equation 7.1.2.3

$$\pi^1(f)^* = \pi^1(f^*).$$

Moreover, Lemma 7.2.2 (2) states $f^* \in S_{n,n}$ if and only if $f \in S_{n,n}$ . This shows that $\pi^1(S_{n,n})$ is closed under the $*$-operation.
By Lemma 7.2.3 we know that the algebra $S_{n,n}$ is commutative for the convolution operator. Hence, for all $f \in S_{n,n}$, the operator $\pi^1(f)$ on $H_n$ satisfies

$$\pi^1(f)^*\pi^1(f) = \pi^1(f^* * f) = \pi^1(f * f^*) = \pi^1(f)\pi^1(f)^*.$$

Similarly, $\pi^1(f)$ and $\pi^1(f)^*$ commute with all $\pi^1(h), h \in S_{n,n}$. Therefore, by Corollary 9 of the Schur Lemma, we have $\pi^1(f) = c_f I_{H_n}$ for some complex number $c_f$ and the identity map $I_{H_n}$ on $H_n$. This is true for all $f \in S_{n,n}$. Thus the (closed) span of any $v \in H_n$ is invariant for $S_{n,n}$. By irreducibility of the representation $\pi^1$ restricted to $S_{n,n}$ (see Lemma 7.2.8), it follows that the span of $v$ is either $\{0\}$ or $H_n$, which completes the proof.

□

## 7.3 Smooth Maps and Lie Derivatives

### 7.3.1 The smooth Structure of $G$

We begin by clarifying the notion of a *smooth* map

$$f : G \to H$$

defined on closed linear groups $G$, $H$.

Recall that the exponential map defined on the Lie Algebra $\mathfrak{g}$ of a closed linear group $G$ gives rise to a local coordinate system in an open neighbourhood of an

element $g \in G$, identifying this neighbourhood with an open set in an Euclidean space: We have that the map

$$\varphi_g(X) = g \exp(X)$$

where $X$ is in a neighbourhood of $0 \in \mathfrak{g}$, is a parametrization near $g \in G$.
We call a function $f : G \to H$ *smooth* (or $C^\infty$) if it is smooth in terms of the coordinates $\varphi_g$, $g \in G$, $\varphi_h$, $h \in H$. More precisely, $f$ is smooth at $x \in G$ if it is continuous at $x$ and if

$$\varphi_{f(x)}^{-1} \circ f \circ \varphi_x$$

is smooth at $\varphi_x^{-1}(x)$, when viewed as a map between euclidean spaces. The definition of $C^k$-maps $f$ on $G$, $k \in \mathbb{N}$ is essentially the same.

We now investigate the smooth structure of the group $G = SL_2(\mathbb{R})$.
Let $\pi : G \to Gl(H)$ denote a representation of $G$ in a Banach space $H$. As in the previous section, we write $\pi^1$ for the induced algebra homomorphism defined on $C_c(G)$.

**Definition 39.** A vector $v \in H$ is called a $C^k$-*vector*, $k \in \mathbb{N}$ if the mapping

$$x \mapsto \pi(x)v, \ x \in G$$

is a $C^k$ map from $G$ into the Banach space $H$. We let $H^k \subseteq H$ denote the subspace of $H$ consisting of $C^k$-vectors.

**Definition 40.** We define the subspace $H^\infty \subseteq H$ to consist of all those vectors $v \in H$ such that the map

$$x \mapsto \pi(x)v, \ x \in G$$

is $C^\infty$. We call $H^\infty$ the space of $C^\infty$-*vectors* or the space of *smooth vectors*.

We have

$$\pi(a)v \in H^\infty$$

for all $a \in G$, $v \in H^\infty$. Moreover, the set of smooth vectors in $H$ is also stable under the action of $\pi^1$ restricted to $C_c^\infty(G)$.
Observe that the space of $C^\infty$ functions in $S_{n,m}$, denoted by $S_{n,m}^\infty$, is dense in $S_{n,m}$. Indeed, as is explained on page 101 in [Lan85], any continuous function with compact support can be uniformly approximated by a $C^\infty$ function whose support is very close, and then the arguments given to prove Lemma 7.2.1, amounting to Fourier series convergence yield $C^\infty$ functions.
The denseness of $S_{q,q}^\infty$ in $S_{q,q}$ shows, that if $\pi^1(S_{q,q})H_q \neq 0$, then

$$\pi^1(S_{q,q}^\infty)H_q \neq 0.$$

**Theorem 13.** Let $\pi : G \to Gl(H)$ be an irreducible unitary representation of $G = SL_2(\mathbb{R})$ on a Hilbert space $H$. Let $n$ be an integer such that $H_n \neq \{0\}$. Then $H_n$ has dimension 1 and any element of $H_n$ is a $C^\infty$ vector.

*Proof.* We know from Theorem 12 that $H_n \neq \{0\}$ has dimension 1. Since $\pi^1(S_{n,n}^\infty)H_n \neq \{0\}$, if $\{v_n\}$ is a basis of $H_n$ over $\mathbb{C}$, then $\pi^1(S_{n,n}^\infty)v_n = \mathbb{C}v_n$ and so $v_n$ is a $C^\infty$ vector. □

### 7.3.2 The Lie Derivative on the Lie Algebra of $SL_2(\mathbb{R})$

We return our attention to the group $G = SL_2(\mathbb{R})$. Let

$$\pi : G \to \mathcal{U}(H(\pi))$$

denote a unitary representation of $G$ on a separable Hilbert space $H(\pi)$. Let us fix a basis $\{W, Q, V\}$ for the Lie Algebra $\mathfrak{g}$ of $G$ defined by

$$W = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\ Q = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\ V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

**Definition 41.** Let $X \in \mathfrak{g}$ and let $v$ be a $C^1$-vector in $H(\pi)$. The *Lie derivative* $L_X v$ is defined by

$$L_X v = \lim_{t \to 0} \frac{\pi(\exp(tX))v - v}{t}.$$

**Lemma 7.3.1.** *The Lie derivative $L_X v$ satisfies*

$$< L_X v, w > = - < v, L_X w >$$

*for all $C^1$-vectors $v, w \in H(\pi)$ and all $X \in \mathfrak{g}$.*

*Proof.* This follows immediately from the definition of the Lie derivative and the fact that $\pi$ is unitary. □

**Definition 42.** The *Casimir operator* $\Omega_\pi$ is defined on $C^2$-vectors in $H(\pi)$ by the formula

$$\Omega_\pi = (L_V^2 + L_Q^2 - L_W^2)/4.$$

**Lemma 7.3.2.** *The Casimir operator $\Omega_\pi$ satisfies*

$$< \Omega_\pi v, w > = < v, \Omega_\pi w >$$

*for all $C^2$-vectors $v, w \in H(\pi)$.*

*Proof.* We have

$$< L_X^2 v, w > = - < L_X v, L_X w > = < v, L_X^2 w >$$

for all $C^2$-vectors $v, w \in H(\pi)$ and all $X \in \mathfrak{g}$, by Lemma 7.3.1. The result then follows from linearity of the inner product. □

**Remark 17.** We will use without proof the fact that $\Omega_\pi$ commutes with all operators $\pi(g), g \in G$ on $H(\pi)$ (for more details see Theorem 5 of Chapter 10 in [Lan85] ). An important consequence of this is the following.

**Theorem 14.** If the unitary representation $\pi$ of $G$ on the separable Hilbert space $H(\pi)$ is irreducible, then the Casimir operator $\Omega_\pi$ is of the form

$$\Omega_\pi v = \lambda v$$

for some $\lambda = \lambda(\pi) \in \mathbb{R}$ and all $C^2$-vectors $v \in H(\pi)$.

Using Remark 17, one might try to prove this Theorem by applying Lemma 7.3.2 and Schur's Theorem 10 to the set of operators $S$ given by $S = \{\pi(g), g \in G\}$. However, a technical issue arises: we need to restrict the functional $\Omega$ to $C^2$-vectors in $H(\pi)$, which prevents the direct application of Schur's Theorem.

To address this, we use the concept of *Sobolev spaces*, which provide a framework for handling functions with certain degree of smoothness. We can construct a Hilbert space structure on the $C^2$-vectors in $H(\pi)$ using Sobolev spaces. The Sobolev norm for $C^2$-vectors $v$ in $H(\pi)$ combines the $L^2$-norm of $v$ and the norms of the first and second Lie-derivatives of $v$ in direction of the subgroup $K$. With this norm, the the space for $C^2$-vectors in $H(\pi)$ becomes a Hilbert space and Schur's Theorem 10 applies. For a thorough treatment of this matter, see Chapter 5 in [EW17].

# Chapter 8

# Quantitative Estimates of Mixing for the Geodesic and the Horocycle Flows

In this chapter, we discuss the rate of convergence for the geodesic and the unstable horocycle flow, defined on the measure space $(X, \mu)$, $X = \Gamma\backslash G$, $G = SL_2(\mathbb{R})$, where $\Gamma \leq G$ is a lattice in $G$ and where $\mu$ denotes the Haar measure on $X$, normalized such that $\mu(X) = 1$. We closely follow the work of Ratner in [Rat87].

We define the geodesic flow $\{p_t,\ t \in \mathbb{R}\}$ and the (unstable) horocycle flow $\{h_t,\ t \in \mathbb{R}\}$ on $X$ as the right multiplication by the respective matrices

$$a(t) := \begin{pmatrix} \mathrm{e}^t & 0 \\ 0 & \mathrm{e}^{-t} \end{pmatrix}, \quad u(t) := \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$$

$t \in \mathbb{R}$. More precisely, we have

$$p_t(\Gamma g) = \Gamma g \begin{pmatrix} \mathrm{e}^t & 0 \\ 0 & \mathrm{e}^{-t} \end{pmatrix}, \quad h_t(\Gamma g) = \Gamma g \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$$

$g \in G, t \in \mathbb{R}$.

**Remark 18.** We slightly changed the definition of these flows to remain consistent with the author's notation. They now describe a *right* group action given by right multiplication of the respective matrices. Moreover, we changed their parametrization given by $t$ (by a factor 2). The reason for the latter is primarily stylistic, as will become apparent.

We know that these flows are mixing on $X$. This means that if $f, \phi \in L^2(X, \mu)$ and $< f, 1 > = 0$, then

$$B_{f,\phi}(t) := < \phi, U_{a(t)}(f) > \to 0$$
$$C_{f,\phi}(t) := < \phi, U_{u(t)}(f) > \to 0$$

as $t \to \infty$ and where $U$ denotes the regular representation of $G$ on $L^2(X, \mu)$. The aim is to give a quantitative estimate for the asymptotic behaviour of the functions $B_{f,\phi}(t)$ and $C_{f,\phi}(t)$, $t \in \mathbb{R}$ above.

## 8.1 An orthonormal Basis consisting of Eigenvectors of K

In this section, we first recall the notion of a direct integral, which was introduced in the last chapter. The goal is to exchange the basis used in the definition of this integral by a basis which will turn out to be essential in various further computations.
We write $K$ for the rotation subgroup of $G$, consisting of matrices

$$r(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \ \theta \in \mathbb{R}.$$

Let $H(T)$ denote a Hilbert space and consider a unitary representation $(T, H(T))$ of $G$, having no non-zero invariant vectors in $H(T)$. Recall from Fact 1 in Section 7.1.5 that $T$ is unitary equivalent to the direct integral $\pi$ we constructed in that section. We may write

$$T = \int_Y \pi_y d\nu(y)$$

where the equality is meant in the sense of unitary equivalence, and where the $\pi_y$, $y \in Y$ denote irreducible representations of $G$ on the complex separable Hilbert spaces $H(y)$, $y \in Y$ with orthonormal basis given by

$$\Psi(y) = \{\psi_n(y), \ n \in \mathbb{Z} \}$$

such that the maps

$$(g, y) \mapsto < \pi_y(g)\psi_n(y), \psi_m(y) >_{H(y)}, \quad n, m \in \mathbb{Z}$$

are measurable for $g \in G$, $y \in Y$.

We call a vector $v \in H(T)$ an eigenvector for the action of $K$ on $H(T)$ if

$$T_{r(\theta)}v = \lambda_v(\theta)v$$

for all $\theta \in \mathbb{R}$ and where $\lambda_v(\theta) \in \mathbb{C}$.

**Remark 19.** Fact 1 in Section 7.1.5 and the definition of the direct integral $\pi$ show that Theorem 11 may be generalized to unitary representations $T$ of $G$ on a Hilbert space $H(T)$ leaving no non-zero vector invariant. That is,

(1) The maps $\lambda_v$ are given by $\lambda_v(\theta) = \mathrm{e}^{in\theta}$ for some $n \in \mathbb{Z}$, depending on $v$.

(2) If we denote

$$H_n(T) = \{v \in H(T) : \ T_{r(\theta)}v = \mathrm{e}^{in\theta}v, \ \theta \in \mathbb{R}\}$$

then $H(T) = \sum_n H_n(T)$ is the orthogonal sum of the subspaces $H_n(T)$. Moreover, if $T$ is irreducible, we have $\dim(H_n) \in \{0, 1\}$, by Theorem 12.

We now aim to replace the orthonormal basis $\Psi(y)$ of $H(y)$, $y \in Y$ in the definition of $\pi$ by a basis consisting of eigenvectors of $K$ in $H(y)$.

