University of Helsinki
Faculty of Science
Master's Programme in Mathematics and Statistics (MAST)



Master's thesis

Patterson-Sullivan measures and the Prime Geodesic Theorem

Alexandre Minetto

Supervisors:
Tuomas Sahlsten
Gaétan Leclerc

November, 2025

Title: Patterson-Sullivan measures and the Prime Geodesic Theorem
Author: Alexandre Minetto
Month and year: November, 2025
Page count: 114 pp.

Abstract:

Following in the footsteps of work initiated by Samuel J. Patterson and Dennis Sullivan, for at least half a century, mathematicians have investigated properties of conformal densities and related measures, such as Patterson-Sullivan measures. The purpose of this project is to present some of the classical developments in hyperbolic dynamics involving such measures. The guiding star of the narrative is the problem of counting geodesics of a given length on a geometrically finite hyperbolic surface. After brief overview chapters on hyperbolic geometry and the length spectrum of a hyperbolic surface, the theory of Patterson-Sullivan measures is introduced. Along the way towards the Prime Geodesic Theorem, a celebrated result showing a striking relationship between geometry, analytic number theory, and complex analysis, two other interconnected topics are discussed. They concern ergodicity and the relationship between Patterson-Sullivan measures and Hausdorff measures. A proof of the ergodicity of the geodesic flow in the geometrically finite setting is included, as well as a proof that the Hausdorff dimension of the limit set $\Lambda(\Gamma)$ is equal to the critical exponent $\delta(\Gamma)$ of the Fuchsian group Γ characterizing a convex cocompact hyperbolic surface. In the last chapter, a sketch of a proof of the Prime Geodesic Theorem based on a zeta function argument is provided, together with a bound on the counting problem that explicitly shows how conformal densities are intimately linked to geometric invariants of hyperbolic surfaces.

Keywords: hyperbolic geometry, hyperbolic dynamics, Patterson-Sullivan theory, ergodicity

Acknowledgments

I would like to thank Tuomas Sahlsten and Gaétan Leclerc for the suggestion of an interesting research topic (and, more generally, research area), support, and supervision. Friday afternoon sessions with Gaétan were a stimulating and comfortable place to ask questions, even silly ones.

I am also grateful to Ville Mustonen and Teemu Kuosmanen for years of collaboration and for hands-on experience on how abstract mathematical concepts, such as ergodicity, have all-too-real implications.

Finally, I thank my parents, José and Marli, for a shared life and everything else that sometimes remains unspoken.

Contents

1 Introduction **5**

2 Brief overview of basic hyperbolic geometry **10**
2.1 Two models of the hyperbolic plane 10
2.2 Stereographic projection 11
2.3 Möbius transformations 11
2.4 Lengths, geodesics, and isometries 12
2.5 Classification of Möbius transformations 13
2.6 Conjugacy of Möbius transformations 13
2.7 Displacement length of a hyperbolic transformation 14
2.8 Fuchsian groups 14
2.9 Hyperbolic surfaces 15
2.10 Limit set of a Fuchsian group 17
2.11 Geometrically finite groups 18
2.12 Cylinders, funnels, and cusps 18
2.13 Nielsen regions 19

3 Closed geodesics and the length spectrum **22**
3.1 Closed geodesics 22
3.2 A general bound on the length counting function 23

4 Patterson-Sullivan measures on the limit set **27**
4.1 Critical exponent 27
4.2 Construction preliminaries 29
4.3 Changing the base point 32
4.4 Changing the exponent 35
4.5 Relationship between limit measures 37
4.6 Patterson-Sullivan measure distortion 39
4.7 Usefulness of the Patterson-Sullivan measure 42
4.8 Patterson-Sullivan measures are atomless 45
4.9 Remarks on conformal densities 53

5 Patterson-Sullivan measures and the ergodicity of the geodesic flow **55**
5.1 The geodesic flow and measure m_{μ_z} 55
5.2 The geodesic flow is conservative 60
5.3 Rudiments of ergodic theory 62
5.3.1 Ergodic theory: discrete-time results 63

5.3.2 Ergodic theory: continuous-time results 67
5.4 Conservation and ergodicity 71
5.5 Uniqueness of Patterson-Sullivan measures and ergodicity 78

6 Hausdorff Measure of the Limit Set **81**
6.1 Hausdorff measure 81
6.2 Sullivan's Shadow Lemma 83
6.3 Hausdorff dimension and convex cocompact groups 87

7 The Prime Geodesic Theorem **93**
7.1 A second bound on the lattice-point counting function 93
7.2 Selberg zeta functions 95
7.3 The zeros and poles of Selberg zeta functions 97
7.4 The Prime Geodesic Theorem 99

8 Some recent developments **107**

Bibliography **110**

Chapter 1

Introduction

The Prime Number Theorem is arguably one of the most celebrated results in mathematics. Although prime numbers have been investigated since ancient times, it was only at the end of the 19th century that a satisfactory answer to the question of their placement along the positive integers was found. The theorem establishes the asymptotic distribution of prime numbers along $\mathbb{Z}_{>0}$. In particular, it determines the rate at which these numbers become rarer as the range of positive integers increases. Its statement is as follows.

Prime Number Theorem (PNT). Let

$$\pi(x) \equiv \#\{p \in \mathbb{Z}_{>0} : p \text{ is prime and } p \leq x\}$$

denote the prime counting function. Then,

$$\lim_{x\to\infty} \frac{\pi(x)}{\frac{x}{\log x}} = 1.$$

In asymptotic notation, the theorem can be restated as

$$\pi(x) \sim \frac{x}{\log x}, \text{ as } x \to \infty.$$

Building on work by (among others) Chebyshev and Riemann, the PNT was originally independently proved by Hadamard and Poussin.[1] More proofs have since appeared. For historical reasons, they have been classified according to their reliance ("non-elementary") or not ("elementary") on methods from complex analysis. The shortest non-elementary proof is due to Newman and it was only discovered in 1980.[2] The first elementary proof is due to work by Selberg [67] and Erdős [21].

Selberg is historically one of the key mathematicians linking this old-school theorem to recent developments in geometry and dynamics. In particular, Selberg pioneered the study of spectral theory of hyperbolic surfaces in the 1950s and some of the methods employed by Selberg have proved fruitful in connecting seemingly different branches of mathematics.

[1] See Ingham ([31], pp. 4-5) for a brief historical appraisal.

[2] For the original proof, see Newman [52]. For additional material, refer to Bak and Newman's book ([3], p. 285) and Zagier [80].

The present project is inspired by Selberg's work and a stunning parallel to the PNT. It is sometimes grouped together with other variants under the umbrella term Prime Geodesic Theorem, or Prime Orbit Theorem. The reason for this label is that statements of the kind, although involving geometrical concepts, reveal asymptotic relationships analogous to that of the PNT.

The thesis focuses on the following formulation of the theorem.

Prime Geodesic Theorem (PGT). For X a geometrically finite non-elementary[3] hyperbolic surface with critical exponent δ,

$$\pi_X(t) \sim \frac{e^{\delta t}}{\delta t} \text{ as } t \to \infty.$$

The function $\pi_X : \mathbb{R}_{>0} \to \mathbb{Z}_{\geq 0}$, formally defined in Chapter 3, is akin to the counting function π specified above for the PNT. However, instead of primes, it counts prime (closed) geodesics (also to be defined) on X of hyperbolic length no greater than a given real number, the input.

Notice that by making the substitution $x(t) = e^{\delta t}$, we obtain the usual asymptotic form that holds for the PNT. That is,

$$\pi_X(t) \sim \frac{x(t)}{\log x(t)} \text{ as } t \to \infty.$$

The first such type of theorem obtained for the geodesic flow was due to work by Huber [30] in the context of compact Riemann surfaces. This result was later improved by Hejhal [27] and Randol [58], still under the compactness requirement. The connection with spectral theory is explicit since the error term in their formulation of the theorem is controlled by an exponent related to the low eigenvalues of the spectrum of the Laplace Beltrami operator on such manifolds. We will later briefly sketch why this situation repeats itself for δ and the more general version of the PGT stated above (see Proposition 4.7.2). For non-compact Riemann surfaces of finite volume, Sarnak [64] obtained similar asymptotic relationships.

The proofs of these preliminary results were based on the Selberg trace formula[4] or the Selberg zeta function (appearing in Chapter 7). The Selberg zeta function for compact or finite-volume hyperbolic surfaces is a meromorphic function constructed from the lengths of (closed) geodesics. This construct encodes information about the length spectrum of a given hyperbolic surface, a geometrical invariant, in much the same way that the Riemann zeta function encodes information about the distribution of prime numbers.

In the case of non-compact hyperbolic surfaces of infinite volume (the general setting of the above PGT), even under the assumption of geometric finiteness (to be discussed) matters are more complicated, and it becomes significantly harder to prove asymptotic relationships as the one described. More so than in the other cases, the relevant spectral quantities are the resonances (i.e., the poles of the meromorphic continuation of the resolvent of the Laplacian) and poor knowledge of the scattering spectrum becomes an obstacle.

[3]This usage of "non-elementary" has a meaning that will be made precise in the next chapter.

[4]We will not pursue the topic of the Selberg trace formula in the present project, but we remark here that it is a particularly important topic when counting geodesics on compact hyperbolic surfaces. For details, the reader should refer to Borthwick [9] and Marklof [45].

Despite the difficulties, progress has still been made. Guillopé [26] and Lalley [36] were the first to prove the shown version of the theorem for convex cocompact hyperbolic surfaces. These surfaces form a subset of hyperbolic surfaces for which a restriction is imposed on the types of surface ends allowed. Guillopé used estimates on heat kernels to prove the PGT conditionally, whereas Lalley applied symbolic dynamics and renewal equations techniques to deduce the result unconditionally. The PGT was eventually extended to higher dimensions (still in the convex cocompact case) by Perry [56], who again used an approach based on the Selberg zeta function, with some remarkable developments having taken place in the last years (see Chapter 8).

Our goal is to study the concepts involved in the formulation of this theorem and some of the steps required in its proof. As the topic touches on several deep areas of mathematics (e.g., negative-curvature geometries, discrete groups, spectral theory, etc.), the proposed appraisal is not meant to be comprehensive. Rather, considerable emphasis is given to studying certain properties of Patterson-Sullivan measures, which are concrete manifestations of conformal densities.

These measures, first introduced by Patterson [55] and generalized by Sullivan [74, 73] in the 1970s and 1980s, have been actively studied ever since.[5] In addition to playing an underlying but significant role in the proof of the PGT, their importance lies in the fact that their introduction enabled an adequate measure-theoretic study of limit sets arising from dynamical systems. By modifying the (absolute) Poincaré series associated with a group of orientation-preserving isometries of the hyperbolic plane (as done in the upcoming Section 4.2), Patterson was able to take advantage of the divergence of this series at its critical exponent, obtaining limit measures that are fully supported on the limit sets of Fuchsian groups acting on the hyperbolic plane (Theorem 4.4.3 in this thesis).