**Lemma 8.1.1.** *For all $y \in Y$ we can find a set of vectors $\Phi_0(y) = \{\phi_n(y)\ n \in \mathbb{Z}\}$ with $\phi_n(y) \neq 0$ if $\dim(H_n(y)) = 1$ and $\phi_n(y) = 0$ if $\dim(H_n(y)) = 0$ such that the set $\Phi(y) = \Phi_0(y) - \{0\}$ forms an orthonormal basis in $H(y)$ and a function $f$ is $\Psi$-measurable if and only if it is $\Phi$-measurable.*

*Proof.* We see from the remark above that $H = H(\pi) = \sum_n H_n(\pi)$. We let $\{\mathrm{e}_k^n : k \in \mathbb{N}\}$, $n \in \mathbb{Z}$ denote an orthonormal basis of $H_n(\pi)$. By definition of $H$ and the direct integral decomposition of $\pi$, this means that the mappings $y \mapsto \mathrm{e}_k^n(y)$, $y \in Y$ are $\Psi$-measurable and such that $\mathrm{e}_k^n(y) \in H_n(y)$ , for all $k \in \mathbb{N}$, $n \in \mathbb{Z}$.
Define

$$\begin{aligned} Y_k^n &= \{y \in Y : \mathrm{e}_k^n(y) \neq 0\} \\ \overline{Y_k^n} &= Y_k^n - \bigcup_{i=1}^{k-1} Y_i^n \\ Y_n &= \bigcup_{k=1}^{\infty} \overline{Y_k^n}. \end{aligned}$$

From the above remarks on the measurability of the functions $\mathrm{e}_k^n$, we see that the set $Y_k^n$ is a measurable subset of $Y$, implying that $\overline{Y_k^n}$, $Y_n$ are so too. We have

$$y \in Y_n \iff \exists k \in \mathbb{N} : \ \mathrm{e}_k^n(y) \neq 0 \ \wedge \ \mathrm{e}_j^n(y) = 0, \ \ \forall 0 \leq j \leq k-1$$

$n \in \mathbb{Z}$. In particular, this shows that $Y_n$ is a disjoint union. It follows from Theorem 11, that the sets $\{\mathrm{e}_k^n(y) : k \in \mathbb{N}\}$ , $n \in \mathbb{Z}$ generate $H(y)$ and so

$$\nu(Y - \bigcup_{n=1}^{\infty} Y_n) = 0$$

Hence, we can assume without loss of generality that $Y = \bigcup_{k=1}^{\infty} Y_n$. For $n \in \mathbb{Z}$, let

$$\phi_n(y) = \begin{cases} \frac{\mathrm{e}_k^n(y)}{\|\mathrm{e}_k^n(y)\|}, & \text{if } y \in \overline{Y_k^n} \\ 0 & \text{if } y \in Y - Y_n \end{cases}$$

By the remarks above, we see that $\phi_n(y) \in H_n(y)$, $n \in \mathbb{Z}$, $y \in Y$ and that the maps $y \mapsto < \phi_n(y), \psi_m(y) >_{H(y)}$ are measurable for all $n, m \in \mathbb{Z}$. Said differently, the functions $\phi_n$, $n \in \mathbb{Z}$, are $\Psi$-measurable. Moreover, we have

$$y \in Y - Y_n \iff e_k^n(y) = 0, \ \forall k = 1, 2, ...$$

and thus $\phi_n(y) = 0$ if $\dim(H_n(y)) = 0$.
Let

$$\Phi_0(y) = \{\phi_n(y) : n \in \mathbb{Z}\}$$

Then the set $\Phi(y) = \Phi_0(y) - \{0\}$ forms an orthonormal basis in $H(y)$. To see this, let $\overline{H(y)}$ be the closed subspace of $H(y)$, spanned by $\Phi(y)$.
It is clear from the definition of the set $\Phi(y)$ that

$$\overline{H(y)} \subseteq \{f(y) : f \in H(\pi)\}.$$

Conversely, a function $f$ on $Y$ with $f(y) \in H(y)$ is in $H(\pi)$ only if it is $\Psi$-measurable. But as the set spanned by $\Phi$ is contained in the set spanned by $\Psi$, a $\Psi$-measurable function $f$ must also be $\Phi$-measurable. Thus $\overline{H(y)} \supseteq \{f(y) : f \in H(\pi)\}$, and we conclude

$$\overline{H(y)} = \{f(y) : f \in H(\pi)\}.$$

We have that

$$(\pi_y(g)(\phi_n(y)) \in \overline{H(y)}$$

for all $y \in Y$, $g \in G$, $n \in \mathbb{Z}$. This shows that $\overline{H(y)}$ is $\pi_y$-invariant. As $\pi_y$ is irreducible, we conclude $\overline{H(y)} = H(y)$.
It follows from

$$\{f(y) : f \in H(\pi)\} = \overline{H(y)} = H(y)$$

that a function $f$ is $\Phi$-measurable if and only if it is $\Psi$-measurable, proving the last assertion of the Lemma.

□

## 8.2 The Set of $K^p$-Vectors

From now on we will always denote by $\pi$ the direct integral of representations $\pi_y$ on $H(y)$, $y \in Y$ defined above and by $\Phi(y) = \{\phi_n(y) : n \in \mathbb{Z}\}$ the basis in $H(y)$ consisting of eigenvectors $\phi_n(y)$ for the action of $K$ on $H(y)$ by $\pi_y$ such that for $n \in \mathbb{Z}$, $y \in Y$

$$\begin{aligned} 0 \neq \phi_n(y) \in H_n(y) \text{ if } \dim H_n(y) = 1, \\ \phi_n(y) = 0 \text{ if } \dim H_n(y) = 0. \end{aligned}$$

In particular, we have for all $n \in \mathbb{Z}$, $y \in Y$, $\theta \in \mathbb{R}$,

$$\pi_y(r(\theta))\phi_n(y) = \mathrm{e}^{in\theta}\phi_n(y).$$

In this section, we study how component vectors in $H(y), y \in Y$ inherit certain differentiability properties of vectors in $H(\pi)$.

**Definition 43.** An element $v \in H(\pi)$ is called a *$K^p$-vector*, $p > 0$, if $\theta \mapsto \pi(r(\theta))v$ is a $C^p$ map from $\mathbb{R}$ to $H(\pi)$. We shall denote by $K(\pi, p)$ the set of all $K^p$-vectors in $H(\pi)$.

For $0 < p < 1$, we mean with $v \in K(\pi, p)$ that $v$ is Hölder continuous in the direction of $K$ with the Hölder exponent $p$, i.e.

$$\sup_{\theta \neq 0} \frac{\|\pi(r(\theta))v - v\|_H}{|\theta|^p} < \infty$$

and for $n < p < n+1$, $n \in \mathbb{Z}^+$, the expression $v \in K(\pi, p)$ means that the $n^{th}$ derviative of of $\pi(r(\theta))v$ exists and belongs to the set $K(\pi, p-n)$.

We want to study the set of $K^p$-vectors in $H(\pi)$. Observe that since the subgroup $K$ is one-dimensional (it is parametrized by $\mathbb{R}$), its Lie-Algebra $\mathfrak{k}$ is generated by exactly one basis element of the Lie-Algebra $\mathfrak{g}$ of $SL_2(\mathbb{R})$. This follows from the fact that $\mathfrak{k}$ is a Lie-Sub-Algebra of $\mathfrak{g}$. We consider once more the basis for $\mathfrak{g}$ given by

$$W = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\ Q = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\ V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

We have

$$\exp(tW) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix} \in K$$

for all $t \in \mathbb{R}$ and so the element $W$ forms a basis for $\mathfrak{k}$. For $f \in K(\pi, 1)$, the Lie-derivative in the direction of $K$ exists (by definition of $K(\pi, 1)$), and is equal to $L_W(f) \in H(\pi)$.
We denote by $L_W^y$ the derivative functions in $H(y)$, $y \in Y$. By the direct integral decomposition of $\pi$, it would be natural to expect that the relation $(L_W(f))(y) = L_W^y(f(y))$ holds. However, as the Lemma below shows, one needs to be careful with their intuition when dealing with uncountable sums of representations such as $\pi$.

**Lemma 8.2.1.** *Let $f \in K(\pi, 1)$. Then for almost every $y \in Y$*

(1) $< (L_W f)(y), \phi_n(y) >_{H(y)} = in < f(y), \phi_n(y) >_{H(y)}$;
(2) $f(y) \in K(\pi_y, 1)$;
(3) $(L_W f)(y) = L_W^y(f(y))$.

*Proof.* Let $Y' \subseteq Y$ be a measurable set with $\nu(Y') > 0$ and let $\phi_n'$, $n \in \mathbb{Z}$ be functions on $Y$ defined by

$$\phi_n'(y) = \begin{cases} \phi_n(y) & \text{if } y \in Y' \\ 0 & \text{if } y \notin Y' \end{cases}$$

Then $\phi_n'$ is $\Psi$-measurable for all $n \in \mathbb{Z}$ by Lemma 8.1.1. Hence, $\phi_n' \in H(\pi)$. We have

$$\left\| \frac{\pi(r(\theta))\phi_n' - \phi_n'}{\theta} - in\phi_n' \right\|^2 = \int_{Y'} \left\| \frac{\pi_y(r(\theta))\phi_n(y) - \phi_n(y)}{\theta} - in\phi_n(y) \right\|_{H(y)}^2 d\nu(y)$$

by the definition of $\pi$. As $\pi_y(r(\theta))\phi_n(y) = \mathrm{e}^{in\theta}\phi_n(y)$, the integral in the above equation is equal to

$$\int_{Y'} \left\| \left( \frac{\mathrm{e}^{in\theta}-1}{\theta} - in \right) \phi_n(y) \right\|^2_{H(y)} d\nu(y) = \int_{Y'} \left| \frac{\mathrm{e}^{in\theta}-1}{\theta} - in \right|^2 \|\phi_n(y)\|^2_{H(y)} d\nu(y)$$

$$= \left| \frac{\mathrm{e}^{in\theta}-1}{\theta} - in \right|^2 \nu(\{y \in Y' : \phi_n(y) \neq 0\}$$

because $\|\phi_n(y)\|_{H(y)} \in \{0,1\}$, $y \in Y$. Since

$$\lim_{\theta \to 0} \frac{\mathrm{e}^{in\theta}-1}{\theta} - in = 0,$$

this implies that

$$L_W(\phi'_n) = \lim_{\theta \to 0} \frac{\pi(r(\theta))\phi'_n - \phi'_n}{\theta} = in\phi'_n$$

$n \in \mathbb{Z}$ and thus $\phi'_n \in K(\pi, 1)$.
From Lemma 7.3.1 and the above, we see

$$< L_W f, \phi'_n >_{H(\pi)} = - < f, L_W \phi'_n >_{H(\pi)} = \ in < f, \phi'_n >_{H(\pi)} .$$

Hence

$$\int_{Y'} < (L_W f)(y), \phi_n(y) >_{H(y)} d\nu(y) = in \int_{Y'} < f(y), \phi_n(y) >_{H(y)} d\nu(y).$$

As $Y' \subset Y$, $\nu(Y') > 0$ was chosen arbitrarily, the terms inside the integral of the above equation hold up to measure zero. Thus

$$< (L_W f)(y), \phi_n(y) >_{H(y)} = in < f(y), \phi_n(y) >_{H(y)}$$

for almost every $y \in Y$ and all $n \in \mathbb{Z}$.
We have

$$\begin{aligned}
\|L_W f\|^2_{H(\pi)} &= \int_Y \|(L_W f)(y)\|^2_{H(y)} \nu(y) \\
&= \int_Y \| \sum_n \phi_n(y) < (L_W f)(y), \phi_n(y) >_{H(y)} \|^2_{H(y)} d\nu(y) \\
&= \int_Y \| \sum_n \phi_n(y) in < f(y), \phi_n(y) >_{H(y)} \|^2_{H(y)} d\nu(y) \\
&= \int_Y \sum_n n^2 | < f(y), \phi_n(y) >_{H(y)} |^2 d\nu(y)
\end{aligned}$$

where we used the first part of the Lemma proven above in the third equality. Since $L_W f \in H(\pi)$, it holds that $\|L_W f\|^2_{H(\pi)} < \infty$ and therefore,

$$\int_Y \sum_n n^2 | < f(y), \phi_n(y) >_{H(y)} |^2 d\nu(y) < \infty.$$

which implies that

$$\sum_n n^2 | < f(y), \phi_n(y) >_{H(y)} |^2 < \infty$$

for almost every $y \in Y$.
For these $y$ we have

$$\begin{aligned} \| L_W^y(f(y)) \|_{H(y)}^2 &= \| \sum_n \phi_n(y) < L_W^y(f(y)), \phi_n(y) >_{H(y)} \|_{H(y)} \\ &= \sum_n n^2 | < f(y), \phi_n(y) >_{H(y)} |^2 \ < \ \infty. \end{aligned}$$

Then $f(y) \in K(\pi_y, 1)$ and

$$(L_W f)(y) = L_W^y(f(y))$$

by definition of $\pi$.

□

**Corollary 10.** Let $f \in K(\pi, m)$ for some integer $m > 0$. Then

$$\| L_W^k f \|_{H(\pi)}^2 = \int_Y \left( \sum_{n \in \mathbb{Z}} n^{2k} | < f(y), \phi_n(y) >_{H(y)} |^2 \right) d\nu(y) < \infty$$

for all $k = 1, ..., m$.

*Proof.* It is clear that $f \in K(\pi, k)$, $1 < k < m$. By Lemma 8.2.1 (1) we have

$$\begin{aligned} < (L_W f)(y), \phi_n(y) > \ &= in < f(y), \phi_n(y) > \\ < (L_W f)^2(y), \phi_n(y) > \ &= in < f(y), L_W(\phi_n(y)) > = (in)^2 < f(y), \phi_n(y) > \\ &\vdots \\ < (L_W f)^k(y), \phi_n(y) > \ &= \ (in)^k < f(y), \phi_n(y) > \end{aligned}$$

for a.e. $y \in Y$. The rest follows from a computation similar to the one we made for $\| L_W f \|_{H(\pi)}$ above. □

Now consider $f \in K(\pi, p)$ for $0 < p < 1$. Then $f$ is Hölder continuous in the direction of $K$ with Hölder exponent $p$, and so

$$\| \pi(r(\theta)) f - f \| \leq C |\theta|^p \tag{8.2.1.1}$$

for some constant $C > 0$ and all $\theta \in \mathbb{R}$. Denote

$$c_{n,f}(y) = < f(y), \phi_n(y) >_{H(y)} .$$

**Lemma 8.2.2.** *Let $f \in K(\pi, p)$ with $0 < p < 1$. Then there is a constant $D > 0$ such that*

$$Q(t) = \int_Y \left( \sum_{|n| \geq t} |c_{n,f}(y)|^2 \right) d\nu(y) \leq Dt^{-2p}$$

*for all $t \geq 1$.*

*Proof.* For $m \in \mathbb{Z}^+ \cup \{0\}$ define

$$\theta_m = \frac{2\pi}{3 \cdot 2^m}.$$

Then for all $n \in \mathbb{Z}$ with $2^m \leq |n| \leq 2^{m+1}$,

$$\frac{2\pi}{3} \leq |n\theta_m| \leq \frac{2^2\pi}{3}.$$

We have $\mathrm{Re}(\mathrm{e}^{in\theta_m}) < -1/2$, and

$$|\mathrm{e}^{in\theta_m} - 1| \geq \sqrt{3}. \tag{8.2.2.1}$$

Moreover,

$$\|\pi(r(\theta_m))f - f\|_H^2 = \int_Y \|\pi_y(r(\theta_m))f(y) - f(y)\|_{H(y)}^2 d\nu(y)$$

$$= \int_Y \|\pi_y(r(\theta_m)) \sum_{n \in \mathbb{Z}} \phi_n(y) c_{n,f}(y) - \sum_{n \in \mathbb{Z}} \phi_n(y) c_{n,f}(y)\|_{H(y)}^2 d\nu(y)$$

$$= \int_Y \| \sum_{n \in \mathbb{Z}} \phi_n(y) c_{n,f}(y) (\mathrm{e}^{in\theta_m} - 1)\|_{H(y)}^2 d\nu(y)$$

since $\pi_y(r(\theta_m))\phi_n(y) = \mathrm{e}^{in\theta_m}\phi_n(y)$. We compute the term inside the norm to get

$$= \int_Y \left( \sum_{n \in \mathbb{Z}} |\mathrm{e}^{in\theta_m} - 1|^2 |c_{n,f}(y)|^2 \right) d\nu(y) \leq C^2 |\theta_m|^{2p} \tag{8.2.2.2}$$

by equation 8.2.1.1. We have using 8.2.2.1 and 8.2.2.2

$$E_m := \int_Y \left( \sum_{2^m \leq |n| \leq 2^{m+1}} |c_{n,f}(y)|^2 \right) d\nu(y)$$

$$\leq \int_Y \left( \sum_{2^m \leq |n| \leq 2^{m+1}} |\mathrm{e}^{in\theta_m} - 1|^2 |c_{n,f}(y)|^2 \right) d\nu(y)$$

$$\leq \|\pi(r(\theta_m))f - f\|_H^2 \leq C^2 \theta_m^{2p} = C^2 \left( \frac{2\pi}{3 \cdot 2^m} \right)^{2p}. \tag{8.2.2.3}$$

This holds for all $m \in \mathbb{Z}^+ \cup \{0\}$. For a given $t \geq 1$ we choose $m_0$ such that

$$2^{m_0} \leq t \leq 2^{m_0+1}.$$

We have $Q(2^{m_0}) = \sum_{m=m_0}^{\infty} E_m$. Hence

$$Q(t) \leq \sum_{m=m_0}^{\infty} E_m \leq C^2 \sum_{m=m_0}^{\infty} \theta_m^{2p} \leq D2^{-2p(m_0+1)} \leq Dt^{-2p}$$

for some $D > 0$ by equation 8.2.2.3, which completes the proof.