A first sign that the measures thus obtained are in some sense the right measures to address various questions was the discovery that they form a unique measure class when the Fuchsian group is geometrically finite,[6] typically a reasonable condition to impose.[7] Another sign is the so-called δ-harmonicity, or δ-conformality (to be approached in Sections 4.6 and 4.9), which was first investigated in depth by Sullivan, who recognized the geometric potential of this property. This investigation led to the so-called Sullivan's Shadow Lemma (Lemma 6.2.1), which established a bridge linking the fractal nature of certain limit sets associated with non-elementary geometrically finite discrete groups of isometries and the Patterson measures. Further associations then followed, to the Hausdorff measure (and Hausdorff dimension), to the packing measure, to the box-counting dimension, and to metrical Diophantine approximations.[8]

The Patterson-Sullivan measures are not only a tool that provides insights

[5] For an overview on Patterson-Sullivan theory, see Quint [57] and the survey by Denker and Stratmann [17]. For other results related to these measures, refer to Chapter 14 in Borthwick [9].

[6] This question will be approached from an ergodic viewpoint, in Section 5.5.

[7] Although geometric finiteness is often a standard hypothesis in many of the theorems related to Patterson-Sullivan measures, we observe here that there have been advances even without assuming geometric finiteness. For example, the resolution of the tameness conjecture for geometrically infinite Kleinian groups (which are higher-dimensional Fuchsian groups). See Agol [1] and Calegari and Gabai [15].

[8] On several of the fractal aspects, see the survey by Stratmann [69]. With respect to the box-counting dimension, see Stratmann and Urbański [70].

into the fractal nature of the limit sets described, but they are also full of dynamical significance. In this regard, most of the implications are related to the ergodicity of geodesic or horocyclic flows, which in the former case can be demonstrated via an *ad hoc* measure constructed directly from the Patterson-Sullivan measures (see Chapter 5). However, this is not all. There are connections to measure-theoretic and topological entropy,[9] to Gibbs measures,[10] to rational maps,[11] and this invariant measure for the geodesic flow can even be shown mixing and Bernoulli.[12]

It is difficult to elaborate a fair presentation of all developments concerning Patterson-Sullivan measures while keeping the exposition somewhat succinct. This is the reason why we opted to focus on particular developments, both on the fractal and dynamical fronts, providing detailed descriptions of a subset of historically significant results that serve to illustrate the importance of Patterson-Sullivan measures beyond their role in the proof of the PGT. This means that we were forced to ignore some of the topics identified above, especially those that are not in some sense classical, while also avoiding comprehensive forays into the spectral theory of hyperbolic surfaces, which is an area that deserves a treatment of its own, *à la* the one found in Borthwick [9].

The structure of the upcoming chapters and the selection of topics are outlined below. As a prerequisite, in this thesis, we assume basic facts from Riemannian geometry, the starting point into the world of non-Euclidean geometries. In Chapter 2, we first delve into the hyperbolic plane and present a brief overview of hyperbolic geometry in a plane setting. For conciseness, the results in this chapter are stated without proof. Chapter 3 introduces the topic of counting prime (closed) geodesics of a given length on a hyperbolic surface. It presents a proof of a first bound on this number, which is later reassessed (in Chapter 7), albeit in a somewhat different setting, via conformal densities. In Chapter 4, we introduce the Patterson-Sullivan measures and go through some of their properties. Chapters 5 and 6 serve to illustrate the power of conformal densities. Specifically, they include fairly complete proofs of, first, the ergodicity of the geodesic flow on geometrically finite hyperbolic surfaces, and, second, the relationship between critical exponents and the Hausdorff dimension of the limit sets associated with (convex cocompact) hyperbolic surfaces. Chapter 7 returns to the PGT and the question of counting the number of geodesics of a given (maximal) length on a geometrically finite hyperbolic surface. We present a proof of the PGT based on Newman's argument concerning the PNT. Finally, Chapter 8 goes through some recent developments of interest that are related to the PGT and other similar problems. The results stated therein do not necessarily rely on the Patterson-Sullivan construction.

Despite its inherent worth, the PGT is most of all an invitation into the world of modern mathematics, in which occasionally seemingly different areas come together to produce results of remarkable beauty. Part of this beauty is what we hope to convey to the reader.

[9] On the topic of entropy, see Sullivan [74, 73].

[10] See Denker and Stratmann [17] on the various parallels between Gibbs measures and the Patterson-Sullivan measures.

[11] On the analogue between Fuchsian groups and rational maps, see Stratmann and Urbański [71] and Sullivan [72].

[12] Refer to Rudolph [62] on the topic of mixing and Nicholls [53] on ergodicity and discrete groups.

On sources

Several sources were consulted during the research and writing of the present manuscript. We mention only a few of them here, those that were frequently used. Beardon [6], Katok [34], and Ratcliffe [59] were essential to make sense of most of the geometric results found in Chapter 2. In terms of conformal densities, ergodicity of the geodesic flow, and the relationship between Patterson-Sullivan measures and the classical Hausdorff measure, the clear account found in Nicholls [53] proved at times to be indispensable. Borthwick [9] includes most of the topics broached in this thesis and much more. Although it assumes some background knowledge from the reader, it served well as a tour guide on the subject.

Chapter 2

Brief overview of basic hyperbolic geometry

In this initial overview chapter, we provide some basic definitions and results from hyperbolic geometry that are required in later chapters. Since the project is restricted to limit sets associated with hyperbolic surfaces, the focus is on plane models and Fuchsian groups (to be discussed). To keep our exposition short, all results in the chapter are presented without proof. Good references for this introductory material are Katok [34] and Beardon [6]. Content-wise, we followed Beardon and Chapter 2 in Borthwick's book ([9], p. 7). Ratcliffe [59] is a general reference on the topic that goes well beyond the hyperbolic plane setting. An acquaintance with the language of Riemannian geometry is also assumed.

2.1 Two models of the hyperbolic plane

The two most common models for studying the hyperbolic plane are the upper half-plane model and the Poincaré disc (or unit disc) model. Both are defined below (as well as their boundaries).

Definition 2.1.1. The upper half-plane model of the hyperbolic plane is defined by $(\mathbb{H}, g_{\mathbb{H}})$, with

$$\mathbb{H} \equiv \{z = x + iy \in \mathbb{C} : y > 0\}, \quad g_{\mathbb{H}} \equiv \frac{dx^2 + dy^2}{y^2}.$$

The tensor metric $g_{\mathbb{H}}$ is generally called the hyperbolic metric.

Definition 2.1.2. The boundary of the upper half-plane model of the hyperbolic plane is defined as $\partial\mathbb{H} \equiv \mathbb{R} \cup \{\infty\}$.

Definition 2.1.3. The unit disc (or Poincaré disc) model of the hyperbolic plane is defined by $(\mathbb{D}, g_{\mathbb{D}})$, where

$$\mathbb{D} \equiv \{z = x + iy \in \mathbb{C} : |z| < 1\}, \quad g_{\mathbb{D}} \equiv 4\frac{dx^2 + dy^2}{\left(1 - |z|^2\right)^2}.$$

Definition 2.1.4. The set $\partial\mathbb{D} = \{z \in \mathbb{C} : |z| = 1\}$ is defined as the boundary of the unit disc model $\mathbb{D}$.

One can switch between models whenever it is convenient to do so. This is usually accomplished with the help of the isometry[1] $h : \mathbb{H} \to \mathbb{D}$ (which also maps $\partial\mathbb{H}$ and $\partial\mathbb{D}$ bijectively, so it is appropriate to write $h : \overline{\mathbb{H}} \to \overline{\mathbb{D}}$) defined as

$$h(z) = \frac{z - i}{iz - 1}.$$

For now, what is meant by an isometry in practice is that the pullback of $g_{\mathbb{D}}$ through h returns the $g_{\mathbb{H}}$ tensor metric.

This is one example of Möbius transformation, a key object in what follows.

2.2 Stereographic projection

Möbius transformations are defined for the extended complex plane: $\hat{\mathbb{C}} \equiv \mathbb{C} \cup \{\infty\}$. The stereographic projection[2] can be extended to a bijection $\hat{\pi} : \hat{\mathbb{C}} \to \mathbb{S}^1$ that, combined with the Euclidean metric, defines the so-called chordal metric on $\hat{\mathbb{C}}$, given by

$$d_\infty(z, w) = \begin{cases} \frac{2|z-w|}{\sqrt{(1+|z|^2)(1+|w|^2)}} & z, w \in \mathbb{C} \\ \frac{2}{\sqrt{1+|z|^2}} & z \in \mathbb{C}, w = \infty. \end{cases}$$

In particular, we are concerned with the one-point compactification of the complex plane, namely the Riemann sphere, denoted by $\left(\hat{\mathbb{C}}, d_\infty\right)$.

2.3 Möbius transformations

Möbius transformations are fractional transformations of the following sort.

Definition 2.3.1. A Möbius transformation of the extended complex plane is a map $\gamma : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ such that

$$\gamma(z) \equiv \frac{az + b}{cz + d},$$

where $a, b, c, d \in \mathbb{C}$ and $ad - bc \neq 0$.

In the case of $\mathbb{H}$, we are particularly interested in the subset of such transformations that have real coefficients, whose elements can also be understood as matrices from the group $\mathrm{PSL}(2, \mathbb{R})$.[3] The following result applies.

Theorem 2.3.2. $\mathrm{PSL}(2, \mathbb{R})$ *acts on* $\mathbb{H}$ *by homeomorphisms.*

Proof. See Theorem 1.1.1. in Katok's book ([34], p. 3). □

[1] This function is often called the Cayley transform.

[2] Details can be found in Ratcliffe's book, including the derivation of d_∞ ([59], p. 104).

[3] The representation of Möbius transformations via matrices in $\mathrm{PSL}(2, \mathbb{R})$ is well explained in Beardon's book ([6], p. 60).

Maps of the form

$$\frac{\alpha z + \beta}{\bar{\beta} z + \bar{\alpha}},\ \alpha, \beta \in \mathbb{C},\ |\alpha|^2 - |\beta|^2 > 0,$$

are the Möbius transformations that preserve $\mathbb{D}$. They form a group under composition, which in turn can also be identified to a topological group of matrices, the special unitary group of degree two (i.e., PSU(2)).

2.4 Lengths, geodesics, and isometries

The existence of tensor metrics $g_{\mathbb{H}}$ and $g_{\mathbb{D}}$ allows us to talk about the lengths of paths that connect any two points in $\mathbb{H}$ or $\mathbb{D}$ (these are connected spaces). The hyperbolic metric $g_{\mathbb{H}}$ is associated with the hyperbolic length $\ell_{\mathbb{H}}$.

Definition 2.4.1. Let $\sigma : [a, b] \to \mathbb{H}$ be a path in the upper half-plane $\mathbb{H}$. The hyperbolic length of σ is obtained by integrating the hyperbolic norm of the tangent vector along σ, that is,

$$\ell_{\mathbb{H}}(\sigma) = \int_a^b |\dot{\sigma}(t)|_{\mathbb{H}}\, dt.$$

For hyperbolic models, distances are defined in the usual way.

Definition 2.4.2. Let $z, z' \in \mathbb{H}$ and assume that $\Sigma(z, z')$ denotes the set of all piecewise smooth paths with endpoints z and z'. We define the hyperbolic distance between z and z' as $d_{\mathbb{H}}(z, z') \equiv \inf\{\ell_{\mathbb{H}}(\sigma) : \sigma \in \Sigma(z, z')\}$.

Similar definitions exist for $\mathbb{D}$.

Geodesics are locally distance-minimizing paths. In the hyperbolic setting, this leads to the following types of geodesic curves.