□

**Corollary 11.** Let $p > 0$ and $f \in K(\pi, p)$. There exists a constant $D_f$ depending only on $f$ such that

$$Q(t) = \int_Y \left( \sum_{|n| \geq t} |c_{n,f}(y)|^2 \right) d\nu(y) \leq D_f t^{-2p}$$

for all $t \geq 1$.

*Proof.* For $0 < p < 1$ this follows from Lemma 8.2.2. We may therefore assume that $p \geq 1$. Denote by $k = k(p)$ the integer part of $p$ and let

$$c_{n,f}^k(y) = < L_W^k f(y), \phi_n(y) >_{H(y)} .$$

It follows from Lemma 8.2.1 (1) that

$$c_{n,f}^k(y) = < L_W^k f(y), \phi_n(y) >_{H(y)} = (in)^k c_{n,f}(y)$$

for a.e. $y \in Y$. Hence

$$\begin{aligned} Q(t) = \int_Y \left( \sum_{|n| \geq t} |c_{n,f}(y)|^2 \right) d\nu(y) &= \int_Y \left( \sum_{|n| \geq t} \frac{|c_{n,f}^k(y)|^2}{n^{2k}} \right) d\nu(y) \\ &\leq t^{-2k} \int_Y \left( \sum_{|n| \geq t} |c_{n,f}^k(y)|^2 \right) d\nu(y). \end{aligned}$$

Since $f \in K(\pi, p)$, we have $L_W^k f \in K(\pi, p-k)$. Suppose $p \neq k$. Then $0 < p - k < 1$ and it follows from Lemma 8.2.2 that

$$\int_Y \left( \sum_{|n| \geq t} |c_{n,f}^k(y)|^2 \right) d\nu(y) \leq D_f t^{-2(p-k)}$$

and

$$Q(t) \leq t^{-2k} D_f t^{-2(p-k)} = D_f t^{-2p}.$$

The case $p = k \geq 1$ is proven by Corollary 10 and a similar computation as above.

□

## 8.3 Rate of Decay for Unitary Irreducible Representations of G

We now assume that the non-trivial unitary representation $T$ of $G = SL_2(\mathbb{R})$ in the Hilbert space $H(T)$ is irreducible. In this case, we know by Theorem 12 and Theorem 11 that there exists an orthonormal basis $\{\phi_n, n \in \mathbb{Z}\} - \{0\}$ of $H(T)$ with $0 \neq \phi_n \in H_n(T)$ if $\dim H_n(T) = 1$ and $\phi_n = 0$ if $\dim H_n(T) = 0$ such that

$$T(r(\theta))\phi_n = \mathrm{e}^{in\theta}\phi_n \tag{8.3.0.1}$$

for all $n \in \mathbb{Z}$, $\theta \in \mathbb{R}$. We consider the following matrix coefficients of $T$ restricted to the geodesic flow, given by

$$< \phi_n, T_{a(t)}(\phi_m) >$$

$n, m \in \mathbb{Z}$. Remark that these coefficients converge to zero as $t \to \pm\infty$, by Theorem 8. The goal of this section is to establish the rate of this convergence. This will yield the rate of convergence for the matrix coefficients of $T$ restricted to the horocycle flow. The idea is to use these estimates in the more general case where the irreducibility assumption on $T$ is dropped. Intuitively speaking, we then use the unitary equivalence to the direct integral decomposition $\pi$ from above to estimate the terms inside the direct integral (which are irreducible) using these results.

We still consider the basis of the Lie algebra $\mathfrak{g}$ of $G$ given by

$$W = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},\ Q = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},\ V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

We defined the Lie derivative $L_X$ at $X \in \mathfrak{g}$ on the set of $C^1$-vectors of $H(T)$ by

$$L_X v = \lim_{t \to 0} \frac{1}{t} \left(T(\exp(tX))v - v\right)$$

$v \in C^1$.
The Casimir operator $\Omega = \Omega_T$ associated to $T$ is defined by

$$\Omega_T = (L_V^2 + L_Q^2 - L_W^2)/4.$$

Theorem 14 states that

$$\Omega v = \lambda v \tag{8.3.0.2}$$

for some $\lambda = \lambda(T) \in \mathbb{R}$ and all $C^2$-vectors $v \in H(T)$.
Recall from Theorem 13, that the vectors $\phi_n$ are $C^\infty$ in $H(T)$. Equation 8.3.0.1 shows that

$$L_W \phi_n = in\phi_n,\ n \in \mathbb{Z}. \tag{8.3.0.3}$$

Moreover, for all $n \in \mathbb{Z}$,

$$\Omega\phi_n = \lambda\phi_n$$

by equation 8.3.0.2.
For $v \in H(T)$, $g \in G$, write

$$v \circ g = T_g v. \tag{8.3.0.4}$$

We denote

$$B_{n,m}(t) = < \phi_n, \phi_m \circ a(t) >$$

$n, m \in \mathbb{Z}$, where

$$a(t) = \begin{pmatrix} \mathrm{e}^t & 0 \\ 0 & \mathrm{e}^{-t} \end{pmatrix},\ t \in \mathbb{R}$$

and $\phi_m \circ a(t) = T_{a(t)}\phi_m$ by 8.3.0.4. We then see that the mapping $t \mapsto B(t)$, $t \in \mathbb{R}$ is in $C^\infty$ and

$$\begin{aligned} B'_{n,m}(t) = d/dt < \phi_n, \phi_m \circ a(t) > = < \phi_n, d/dt(T_{a(t)}\phi_m) > \\ = < \phi_n, (L_Q\phi_m) \circ a(t) > \end{aligned}$$

since $\exp tQ = a(t)$ and by rules of differentiation. Similarly,

$$B''_{n,m}(t) = < \phi_n, (L_Q^2\phi_m) \circ a(t) > .$$

**Lemma 8.3.1.** *For $t \geq 1$, the function $B_{n,m}(t)$ satisfies the following differential equation:*

$$y'' + 2y' - 4\lambda y = f_1(t) + f_2(t) \tag{8.3.1.1}$$

*where*

$$\begin{aligned} f_1(t) &= (2e^{2t}\sinh(2t))^{-1}y'(t) = 1/(e^{4t}-1)y'(t) \\ f_2(t) &= y(t)\left[\frac{2m(n-me^{-2t})}{\sinh(2t)} - \frac{(n-me^{-2t})^2}{\sinh^2(2t)}\right]. \end{aligned} \tag{8.3.1.2}$$

*Proof.* Let $y(t) = B_{n,m}(t)$. We have

$$\lambda y(t) = < \phi_n, (\Omega\phi_m) \circ a(t) > \tag{8.3.1.3}$$

by 8.3.0.2. Write

$$N = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix},\ U = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in \mathfrak{g}.$$

A computation shows that

$$V = 2N + W,\ [W, N] = Q$$

and by linearity of the Lie derivative,

$$\begin{aligned} L_V^2 &= 4L_N^2 + 2L_NL_W + 2L_WL_N + L_W^2 \\ L_WL_N &= L_{[W,N]} + L_NL_W = L_Q + L_NL_W. \end{aligned} \tag{8.3.1.4}$$

Since $\Omega = \frac{1}{4}(L_V^2 + L_Q^2 - L_W^2)$, we have

$$\begin{aligned} L_V^2 &= 4L_N^2 + 4L_NL_W + 2L_Q + L_W^2 \\ 4\Omega &= 4L_N^2 + 4L_NL_W + 2L_Q + L_Q^2. \end{aligned}$$

Plugging the expressions we obtained for $B'_{n,m}(t) = y'(t)$ and $B''_{n,m}(t) = y''(t)$ into equation 8.3.1.1, we get

$$< \phi_n, (L_Q^2\phi_m) \circ a(t) > +2 < \phi_n, (L_Q\phi_m) \circ a(t) > -4\lambda y(t) = q_1(t) + q_2(t)$$

where $q_1(t)$ and $q_2(t)$ are given by, via 8.3.1.3

$$\begin{aligned} q_1(t) &= 4 < \phi_n, (L_N^2\phi_m) \circ a(t) > \\ q_2(t) &= 4 < \phi_n, (L_N L_W \phi_m) \circ a(t) >= -4im < \phi_n, (L_N\phi_m) \circ a(t) > \end{aligned}$$

Or equivalently,

$$y'' + 2y' - 4\lambda y = q_1(t) + q_2(t). \tag{8.3.1.5}$$

We now want to find the functions $q_1$ and $q_2$ explicitly. First, let us show that

$$L_X(v \circ g) = (L_{g^{-1}Xg}v) \circ g \tag{8.3.1.6}$$

for every $g \in G$, $X \in \mathfrak{g}$ and every $C^1$-vector $v \in H(T)$. This just follows from the definition of the map $L_X$, and the fact that $T$ is a group homomoprhism. Indeed,

$$\begin{aligned} L_X(v \circ g) &= \lim_{t\to 0} \frac{T(\exp tX)T(g)v - T(g)v}{t} = \lim_{t\to 0} \frac{T(g)T(g^{-1}\exp tXg)v - T(g)v}{t} \\ &= T(g) \lim_{t\to 0} \frac{T(\exp(g^{-1}tXg))v - v}{t} \\ &= (L_{g^{-1}Xg}v) \circ g. \end{aligned}$$

Since $L_w\phi_n = in\phi_n$ by 8.3.0.3, we get from 8.3.1.6

$$\begin{aligned} iny(t) &= < L_W\phi_n, \phi_m \circ a(t) > = - < \phi_n, L_W(\phi_m \circ a(t)) > \\ &= - < \phi_n, (L_{a(t)^{-1}Wa(t)}\phi_m) \circ a(t) > . \end{aligned}$$

We have $W = U - N$, and $a(t)^{-1}Ua(t) = \mathrm{e}^{-2t}U$, $-a(t)^{-1}Na(t) = -\mathrm{e}^{2t}N$. Hence,

$$a(t)^{-1}Wa(t) = -\mathrm{e}^{2t}N + \mathrm{e}^{-2t}U = -2\sinh(2t)N + \mathrm{e}^{-2t}W.$$

Therefore, by linearity of the Lie derivative $L_X$, $X \in \mathfrak{g}$, and by 8.3.0.3

$$iny(t) = 2\sinh(2t) < \phi_n, (L_N\phi_m) \circ a(t) > +im\mathrm{e}^{-2t} < \phi_n, \phi_m \circ a(t) > \tag{8.3.1.7}$$

$$= -\frac{1}{2im}\sinh(2t)q_2(t) + im\mathrm{e}^{-2t}y(t)$$

and so

$$q_2(t) = \frac{2y(t)m(n - m\mathrm{e}^{-2t})}{\sinh(2t)}.$$

Write

$$z(t) = < \phi_n, (L_N\phi_m) \circ a(t) > .$$

Applying the arguments above to $z(t)$, we have

$$\begin{aligned}
-inz(t) = & - < L_W\phi_n, (L_N\phi_m)\circ a(t) > \\
= & < \phi_n, L_W((L_N\phi_m)\circ a(t)) > \\
= & < \phi_n, (L_{a(t)^{-1}Wa(t)}L_N\phi_m)\circ a(t)) > \\
= & < \phi_n, (-2\sinh 2tL_N^2\phi_m)\circ a(t) > \\
& + < \phi_n, \mathrm{e}^{-2t}(L_WL_N\phi_m)\circ a(t) > \\
= & -2\sinh(2t) < \phi_n, (L_N^2\phi_m)\circ a(t) > \\
& + \mathrm{e}^{-2t} < \phi_n, (L_WL_N\phi_m)\circ a(t) > . \qquad (8.3.1.8)
\end{aligned}$$

Recall that $WN - NW = [W, N] = Q$ and

$$y'(t) = B'_{n,m}(t) = < \phi_n, (L_Q\phi_m)\circ a(t) > .$$

We have

$$< \phi_n, (L_WL_N\phi_m)\circ a(t) > = y'(t) - im < \phi_n, (L_N\phi_m)\circ a(t) > = y'(t) - imz(t).$$

Thus we conclude, together with 8.3.1.7 and 8.3.1.8,

$$q_1(t) = \frac{2iz(t)(m\mathrm{e}^{-2t} - n)}{-2\sinh(2t)} + \mathrm{e}^{-2t}\frac{y'(t)}{2\sinh(2t)}$$

or

$$q_1(t) = -\frac{y(t)(m\mathrm{e}^{-2t} - n)^2}{\sinh^2(2t)} + \frac{\mathrm{e}^{-2t}y'(t)}{2\sinh(2t)}.$$

This and 8.3.1.5 completes the proof.

□

From the Howe-Moore Theorem (Theorem 8) we know that

$$B_{n,m}(t) \to 0, \text{ when } t \to \infty$$

$n, m \in \mathbb{Z}$, since $T$ has no invariant vectors in $H(T)$. We want to estimate for large $t$ the solution of 8.3.1.1 satisfying

$$\begin{aligned}
& y(t) \to 0 \text{ when } t \to \infty \\
& y(1) = B_{n,m}(1) \\
& y'(1) = B'_{n,m}(1).
\end{aligned}$$

We have for $n, m \in \mathbb{Z}$, $t \in \mathbb{R}$

$$|B_{n,m}(t)|^2 \leq \|\phi_n\|^2\|T_{a(t)}\phi_m\|^2 \leq 1$$

by the Cauchy Schwarz inequality and the fact that $T$ is unitary. Similarly,

$$|B'_{n,m}(t)|^2 \leq \|L_Q\phi_m\|^2.$$

Moreover, as $\Omega = \frac{1}{4}(L_V^2 + L_Q^2 - L_W^2)$ we have by 8.3.0.2

$$< (L_V^2 + L_Q^2 - L_W^2)\phi_m, \phi_m > = 4\lambda$$

if $\phi_m \neq 0$. Since $| < \phi_m, L_X^2\phi_m > | = | < L_X\phi_m, L_X\phi_m > | = \|L_X\phi_m\|^2$, $X \in \mathfrak{g}$, this implies for $\phi_m \neq 0$,

$$\|L_V\phi_m\|^2 + \|L_Q\phi_m\|^2 = m^2 - 4\lambda$$

and therefore

$$m^2 - 4\lambda \geq 0$$
$$|B'_{n,m}(t)| \leq \|L_Q\phi_m\| \leq \sqrt{m^2 - 4\lambda} \tag{8.3.1.9}$$

for all $t \in \mathbb{R}$, $n, m \in \mathbb{Z}$, $\phi_m \neq 0$.
The characteristic equation

$$k^2 + 2k - 4\lambda = 0$$

has roots given by

$$r_1 = r_1(\lambda) = -1 + \sqrt{1 + 4\lambda},\ r_2 = r_2(\lambda) = -1 - \sqrt{1 + 4\lambda}$$

We shall henceforth denote by $\sqrt{a}$ the positive square root if $a > 0$ and $i\sqrt{|a|}$ if $a < 0$. Remark that $\mathrm{Re}(r_1) \geq -1$ as $\lambda$ is real. We denote by $D$ the differentiation operator. Then equation 8.3.1.1 can be written in the form

$$(D - r_1)(D - r_2)y = f_1(t) + f_2(t) = f(t)$$

The proof of the next Lemma is similar to the computations on page 7 of [Mat13].

**Lemma 8.3.2.** *For all $t \geq 1$, it holds that*

$$|f_1(t)| \leq C_1 e^{-4t}\sqrt{m^2 - 4\lambda}$$
$$|f_2(t)| \leq C_2(m^2 + n^2)e^{-2t}$$

*for some constants $C_1, C_2 > 0$.*

*Proof.* We have that $|y'(t)| = |B'_{n,m}(t)| \leq \sqrt{m^2 - 4\lambda}$ by equation 8.3.1.9. Hence

$$|f_1(t)| \leq \left|\frac{1}{\mathrm{e}^{4t} - 1}\right| \sqrt{m^2 - 4\lambda} = \left|\frac{\mathrm{e}^{-4t}}{1 - \mathrm{e}^{-4t}}\right| \sqrt{m^2 - 4\lambda} \leq \frac{\mathrm{e}^{-4t}}{1 - \mathrm{e}^{-4}}\sqrt{m^2 - 4\lambda}.$$

The bound on $|f_2(t)|$ is obtained similarly, by using $|y(t)| = |B_{n,m}(t)| \leq 1$ and that $|mn| \leq m^2 + n^2$, $m, n \in \mathbb{Z}$. □

By the product rule for differentiation, one verifies that

$$(D - r_2)y = \mathrm{e}^{r_2 t}D(\mathrm{e}^{-r_2 t}y)$$

$$D(\mathrm{e}^{-r_1 t}(D-r_2)y) = \mathrm{e}^{-r_1 t} f(t)$$

which gives, by integrating both sides,

$$\int_1^t D(\mathrm{e}^{(r_2-r_1)s} D(\mathrm{e}^{-r_2 s} y))ds = \int_1^\infty \mathrm{e}^{-r_1 s} f(s)ds - \int_t^\infty \mathrm{e}^{-r_1 s} f(s)ds.$$

Equivalently,

$$\mathrm{e}^{(r_2-r_1)t} D(\mathrm{e}^{-r_2 t} y) = -\int_t^\infty \mathrm{e}^{-r_1 s} f(s)ds + P_1 \tag{8.3.2.1}$$

where

$$\begin{aligned} P_1 &= \int_1^\infty \mathrm{e}^{-r_1 s} f(s)ds - \mathrm{e}^{(r_2-r_1)} D(\mathrm{e}^{-r_2 t} y)(1) \\ &= \int_1^\infty \mathrm{e}^{-r_1 s} f(s)ds - r_2 \mathrm{e}^{-r_1} y(1) + \mathrm{e}^{-r_1} y'(1) \end{aligned} \tag{8.3.2.2}$$

is a constant.
Since $\mathrm{Re}(r_1+2) \geq 1$, we have for $s \geq 1$

$$\begin{aligned} |\mathrm{e}^{-r_1 s} f(s)| &\leq \mathrm{e}^{-\mathrm{Re}(r_1)s}(|f_1(s)| + |f_2(s)|) \\ &\leq \mathrm{e}^{-3s} C_1' + \mathrm{e}^{-s} C_2' \end{aligned}$$

by Lemma 8.3.2, where $C_1' = C_1\sqrt{m^2-4\lambda}$, $C_2' = C_2(m^2+n^2)$. Therefore, the integral on the right side of 8.3.2.1 converges. Moreover, together with this Lemma, $\mathrm{Re}(r_1+2) \geq 1$ implies that

$$\begin{aligned} \left|\int_t^\infty \mathrm{e}^{-r_1 s} f_1(s)ds\right| &\leq C_1\sqrt{m^2-4\lambda}\int_t^\infty \mathrm{e}^{-(\mathrm{Re}(r_1)+4)s}ds \\ &\leq C_1\sqrt{m^2-4\lambda}\int_t^\infty \mathrm{e}^{-3s}ds \\ &\leq C_1\sqrt{m^2-4\lambda}\mathrm{e}^{-3t} \\ \left|\int_t^\infty \mathrm{e}^{-r_1 s} f_2(s)ds\right| &\leq C_2(m^2+n^2)\int_t^\infty \mathrm{e}^{-(\mathrm{Re}(r_1)+2)s}ds \\ &\leq C_2(m^2+n^2)\int_t^\infty \mathrm{e}^{-s}ds \\ &\leq C_2(m^2+n^2)\mathrm{e}^{-t}. \end{aligned} \tag{8.3.2.3}$$

In case if $r_1 = r_2$, then $r_1 = r_2 \equiv -1$ and we get from equation 8.3.2.1

$$\begin{aligned} D(\mathrm{e}^{-r_2 t} y) &= \mathrm{e}^{-(r_2-r_1)t}\left(-\int_t^\infty \mathrm{e}^{-r_1 s} f(s)ds + P_1\right) = -\int_t^\infty \mathrm{e}^{u} f(u)du + P_1 \\ &= D(\mathrm{e}^t y) \end{aligned}$$

Or, equivalently

$$\mathrm{e}^t y = -\int_1^t \left( \int_s^\infty \mathrm{e}^u f(u)du + P_1 \right) ds + y(1)\mathrm{e}.$$

We obtain

$$y = y(t) = -\mathrm{e}^{-t} \int_1^t \left( \int_s^\infty \mathrm{e}^u f(u)du \right) ds + P_1 t\mathrm{e}^{-t} + P_2\mathrm{e}^{-t}$$

where $P_2 = y(1)\mathrm{e}$.

If $r_1 \neq r_2$ then similarly

$$\begin{aligned} y(t) &= -\mathrm{e}^{r_2 t} \int_1^t \mathrm{e}^{-(r_2-r_1)s} \left( \int_s^\infty \mathrm{e}^{-r_1 u} f(u)du + P_1 \right) ds + \mathrm{e}^{r_2 t} P_2 \\ &= \mathrm{e}^{r_2 t} A(t) + \mathrm{e}^{r_2 t} \left( P_1 \int_1^t \mathrm{e}^{(r_1-r_2)s} ds + P_2 \right) \qquad (8.3.2.4) \\ &= \mathrm{e}^{r_2 t} A(t) + \frac{P_1}{2\sqrt{1+4\lambda}} \mathrm{e}^{r_1 t} + \left( P_2 - \frac{\mathrm{e}^{2\sqrt{1+4\lambda}} P_1}{2\sqrt{1+4\lambda}} \right) \mathrm{e}^{r_2 t} \qquad (8.3.2.5) \end{aligned}$$

where

$$A(t) = A_1(t) + A_2(t)$$

$$A_i(t) = \int_1^t \mathrm{e}^{(r_1-r_2)s} \left( \int_s^\infty \mathrm{e}^{-r_1 u} f_i(u)du \right) ds \quad i = 1, 2$$

and $P_1$ is as in 8.3.2.2, $P_2 = y(1)\mathrm{e}^{-r_2}$. Observe that

$$\left| \mathrm{e}^{r_2 t} \int_1^t \mathrm{e}^{(r_1-r_2)s} ds \right| \leq t\mathrm{e}^{t\mathrm{Re}(r_1)}. \qquad (8.3.2.6)$$

We then get from equation 8.3.2.4

$$|y(t)| \leq \mathrm{e}^{t\mathrm{Re}(r_2)}|A(t)| + t\mathrm{e}^{t\mathrm{Re}(r_1)}|P_1| + \mathrm{e}^{t\mathrm{Re}(r_2)}|P_2| \qquad (8.3.2.7)$$

Using Lemma 8.3.2 and equation 8.3.2.3 we see that

$$\begin{aligned} |\mathrm{e}^{r_2 t} A_1(t)| &\leq \mathrm{e}^{t\mathrm{Re}(r_2)} \int_1^t \mathrm{e}^{(\mathrm{Re}(r_1)-\mathrm{Re}(r_2))s} \left| \int_s^\infty \mathrm{e}^{-r_1 u} f_1(u)du \right| ds \\ &\leq C_1 \sqrt{m^2-4\lambda}\, \mathrm{e}^{t\mathrm{Re}(r_2)} \int_1^t \mathrm{e}^{(\mathrm{Re}(r_1)-\mathrm{Re}(r_2))s} \int_s^\infty \mathrm{e}^{-(\mathrm{Re}(r_1)+2)u} du ds \\ &\leq C_1 \sqrt{m^2-4\lambda}\, \mathrm{e}^{t\mathrm{Re}(r_2)} \int_1^t \mathrm{e}^{(\mathrm{Re}(r_1)-\mathrm{Re}(r_2))s} \mathrm{e}^{-(\mathrm{Re}(r_1)+2)s} ds \\ &\leq C_1 \sqrt{m^2-4\lambda}\, \mathrm{e}^{t\mathrm{Re}(r_2)} \mathrm{e}^{(\mathrm{Re}(r_1)-\mathrm{Re}(r_2))t} \int_1^t \mathrm{e}^{-(\mathrm{Re}(r_1)+2)s} ds \\ &\leq \bar{C}_1 \sqrt{m^2-4\lambda}\, \mathrm{e}^{t\mathrm{Re}(r_1)} \qquad (8.3.2.8) \end{aligned}$$

since $\mathrm{Re}(r_1) - \mathrm{Re}(r_2) \geq 0$ and $\mathrm{Re}(r_1) + 2 \geq 1$. A similar computation shows

$$|\mathrm{e}^{r_2 t} A_2(t)| \leq \bar{C}_2(m^2 + n^2)\mathrm{e}^{t\mathrm{Re}(r_1)} \tag{8.3.2.9}$$

for some constant $\bar{C}_2 > 0$ and all $t \geq 1$.

For the proof of the following Lemma, we borrow some ideas from the proof of Lemma 4.1 in [Mat13].

**Lemma 8.3.3.** *Let $\lambda \in \mathbb{R}$ such that $\Omega v = \lambda v$ for all $C^2$-vectors $v \in H(T)$. For $t \geq 1$ and $n, m \in \mathbb{Z}$*

$$|B_{n,m}(t)| \leq Kb(t)$$

*where*

1. $b(t) = \min\{te^{-t}, e^{-t}(1 + |1 + 4\lambda|^{-1/2})\}$, *if* $\lambda \leq -\frac{1}{4}$;
2. $b(t) = \min\{te^{r_1 t}, e^{r_1 t}|1 + 4\lambda|^{-1/2}\}$, *if* $\lambda \in (-1/4, 0)$;
3. $b(t) = te^{-2t}$, *if* $\lambda \geq 0$;

*and $K = \bar{K}(m^2 + n^2) + \tilde{K}, \bar{K}, \tilde{K} > 0$ are independent of $\lambda$.*

*Proof.* Consider the following possibilities for $\lambda$.

*Case* $\lambda \in [-\frac{1}{2}, -\frac{1}{4})$.
We have $r_1 \neq r_2$, $\mathrm{Re}(r_1) = \mathrm{Re}(r_2) = -1$ and $|\sqrt{1 + 4\lambda}| \leq 1$. Moreover, $|y'(1)| \leq \sqrt{m^2 - 4\lambda} \leq \sqrt{m^2 + 4}$ by 8.3.1.9. Equations 8.3.2.8 and 8.3.2.9 show that

$$|\mathrm{e}^{tr_2} A(t)| \leq \mathrm{e}^{-t}(\bar{C}_1\sqrt{m^2 + 4} + \bar{C}_2(m^2 + n^2)). \tag{8.3.3.1}$$

To estimate $P_1$ and $P_2$, first observe

$$|P_2| = |y(1)\mathrm{e}^{-r_2}| \leq \mathrm{e}$$

since $|y(t)| \leq 1$. Furthermore,

$$\left|\int_1^\infty \mathrm{e}^{-r_1 s} f(s) ds\right| \leq \int_1^\infty \mathrm{e}^s |f(s)| ds \leq \mathrm{e}^{-1}(C_1\sqrt{m^2 + 4} + C_2(m^2 + n^2))$$

by Lemma 8.3.2, and equation 8.3.2.3. Moreover, $|r_2\mathrm{e}^{-r_1} y(1)| \leq 2\mathrm{e}$. Hence,

$$|P_1| \leq \mathrm{e}^{-1}(C_1\sqrt{m^2 + 4} + C_2(m^2 + n^2)) + \mathrm{e}(2 + \sqrt{m^2 + 4}).$$

**Remark 20.** In particular, we see that the bounds on $|\mathrm{e}^{tr_2} A(t)|$, $|P_1|$ and $|P_2|$ depend only on $K$ of the form as in the Lemma.