Theorem 2.4.3. *The geodesics of $\mathbb{H}$ are precisely the arcs of circles that intersect $\partial\mathbb{H}$ orthogonally, that is, segments of vertical straight lines or arcs of semicircles orthogonal to the real axis. (Similarly, the geodesics of $\mathbb{D}$ are arcs of circles that intersect $\partial\mathbb{D}$ orthogonally.)*

Proof. See Theorem 1.2.1. in Katok's book ([34], p. 4). □

Bijective maps from a space onto itself that preserve distances are called isometries (and this is equivalent to preserving the metric tensor under the pullback). There is a deep relationship between the Möbius transformations that preserve $\mathbb{H}$ and the set of isometries of $\mathbb{H}$.

Theorem 2.4.4. *The group of orientation-preserving isometries of $\mathbb{H}$ is the matrix group $\mathrm{PSL}(2, \mathbb{R})$, or equivalently, the group of Möbius transformations that preserve the upper half-plane.*

Proof. See Theorem 1.1.2. and Theorem 1.3.1. in Katok's book ([34], p. 4; pp. 8-9). □

Likewise, PSU(2) encodes the orientation-preserving isometries of $\mathbb{D}$.

2.5 Classification of Möbius transformations

Möbius transformations can be classified according to their fixed points.

Theorem 2.5.1. *Let γ be a Möbius transformation that preserves $\mathbb{H}$ and suppose that γ is not the identity. Then γ has either:*

1. *no fixed points in $\partial\mathbb{H}$ and one in $\mathbb{H}$;*
2. *one fixed point in $\partial\mathbb{H}$ and none in $\mathbb{H}$;*
3. *two fixed points in $\partial\mathbb{H}$ and none in $\mathbb{H}$.*[4]

Proof. For such γ, the classification can be deduced by studying the equation $\gamma(z) = z$.[5] □

The three different classes have received specific names in the literature. Using the bijection between the set of Möbius transformations that preserve $\mathbb{H}$ and $\mathrm{PSL}(2,\mathbb{R})$, the distinctions equivalently depend on the trace of the matrix in question.

Definition 2.5.2. A transformation $T \in \mathrm{PSL}(2,\mathbb{R})$ is:

1. elliptic if $\operatorname{tr} T < 2$, implying one fixed point within $\mathbb{H}$ (with a matching point in the lower half-plane by complex conjugation) and none in $\partial\mathbb{H}$;
2. parabolic if $\operatorname{tr} T = 2$ (and $T \neq I$), with a single degenerate fixed point in $\partial\mathbb{H}$;
3. hyperbolic if $\operatorname{tr} T > 2$, yielding two distinct fixed points in $\partial\mathbb{H}$, one attracting and one repelling.

When working in the Poincaré disc $\mathbb{D}$, we have the same classification and arrangement, but the boundary is, of course, $\partial\mathbb{D}$. From any type of transformation in $\mathbb{H}$ we can obtain the same type of transformation in $\mathbb{D}$ and vice versa.

Later we will need the definition of the axis of a hyperbolic transformation, which we can readily state.

Definition 2.5.3. Given a hyperbolic transformation $T \in \mathrm{PSL}(2,\mathbb{R})$ (or $\mathrm{PSU}(2)$) with fixed points $p_-, p_+ \in \partial\mathbb{H}$ (respectively, $\partial\mathbb{D}$), its axis $\alpha(T)$ is defined as the unique geodesic in $\mathbb{H}$ (respectively, $\mathbb{D}$) joining p_- and p_+.

2.6 Conjugacy of Möbius transformations

Before moving on to Fuchsian groups, we introduce the notion of conjugacy of Möbius transformations.

Definition 2.6.1. Let γ, σ be two Möbius transformations that preserve $\mathbb{H}$. We say that γ and σ are conjugate if there exists another such Möbius transformation g such that $\gamma = g^{-1} \circ \sigma \circ g$.

[4] One of the fixed points p is repelling, often denoted p_-, while the other is attracting, thus denoted p_+.

[5] This is done in detail in Walkden's lecture notes ([77], pp. 56-57), for example.

Conjugacy is a useful tool due to the next theorem.

Theorem 2.6.2. *Suppose γ is a Möbius transformation that preserves $\mathbb{H}$. Then, the following equivalences hold:*

1. *γ is parabolic if and only if γ is conjugate to a translation $z \mapsto z + b$, for $b \in \mathbb{R}$;*
2. *γ is hyperbolic if and only if γ is conjugate to a dilation $z \mapsto kz$, where $k \in \mathbb{R}$ and $k > 0$;*
3. *γ is elliptic if and only if γ is conjugate to a rotation.*

Proof. See Chapters 10 and 11 of Walkden [77]. □

Similar notions hold for $\mathbb{D}$.[6]

2.7 Displacement length of a hyperbolic transformation

By conjugation, the repelling fixed point of a hyperbolic transformation can be sent to zero and the attracting point to ∞. The resulting map is then a dilation $z \mapsto e^{\ell} z$ with $\ell > 0$. This dilation factor ℓ is especially important.

Definition 2.7.1. Given a hyperbolic transformation $T \in \mathrm{PSL}(2, \mathbb{R})$ (or $\mathrm{PSU}(2)$), the positive number $\ell = \ell(T)$ associated with T (via conjugation) is called the displacement length of T.

Conjugation gives rise to an equivalence relation such that any two hyperbolic transformations that are conjugate to a dilation have either the same dilation factor or not. Thus, the conjugacy classes of hyperbolic elements of an arbitrary group of orientation-preserving isometries of $\mathbb{H}$ (or $\mathbb{D}$) can be indexed by the dilation factor ℓ. The displacement length $\ell(T)$ is also the distance by which the points along the axis $\alpha(T)$ are translated.

2.8 Fuchsian groups

We wish to study the action of discrete subgroups Γ of $\mathrm{PSL}(2, \mathbb{R})$ or $\mathrm{PSU}(2)$ (or their corresponding groups of maps) on $\mathbb{H}$ or $\mathbb{D}$, respectively. These groups have a special name.

Definition 2.8.1. A Fuchsian group is a discrete subgroup of $\mathrm{PSL}(2, \mathbb{R})$ (or $\mathrm{PSU}(2)$).

Discreteness is a crucial ingredient in several of the forthcoming results. For example, it is responsible for making Fuchsian groups countable.[7]

[6] Because of symmetry, it is more natural to think about rotations in the unit disc model. In $\mathbb{D}$, γ being elliptic and thus conjugate to a rotation means that it can be written as $z \mapsto ze^{i\theta}$. We can then use the Cayley transform h to find how such rotations are written in $\mathbb{H}$.

[7] See Beardon's book ([6], p. 15).

It is often helpful to think of discreteness in terms of (properly) discontinuous action,[8] since the following result holds.[9]

Theorem 2.8.2. *A subgroup $\Gamma \subset \mathrm{PSL}(2,\mathbb{R})$ acts (properly) discontinuously on $\mathbb{H}$ if and only if it is Fuchsian.*

Proof. See Theorem 2.2.6. in Katok's book ([34], p. 32). □

In our context, the definition of (properly) discontinuous action is as follows.

Definition 2.8.3. Given a subgroup Γ of $\mathrm{PSL}(2,\mathbb{R})$, we say that Γ acts (properly) discontinuously on $\mathbb{H}$ if the Γ-orbit of any point $z \in \mathbb{H}$ is locally finite: for every compact subset $K \subset \mathbb{H}$ (in the topology induced by $d_{\mathbb{H}}$), $\{T \in \Gamma : Tz \in K\}$ is finite.

A similar definition applies to $\mathbb{D}$, so an analogous theorem holds for subgroups $\Gamma \subset \mathrm{PSU}(2)$.

To fully understand it, note that given a group Γ acting on a set X, we can define the orbit of an arbitrary point $x \in X$.

Definition 2.8.4. Let Γ and X be as above. The subset $\Gamma x = \{\gamma x : \gamma \in \Gamma\}$ of X is called the Γ-orbit of $x \in X$.

The set of all such Γ-orbits, named the orbit space of X, is of special relevance.

Definition 2.8.5. Let Γ and X be as above. The set of all Γ-orbits is given by the quotient $\Gamma\backslash X = \{\Gamma x : x \in X\}$, with the natural projection given by $\pi : X \to \Gamma\backslash X$, $\pi(z) = \Gamma z$.

2.9 Hyperbolic surfaces

In essence, general hyperbolic surfaces can be thought of as manifestations of the hyperbolic plane via quotient structures. The next definitions are standard.

Definition 2.9.1. A surface is a connected, orientable 2-dimensional smooth manifold without boundary (unless otherwise specified).

Definition 2.9.2. A hyperbolic surface is a surface equipped with a complete Riemannian metric of constant Gaussian curvature -1.

We will focus on hyperbolic surfaces that may have infinite area, while still being, in a topological sense, finite.

Definition 2.9.3. A surface is topologically finite if it is homeomorphic to a compact surface with finitely many points excised.

[8] A word of caution. This terminology is particularly unfortunate and varies widely between authors. Refer to Lee's book on topological manifolds for some useful remarks on this matter ([38], p. 312).

[9] We remark that the theorem is not true in a more general setting, that of the extended complex plane. Specifically, if Γ acts (properly) discontinuously in some non-empty open subset of $\hat{\mathbb{C}}$, then Γ is discrete, but the converse does not hold. For a counterexample, we refer the reader to Example 5.3.4. in Beardon's book ([6], pp. 95-96).

In other words, a topologically finite surface is homeomorphic to a compact surface with a finite number of punctures. We will later clarify the geometric and algebraic consequences of this topological finiteness. As an extension of the widely known Classification Theorem for Compact Surfaces, which establishes that hyperbolic surfaces are homeomorphic to a connected sum of tori, it can be shown that topologically finite hyperbolic surfaces M are classified up to diffeomorphism by their genus g and their number of ends n, with the corresponding value of their Euler characteristic being given by $\chi(M) = 2 - 2g - n$.[10]

One often thinks about hyperbolic surfaces in terms of orbit spaces related to a particular Fuchsian group. The reason for this is given by the theorem below and its variants.

Theorem 2.9.4. *For any hyperbolic surface X there is a Fuchsian group without elliptic elements Γ and a Γ-invariant Riemannian covering map $\pi : \mathbb{H} \to X$ that realizes the isometry $X \cong \Gamma\backslash\mathbb{H}$.*

Proof. See Theorem 2.8 in Borthwick's book ([9], p. 14) or, for a deeper version thereof, Corollary 12.5 in Lee's reference work on Riemannian geometry ([37], p. 349). □

The importance of the previous theorem lies in the fact that it connects hyperbolic surfaces to Fuchsian groups. In a sense, it means that we can understand surfaces by studying how Fuchsian groups act on $\mathbb{H}$ (or $\mathbb{D}$, since analogous results also hold for the unit disc).

Another related basic concept is that of a fundamental region/domain.

Definition 2.9.5. A fundamental region $\mathcal{F} \subset \mathbb{H}$ for a Fuchsian group Γ is a closed region (i.e., a closure of the interior of $\mathcal{F}$, a non-empty open set) such that

$$\Gamma\mathcal{F} \equiv \bigcup_{T\in\Gamma} T\mathcal{F} = \mathbb{H},$$

and for each $T \in \Gamma$,$T \neq Id$, the interiors $\text{int}(\mathcal{F})$ and $T(\text{int}(\mathcal{F}))$ do not intersect.

Given a Fuchsian group Γ, one can construct a convenient fundamental domain.