Since

$$\left|\frac{P_1}{2\sqrt{1 + 4\lambda}}\mathrm{e}^{r_1 t}\right| \leq \frac{\mathrm{e}^{-t}}{2\sqrt{|1 + 4\lambda|}}|P_1|$$

$$\left|\left(P_2 - \frac{\mathrm{e}^{2\sqrt{1+4\lambda}}P_1}{2\sqrt{1+4\lambda}}\right)\mathrm{e}^{r_2 t}\right| \leq \mathrm{e}^{-t}\left(\mathrm{e} + \frac{|P_1|}{2\sqrt{|1+4\lambda|}}\right)$$

we have

$$\left|\frac{P_1}{2\sqrt{1+4\lambda}}\mathrm{e}^{r_1 t}\right| + \left|\left(P_2 - \frac{\mathrm{e}^{2\sqrt{1+4\lambda}}P_1}{2\sqrt{1+4\lambda}}\right)\mathrm{e}^{r_2 t}\right| \leq \mathrm{e}^{-t}\left(\mathrm{e} + \frac{|P_1|}{\sqrt{|1+4\lambda|}}\right).$$

This, Remark 20, equations 8.3.3.1 and 8.3.2.5 show that $|B_{n,m}(t)| \leq K\mathrm{e}^{-t}(1 + |1+4\lambda|^{-1/2})$ for $\lambda \in [-\frac{1}{2}, -\frac{1}{4})$ and some $K = \bar{K}(m^2+n^2) + \tilde{K}$ independent of $\lambda$.
To obtain the bound $Kb(t)$ with $b(t)$ as in the first assertion of the Lemma, it remains to check what happens if $t < (1 + |1+4\lambda|^{-1/2})$. In this case, we have via equation 8.3.2.7

$$|y(t)| \leq \mathrm{e}^{-t}|A(t)| + t\mathrm{e}^{-t}|P_1| + \mathrm{e}^{-t}|P_2| \leq t\mathrm{e}^{-t}(|A(t)| + |P_1| + |P_2|)$$

$t \geq 1$. This and Remark 20 show $|B_{n,m}(t)| \leq Kb(t)$ with $b(t) = t\mathrm{e}^{-t}$ when $t < (1 + |1+4\lambda|^{-1/2})$.

*Case* $\lambda = -\frac{1}{4}$.
We have $r_1 = r_2 = -1$ and we know that $y(t)$ is of the form

$$y(t) = -\mathrm{e}^{-t}\int_1^t \left(\int_s^\infty \mathrm{e}^u f(u)du\right)ds + P_1 t\mathrm{e}^{-t} + P_2\mathrm{e}^{-t}.$$

We apply 8.3.2.3 to obtain

$$\begin{aligned}
|y(t)| &\leq \mathrm{e}^{-t}C_1\sqrt{m^2+1}\int_1^t \mathrm{e}^{-3s}ds + \mathrm{e}^{-t}C_2(m^2+n^2)\int_1^t \mathrm{e}^{-s}ds + (t|P_1| + |P_2|)\mathrm{e}^{-t} \\
&\leq \mathrm{e}^{-t}(C_1\sqrt{m^2+1} + C_2(m^2+n^2) + t|P_1| + |P_2|) \\
&\leq t\mathrm{e}^{-t}(C_1\sqrt{m^2+1} + C_2(m^2+n^2) + |P_1| + |P_2|)
\end{aligned}$$

since $t \geq 1$. Moreover,

$$\begin{aligned}
|P_1| &= \left|\int_1^\infty \mathrm{e}^s f(s)ds + \mathrm{e}y(1) + \mathrm{e}y'(1)\right| \\
&\leq \mathrm{e}^{-1}(C_1\sqrt{m^2+1} + C_2(m^2+n^2)) + \mathrm{e} + \mathrm{e}\sqrt{m^2+1}
\end{aligned}$$

and $|P_2| \leq \mathrm{e}$, by Lemma 8.3.2, equations 8.3.2.2, 8.3.2.3 and since $|y(1)| \leq 1$, $|y'(1)| \leq \sqrt{m^2+4\lambda} = \sqrt{m^2+1}$ via equation 8.3.1.9.
Now, we see that $|y(t)|$ is bounded by the desired function in the Lemma.

*Case* $-\frac{1}{4} < \lambda < 0$.
We have $0 < \sqrt{1+4\lambda} < 1$, $\mathrm{Re}(r_1) = r_1 = -1 + \sqrt{1+4\lambda} \in (-1, 0)$, $\mathrm{Re}(r_2) = r_2 = -1 - \sqrt{1+4\lambda} \in (-2, -1)$, $|r_2| = 1 + \sqrt{1+4\lambda} \in (1, 2)$ and $\sqrt{m^2-4\lambda} \leq$

$\sqrt{m^2+1} \leq m^2+n^2+1$. We use equation 8.3.2.3 and put this into 8.3.2.2 to get

$$|P_1| \leq \frac{C_1}{3\mathrm{e}^3}(m^2+n^2+1) + \frac{C_2}{\mathrm{e}}(m^2+n^2) + 2\mathrm{e} + \mathrm{e}(m^2+n^2+1)$$

$$|P_2| \leq \mathrm{e}^2.$$

Using equations 8.3.2.8 and 8.3.2.9 we have in 8.3.2.5

$$\begin{aligned} |y(t)| &\leq \bar{C_1}(m^2+n^2+1)\mathrm{e}^{tr_1} + \bar{C_2}(m^2+n^2)\mathrm{e}^{tr_1} \\ &+ \frac{|P_1|}{2\sqrt{1+4\lambda}}\mathrm{e}^{tr_1} + \left(\mathrm{e}^2 + \frac{\mathrm{e}^2|P_1|}{2\sqrt{1+4\lambda}}\right)\mathrm{e}^{tr_1} \end{aligned} \tag{8.3.3.2}$$

and so

$$\begin{aligned} |y(t)| &\leq (\bar{C_1}(m^2+n^2+1) + \bar{C_2}(m^2+n^2))\mathrm{e}^{tr_1} \\ &+ (|P_1|(1+\mathrm{e}^2) + \mathrm{e}^2)\mathrm{e}^{tr_1}|1+4\lambda|^{-1/2} \end{aligned}$$

for all $t \geq 1$, since $0 < \sqrt{1+4\lambda} < 1$. We see from the estimates on $|P_1|$ from above that $|y(t)| \leq K\mathrm{e}^{r_1 t}|1+4\lambda|^{-1/2}$ for a constant $K$ of the form as in the Lemma. If $t < |1+4\lambda|^{-1/2}$, then we proceed analogously as in case $\lambda \in [-1/2, -1/4)$ and $t < 1 + |1+4\lambda|^{-1/2}$, by using equation 8.3.2.7. This proves case 2 of the Lemma.

*Case* $\lambda < -\frac{1}{2}$.
Now, $\sqrt{m^2-4\lambda}$ is unbounded and we cannot estimate $\mathrm{e}^{r_2 t}A_1(t)$ using equation 8.3.2.8. We may proceed as follows. Recall that $f_1(t)$ in 8.3.1.2 is given by

$$f_1(t) = (2\mathrm{e}^{2t}\sinh(2t))^{-1}y'(t)$$

and that

$$A_1(t) = \int_1^t \mathrm{e}^{(r_1-r_2)s}\left(\int_s^\infty \mathrm{e}^{-r_1 u}f_1(u)du\right)ds.$$

Set

$$I(s) = \int_s^\infty \mathrm{e}^{-r_1 u}f_1(u)du = 2\int_s^\infty \mathrm{e}^{-(r_1+2)u}\frac{y'(u)}{\sinh(2u)}du$$

and let

$$J(u) = \int_u^\infty \frac{y'(v)}{\sinh(2v)}dv.$$

Integration by parts yields

$$\begin{aligned} J(u) &= \left[\frac{y(v)}{\sinh(2v)}\right]_u^\infty - \int_u^\infty y(v)(\sinh(2v)^{-1})'dv \\ &= \frac{y(u)}{\sinh(2u)} + 2\int_u^\infty \frac{y(v)\cosh(2v)}{\sinh^2(2v)}dv. \end{aligned}$$

We claim that there is a constant $Q_1 > 0$ with

$$|J(u)| \leq Q_1 \mathrm{e}^{-2u}. \tag{8.3.3.3}$$

for all $u \geq 1$. Indeed, the function

$$\frac{\cosh(2v)}{\sinh(2v)} = \frac{1+\mathrm{e}^{-4v}}{1-\mathrm{e}^{-4v}}$$

attains its maximum at $v = 1$. Therefore

$$\begin{aligned}
|J(u)| &\leq \left|\frac{y(u)}{\sinh(2u)}\right| + 2\left(\frac{1+\mathrm{e}^{-4}}{1-\mathrm{e}^{-4}}\right)\int_u^\infty \frac{y(v)}{\sinh(2v)}dv \\
&\leq \left|\frac{1}{\sinh(2u)}\right| + 2\left(\frac{1+\mathrm{e}^{-4}}{1-\mathrm{e}^{-4}}\right)\int_u^\infty \frac{1}{\sinh(2v)}dv \\
&\leq \mathrm{e}^{-2u}\frac{2}{1-\mathrm{e}^{-4}} + 2\left(\frac{1+\mathrm{e}^{-4}}{1-\mathrm{e}^{-4}}\right)\int_u^\infty \frac{1}{\sinh(2v)}dv \\
&\leq \left(\frac{2}{1-\mathrm{e}^{-4}} + 2\left(\frac{1+\mathrm{e}^{-4}}{1-\mathrm{e}^{-4}}\right)\right)\mathrm{e}^{-2u}.
\end{aligned}$$

since $|y(v)| \leq 1$, $\frac{2}{\sinh(2u)} = \frac{2\mathrm{e}^{-2u}}{1-\mathrm{e}^{-4u}} \leq \frac{2\mathrm{e}^{-2u}}{1-\mathrm{e}^{-4}}$ when $u \geq 1$.
Integrating by parts once more, we get

$$\begin{aligned}
I(s) &= 2\int_s^\infty \mathrm{e}^{-(r_1+2)u}J'(u)du \\
&= 2\left(\mathrm{e}^{-(r_1+2)s}J(s) + (r_1+2)\int_s^\infty \mathrm{e}^{-(r_1+2)u}J(u)du\right)
\end{aligned}$$

and so

$$\begin{aligned}
A_1(t) &= 2\left[\int_1^t \mathrm{e}^{-(r_2+2)s}J(s)ds + (r_1+2)\int_1^t \mathrm{e}^{(r_1-r_2)s}\left(\int_s^\infty \mathrm{e}^{-(r_1+2)u}J(u)du\right)ds\right] \\
&= 2(F_1(t)+F_2(t))
\end{aligned}$$

where

$$F_1(t) = \int_1^t \mathrm{e}^{-(r_2+2)s}J(s)ds$$

and

$$F_2(t) = (r_1+2)\int_1^t \mathrm{e}^{(r_1-r_2)s}\left(\int_s^\infty \mathrm{e}^{-(r_1+2)u}J(u)du\right)ds.$$

By equation 8.3.3.3 we have

$$|\mathrm{e}^{r_2 t}F_1(t)| \leq \mathrm{e}^{\mathrm{Re}(r_2)t}\int_1^t |\mathrm{e}^{-(r_2+2)s}J(s)|ds \leq \mathrm{e}^{-t}Q_1\int_1^t \mathrm{e}^{-3s}ds \leq Q_2\mathrm{e}^{-t} \tag{8.3.3.4}$$

since $\mathrm{Re}(r_2) = -1$.

To estimate $\mathrm{e}^{r_2 t}F_2(t)$, we integrate by parts:

$$\begin{aligned}\mathrm{e}^{r_2 t}F_2(t) &= \mathrm{e}^{r_2 t}(r_1+2)\left[\frac{\mathrm{e}^{(r_1-r_2)s}}{r_1-r_2}\int_s^\infty \mathrm{e}^{-(r_1+2)u}J(u)du\right]_1^t \\ &\quad + \frac{\mathrm{e}^{r_2 t}(r_1+2)}{r_1-r_2}\int_1^t \mathrm{e}^{(r_1-r_2)s}\mathrm{e}^{-(r_1+2)s}J(s)ds \\ &= \frac{\mathrm{e}^{r_2 t}(r_1+2)}{r_1-r_2}\left(\left[\mathrm{e}^{(r_1-r_2)s}\int_s^\infty \mathrm{e}^{-(r_1+2)u}J(u)du\right]_1^t + F_1(t)\right).\end{aligned}$$

We have by 8.3.3.3

$$\begin{aligned}\left|\left[\mathrm{e}^{(r_1-r_2)s}\int_s^\infty \mathrm{e}^{-(r_1+2)u}J(u)du\right]_1^t\right| &\le Q_1\left|\int_t^\infty \mathrm{e}^{-3u}du\right| + Q_1\left|\int_1^\infty \mathrm{e}^{-3u}du\right| \\ &\le \frac{2\cdot Q_1}{3\cdot \mathrm{e}^3}\mathrm{e}^{-3t} \qquad (8.3.3.5)\end{aligned}$$

Observe that

$$\left|\frac{r_1+2}{r_1-r_2}\right| = \left|\frac{1+\sqrt{1+4\lambda}}{2\sqrt{1+4\lambda}}\right| \le 1$$

since $\lambda < -\frac{1}{2}$ implies $|\sqrt{1+4\lambda}| > 1$. Together with 8.3.3.4 and 8.3.3.5 this gives

$$|\mathrm{e}^{r_2 t}F_2(t)| \le \frac{2\cdot Q_1}{3\cdot \mathrm{e}^3}\mathrm{e}^{-3t} + Q_2\mathrm{e}^{-t} \le Q_3\mathrm{e}^{-t}. \qquad (8.3.3.6)$$

Putting together what we found in equations 8.3.2.9, 8.3.3.4 and 8.3.3.6, we see that for the term $\mathrm{e}^{r_2 t}A(t)$ in 8.3.2.5 we have

$$|\mathrm{e}^{r_2 t}A(t)| \le 2\mathrm{e}^{-t}(Q_2+Q_3)+\bar{C_2}(m^2+n^2)\mathrm{e}^{\mathrm{Re}(r_1)t} = 2\mathrm{e}^{-t}(Q_2+Q_3+\bar{C_2}(m^2+n^2))$$

since $\mathrm{Re}(r_1) = -1$, for all $t \ge 1$. We set $\tilde{Q} = 2(Q_2+Q_3)$ and $\bar{Q} = \bar{C_2}$. We now want to estimate the quantity $|P_1/(2\sqrt{1+4\lambda})|$ in 8.3.2.5. We follow the computations in [Mat13] on page 14. We have $r_1 = -1+i\sqrt{|1+4\lambda|}$, $|y(1)| \le 1$ and $|y'(1)| \le \sqrt{m^2-4\lambda}$. Thus

$$\begin{aligned}\left|\frac{P_1}{2\sqrt{1+4\lambda}}\right| &\le \frac{1}{2\sqrt{|1+4\lambda|}}\left(\left|\int_1^\infty \mathrm{e}^{-r_1 s}f_1(s)ds\right| + \left|\int_1^\infty \mathrm{e}^{-r_1 s}f_2(s)ds\right|\right) \\ &\quad + \frac{1}{2\sqrt{|1+4\lambda|}}(|r_2\mathrm{e}^{-r_1}y(1)| + |\mathrm{e}^{-r_1}y'(1)| \\ &\le \frac{1}{2\sqrt{|1+4\lambda|}}\left(\frac{C_1}{3\mathrm{e}^3}\sqrt{m^2-4\lambda} + \frac{C_2}{\mathrm{e}}(m^2+n^2)\right) \\ &\quad + \frac{1}{2\sqrt{|1+4\lambda|}}(\sqrt{1+|1+4\lambda}\cdot\mathrm{e} + \sqrt{m^2-4\lambda}\cdot\mathrm{e})\end{aligned}$$

where we used Lemma 8.3.2 for the estimate of the first two terms, and the fact that $|r_2\mathrm{e}^{-r_1}y(1)| \le |r_2|\mathrm{e} = |-1-i\sqrt{|1-4\lambda|}\ |\ \mathrm{e}$, $|\mathrm{e}^{-r_1}y'(1)| \le \sqrt{m^2-4\lambda}\ \mathrm{e}$.