Definition 2.9.6. Let Γ be a Fuchsian group. The Dirichlet domain for Γ of a point $w \in \mathbb{H}$ is defined by

$$\mathcal{D}_w(\Gamma) \equiv \{z \in \mathbb{H} : d_{\mathbb{H}}(z, w) \leq d_{\mathbb{H}}(z, Tw) \text{ for all } T \in \Gamma\}.$$

We often denote the Dirichlet domain $\mathcal{D}_w(\Gamma)$ simply as $\mathcal{D}_w$, since the specific Fuchsian group Γ used in the derivations is clear from context.

Not only is the Dirichlet domain (possibly) a fundamental region, but much more holds.

Lemma 2.9.7. *If w is not the fixed point of an elliptic element of Γ, then the Dirichlet domain $\mathcal{D}_w(\Gamma)$ is a fundamental domain for Γ. Moreover, $\mathcal{D}_w(\Gamma)$ is convex and bounded by a union of geodesics.*

Proof. See Lemma 2.10 in Borthwick's book ([9], p. 16). □

[10] See Massey (Chapter 1 in [46]) for a discussion on the Euler characteristic and classification.

Lemma 2.9.8. *The tessellation* $\{T\mathcal{D}_w(\Gamma) : T \in \Gamma\}$ *is locally finite, meaning that any compact region of* $\mathbb{H}$ *meets only finitely many copies of* $\mathcal{D}_w(\Gamma)$*, and contains only finitely many of its vertices and sides.*

Proof. See Lemma 2.11 in Borthwick's book ([9], p. 16). □

As $\mathbb{H}$ (or $\mathbb{D}$) can be completely tessellated by a fundamental region, hyperbolic surfaces (as manifestations of the hyperbolic plane) should be related in some way to fundamental domains. This is indeed the case, as the following theorem specifies (through the orbit space).

Theorem 2.9.9. *If* $\mathcal{F}$ *is a fundamental region for a discontinuous group*[11] Γ *of isometries of a metric space* X*, then the inclusion* $\iota : \mathcal{F} \to X$ *induces a continuous bijection* $\kappa : \Gamma\backslash\mathcal{F} \to \Gamma\backslash X$*, and* κ *is a homeomorphism if and only if* $\mathcal{F}$ *is locally finite.*

Proof. For a proof, see Ratcliffe's Theorem 6.6.7. (in [59], pp. 235-237). □

2.10 Limit set of a Fuchsian group

To some extent, the entire theory of the upcoming chapters is centered on the study of the limit set of a Fuchsian group.

Definition 2.10.1. For a Fuchsian group Γ, the limit set $\Lambda(\Gamma) \subset \partial\mathbb{H}$ is the set of limit points (in the Riemann sphere topology induced by d_∞) of all orbits Γz for $z \in \mathbb{H}$. The complement of the limit set in $\partial\mathbb{H}$ is the set of ordinary points.

This definition means that if $\lambda \in \Lambda(\Gamma)$, then $\lambda \in \partial\mathbb{H}$ and λ is a limit point of some orbit Γz, $z \in \mathbb{H}$. So, for every neighborhood U_λ, open in the Riemann sphere topology, $\Gamma z \cap (U_\lambda \backslash \{\lambda\}) \neq \emptyset$.

In most situations of interest, to find $\Lambda(\Gamma)$ we need only to consider one orbit. That is, the ensuing result is valid.

Lemma 2.10.2. *If* $w \in \mathbb{H}$ *is not an elliptic fixed point of* Γ,[12] *then* $\Lambda(\Gamma)$ *is the set of limit points of the single orbit* Γw*. It follows that* $\Lambda(\Gamma)$ *is closed and invariant under* Γ*.*

Proof. For a proof, see Lemma 2.12 in Borthwick's book ([9], p. 17). □

There exists a classification that differentiates between three types of limit set that one might encounter when studying the orbit space of the action of a Fuchsian group on the upper half-plane or the Poincaré disc.

Theorem 2.10.3 (Poincaré, Fricke-Klein)**.** *The possibilities for the limit set of a Fuchsian group* Γ *are:*

1. $\Lambda(\Gamma)$ *is elementary, meaning it contains zero, one, or two points.*
2. $\Lambda(\Gamma)$ *is non-elementary of the first kind, meaning* $\Lambda(\Gamma) = \partial\mathbb{H}$ *(or* $\Lambda(\Gamma) = \partial\mathbb{D}$*).*

[11] That is, a group that acts (properly) discontinuously on X.

[12] This means that there exists no transformation in $\gamma \in \Gamma$ such that w is an elliptic fixed point of γ. In fact, the result holds even without this requirement. See Theorem 5.3.9. in Beardon ([6], p. 98)

3. $\Lambda(\Gamma)$ *is non-elementary of the second kind, meaning it is a perfect nowhere dense subset of* $\partial\mathbb{H}$ *(or* $\partial\mathbb{D}$*).*

Proof. See Theorem 2.13 in Borthwick's book ([9], pp. 17-18). □

Notice that cyclic Fuchsian groups (i.e., Fuchsian groups that are generated by a single element) are elementary, with $\Lambda(\Gamma)$ comprising zero, one, and two points in the elliptic, parabolic, and hyperbolic cases, respectively. The most interesting case is that of non-elementary Fuchsian groups of the second kind, and this is our focus once we turn to the question of counting geodesic lengths on hyperbolic surfaces.

2.11 Geometrically finite groups

The counting of geodesic lengths, among other questions relating to hyperbolic surfaces, is generally only approachable in the context of a topologically finite hyperbolic surface. The condition of topological finiteness of the hyperbolic surface $\Gamma\backslash\mathbb{H}$ (or $\Gamma\backslash\mathbb{D}$) is often best approached through an equivalent geometric or algebraic criterion.

Definition 2.11.1. A Fuchsian group Γ is said to be geometrically finite if it admits a fundamental domain that is a finite-sided convex polygon.

Definition 2.11.2. The Fuchsian group Γ is finitely generated if there exists a finite subset of transformations that generate the group.

Based on the previous definitions, the following interchangeability result holds.

Theorem 2.11.3 (Topological, Algebraic, and Geometric Finiteness)**.** *For a Fuchsian group* Γ *the following are equivalent:*

1. $\Gamma\backslash\mathbb{H}$ *(or* $\Gamma\backslash\mathbb{D}$*) is topologically finite (i.e., finite Euler characteristic).*
2. Γ *is finitely generated.*
3. Γ *is geometrically finite.*

Proof. See Theorem 10.1.2. in Beardon's book ([6], p. 254), which includes a proof for the model $\mathbb{D}$, or Theorem 2.16 in Borthwick's work ([9], pp. 20-23), a proof for $\mathbb{H}$. □

Geometric finiteness imposes restrictions on the ends of a hyperbolic surface. We will see that the only possibilities, beyond the hyperbolic plane itself, are the ends of cylinders, that is, quotients of $\mathbb{H}$ or $\mathbb{D}$ by hyperbolic and parabolic cyclic groups.

2.12 Cylinders, funnels, and cusps

A hyperbolic transformation T generates a cyclic hyperbolic group denoted by $\langle T\rangle$. This group defines a hyperbolic cylinder.

Definition 2.12.1. For a hyperbolic transformation $T \in \mathrm{PSL}(2,\mathbb{R})$, the quotient $C_\ell \equiv \langle T\rangle\backslash\mathbb{H}$ is a hyperbolic cylinder of diameter $\ell = \ell(T)$.

We can now talk about funnels.

Definition 2.12.2. A funnel is a closed half of a hyperbolic cylinder, with boundary given by the central geodesic. In other words, if T is a hyperbolic transformation of $\mathbb{H}$ with displacement length ℓ, and if H denotes one of the open half-planes of $\mathbb{H}$ bounded by the axis $\alpha(T)$, then

$$F_\ell \equiv \langle T\rangle\backslash\overline{H}$$

is a funnel of diameter $\ell(T)$, where $\overline{H} = H \cup \alpha(T)$.

Parabolic cylinders are similar to hyperbolic ones, with the exception that the defining transformation is, of course, parabolic and not hyperbolic.

Definition 2.12.3. For a parabolic $T \in \mathrm{PSL}(2,\mathbb{R})$, the quotient $\langle T\rangle\backslash\mathbb{H}$ is called a parabolic cylinder.

In order to define cusps, the important notion of a horocycle needs to be addressed.

Definition 2.12.4. A horocycle, in the context of the upper half-plane model, is a circle that lies in $\mathbb{H}$ and is tangent to $\partial\mathbb{H}$.

Definition 2.12.5. A cusp is the small end of a parabolic cylinder, which has the unique closed horocycle of unit hyperbolic length as the boundary.

All definitions above have analogous counterparts in the unit disc setting. As a final remark, we observe that using the volume/area form induced by the appropriate tensor metric, $g_{\mathbb{H}}$ or $g_{\mathbb{D}}$, a simple calculation shows that funnels have infinite area, whereas cusps have finite area.

2.13 Nielsen regions

Let us focus on the case of a hyperbolic surface $X \cong \Gamma\backslash\mathbb{H}$ where Γ is Fuchsian and non-elementary: Γ is of the first kind if $\Lambda(\Gamma) = \partial\mathbb{H}$, or of the second kind if $\Lambda(\Gamma)$ is perfect and nowhere dense (see Theorem 2.10.3 and accompanying definition).

If Γ is of the second kind, $\Lambda(\Gamma)$ is perfect and thus closed, which means that its complement $\partial\mathbb{H} - \Lambda(\Gamma)$ is open. Therefore, it can be written as a countable union of open intervals I_j. For each j, suppose that γ_j is the geodesic whose endpoints are the endpoints that determine the open interval I_j, and let H_j be the open half-plane bounded between γ_j and I_j. The geodesics are semicircles centered on the real line, when the open intervals have finite endpoints, or vertical lines, in the case where the open interval has infinity as one of its endpoints. In the last case, we can choose either one of the sides of the vertical line as H_j.

On the other hand, if Γ is of the first kind, that is, if $\Lambda(\Gamma) = \partial\mathbb{H}$, then $\partial\mathbb{H} - \Lambda(\Gamma) = \emptyset$ and we agree that $\{H_j\} = \emptyset$.

Now we are ready to introduce the notion of a Nielsen region.

Definition 2.13.1. Given a collection of open half-planes $\{H_j\}$ defined as indicated above, the Nielsen region of a Fuchsian group Γ is defined as the set $\widetilde{N} \equiv \mathbb{H} - \bigcup H_j$.

The Nielsen region is also equivalently defined as the convex hull of

$$\bigcup_{q_-, q_+ \in \Lambda(\Gamma)} [q_-, q_+],$$

where $[q_-, q_+]$ denotes the geodesic arc between endpoints q_- and q_+. In other words, the Nielsen region is the smallest closed convex[13] set containing all geodesics whose defining endpoints lie both in the limit set $\Lambda(\Gamma)$.

Starting from a Nielsen region, one can talk about the convex core of a hyperbolic surface.

Definition 2.13.2. Given a Nielsen region $\widetilde{N}$, the quotient $N \equiv \Gamma\backslash\widetilde{N}$ is called the convex core of the hyperbolic surface $X \cong \Gamma\backslash\mathbb{H}$.

If Γ is of the first kind, then $\widetilde{N} \equiv \mathbb{H} - \emptyset$ and $N \equiv \Gamma\backslash\widetilde{N} = \Gamma\backslash\mathbb{H} \cong X$.

These definitions are readily transferable to the Poincaré disc setting, which is the model that lends itself more naturally to visualization purposes.

Given a parabolic fixed point $p \in \partial\mathbb{H}$, let Γ_p be the parabolic cyclic subgroup of Γ fixing p. Let σ_p be the unique horocycle tangent to $\partial\mathbb{H}$ at p such that the translations along the horocycle have unit hyperbolic length. Then let O_p be the open region bounded by σ_p, so that $\Gamma_p\backslash\bar{O}_p$ is a cusp, according to the convention introduced in the previous section. The following result summarizes some relevant properties of horocyclic or cusp regions.

Lemma 2.13.3. *For σ_p and O_p defined as above, the following statements hold:*

1. *If two points of O_p are related by $T \in \Gamma$, then $T \in \Gamma_p$.*
2. *The horocycles σ_p for different parabolic fixed points do not intersect.*
3. *The horocycles σ_p do not intersect the half-planes H_j that define the Nielsen region, so for each $\bar{O}_p$, we have $\bar{O}_p \subset \widetilde{N}$.*

Proof. For a proof, refer to Lemma 2.21 in Borthwick's book ([9], p. 26). □

Now that we know more about cusp regions (particularly, that they do not intersect each other nor the open half-planes H_j), we can redefine the Nielsen region, isolating not only funnels, but also cusps. This is what is called a truncated Nielsen region.

Definition 2.13.4. Given a Nielsen region $\widetilde{N}$, its truncated Nielsen region is defined as $\widetilde{K} \equiv \widetilde{N} - \bigcup O_p$ with the union running over all parabolic fixed points p of Γ.

As in the case of the convex core of a hyperbolic surface, we can talk about the compact core of a geometrically finite hyperbolic surface.

Definition 2.13.5. If Fuchsian Γ defines a geometrically finite hyperbolic surface $X \cong \Gamma\backslash\mathbb{H}$ with a truncated Nielsen region $\widetilde{K}$, the quotient region $K \equiv \Gamma\backslash\widetilde{K}$ is called the compact core of X.

[13] This convexity requirement should be understood in the sense of geodesic segments.

Apart from funnels and cusps, the next result establishes that there are no other possibilities concerning hyperbolic endings in a surface setting.

Theorem 2.13.6 (Classification of Hyperbolic Ends)**.** *Let $X \cong \Gamma\backslash\mathbb{H}$ be a non-elementary geometrically finite hyperbolic surface. Then the compact core K (as defined above) is compact and $X - K$ is a finite disjoint union of cusps and funnels.*

Proof. See Theorem 2.23 in Borthwick's book ([9], p. 28) □

It can be shown that the compactness of K is equivalent to the geometric finiteness of a hyperbolic surface.[14] The compact core K of a hyperbolic surface can be further divided into what is called a pants decomposition.[15]

If the convex core of a hyperbolic surface $X \cong \Gamma\backslash\mathbb{H}$ is compact, the Fuchsian group receives a special name.

Definition 2.13.7. If the convex core of a hyperbolic surface $X \cong \Gamma\backslash\mathbb{H}$ is compact, that is, $N = K$ (following the notation introduced in the present section), then Γ is said to be convex cocompact.

Observe that this situation is equivalent to the surface being geometrically finite with no cusps.

[14] Refer to Fenchel and Nielsen's book (§15.1 in [23], p. 115) for a proof.
[15] See Section 2.7 in Borthwick's book ([9], p. 40).

Chapter 3

Closed geodesics and the length spectrum

In this short preparatory chapter, we start our discussion on lengths of closed geodesics, which are also examples of geometric invariants of hyperbolic surfaces, as is the case of the Euler characteristic, genus, and number of terminations (i.e., number of funnels and cusps). The aim of the chapter is to prove a first bound on the length counting function (to be defined). For space reasons, all other results are presented without proof. The chapter follows Borthwick (Chapter 2 in [9]), though Buser's book (Chapter 1 in [14]) is also useful.

3.1 Closed geodesics

For compact surfaces of negative curvature, each closed geodesic belongs to a unique free homotopy class of closed curves. The intuition is that each class contains loops that can be deformed into each other. The closed geodesic of a particular class is then the representative of minimum hyperbolic length. For a non-compact surface, matters are not as simple, as the horocycle bounding a cusp has no geodesic within its homotopy class. One can circumvent this problematic case and prove that any curve that is not freely homotopic to a horocycle is then homotopic to a closed geodesic. With this aim, we make a distinction between the cases.

Definition 3.1.1. A curve is cuspidal if it is freely homotopic to the horocyclic boundary of a cusp. Otherwise, it is non-cuspidal.

We are only interested in the lengths of the closed curves that cannot be deformed to a point, called homotopically nontrivial curves, since those that can be deformed to a point have minimum hyperbolic length given by the hyperbolic length of a point, which is, of course, zero.

The distinction introduced above grants the following result.

Proposition 3.1.2. *Let η be a homotopically nontrivial curve on a hyperbolic surface $X \cong \Gamma\backslash\mathbb{H}$. If η is non-cuspidal, then there exists a unique closed geodesic γ which is the shortest closed curve freely homotopic to η.*

Proof. See Proposition 2.24 in Borthwick's book ([9], p. 30). □

The lengths of closed geodesics are invariants of a given hyperbolic surface. As seen, each hyperbolic surface is encoded by a Fuchsian group Γ acting on $\mathbb{H}$ or $\mathbb{D}$, containing a countable number of hyperbolic transformations. These transformations can be conjugated to a dilation that fixes zero and infinity, and they can then be ordered according to how far they send i along the imaginary axis. The sequence of lengths thus obtained is always the same as long as we consider sequences whose elements are from the same conjugacy class indexed by the displacement length. The intuition is that the given hyperbolic surface does not allow any other closed curves whose length does not appear in the described sequence to be taken as "straight" (i.e., geodesic). In other words, the next proposition is also true.

Proposition 3.1.3. *There is a one-to-one correspondence between the closed oriented geodesics of a hyperbolic surface $X \cong \Gamma\backslash\mathbb{H}$ and the conjugacy classes of hyperbolic elements of Γ. The length of the geodesic corresponding to the conjugacy class $[T]$ is the displacement length $\ell(T)$.*

Proof. See Proposition 2.25 in Borthwick's book ([9], p. 30). □

Once we are given a closed geodesic, we can generate a family of iterates which move along the same geodesic path multiple times. Based on this fact, we introduce the following definition.

Definition 3.1.4. A primitive closed geodesic is the root element of a family of iterates, that is, a closed geodesic that is not an iterate of a shorter closed geodesic.

We have an analogous definition for the elements of a Fuchsian group Γ.

Definition 3.1.5. An element of Γ is called primitive if it is not the power of some other element of Γ.

In the curve setting, note that an oriented closed geodesic γ can be uniquely represented as the iterate of a primitive oriented closed geodesic σ. To find σ, if we start from $\gamma(0)$, we simply follow the geodesic and find the first value of $t > 0$ such that $\gamma(t) = \gamma(0)$ and $\gamma'(t) = \gamma'(0)$. This is the primitive curve σ we are looking for.

In the group setting, there exists a corresponding result.

Lemma 3.1.6. *Given a Fuchsian group Γ, each element $S \in \Gamma$ can be written uniquely as a power T^k where $T \in \Gamma$ is primitive and $k \geq 1$. The centralizer Z_S of S in Γ is the cyclic group $\langle T \rangle$.*

Proof. See Lemma 2.26 in Borthwick's book ([9], p. 31). □

3.2 A general bound on the length counting function

Even when considering only primitive curves, it is still possible for two distinct primitive oriented closed geodesics to have the same length (e.g., consider the same curve with the opposite orientation). The counting function should not

ignore such cases, and this is why the notion of a multiset[1] is used in the following definition.

Definition 3.2.1. The (primitive) length spectrum of a hyperbolic surface X is the multiset

$$\mathcal{L}_X \equiv \{\ell(\gamma) : \gamma \text{ is a primitive oriented closed geodesic on } X\},$$

with lengths repeated according to multiplicity.

Observe that $\mathcal{L}_X$ as defined becomes sensitive not only to the orientation of primitive closed geodesics, but also to the size of conjugacy classes of hyperbolic elements of Γ. This happens because all elements in a class can be conjugated to a dilation by $e^\ell, \ell > 0$, but each preserves different primitive oriented closed geodesics. The integer multiplicity of ℓ then counts how many such elements there are in its class.

Given the spectrum, we can now define a length counting function.

Definition 3.2.2. The length counting function $\pi_X : \mathbb{R}_{>0} \to \mathbb{Z}_{\geq 0}$ of a hyperbolic surface X is defined as

$$\pi_X(t) \equiv \#\left\{\ell \in \mathcal{L}_X : \ell \leq t\right\}.$$

As stated in the Introduction (Chapter 1), one of our goals is to find an asymptotic formula for $\pi_X(t)$. For the moment, we can already state and prove a basic bound.

Proposition 3.2.3. *For X a geometrically finite hyperbolic surface,*

$$\pi_X(t) \leq Ce^t,$$

for some constant $C > 0$.

Proof. Given an isometry $X \cong \Gamma\backslash\mathbb{H}$, let $K = \Gamma\backslash\widetilde{K}$ be the compact core of X, where $\widetilde{K}$ is the truncated Nielsen region of X. According to the definition of the open half-planes H_j (in the notation introduced in Section 2.13), the open intervals along $\partial\mathbb{H}$ used to define them cannot contain limit points of $\mathbb{H}$ under the action of Γ. This implies that all geodesics in the upper half-plane model whose defining endpoints are both in $\Lambda(\Gamma)$ must be contained in the Nielsen region $\widetilde{N} = \mathbb{H} - \bigcup H_j$, since this region is equivalently defined as the convex hull of the union of all such geodesics (and thus contains this union). This is the reason why closed geodesics on a hyperbolic surface are contained in its convex core $N = \Gamma\backslash\widetilde{N}$.[2] They are not necessarily contained in the compact core of the

[1] A multiset is a set that admits repetitions of its elements.

[2] Due to the nature of Fuchsian Γ, the projection $p : \mathbb{H} \to X \cong \Gamma\backslash\mathbb{H}$, such that $z \mapsto \Gamma z$, is not just an ordinary projection, but is, in fact, a Riemannian covering. As such, it entails a local isometry that implies that projections of geodesics of $\mathbb{H}$ are geodesics in $\Gamma\backslash\mathbb{H}$ and lifts of geodesics in $\Gamma\backslash\mathbb{H}$ are geodesics in the upper half-plane. Recall that being a geodesic is a local property, thus preserved by local isometries (see Corollary 5.14 in [37], p. 125). This is why we talk freely about geodesics in the upper half-plane and surface geodesics. Now, if we take the lift of a closed surface geodesic, we obtain a geodesic in $\mathbb{H}$ with an infinite number of points identified via the projection p. Therefore, the geodesic is such that its defining endpoints are limit points in $\Lambda(\Gamma)$. Hence, the geodesic is in the Nielsen region $\widetilde{N}$, implying that its (closed) projection is in N.

hyperbolic surface, because there could still be geodesics that have at least one of its fixed points as the limit point that defines a horocyclic region. However, even in this case, the geodesic cannot be fully contained in the horocyclic region, because it has another limit point, and even if this second fixed point lies in a horocyclic region, these regions cannot intersect (see Lemma 2.13.3), which shows that every geodesic whose endpoints lie in the limit set $\Lambda(\Gamma)$ has to cross the truncated Nielsen region $\widetilde{K}$. In other words, any closed geodesic must pass through the compact core K of a hyperbolic surface.