Since $\lambda < -1/2$, we have $|1+4\lambda| = |4\lambda| - 1$ and so $m^2 - 4\lambda = m^2 + 4|\lambda| = m^2 + |1+4\lambda| + 1$. By continuing the above computations, we get

$$\begin{aligned}
&= \frac{1}{2\sqrt{|1+4\lambda|}} \left( \frac{C_1}{3\mathrm{e}^3} \sqrt{m^2 + |1+4\lambda| + 1} + \frac{C_2}{\mathrm{e}} (m^2+n^2) \right) \\
&+ \frac{\mathrm{e}}{2\sqrt{|1+4\lambda|}} (\sqrt{1+|1+4\lambda|} + \sqrt{m^2+|1+4\lambda|+1}) \\
&\leq \frac{1}{2\sqrt{|1+4\lambda|}} \left( \frac{C_1}{3\mathrm{e}^3} + \frac{C_2}{\mathrm{e}} + \mathrm{e} \right) (m^2+n^2) \\
&+ \frac{\sqrt{1+|1+4\lambda|}}{2\sqrt{|1+4\lambda|}} \left( \frac{C_1}{3\mathrm{e}^3} + 2\mathrm{e} \right) \\
&\leq \frac{1}{2} \left( \frac{C_1}{3\mathrm{e}^3} + \frac{C_2}{\mathrm{e}} + \mathrm{e} \right) (m^2+n^2) + \frac{1}{\sqrt{2}} \left( \frac{C_1}{3\mathrm{e}^3} + 2\mathrm{e} \right)
\end{aligned}$$

where we used that $|1+4\lambda| > 1$ (since $\lambda < -1/2$) and $\sqrt{(1+x)/x} < \sqrt{2}$ whenever $x > 1$. We similarly get

$$\left| P_2 - \frac{\mathrm{e}^{2\sqrt{1+4\lambda}} P_1}{2\sqrt{1+4\lambda}} \right| \leq |P_2| + \left| \frac{P_1}{2\sqrt{1+4\lambda}} \right| \leq \bar{Q}_1 (m^2+n^2) + \tilde{Q}_1$$

where $\bar{Q}_1 = \frac{1}{2}(\frac{C_1}{3\mathrm{e}^3} + \frac{C_2}{\mathrm{e}} + \mathrm{e})$ and $\tilde{Q}_1 = \frac{1}{\sqrt{2}}(\frac{C_1}{3\mathrm{e}^3} + 2\mathrm{e}) + \mathrm{e}$. This completes case (1) of the Lemma.

*Case* $\lambda \geq 0$.
We have $r_1 \geq 0$ and we therefore cannot use 8.3.2.8 and 8.3.2.9. We begin by estimating $|\mathrm{e}^{r_2 t} A_2(t)|$ and $|\mathrm{e}^{r_2 t} A_1(t)|$ by using Lemma 8.3.2 and $r_1 = 1 + \sqrt{1+4\lambda} \geq 0$, $r_2 = -1 - \sqrt{1+4\lambda} \leq -2$. We have

$$\begin{aligned}
|\mathrm{e}^{r_2 t} A_2(t)| &\leq C_2 (m^2+n^2) \mathrm{e}^{r_2 t} \int_1^t \mathrm{e}^{(r_1 - r_2)s} \int_s^\infty \mathrm{e}^{-r_1 u} \mathrm{e}^{-2u} du ds \\
&\leq \frac{C_2}{2} (m^2+n^2) \mathrm{e}^{r_2 t} \int_1^t \mathrm{e}^{(-r_2 - 2)s} ds \qquad (8.3.3.7) \\
&\leq \frac{C_2}{2} (m^2+n^2) t \mathrm{e}^{-2t}
\end{aligned}$$

and

$$\begin{aligned}
|\mathrm{e}^{r_2 t} A_1(t)| &\leq \frac{C_1}{2} \sqrt{m^2 - 4\lambda} \mathrm{e}^{r_2 t} \int_1^t \mathrm{e}^{-(r_2+2)s} ds \leq \frac{C_1}{2} |m| t \mathrm{e}^{-2t} \\
&\leq \frac{C_1}{2} (m^2+n^2) t \mathrm{e}^{-2t}.
\end{aligned}$$

where we used that $m^2 - 4\lambda \geq 0$ if $\phi_m \neq 0$ by 8.3.1.9. Therefore,

$$|\mathrm{e}^{r_2 t} A(t)| \leq |\mathrm{e}^{r_2 t} A_1(t)| + |\mathrm{e}^{r_2 t} A_2(t)| \leq \frac{C_1 + C_2}{2} (m^2+n^2) t \mathrm{e}^{-2t} \qquad (8.3.3.8)$$

for all $t \geq 1$. We set $\bar{C} = \frac{C_1+C_2}{2}$ and $\tilde{C} = 0$.
Since $y(t) \to 0$ when $t \to \infty$ and $r_1 \geq 0$ we have (by 8.3.2.5 and 8.3.3.8) $P_1 = 0$. Thus

$$y(t) = \mathrm{e}^{r_2 t} A(t) + y(1)\mathrm{e}^{-r_2}\mathrm{e}^{r_2 t}$$

by 8.3.3.7 and 8.3.2.2. Hence

$$|y(t)| \leq (\bar{C}_1(m^2+n^2) + \tilde{C}_1)t\mathrm{e}^{-2t}$$

for all $t \geq 1$ by 8.3.3.8. This completes the proof of the Lemma.

□

**Lemma 8.3.4.** *Let* $C_{n,m}(t) = < \phi_n, \phi_m \circ u(t) >$ *where*

$$u(t) = \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}$$

*Then for all* $t \geq 1$, $n, m \in \mathbb{Z}$

$$|C_{n,m}(t)| \leq Kb(\ln t)$$

*where* $b(t)$ *and* $K$ *are as in Lemma 8.3.3.*

*Proof.* We first show that there exists $\theta_1 = \theta_1(t)$, $\theta_2 = \theta_2(t) \in (-\pi, \pi]$ with

$$u(t) = r(\theta_2)a(\ln t)r(\theta_1).$$

In fact, any $g \in G$ can be written as above. This decomposition is known as the *Cartan decomposition* of $G$. We follow the draft of a geometrical proof from [httb] for such a decomposition of the element $u(t)$ above: Consider the action of $G$ as a Möbius Transformation on the hyperbolic plane $\mathbb{H}$. Recall that the subgroup $K$ of $G$ is the stabilizer of the point $i \in \mathbb{H}$ and that $G$ acts simply transitively on the unit tangent bundle $\mathrm{T}^1\mathbb{H}$. The point $i$ can be mapped to $u(t)(i) = \frac{i}{ti+1}$ by first doing a stretching by $d(i, u(t)(i)) = t^2$ along the imaginary axis in direction to $\infty$, which is given by $a(\ln t)i$. The geodesic curve determined by the points $i$ and $u(t)(i)$, whose center $c \in \mathbb{R}$ lies on the real axis, intersects the imaginary axis at an angle which we denote as $\theta_2$. The center $c$ is the solution of

$$|c - i| = |c - u(t)(i)|$$

which is a quadratic equation. Once $c$ is determined, we know the value of $\theta_2$: We may draw the triangle connecting the points $c, 0, i$ to see that $\theta_2 = \cot^{-1} c$. We rotate the point $a(\ln t)(i)$ by the angle $\theta_2$. Then $u(t)(i) = r(\theta_2)a(\ln t)(i)$.
Now, we use that the action of $K$ stabilizes the point $i$. By the simple transitivity of the action of $SL_2(\mathbb{R})$ on $\mathrm{T}^1\mathbb{H}$, we may choose the unique $\theta_1$ such that $u(t)(i, v) = r(\theta_2)a(\ln t)r(\theta_1)(i, v)$ for all $v$ in the unit tangent bundle $\mathrm{T}^1_i\mathbb{H}$. This shows $u(t) = r(\theta_2)a(\ln t)r(\theta_1)$.

Let us now return to the statement of the Lemma. We have

$$\begin{aligned} C_{n,m}(t) &= <\phi_n, \phi_m \circ u(t)> = <\phi_n, \phi_m \circ r(\theta_1) \circ a(\ln t) \circ r(\theta_2)> \\ &= <\phi_n \circ r(-\theta_2), \phi_m \circ r(\theta_1) \circ a(\ln t)> \\ &= \mathrm{e}^{-in\theta_2}\mathrm{e}^{im\theta_1} <\phi_n, \phi_m \circ a(\ln t)> \end{aligned}$$

and so

$$|C_{n,m}(t)| = |B_{n,m}(\ln t)| \leq Kb(\ln t)$$

by Lemma 8.3.3. □

**Theorem 15.** Let $T$ be a non-trivial irreducible unitary representation of $G$ on $H(T)$ and let $\lambda = \lambda(T) \in \mathbb{R}$ be such that $\Omega_T v = \lambda v$ for all $C^2$-vectors $v \in H(T)$. Let $v, w \in K(T,p)$, $p > 0$, $B(t) = <v, w \circ a(t)>$, $C(t) = <v, w \circ u(t)>$ and $r(\lambda) = -1 + \sqrt{1+4\lambda}$, $\lambda \in (-1/4, 0)$. Then for all $|t| \geq 1$

$$|B(t)| \leq C(b_\lambda(|t|))^{\alpha(p)}, \quad |C(t)| \leq D(b_\lambda(\ln|t|))^{\alpha(p)}$$

where

$$\begin{aligned} &(1)\ b_\lambda(t) = \min\{t\mathrm{e}^{-t}, \mathrm{e}^{-t}(1+|1+4\lambda|^{-1/2}\},\ \text{if } \lambda \leq -1/4; \\ &(2)\ b_\lambda(t) = \min\{t\mathrm{e}^{r(\lambda)t}, \mathrm{e}^{r(\lambda)t}|1+4\lambda|^{-1/2}\}\ \text{ if } \lambda \in (-1/4, 0); \\ &(3)\ b_\lambda(t) = t\mathrm{e}^{-2t}\ \text{ if } \lambda \geq 0; \end{aligned}$$

and

$$\alpha(p) = \begin{cases} 1 & p \geq 3 \\ \frac{2p}{2p+1} & 2 \leq p < 3 \\ \frac{2p}{2p+3} & 1 \leq p < 2 \\ \frac{p}{p+3} & 0 < p < 1 \end{cases}$$

with $C, D > 0$ depending only on $v, w$ and $p$ (not on $\lambda$).

*Proof.* We follow the proofs of Theorem 1 in [Rat87] and of Theorem 3 in [Mat13]. Let $T$ be a non-trivial irreducible unitary representation of $G$ in $H(T)$, let $\{\phi_n,\ n \in \mathbb{Z}\}$ denote the basis considered before and let $v, w \in K(T,p)$, $p > 0$. Let $c_n = <v, \phi_n>$, $d_n = <w, \phi_n>$, $n \in \mathbb{Z}$. Then

$$v = \sum_{n\in\mathbb{Z}} c_n\phi_n, \quad \sum_{n\in\mathbb{Z}} d_n\phi_n.$$

*Case 1*: $p \geq 3$.
We have

$$B(t) = <v, w \circ a(t)> = \sum_{n,m\in\mathbb{Z}} c_n d_m <\phi_n, \phi_m \circ a(t)> = \sum_{n,m\in\mathbb{Z}} c_n d_m B_{n,m}(t).$$

Hence

$$|B(t)| \leq \sum_{n,m\in\mathbb{Z}} |c_n d_m B_{n,m}(t)| \leq b(t) \sum_{n,m\in\mathbb{Z}} |c_n||d_m|(\bar{K}(m^2+n^2) + \tilde{K})$$

by Lemma 8.3.3. Set $S = \sum_{n,m\in\mathbb{Z}} |c_n||d_m|(\bar{K}(m^2+n^2)+\tilde{K})$. We claim that $S$ is finite. We have

$$\begin{aligned} S &= \bar{K}\left(\sum_{n,m\in\mathbb{Z}} |c_n||d_m|m^2 + \sum_{n,m\in\mathbb{Z}} |c_n||d_m|n^2\right) + \tilde{K}\sum_{n,m\in\mathbb{Z}} |c_n||d_m| \\ &\leq \bar{K}\left(\sum_{n,m\in\mathbb{Z}} |c_n||d_m|m^2 + \sum_{n,m\in\mathbb{Z}} |c_n||d_m|n^2\right) + \tilde{K}\sum_{n,m\in\mathbb{Z}} |c_n||d_m|(m^2+n^2) \\ &\leq K'\left(\sum_{n,m\in\mathbb{Z}} |c_n||d_m|m^2 + \sum_{n,m\in\mathbb{Z}} |c_n||d_m|n^2\right) \end{aligned}$$

for some $K' > 0$. We proceed by estimating $\sum_{n\in\mathbb{Z}} |c_n|n^2$ as follows. By the Cauchy-Schwarz inequality, it holds that

$$\sum_{n\in\mathbb{Z}} |c_n| = |c_0| + \sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{n^3}(|c_n|n^3) \leq |c_0| + \left(\sum_{n\in\mathbb{Z}\setminus\{0\}} n^6|c_n|^2\right)^{1/2}\left(\sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{n^6}\right)^{1/2}$$

We have that $L_W^3\phi_n = -in^3\phi_n$ for all $n \in \mathbb{Z}$. Thus

$$|< L_W^3 v, \phi_n >| = |< v, L_W^3\phi_n >| = n^3|c_n|$$

and

$$\left(\sum_{n\in\mathbb{Z}\setminus\{0\}} n^6|c_n|^2\right)^{1/2} = \|L_W^3 v\| < \infty$$

by Corollary 10, since $p \geq 3$. Therefore

$$\sum_{n\in\mathbb{Z}} |c_n| \leq \|v\| + \sqrt{2\zeta(6)}\|L_W^3 v\|$$

where $\zeta$ denotes the *Zeta-function*, which for $s > 1$ is defined by $\zeta(s) = \sum_{n\in\mathbb{N}\setminus\{0\}} \frac{1}{n^s}$. This implies

$$\begin{aligned} \sum_{n\in\mathbb{Z}} |c_n|n^2 &\leq \left(\sum_{n\in\mathbb{Z}\setminus\{0\}} n^6|c_n|^2\right)^{1/2}\left(\sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{n^2}\right)^{1/2} \\ &\leq \sqrt{2\zeta(2)}\|L_W^3 v\| \end{aligned}$$

by the Cauchy-Schwarz inequality.
A similar argument to the one above shows that

$$\sum_{m\in\mathbb{Z}} |d_m| \leq \|w\| + \sqrt{2\zeta(6)}\|L_W^3 w\|,$$

$$\sum_{m\in\mathbb{Z}} |d_m| m^2 \leq \sqrt{2\zeta(2)} \|L_W^3 w\|.$$

We conclude that $|B_{n,m}(t)|$ for $v, w \in K(T,3)$ is bounded as in the Theorem with $\alpha(p) = 1$.
For the bound on $|C(t)|$, $v, w \in K(T,3)$ we have

$$C(t) = < v, w \circ u(t) > = \sum_{n,m\in\mathbb{Z}} c_n d_m < \phi_n, \phi_m \circ u(t) > = \sum_{n,m\in\mathbb{Z}} c_n d_m C_{n,m}(t).$$

Hence

$$|C(t)| \leq \sum_{n,m\in\mathbb{Z}} |c_n d_m C_{n,m}(t)| \leq b(\ln t) \sum_{n,m\in\mathbb{Z}} |c_n||d_m|(\bar{K}(m^2+n^2) + \tilde{K})$$

by Lemma 8.3.4. The final estimate of $\sum_{n,m\in\mathbb{Z}} |c_n||d_m|(\bar{K}(m^2+n^2)+\tilde{K})$ is then the same as for the function $|B(t)|$. We may thus omit the computations on $|C(t)|$ in the rest of this proof.