Assume that Γ does not have elliptic elements (i.e., X is a smooth hyperbolic surface). For some $w \in \widetilde{K}$, consider the Dirichlet domain $\mathcal{D}_w$ and let Z denote the compact region $\widetilde{K} \cap \mathcal{D}_w$. Take a primitive closed geodesic γ on X with displacement length ℓ. Because γ necessarily traverses K it can be covered by a geodesic $\tilde{\gamma}$ in $\mathbb{H}$ that passes through some point $q \in Z$.

This geodesic $\tilde{\gamma}$ is the axis of some primitive $T \in \Gamma$ (here we use Lemma 3.1.6), such that $\ell(T) = \ell$. Let $a = \operatorname{diam}_{\mathbb{H}}(Z) \equiv \sup_{z_1, z_2 \in Z} d_{\mathbb{H}}(z_1, z_2)$. Then,

$$\begin{aligned} \text{(I)} \quad d_{\mathbb{H}}(w, Tw) &\leq d_{\mathbb{H}}(w, q) + d_{\mathbb{H}}(q, Tq) + d_{\mathbb{H}}(Tq, Tw) \\ &= a + \ell + d_{\mathbb{H}}(q, w) \\ &= \ell + 2a. \end{aligned}$$

Therefore, for each primitive closed geodesic of displacement length ℓ, we have shown that there exists a hyperbolic transformation T that maps w to a point of hyperbolic distance less than $\ell + 2a$ away, for some fixed a (depending on the hyperbolic surface). We can thus bound $\pi_X(t)$ by the number of images of w within distance $t + 2a$ of the point w. Observe that if we count the number of images of w in a closed hyperbolic disc of radius $t + 2a$, we can ignore all transformations that do not move w along a geodesic, because then w and its image will necessarily have a hyperbolic length greater than $t + 2a$. Moreover, since this disc is compact, the number of images of w within $\overline{B(w, t + 2a)}$ cannot be infinite due to the discreteness of Γ. Therefore, we can simply bound this number by the area of $B(w, t+2a)$, since this represents the case where each point of $B(w, t + 2a)$ is seen as an image of w under some hyperbolic transformation (i.e., recall that between any two points in the upper half-plane, we can find a geodesic that joins them, which in itself is associated to some hyperbolic transformation that can be applied to w).

To find the area of the disc, we can use the hyperbolic area form[3] $\operatorname{Area}(z) = \frac{dxdy}{y^2}$ and change coordinates to polar coordinates in order to obtain

$$\begin{aligned} \operatorname{area}(B(w, t + 2a)) &= \int_{B(w, t+2a)} \operatorname{Area}(z) \\ &= \int_0^{\pi} \int_0^{t+2a} 2 \sinh r dr d\phi \\ &= 2\pi(\cosh(t + 2a) - 1. \end{aligned}$$

As a consequence of the above argument, we have

$$\pi_X(t) \leq \operatorname{area}(B(w, t + 2a)) = 2\pi(\frac{e^{t+2a} - e^{-(t+2a)}}{2} - 1) \leq \pi e^{2a} e^t,$$

[3]For the derivation of the hyperbolic area form for the upper half-plane model, refer to the fully general result given by Theorem 4.6.7. in Ratcliffe ([59], p. 131).

which proves the desired asymptotic relationship. □

Chapter 4

Patterson-Sullivan measures on the limit set

To address the problem of assigning a measure to the limit set $\Lambda(\Gamma)$ of a discrete group Γ, Patterson [55] and Sullivan [74] developed a set of finite measures that are fully supported on $\Lambda(\Gamma)$. These special measures became known as Patterson-Sullivan measures and are of importance in all the upcoming chapters. To a certain degree, one can say that Patterson-Sullivan measures are the right measures for studying different kinds of dynamical questions in hyperbolic models, and if nothing else, the next chapters serve as an illustration of their usefulness. Recall also similar remarks made in the Introduction (Chapter 1). The focus of this initial chapter is mainly on the construction of Patterson-Sullivan measures for the unit disc $\mathbb{D}$ and on establishing some of their key properties. For that, the treatment herein follows mostly Nicholls [53] and Borthwick [9]. We start by discussing the critical exponent of a Fuchsian group.

4.1 Critical exponent

One of the core concepts in the upcoming results is that of the convergence exponent, also called the critical exponent, of a Fuchsian group Γ.

To some extent, the study of the rate at which the Γ-orbits of $z \in \mathbb{D}$ tend to $\partial\mathbb{D}$ can be approached through the series

$$\sum_{T\in\Gamma} (1 - |Tz|)^s \,.$$

However, since its convergence or divergence (which does not depend on z) is related to that of the (absolute) Poincaré series,[1] the latter is usually the series of choice in what concerns the development of Patterson-Sullivan theory.

[1] This is due to the hyperbolic formula

$$d_{\mathbb{D}}(0, \gamma(0)) = \ln\left(\frac{1 + |\gamma(0)|}{1 - |\gamma(0)|}\right),$$

for γ an orientation-preserving isometry of $\mathbb{D}$. See Section 1.6 in Nicholls' book ([53], p. 20). For the formula, see Lemma 4.3.2 below.

Definition 4.1.1. For a Fuchsian group Γ, the function (dependent on s, z, and w) given by summing $e^{-sd_{\mathbb{D}}(z,Tw)}$ over $T \in \Gamma$ is called the (absolute) Poincaré series for Γ.

Sometimes we abbreviate the (absolute) Poincaré series as

$$g_s(z,w) \equiv \sum_{T\in\Gamma} e^{-sd_{\mathbb{D}}(z,Tw)}.$$

This series converges for all s above some threshold value, defined as the convergence exponent of Γ.

Definition 4.1.2. The critical exponent of a Fuchsian group Γ is defined as

$$\delta(\Gamma) \equiv \inf\left\{ s \in \mathbb{R}_{\geq 0} : \sum_{T\in\Gamma} e^{-sd_{\mathbb{D}}(z,Tw)} < \infty \right\},$$

for $z, w \in \mathbb{D}$.

We emphasize here that although for the most part we assume $\delta(\Gamma)$ to be a real number, in some situations it is worthwhile to adopt the complex counterpart to this definition (as Borthwick does in [9]). This change is necessary, for instance, when discussing the analytic continuation of the Selberg zeta function.

In general, if the Fuchsian group Γ is clear from context, we simply let δ denote $\delta(\Gamma)$. Furthermore, because the triangle inequalities

$$d_{\mathbb{D}}(z,Tw) \leq d_{\mathbb{D}}(z,w) + d_{\mathbb{D}}(w,Tw)$$

and

$$d_{\mathbb{D}}(z,Tw) \geq d_{\mathbb{D}}(w,Tw) - d_{\mathbb{D}}(z,w)$$

yield

$$e^{-sd_{\mathbb{D}}(z,w)} g_s(w,w) \leq g_s(z,w) \leq e^{sd_{\mathbb{D}}(z,w)} g_s(w,w),$$

the definition does not depend on the choice of $z, w \in \mathbb{D}$.[2]

To better grasp the behavior of the (absolute) Poincaré series, a first bound on δ is shown. An intuitive argument using a modified version of the orbit-counting function from Chapter 3 establishes a general upper bound on δ. This adaptation of $\pi_X(t)$ is given by

$$\pi_X(t,x,y) \equiv \#\{T \in \Gamma : d_{\mathbb{D}}(x,Ty) < t\},$$

for $x, y \in \mathbb{D}$ and $t > 0$.[3]

[2]That is, if $g_s(z,w) < \infty$ for some $z, w \in \mathbb{D}$ and $s \in \mathbb{R}_{>0}$, then for any $z', w' \in \mathbb{D}$, the chain of implications

$$e^{-sd_{\mathbb{D}}(z,w)} g_s(w,w) \leq g_s(z,w) < \infty \Rightarrow g_s(w,w) < \infty$$
$$g_s(z',w) \leq e^{sd_{\mathbb{D}}(z',w)} g_s(w,w) < \infty \Rightarrow g_s(z',w) < \infty$$
$$e^{-sd_{\mathbb{D}}(w,z')} g_s(z',z') \leq g_s(w,z') < \infty \Rightarrow g_s(z',z') < \infty$$
$$g_s(z',w') \leq e^{sd_{\mathbb{D}}(z',w')} g_s(z',z') < \infty \Rightarrow g_s(z',w') < \infty$$

demonstrates that if the (absolute) Poincaré series converges for a given s, the choice of z or w does not impact the infimum as defined.

[3]This function is sometimes called the lattice-point counting function (see Borthwick in [9], p. 352).

It is helpful to recall Proposition 3.2.3 and observe that it can be altered so that there exists a constant C (depending on Γ) and $y \in \mathbb{D}$ such that for any $x \in \mathbb{D}$ the bound

$$\pi_X(t, x, y) \overset{(*)}{<} Ce^t$$

holds.[4]

This bound can be used as follows.

Proposition 4.1.3. *Let Γ be a Fuchsian group preserving $\mathbb{D}$. If $s > 1$, the series $\sum_{T\in\Gamma} e^{-sd_{\mathbb{D}}(0,T0)}$ converges.*

Proof. Using integration by parts,[5] we can write the partial sum of the series as

$$\begin{aligned}\sum_{\substack{T\in\Gamma \\ d_{\mathbb{D}}(0,T0)<R}} e^{-sd_{\mathbb{D}}(0,T0)} &= \int_0^R e^{-st} d\pi_X(t,0,0) \\ &= \pi_X(R,0,0)e^{-sR} + s\int_0^R \pi_X(t,0,0)e^{-st}dt\end{aligned}$$

Then, using the above bound $(*)$, for a constant C that depends on Γ, we obtain the desired convergence. □

An immediate consequence of this proposition is that $\delta \leq 1$. Later we will prove specific bounds on δ that hold when the Fuchsian group is of the second kind (see Theorem 4.7.1 below). For completeness, we mention two more bounds. If Γ is elementary, it is clear that $\delta = 0$, since the (absolute) Poincaré series converges for any $s_n = \frac{1}{n}, n \in \mathbb{Z}_{>0}$. On the other hand, if Γ is of the first kind (i.e., the hyperbolic surface $\Gamma\backslash\mathbb{D}$ has finite area), then $\delta = 1$.[6]

One can ask whether the (absolute) Poincaré series diverges or not at the exponents $s = 1$ and $\delta < 1$. It is not difficult to prove that if Γ is of the second kind, then the series converges at the exponent $s = 1$.[7] Some notions of spectral theory can be used to answer the second question. It turns out that, for non-elementary geometrically finite hyperbolic surfaces, the (absolute) Poincaré series diverges at δ.[8] As will become clear in the following, this divergence relatively eases the construction of the Patterson-Sullivan measures, at least for the types of discrete groups usually studied (i.e., the non-elementary ones). Nevertheless, the construction proposed in the next section is quite general insofar as it also deals with the case of a Fuchsian group that converges at δ.

4.2 Construction preliminaries

When discussing the convergence exponent in the last section, a basic distinction became apparent: between Fuchsian groups Γ of convergence type (i.e., the

[4]For details on the modified proof, see Theorem 1.5.1 in Nicholls' book ([53], p. 18).

[5]Two observations. The number of Γ-images whose hyperbolic distance from zero is less than some real R is finite, due to discreteness. Thus, we can write the sum as a simple function. The second observation is that integration by parts does not require that π_X be continuous, only that it be Lebesgue-integrable and that e^{-st} be absolutely continuous (which it is).

[6]See Theorem 1.6.3 in Nicholls' book ([53], p. 22).

[7]See Theorem 1.6.2 in Nicholls' book ([53], p. 21).

[8]Borthwick presents a proof in Proposition 14.17. (in [9], p. 341)

(absolute) Poincaré series converges at the critical exponent $\delta(\Gamma)$) and of divergence type (i.e., said series diverges at the critical exponent). This distinction is important in the construction of the Patterson-Sullivan measures.

How to construct a measure on the limit set $\Lambda(\Gamma)$? Given $x, y \in \mathbb{D}$ and a Fuchsian group Γ, with $s > \delta(\Gamma) \equiv \delta$, the measure is constructed by placing a Dirac mass of weight $\frac{e^{-sd_{\mathbb{D}}(x,\gamma y)}}{g_s(y,y)}$ at each point $\gamma y, \gamma \in \Gamma$. Then a theorem by Helly or Prokhorov (to be discussed) can be used to obtain a limit measure in the limit $s \to \delta^+$. It is clear that if Γ is of divergence type, then all the mass will be pushed to the limit set, since $g_\delta(y, y) = \infty$. The construction would be complete if this were the only possible case. However, as anticipated, it is not. If Γ is of convergence type, then the limit measure has point masses on the orbit of y. To remedy this case, we multiply the point masses by a special factor $h\left(e^{d_{\mathbb{D}}(x,\gamma y)}\right)$ with the following behavior: it does not alter δ, but ensures divergence of the (absolute) Poincaré series at δ. The existence of such a function with the described behavior is established by the following lemma.

Lemma 4.2.1. *Let Γ be a Fuchsian group with critical exponent δ. There exists a function $h : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ which is continuous, non-decreasing, and*

1. *the series $\sum_{\gamma\in\Gamma} e^{-sd_{\mathbb{D}}(x,\gamma y)} h\left(e^{d_{\mathbb{D}}(x,\gamma y)}\right)$ converges for $s > \delta$ and diverges for $s \leq \delta$ and*
2. *if $\epsilon > 0$ is given there is r_0 such that for $r > r_0, t > 1, h(rt) \leq t^\epsilon h(r)$.*

Proof. As discrete groups are countable (see Section 2.8), let

$$\Gamma = \{\gamma_n : n = 1, 2, \ldots\}$$

be ordered in such a way that $a_n = e^{d_{\mathbb{D}}(x,\gamma_n y)}$ increases to infinity. Take a sequence $\{\epsilon_n\}$ of decreasing positive numbers going to zero. We will define an inductive sequence $\{X_n\}$, with $X_n \to \infty$, of intervals $[X_n, X_{n+1}]$, such that h is defined on each interval in the following way. Let $X_1 = 1$ and set $h(x) = 1$ on $[0, 1]$. By the induction step, h is defined on $[0, X_n]$. Then choose X_{n+1} so that

$$\frac{h\left(X_n\right)}{X_n^{\epsilon_n}} \sum_{X_n < a_p \leq X_{n+1}} a_p^{-(\delta-\epsilon_n)} \geq 1.$$

By the definition of the critical exponent δ, there cannot be any other smaller exponent δ' such that the series $\sum a_p^{-\delta'}$ converges, when we are free to choose any $a_p > X_n$. Thus, since $\sum a_n^{-(\delta-\epsilon_n)}$ diverges under these circumstances, we can always find the next X_{n+1} such that the above inequality holds. Given this bound, for $x \in [X_n, X_{n+1}]$ we define

$$h(x) = h\left(X_n\right)\left(\frac{x}{X_n}\right)^{\epsilon_n}.$$

Now we prove that h is the function we are looking for.

Step 1) Divergence of the modified Poincaré series at δ and $s < \delta$.

Notice that

$$\sum_{\gamma\in\Gamma} e^{-\delta d_{\mathbb{D}}(x,\gamma y)} h\left(e^{d_{\mathbb{D}}(x,\gamma y)}\right) = \sum_{\gamma_n\in\Gamma} e^{-\delta d_{\mathbb{D}}(x,\gamma_n y)} h\left(e^{d_{\mathbb{D}}(x,\gamma_n y)}\right) = \sum_{n=1}^{\infty} h\left(a_n\right)\cdot a_n^{-\delta}$$

diverges, because

$$\sum_{n=1}^{\infty} h\left(a_n\right) a_n^{-\delta} = \sum_{n=1}^{\infty} \sum_{a_p\in[X_n,X_{n+1}]} h\left(X_n\right)\left(\frac{a_p}{X_n}\right)^{\epsilon_n} a_p^{-\delta} \geq \sum_{n=1}^{\infty} 1,$$

given the construction of the function h.

For $s<\delta$, since h is increasing (see below) and $a_n>1, \forall n$, we have $\sum_{n=1}^{\infty} h\left(a_n\right) a_n^{-s} \geq \sum_{n=1}^{\infty} h\left(a_n\right) a_n^{-\delta}$. Thus, the above argument also proves the divergence of the modified Poincaré series for $s<\delta$.

Step 2) 2nd Statement.

By the definition of h and the intervals $[X_n, X_{n+1}]$, with $X_1=1$, h is positive and increasing. For an arbitrary $\epsilon>0$, let n be sufficiently large so that $\epsilon\geq\epsilon_n$. For $y\in(X_n, X_n+1]$ and $x\in[X_n, X_{n+1}], x>1$, we have

$$\begin{aligned}\log h(xy)-\log h(y) &= \log h(X_n)+\epsilon_n(\log x+\log y-\log X_n)\\ &\quad -\log h(X_n)-\epsilon_n(\log y-\log X_n)\\ &= \epsilon_n\log x\\ &\leq \epsilon\log x,\end{aligned}$$

with the inequality holding more generally for

$$x, y\in[X_{n+1}, X_{n+2}], [X_{n+2}, X_{n+3}], \ldots.$$

Raising both sides by e gives the desired inequality.

Step 3) Convergence for $s>\delta$.

We conclude the lemma by showing that $\sum h\left(a_p\right)\cdot a_p^{-s}$ converges if $s>\delta$. Let $\epsilon>0$ be such that $\delta+\epsilon<s$. By what was proved in the last step, there exists r_0 such that for $r>r_0$, $r>1$, and $a_p>X_1>1$, we have

$$h(a_p)\leq h(ra_p)\leq a_p^{\epsilon}h(r),$$

where we have used (in the left hand-side inequality) the fact that h is increasing. Then,

$$\sum h(a_p)a_p^{-s}\leq\sum h(r)a_p^{\epsilon-s}=h(r)\sum a_p^{-(s-\epsilon)}.$$

Since $s-\epsilon>\delta$, $\sum a_p^{-(s-\epsilon)}$ cannot diverge, otherwise δ would not be the critical exponent of the (absolute) Poincaré series. Thus, it converges, and by comparison, $\sum h(a_p)a_p^{-s}$ converges. □

The second property of Lemma 4.2.1 implies that for t in a bounded interval of $\mathbb{R}$, $\frac{h\left(e^{r+t}\right)}{h(e^r)}\to 1$ uniformly as $r\to\infty$.

Now, denote the modified $g_s(y,y)$ as

$$g_s^*(y,y) \equiv \sum_{\gamma\in\Gamma} e^{-sd_{\mathbb{D}}(y,\gamma y)} h\left(e^{d_{\mathbb{D}}(y,\gamma y)}\right)$$

and form the measure

$$\mu_{x,y,s} \equiv \frac{1}{g_s^*(y,y)} \sum_{\gamma\in\Gamma} e^{-sd_{\mathbb{D}}(x,\gamma y)} h\left(e^{d_{\mathbb{D}}(x,\gamma y)}\right) \delta_{\gamma y},$$

for $x, y \in \mathbb{D}, s > \delta$, where we have used $\delta_{\gamma y}$ to denote the Dirac mass of weight one at γy.

Using the definition of the Dirac mass, the $\mu_{x,y,s}$ measure of a Borel set E in $\bar{\mathbb{D}}$ is given by

$$\mu_{x,y,s}(E) = \frac{1}{g_s^*(y,y)} \sum_{\gamma\in\Gamma} e^{-sd_{\mathbb{D}}(x,\gamma y)} h\left(e^{d_{\mathbb{D}}(x,\gamma y)}\right) 1_E(\gamma y),$$

where 1_E denotes the characteristic function of E.

Having proved the existence of h with the desired properties, notice that if the (absolute) Poincaré series is of divergence type, we can simply redefine h as having constant unit value. In this case, the measure is simply written

$$\mu_{x,y,s}(E) = \frac{1}{g_s(y,y)} \sum_{\gamma\in\Gamma} e^{-sd_{\mathbb{D}}(x,\gamma y)} 1_E(\gamma y).$$

Although we are most interested in Fuchsian groups for which one can simply assume $h = 1$, at least in the initial sections of the present chapter h is taken to be a function in the vein of Lemma 4.2.1.

In the next two sections, we detail how changing the base point x in $\mu_{x,y,s}$, or the exponent s, affects the measure. Changing the exponent s, and in particular sending it to δ^+, is crucial to obtaining a measure on the limit set $\Lambda(\Gamma)$.

4.3 Changing the base point

We are interested in changing the point x in $\mu_{x,y,s}$ (as defined above), with y and $s > \delta$ fixed. The y-dependence of the measure is not so interesting, so it is kept implicit throughout. In other words, we write $\mu_{x,y,s} \equiv \mu_{x,s}$. Given $x' \in \mathbb{D}$, we will need to find a way to compare terms $e^{-sd_{\mathbb{D}}(x,y)}$ and $e^{-sd_{\mathbb{D}}(x',y)}$.

Before this is accomplished, let us introduce the Poisson kernel.

Definition 4.3.1. The Poisson kernel for $\mathbb{D}$ is defined as

$$P(z,q) \equiv \frac{1-|z|^2}{|z-q|^2},$$

where $z \in \mathbb{D}$ and $q \in \partial\mathbb{D}$.

Observe that there is nothing wrong with writing $P(z_1, z_2)$ for $z_1, z_2 \in \mathbb{D}$, as long as $z_1 \neq z_2$.

Base points and Poisson kernels are related via the next lemma.