*Case 2*: $2 \leq p < 3$.
We let $0 < \delta < 1$ be chosen later. Denote

$$\bar{v_t} = \sum_{|n|\leq (b(t))^{-\delta}} c_n \phi_n, \quad \hat{v_t} = \sum_{|n|>(b(t))^{-\delta}} c_n \phi_n$$

It follows from Corollarly 11 that

$$\|\hat{v_t}\| = \left( \sum_{|n|>(b(t))^{-\delta}} |c_n|^2 \right)^{1/2} \leq D(b(t))^{\delta p} \tag{8.3.4.1}$$

Remark that

$$B(t) = < v, w \circ a(t) > = < \bar{v_t}, w \circ a(t)) > + < \hat{v_t}, w \circ a(t) > =: \bar{V} + \hat{V}$$

and

$$\bar{V} = < \bar{v_t}, \bar{w_t} \circ a(t) > + < \bar{v_t}, \hat{w_t} \circ a(t) > =: \bar{W} + \hat{W}.$$

We have

$$|\hat{V}| \leq \|\hat{v_t}\| \cdot \|w\| \leq D_1 \|w\| (b(t))^{\delta p} \tag{8.3.4.2}$$

by 8.3.4.1. Similarly,

$$|\hat{W}| \leq \|v\| \cdot \|\hat{w_t}\| \leq D_2 \|v\| (b(t))^{\delta p} \tag{8.3.4.3}$$

Moreover

$$\begin{aligned} |\bar{W}| &\leq \sum_{|m|,|n|\leq (b(t))^{-\delta}} |c_n||d_m|| < \phi_n, \phi_m \circ a(t) > | \\ &= \sum_{|m|,|n|\leq (b(t))^{-\delta}} |c_n||d_m||B_{n,m}(t)| \\ &\leq b(t) K' \left( \sum_{|m|,|n|\leq (b(t))^{-\delta}} |c_n||d_n| \cdot m^2 + \sum_{|m|,|n|\leq (b(t))^{-\delta}} |c_n||d_n| \cdot n^2 \right) \end{aligned} \tag{8.3.4.4}$$

for some $K' > 0$ by Lemma 8.3.3. Since $p \geq 2$, we have by Corollary 10 and the Cauchy-Schwarz inequality

$$\begin{aligned}\sum_{|m|\leq (b(t))^{-\delta}} |d_m|m^2 &= \sum_{m\in\mathbb{Z}} \mathbb{1}_{|m|\leq (b(t))^{-\delta}}(m)|d_m|m^2 \\ &\leq (b(t))^{-\delta/2}\left(\sum_{m\in\mathbb{Z}} |d_m|^2 m^4\right)^{1/2} \\ &= (b(t))^{-\delta/2}\|L_W^2 w\| \qquad (8.3.4.5)\\ \sum_{|n|\leq (b(t))^{-\delta}} |c_n| &\leq \left(\sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{n^4}\right)^{1/2}\left(\sum_{n\in\mathbb{Z}\setminus\{0\}} n^4|c_n|^2\right)^{1/2} + |c_0| \\ &= \sqrt{2\zeta(4)}\|L_W^2 v\| + |c_0|.\end{aligned}$$

Therefore

$$\begin{aligned}|\bar{W}| &\leq b(t)K'(b(t))^{-\delta/2}\|L_W^2 w\|(\sqrt{2\zeta(4)}\|L_W^2 v\| + |c_0|) \\ &\quad + b(t)K'(b(t))^{-\delta/2}\|L_W^2 v\|(\sqrt{2\zeta(4)}\|L_W^2 w\| + |c_0|) \\ &\leq Q_2(b(t))^{1-\delta/2}\end{aligned}$$

by 8.3.4.4 and

$$\begin{aligned}|B(t)| &\leq |\bar{W}| + |\hat{W}| + |\hat{V}| \\ &\leq Q_2(b(t))^{1-\delta/2} + D_2\|v\|(b(t))^{\delta p} + D_1\|w\|(b(t))^{\delta p} \\ &\leq Q_3[(b(t))^{\delta p} + (b(t))^{1-\delta/2}]\end{aligned}$$

by 8.3.4.2 and 8.3.4.3. Observe that $Q_2, Q_3 > 0$ depend only on $v, w$.
We now see that for $\delta = 2p/(2p+1)$

$$|B(t)| \leq 2Q_3(b(t))^{2p/(2p+1)}$$

and so $\alpha(p) = 2p/(2p+1)$ for $2 \leq p < 3$.

*Case 3*: $1 \leq p < 2$.
We proceed as in the previous case, adjusting the estimates in 8.3.4.5 as follows. Since $p \geq 1$, we have by Corollary 10 and by the Cauchy-Schwarz inequality

$$\begin{aligned}\sum_{|m|\leq (b(t))^{-\delta}} |d_m|m^2 &\leq \left(\sum_{|m|\leq (b(t))^{-\delta}} m^2\right)^{1/2}\left(\sum_{m\in\mathbb{Z}} |d_m|^2 m^2\right)^{1/2} \\ &= \left(\frac{(b(t))^{-\delta}((b(t))^{-\delta}+1)(2(b(t))^{-\delta}+1)}{3}\right)^{1/2}\left(\sum_{m\in\mathbb{Z}} |d_m|^2 m^2\right)^{1/2} \\ &\leq L_1(b(t))^{-3\delta/2}\|L_W w\|.\end{aligned}$$

Similarly

$$\sum_{|n|\leq(b(t))^{-\delta}} |c_n| \leq \left(\sum_{n\in\mathbb{Z}\setminus\{0\}} \frac{1}{n^2}\right)^{1/2} \left(\sum_{n\in\mathbb{Z}\setminus\{0\}} n^2|c_n|^2\right)^{1/2} + |c_0|$$
$$= \sqrt{\zeta(2)}\|L_W v\| + |c_0|.$$

Thus

$$|\bar{W}| \leq L_2(b(t))^{1-3/2\delta}$$

and

$$|B(t)| \leq L_3[(b(t))^{\delta p} + (b(t))^{1-3/2\delta}].$$

For $\delta = 2p/(2p+1)$ we have $\delta p = 1 - 3/2\delta$ and so

$$|B(t)| \leq L_4(b(t))^{\alpha(p)}.$$

for $\alpha(p) = 2p/(2p+1)$ if $1 \leq p < 2$.

*Case 4*: $0 < p < 1$. We use Lemma 8.2.2 for $p$ to get

$$\|\hat{v_t}\| \leq (b(t))^{\delta p}$$

and therefore

$$|\hat{V}| \leq \|\hat{v_t}\|\|w\| \leq D_1(b(t))^{\delta p}$$
$$|\hat{W}| \leq \|v\|\|\hat{w_t}\| \leq D_2(b(t))^{\delta p}$$

For $\bar{W}$ we get by the Cauchy-Schwarz inequality

$$\sum_{|m|\leq(b(t))^{-\delta}} |d_m|m^2 \leq \left(\sum_{|m|\leq(b(t))^{-\delta}} m^4\right)^{1/2} \left(\sum_{m\in\mathbb{Z}} |d_m|^2\right)^{1/2}$$
$$\leq (b(t))^{-5\delta/2} \left(\sum_{m\in\mathbb{Z}} |d_m|^2\right)^{1/2}$$
$$\leq P_1(b(t))^{-5\delta/2}\|w\|$$

$$\sum_{|n|\leq(b(t))^{-\delta}} |c_n| \leq (b(t))^{-\delta/2} \left(\sum_{n\in\mathbb{Z}\setminus\{0\}} |c_n|^2\right)^{1/2} = (b(t))^{-\delta/2}\|v\|.$$

Therefore, by 8.3.4.4

$$\bar{W} \leq P_2(b(t))^{1-3\delta}$$

and

$$|B(t)| \leq |\bar{W}| + |\hat{W}| + |\hat{V}| \leq P_3[(b(t))^{\delta p} + (b(t))^{1-3\delta}].$$

Setting $\delta = 1/(p+3)$ we get that $\alpha(p) = p/(p+3)$.
We see that the proof of the desired statements in the Theorem is complete.

□

## 8.4 Rate of Mixing for the Geodesic and the Horocycle Flows on $L^2(X, \mu)$

The hyperbolic plane $\mathbb{H}$ has been shown (see Lemma 1.2.1) to be diffeomorphic to $G/K$ via the map $\phi : G/K \to \mathbb{H}$, defined by

$$\phi(gK) = g(i)$$

$g \in G$. The coordinate system and the hyperbolic Riemannian metric on $\mathbb{H}$ are endowed on $G/K$ by the map $\phi^{-1}$. Then, in this metric, the surface $S = \Gamma \setminus (G/K) \cong \Gamma \setminus \mathbb{H}$ is of constant negative curvature, since the curvature of $\mathbb{H}$ is equal to -1 at any point $z \in \mathbb{H}$.
Define

$$H' = \{f \in L^2(X, \mu) : \ f \text{ is constant on orbits of } K\}$$

where as before, we write $X = \Gamma \setminus G$. Notice that a function on $H'$ is in fact a function on $S$. Hence, $H' \subseteq L^2(X, \mu)$ can be identified with $L^2(S)$. The Laplace operator $\Delta$ on $L^2(S)$, defined in coordinates $(x, y)$ is of the form

$$\Delta = y^2 \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right).$$

It turns out that the Laplace operator plays a crucial role in our upcoming computations. As is explained in Chapter 10 of [Lan85], the Casimir operator $\Omega$ associated to the regular representation of $G$ on $L^2(X, \mu)$, is, in terms of coordinates given by the Iwasawa decomposition of $G$, given by

$$\Omega = 4y^2 \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) - 4y \frac{\partial^2}{\partial x \, \partial \theta}.$$

This tells us that, when restricted to $C^2$-vectors in $H'$, the operator $\Omega$ coincides, up to a constant multiple, with the Laplace operator $\Delta$ on $L^2(S)$.
Denote by $\Lambda(\Omega)$ the spectrum of the operator $\Omega$ on $L^2(X, \mu)$ and by $\Lambda(\Delta)$ the spectrum of $\Delta$ on $L^2(S)$. Write

$$\begin{aligned} A(\Gamma) &= \Lambda(\Delta) \cap (-1/4, 0) \\ \beta(\Gamma) &= \sup A(\Gamma), \ A(\Gamma) \neq \emptyset \\ \sigma(\Gamma) &= r(\beta(\Gamma)) = -1 + \sqrt{1 + 4\beta(\Gamma)}. \end{aligned}$$

It is a fact (see Chapter 10 in [Lan85]) that

$$\lambda \in (-1/4, 0) \cap \Lambda(\Omega) \iff \lambda \in (-1/4, 0) \cap \Lambda(\Delta)$$

and if $\Gamma \leq G$ is a lattice in $G$, then $A(\Gamma)$ is finte.
We are now in position to state the final result, proving an exponential decay for mixing of the flows in consideration. For convenience, we may continue the notation

$$v \circ g = U_g v$$

$v \in L^2(X, \mu)$, $g \in G$, and where $U$ is the regular representation of $G$ on $L^2(X, \mu)$.

**Theorem 16.** Let $\Gamma$ be a lattice in $G$, $X = \Gamma \setminus G$ and let $U$ be the regular representation of $G$ on $L^2(X,\mu)$. Let $v, w \in K(U,p)$, $p \geq 1$, $< w, 1 > = 0$ and $B(t) = < v, w \circ a(t) >$, $C(t) = < v, w \circ u(t) >$. Then there exists $t_0 = t_0(\Gamma) > 0$ such that for all $|t| \geq t_0$ and some constants $C, D > 0$

$$|B(t)| \leq C(b(|t|))^{\alpha(p))}, \quad |C(t)| \leq D(b(\ln|t|))^{\alpha(p)}$$

where

1. $b(t) = \mathrm{e}^{\sigma(\Gamma)t}$ if $A(\Gamma) \neq \emptyset$;
2. $b(t) = \mathrm{e}^{-t}$ if $A(\Gamma) = \emptyset$, $\sup(\Lambda(\Omega) \cap (-\infty, -1/4)) < -1/4$ and $-1/4$ is not an eigenvalue of $\Omega$;
3. $b(t) = t\mathrm{e}^{-t}$ if $A(\Gamma) = \emptyset$ and either $\sup(\Lambda(\Omega) \cap (-\infty, -1/4)) = -1/4$ or $-1/4$ is an eigenvalue of $\Omega$;

and $\alpha(p)$ is as in Theorem 15.

*Proof.* We are going to prove the estimates for the geodesic flow only, as the ones for the horocycle flow are proven similarly.
The regular representation $U$ of $G$ leaves no non-zero vector invariant if restricted to the space $L^2_0(X,\mu) \subseteq L^2(X,\mu)$ consisting of functions with zero mean, because the action of $G$ on $X$ is mixing. Thus, $U$ restricted to $L^2_0(X,\mu)$ is unitary equivalent to a direct integral $\pi = \int_Y \pi_y d\nu(y)$ as described at the beginning of this Chapter. Thus we have, up to a unitary isomorphism

$$U = \int_Y \pi_y d\nu(y).$$

Let $v, w \in K(U,p)$, $p \geq 1$. We have the decomposition

$$v = \int_Y v_y d\nu(y), \quad w = \int_Y w_y d\nu(y)$$

and so

$$\begin{aligned} |B(t)| = \ | < v, w \circ a(t) > | &= \left| \int_Y \ < v_y, \pi_y(a(t)) w_y > \ d\nu(y) \right| \\ &\leq \int_Y \ | < v_y, \pi_y(a(t)) w_y > | \ d\nu(y) \end{aligned}$$

We now want to estimate the term inside the integral by using Theorem 15.

*Case 1*:
Assume $A(\Gamma) \neq \emptyset$. When restricted to the interval $(-1/4, 0)$, the spectra $\Lambda(\Delta)$ and $\Lambda(\Omega)$ coincide. Thus, $A(\Gamma) \neq \emptyset$ implies that there exists $y \in Y$ such that the Casimir operator $\Omega_y$ associated to the irreducible representation $\pi_y$ is of the form

$$\Omega_y = \lambda_y I$$

for some $\lambda_y \in (-1/4, 0)$, by Theorem 14. It follows that

$$| < v_y, \pi_y(a(t))w_y > | \leq C_{v_y,w_y,p}(b_{\lambda_y}(|t|))^{\alpha(p)}$$

where $b_{\lambda_y}(t)$ is given by case (2) of Theorem 15. As $\Gamma$ is a lattice in $G$, we know that $A(\Gamma)$ is finite. Thus, there are only finitely many $y \in Y$ with $\Omega_y = \lambda_y I$, $\lambda_y \in (-1/4, 0)$ and we shall choose $t_0^1(\Gamma)$ in the statement of the Theorem such that

$$t_0^1(\Gamma) = \max_{y:\ \lambda_y \in (-\frac{1}{4}, 0)} \left(1 + |1 + 4\lambda_y|^{-1/2}\right).$$

Observe that the bound in case (2) of Theorem 15 is the largest (and thus the worst) possible of all cases and it is applicable to almost all terms inside the integral by Lemma 8.2.1. Hence, since the set $\{y \in Y\}$ on which this estimate does not hold is of $\nu$-measure zero, we can estimate the integral as follows. Set $\lambda = \max_y\{\lambda_y, \lambda_y \in (-1/4, 0) \cap \Lambda(\Omega)\}$ and denote by $b_\lambda(t)$ the corresponding function from case (2) of Theorem 15. Then for all $|t| \geq t_0^1(\Gamma)$

$$\begin{aligned} |B(t)| &\leq (b_\lambda(|t|))^{\alpha(p)} \int_Y C_{v_y,w_y,p}\ d\nu(y) \\ &= t_0(\Gamma)(\mathrm{e}^{\sigma(\Gamma)|t|})^{\alpha(p)} \int_Y C_{v_y,w_y,p}\ d\nu(y) \end{aligned}$$

We are now left to find a uniform estimate $C_{v,w,p}$ for the constant $\int_Y C_{v_y,w_y,p}\, d\nu(y)$. This is done by using the proof of Theorem 15 and Lemma 8.2.1. We recall the main ideas in the case where $p \geq 3$, which lead us to the desired bound. The other cases are proven analogously.
Lemma 8.2.1 (2) states that if $v, w \in K(U, 3)$, then $v(y) = v_y, w(y) = w_y \in K(\pi_y, 3)$, for almost every $y \in Y$. For these $y$, we have

$$|B_y(t)| := | < v_y, \pi_y(a(t))w_y > | \leq t_0(\Gamma)(\mathrm{e}^{\sigma(\Gamma)|t|})C_{v_y,w_y,3}.$$

Let $c_n^y = < v_y, \phi_n(y) >$, $d_n^y = < w_y, \phi_n(y) >$, where $\phi_n(y)$, $n \in \mathbb{Z}$ is the orthonormal basis of $H(y)$, consisting of $K$-eigenvectors with respect to the representation $\pi_y$. Then

$$v_y = \sum_{n\in\mathbb{Z}} c_n^y \phi_n(y), \quad w_y = \sum_{n\in\mathbb{Z}} d_n^y \phi_n(y).$$

We have

$$\begin{aligned} |B_y(t)| &= \| \sum_{n,m\in\mathbb{Z}} c_n^y d_m^y B_{n,m}^y(t) \| \\ &\leq b_\lambda^y(t) \sum_{n,m\in\mathbb{Z}} |c_n^y||d_m^y|(\bar{K}(m^2 + n^2) + \tilde{K}) \end{aligned}$$

where $b^y_\lambda(t)$, $\bar{K}$, $\tilde{K}$, are determined by Lemma 8.3.3 and where $B^y_{n,m}(t) =< \phi_n(y), \phi_m(y) \circ a(t) >$. We established in in the proof of Theorem 15 that

$$\sum_{n\in\mathbb{Z}} |c^y_n| \leq \|v_y\| + \sqrt{2\zeta(6)}\|(L^y_W)^3 v_y\|$$
$$\sum_{m\in\mathbb{Z}} |d^y_m| \leq \|w_y\| + \sqrt{2\zeta(6)}\|(L^y_W)^3 w_y\|$$

and

$$\sum_{n\in\mathbb{Z}} |c^y_n| n^2 \leq \sqrt{2\zeta(2)}\|(L^y_W)^3 v_y\|$$
$$\sum_{m\in\mathbb{Z}} |d^y_m| m^2 \leq \sqrt{2\zeta(2)}\|(L^y_W)^3 w_y\|.$$

This shows that

$$\begin{aligned} C_{v_y,w_y,3} &= \sqrt{2\zeta(2)}\bar{K}\|(L^y_W)^3 v_y\|(\|w_y\| + \sqrt{2\zeta(6)}\|(L^y_W)^3 w_y\|) \\ &+ \sqrt{2\zeta(2)}\bar{K}(\|v_y\| + \sqrt{2\zeta(6)}\|(L^y_W)^3 v_y\|)\|(L^y_W)^3 w_y\| \\ &+ \bar{K}(\|v_y\| + \sqrt{2\zeta(6)}\|(L^y_W)^3 v_y\|)(\|w_y\| + \sqrt{2\zeta(6)}\|(L^y_W)^3 w_y\|). \end{aligned}$$

The crucial point is that

$$C_{v,w,3} = \int_Y C_{v_y,w_y,3} d\nu(y)$$

now depends on a finite linear combination of $\|v\|$, $\|w\|$, $\|L^3_W v\|$, $\|L^3_W w\|$ by Lemma 8.2.1 (3).

*Case 2*:
Assume $A(\Gamma) = \emptyset$, $\sup(\Lambda(\Omega) \cap (-\infty, -1/4)) < -1/4$ and $-1/4$ is not an eigenvalue of $\Omega$. The condition $A(\Gamma) = \emptyset$ implies that there is no $y \in Y$ such that $\Omega_y = \lambda I$ for $\lambda \in (-1/4, 0)$. Therefore, almost every $| < v_y, \pi_y(a(t))w_y > |$, $y \in Y$ is bounded using the function from case (1) or case (3) of Theorem 15, of which (3) is the sharper bound. From the assumptions we know

$$\lambda_0 := \sup_{\lambda\in\Lambda(\Omega)} (1 + |1 + 4\lambda|^{-1/2}) < \infty$$

and we may choose $t_0^2(\Gamma) > \lambda_0$. For similar reasons as in the previous case, we can bound $|B(t)|$, $|t| \geq t_0^2(\Gamma)$ by using the function $b(t) = \mathrm{e}^{-t}$, and the coefficients $C_{v,w,p}$ arising from the above discussion. This proves case (2) of the Theorem.

*Case 3*:
Assume $A(\Gamma) = \emptyset$, and either $\sup(\Lambda(\Omega) \cap (-\infty, -1/4)) = -1/4$ or $-1/4$ is an eigenvalue of $\Omega$. Then again, almost every $| < v_y, \pi_y(a(t))w_y > |$, $y \in Y$ is

bounded using the functions from case (1) or case (3) of Theorem 15. Assume first that $\Lambda(\Omega) \cap (-\infty, -1/4)) < -1/4$ so that $-1/4$ is an eigenvalue of $\Omega$. Set $t_0(\Gamma)^3$ as in case 2. There exists $y \in Y$ such that

$$| < v_y, \pi_y(a(t))w_y > | \leq C_{v_y,w_y,p}\ t\mathrm{e}^{-t}$$

for $|t| \geq 1$. It follows that

$$|B(t)| \leq C_{v,w,p}\ t\mathrm{e}^{-t}$$

for $|t| \geq t_0(\Gamma)$. If $\sup(\Lambda(\Omega) \cap (-\infty, -1/4)) = -1/4$ then

$$\sup_t\{\lambda(\pi_y) < t,\ y \in Y\} < L$$

for some constant $L < \infty$. Similarly as above,

$$|B(t)| \leq C_{v,w,p}\ t\mathrm{e}^{-t}$$

for all $|t| > L$.

Finally, set $t_0(\Gamma) = \max\{t_0^1(\Gamma), t_0^2(\Gamma)\}$ to complete the proof.

□

# Bibliography


[Rat87] Marina Ratner. "The rate of mixing for geodesic and horocycle flows". In: *Ergodic Theory and Dynamical Systems* 7 (1987), pp. 267–288. URL: https://api.semanticscholar.org/CorpusID:122050247.

[EFS09] Manfred Einsiedler, Lior Fishman, and Uri Shapira. "Diophantine Approximations on Fractals". In: *Geometric and Functional Analysis* 21 (2009), pp. 14–35. URL: https://api.semanticscholar.org/CorpusID:18360593.

[Dan85] S. G. Dani. "Divergent trajectories of flows on homogeneous spaces and Diophantine approximation." In: *Journal für die reine und angewandte Mathematik (Crelles Journal)* 1985 (1985), pp. 55–89. URL: https://api.semanticscholar.org/CorpusID:117854388.

[Dyn23] Simons Semester on Dynamics. *Reynold Fregoli (Universität Zürich), lecture 2.* 9.6.2024. 2023. URL: https://youtu.be/lbOcNMEC5UA?si=wG-3GSX91VUutbBa.

[GGN14] Anish Ghosh, Alexander Gorodnik, and Amos Nevo. "Diophantine approximation exponents on homogeneous varieties". In: *arXiv: Number Theory* (2014). URL: https://api.semanticscholar.org/CorpusID:119635496.

[Gor18] Alexander Gorodnik. "Higher Order Correlations for Group Actions". In: *Infosys Science Foundation Series* (2018). URL: https://api.semanticscholar.org/CorpusID:263792649.

[GN13] Alexander Gorodnik and Amos Nevo. "Quantitative ergodic theorems and their number-theoretic applications". In: *Bulletin of the American Mathematical Society* 52 (2013), pp. 65–113. URL: https://api.semanticscholar.org/CorpusID:9735163.

[Mar08] G. A. Margulis. "HOMOGENEOUS DYNAMICS AND NUMBER THEORY". In: 2008. URL: https://api.semanticscholar.org/CorpusID:216153144.

[EW10] M. Einsiedler and T. Ward. *Ergodic Theory: with a view towards Number Theory.* Graduate Texts in Mathematics. Springer London, 2010. ISBN: 9780857290212. URL: https://books.google.ch/books?id=PiDET2fS7H4C.

[Isa17] Alexander Isaev. "Conformal Maps (Continued). Möbius Transformations". In: 2017. URL: https://api.semanticscholar.org/CorpusID:125460787.

[htta] (https://math.stackexchange.com/users/833796/red-whisker). *Generator of $SL(2, R)$*. Mathematics Stack Exchange. eprint: https://math.stackexchange.com/q/3953493. URL: https://math.stackexchange.com/q/3953493.

[BM00] M.B. Bekka and M. Mayer. *Ergodic Theory and Topological Dynamics of Group Actions on Homogeneous Spaces*. Lecture note series / London mathematical society. Cambridge University Press, 2000. ISBN: 9780521660303. URL: https://books.google.ch/books?id=T5NgCaa1j-4C.

[Lab22] Daniel Labardini-Fragoso Labardini's Math. *Hyperbolic Geometry 5.6. Fuchsian groups: the Dirichlet polygon*. 10.3.2024. 2022. URL: https://youtu.be/OU3ah0sKGEo?si=rSroM3MEKFfAlvan.

[FK12] H.M. Farkas and I. Kra. *Riemann Surfaces*. Graduate Texts in Mathematics. Springer New York, 2012. ISBN: 9781468499308. URL: https://books.google.com.na/books?id=6pXuBwAAQBAJ.

[Mau57] F. I. Mautner. "GEODESIC FLOWS ON SYMMETRIC RIEMANN SPACES". In: *Annals of Mathematics* 65 (1957), p. 416. URL: https://api.semanticscholar.org/CorpusID:123754023.

[Moo80] Calvin Cooper Moore. "The Mautner Phenomenon for General Unitary Representations". In: *Pacific Journal of Mathematics* 86 (1980), pp. 155–169. URL: https://api.semanticscholar.org/CorpusID:54923183.

[HM79] Roger E. Howe and Calvin C. Moore. "Asymptotic properties of unitary representations". In: *Journal of Functional Analysis* 32.1 (1979), pp. 72–96. ISSN: 0022-1236. DOI: https://doi.org/10.1016/0022-1236(79)90078-8. URL: https://www.sciencedirect.com/science/article/pii/0022123679900788.

[Lan85] S. Lang. *SL2(R)*. Graduate texts in mathematics. Springer, 1985. URL: https://books.google.ch/books?id=GUxivQEACAAJ.

[Neu49] John von Neumann. "On Rings of Operators. Reduction Theory". In: *Annals of Mathematics* 50 (1949), p. 401. URL: https://api.semanticscholar.org/CorpusID:124439084.

[Mac76] George W. Mackey. "The theory of unitary group representations". In: 1976. URL: https://api.semanticscholar.org/CorpusID:118835363.

[EW17] M. Einsiedler and T. Ward. *Functional Analysis, Spectral Theory, and Applications*. Graduate Texts in Mathematics. Springer International Publishing, 2017. ISBN: 9783319585406. URL: https://books.google.ch/books?id=ROs_DwAAQBAJ.

[Mat13] Carlos Matheus. "Some quantitative versions of Ratner's mixing estimates". In: *Bulletin of the Brazilian Mathematical Society, New Series* 44 (2013), pp. 469–488. URL: https://api.semanticscholar.org/CorpusID:54613769.

[httb] anon (https://math.stackexchange.com/users/11763/anon). *KAK decomposition of $SL(2,\mathbb{R})$*. Mathematics Stack Exchange. eprint: https://math.stackexchange.com/q/3721760. URL: https://math.stackexchange.com/q/3721760.