Lemma 4.3.2. *Let $z, z', w \in \mathbb{D}$ and $q \in \partial\mathbb{D}$. Then*

$$\frac{e^{d_{\mathbb{D}}(z,w)}}{e^{d_{\mathbb{D}}(z',w)}} \to \frac{P(z',q)}{P(z,q)} \quad \text{as} \quad w \to q.$$

Proof. Using the 1-1 map from $\mathbb{H}$ onto $\mathbb{D}$ defined previously (see Section 2.1), it is possible to convert widely known expressions that are valid for $\mathbb{H}$ (e.g., the formula $\cosh d_{\mathbb{H}}(z,z')$) into formulas for $\mathbb{D}$.[9] We can then obtain, for example,

$$\sinh^2(d_{\mathbb{D}}(z,w)/2) = \frac{|z-w|^2}{(1-|z|^2)(1-|w|^2)},$$

and

$$d_{\mathbb{D}}(z,w) = \ln \frac{|1-z\bar{w}| + |z-w|}{|1-z\bar{w}| - |z-w|}.$$

Using the first formula for $q \in \partial\mathbb{D}$, observe that

$$\begin{aligned}
\frac{e^{d_{\mathbb{D}}(z,w)}}{e^{d_{\mathbb{D}}(z',w)}} \left(\frac{1 - \frac{2}{e^{d_{\mathbb{D}}(z,w)}} + \frac{1}{e^{2d_{\mathbb{D}}(z,w)}}}{1 - \frac{2}{e^{d_{\mathbb{D}}(z',w)}} + \frac{1}{e^{2d_{\mathbb{D}}(z',w)}}} \right) &= \frac{e^{d_{\mathbb{D}}(z,w)} - 2 + e^{-d_{\mathbb{D}}(z,w)}}{e^{d_{\mathbb{D}}(z',w)} - 2 + e^{-d_{\mathbb{D}}(z',w)}} \\
&= \frac{\sinh^2(d_{\mathbb{D}}(z,w)/2)}{\sinh^2(d_{\mathbb{D}}(z',w)/2)} \\
&= \frac{\frac{|z-w|^2}{(1-|z|^2)(1-|w|^2)}}{\frac{|z'-w|^2}{(1-|z'|^2)(1-|w|^2)}}.
\end{aligned}$$

Thus, taking the limit $w \to q$, we have

$$\begin{aligned}
&\lim_{w\to q} \frac{e^{d_{\mathbb{D}}(z,w)}}{e^{d_{\mathbb{D}}(z',w)}} \\
&= \lim_{w\to q} \left[\left(\frac{1 - \frac{2}{e^{d_{\mathbb{D}}(z,w)}} + \frac{1}{e^{2d_{\mathbb{D}}(z,w)}}}{1 - \frac{2}{e^{d_{\mathbb{D}}(z',w)}} + \frac{1}{e^{2d_{\mathbb{D}}(z',w)}}} \right)^{-1} \frac{|z-w|^2}{(1-|z|^2)} \frac{\left(1-|z'|^2\right)}{|z'-w|^2} \right] \\
&= \frac{P(z',q)}{P(z,q)},
\end{aligned}$$

where we have used $\lim_{w\to q} \left(\frac{1 - \frac{2}{e^{d_{\mathbb{D}}(z,w)}} + \frac{1}{e^{2d_{\mathbb{D}}(z,w)}}}{1 - \frac{2}{e^{d_{\mathbb{D}}(z',w)}} + \frac{1}{e^{2d_{\mathbb{D}}(z',w)}}} \right)^{-1} = 1$, the properties of limits, and continuity. □

We can compare the size of $\mu_{x,s}(E), \mu_{x',s}(E)$, for a Borel set E contained in a small neighborhood of $\xi \in \partial\mathbb{D}$, using the next result.

Theorem 4.3.3. *Let Γ be a Fuchsian group with critical exponent δ and suppose $s > \delta$. Choose $x, x' \in \mathbb{D}$ and $\xi \in \partial\mathbb{D}$. Let E be a Borel subset of $\bar{\mathbb{D}}$ contained in a small neighborhood of ξ and, for $t > 0$, let $E(t)$ be the part of E within a Euclidean distance t of ξ. For $\epsilon > 0$ there exists $t(\epsilon)$ such that if $t < t(\epsilon)$, the inequalities*

$$\begin{aligned}
\left[\left(\frac{P(x,\xi)}{P(x',\xi)} \right)^s - \epsilon \right] (1-\epsilon)\mu_{x',s}(E(t)) &\le \mu_{x,s}(E(t)) \\
&\le \left[\left(\frac{P(x,\xi)}{P(x',\xi)} \right)^s + \epsilon \right] (1+\epsilon)\mu_{x',s}(E(t))
\end{aligned}$$

[9] Beardon does this for Theorem 7.2.1. (in [6], pp. 130-131).

hold.

Proof. Suppose that $U(t)$ is the part of $\bar{\mathbb{D}}$ within a Euclidean distance t of ξ. Then, $E(t) \subseteq U(t)$, which by definition of the measure leads to

$$\mu_{x,s}(E(t)) = \frac{1}{g_s^*(y,y)} \sum_{\gamma\in\Gamma,\gamma y\in U(t)} e^{-sd_{\mathbb{D}}(x,\gamma y)} h\left(e^{d_{\mathbb{D}}(x,\gamma y)}\right) 1_{E(t)}(\gamma y).$$

This can be rewritten as

$$\frac{1}{g_s^*(y,y)} \sum_{\gamma\in\Gamma,\gamma y\in U(t)} \frac{e^{-sd_{\mathbb{D}}(x,\gamma y)}}{e^{-sd_{\mathbb{D}}(x',\gamma y)}} \frac{h\left(e^{d_{\mathbb{D}}(x,\gamma y)}\right)}{h\left(e^{d_{\mathbb{D}}(x',\gamma y)}\right)} e^{-sd_{\mathbb{D}}(x',\gamma y)} h\left(e^{d_{\mathbb{D}}(x',\gamma y)}\right) 1_{E(t)}(\gamma y).$$

For a given $\epsilon > 0$, we can use both the observation made after Lemma 4.2.1 and Lemma 4.3.2 to find a $t(\epsilon)$ such that if $t < t(\epsilon)$ and $\gamma y \in U(t)$, we have

$$\left|\frac{h\left(e^{d_{\mathbb{D}}(x,\gamma y)}\right)}{h\left(e^{d_{\mathbb{D}}(x',\gamma y)}\right)} - 1\right| < \epsilon \quad \text{and} \quad \left|\left(\frac{e^{d_{\mathbb{D}}(x,\gamma y)}}{e^{d_{\mathbb{D}}(x',\gamma y)}}\right)^{-s} - \left(\frac{P(x',\xi)}{P(x,\xi)}\right)^{-s}\right| < \epsilon.$$

Substituting these estimates in the expression for $\mu_{x,s}(E(t))$ yields the result. □

As a corollary, we get the following limit.

Corollary 4.3.4. *With x, x' and $E(t)$ as in the previous theorem, we have*

$$\lim_{t\to 0} \frac{\mu_{x,s}(E(t))}{\mu_{x',s}(E(t))} = \left(\frac{P(x,\xi)}{P(x',\xi)}\right)^s.$$

Proof. Using the estimate in Theorem 4.3.3, divide all sides of the inequality by $\mu_{x',s}(E(t))$ and send ϵ to zero. Notice that by the definition of $\mu_{x',s}$, we cannot have $\mu_{x',s}(E(t)) = 0$, regardless of how small $E(t)$ is, since the weights are exponential functions (i.e., never zero) and the Dirac mass does not vanish if one goes sufficiently far in the direction of the boundary at infinity (i.e., the limit point ξ is in $E(t)$). □

The measures $\mu_{x,s}$ and $\mu_{x',s}$ still do not have any mass on $\partial\mathbb{D}$, as we have not yet taken the limit $s \to \delta^+$. In this sense, it is too early to speak of $\frac{d\mu_{x,s}}{d\mu_{x',s}}(\xi)$ as a Radon-Nikodym derivative. Nonetheless, the previous corollary gives an idea of the relative sizes of $\mu_{x,s}$ and $\mu_{x',s}$ near ξ. Later, the quotient $\frac{P(x,\xi)}{P(x',\xi)}$ will reappear in the context of a genuine Radon-Nikodym derivative.

There exists an invariance property concerning the measures $\mu_{x,s}$ and $\mu_{x',s}$ when x, x' are Γ-equivalent (i.e., when $\exists\gamma \in \Gamma$ such that $\gamma x = x'$). Given a positive finite measure μ on $\bar{\mathbb{D}}$, a Borel set E in $\bar{\mathbb{D}}$, and any Möbius transform γ, we can define a new pullback measure[10] $\gamma^*\mu$ through

$$\gamma^*\mu(E) \equiv \mu(\gamma(E)).$$

[10]Although a common practice in part of the literature (e.g., in Borthwick [9] and Nicholls [53]), pullback measures generated by Möbius transforms $\gamma \in \Gamma$ are not as intuitive as pushforward measures. If using the pushforward measure, we have

$$\gamma_*\mu(E) \equiv \mu(\gamma^{-1}(E)).$$

Observe in Theorem 4.3.5 that the pullback by γ assumes that the inverse γ^{-1} exists, whereas γ^{-1} in the definition of the pushforward measure above denotes the preimage, which always exists. The non-existence of an inverse γ^{-1} is not a problem in our setting, since Γ is assumed to be a group, as seen. Thus, we chose to maintain the pullback practice, despite its recognized inelegance.

Theorem 4.3.5. *Let Γ be a Fuchsian group with critical exponent $\delta, s > \delta$ and $x \in \mathbb{D}$. Then, for any $\gamma \in \Gamma$,*

$$\gamma^* \mu_{x,s} = \mu_{\gamma^{-1}x,s}.^{11}$$

Proof. Assume that γ is a Möbius transformation that preserves $\mathbb{D}$ and that $\gamma \in \Gamma$. By definition of the pullback measure and the definition of the measure,

$$\gamma^* \mu_{x,s}(E) \equiv \mu_{x,s}(\gamma(E)) = \frac{1}{g_s^*(y,y)} \sum_{T \in \Gamma} e^{-s d_{\mathbb{D}}(x,Ty)} h\left(e^{d_{\mathbb{D}}(x,Ty)}\right) 1_{\gamma(E)}(Ty).$$

Let $\eta = \gamma^{-1}T$. Then $Ty \in \gamma(E)$ if and only if $\gamma^{-1}Ty \in E$, that is, if and only if $\eta y \in E$. In other words, we can change the characteristic function. This does not affect the sum, because if $\gamma \in \Gamma$, then as T runs over Γ in the original sum, η runs over Γ by definition. Thus, we can write

$$\begin{aligned}\gamma^* \mu_{x,s}(E) \equiv \mu_{x,s}(\gamma(E)) &= \frac{1}{g_s^*(y,y)} \sum_{\eta \in \Gamma} e^{-s d_{\mathbb{D}}(\gamma^{-1}x,\eta y)} h\left(e^{d_{\mathbb{D}}(\gamma^{-1}x,\eta y)}\right) 1_E(\eta y) \\ &= \mu_{\gamma^{-1}x,s}(E).\end{aligned}$$

□

4.4 Changing the exponent

Now we study the behavior of the measure $\mu_{x,s}$ when x is fixed and s changes. Specifically, we are interested in what happens to $\mu_{x,s}$ as s approaches δ^+. By construction, the measures $\mu_{x,s}$ are probability measures on $\bar{\mathbb{D}}$. Therefore, Prokhorov's Theorem (sometimes under different names) is invoked with respect to the collection $\{\mu_{x,s} : \delta < s < \delta + 1\}$, in order to obtain a weakly convergent subsequence of probability measures.

Theorem 4.4.1 (Prokhorov's Theorem)**.** *Let X be a complete separable metric space and let $\mathcal{M}$ be a family of Borel measures on X. Then the following conditions are equivalent:*

1. *every sequence $\{\mu_n\} \subset \mathcal{M}$ contains a weakly convergent subsequence;*
2. *the family $\mathcal{M}$ is (uniformly) tight and uniformly bounded in the variation norm.*

Proof. See Theorem 8.6.2. in Bogachev ([7], p. 202). □

Note that $\bar{\mathbb{D}}$ is a complete separable metric space. Moreover, since $\bar{\mathbb{D}}$ has a metric and is compact, any sequence of measures on $\bar{\mathbb{D}}$ is (uniformly) tight. Thus, only the requirement of uniform boundedness in variation norm needs to be checked to deduce the desired weak convergence. Instead of delving into the matter of total variation of measures, we simply show the following fact.

[11] Using the pushforward, this result would read

$$\gamma_* \mu_{x,s} = \mu_{\gamma x,s}.$$

Lemma 4.4.2. *Let Γ be a Fuchsian group with critical exponent δ. For $x \in \mathbb{D}$ fixed, the family of measures $\{\mu_{x,s} : \delta < s < \delta + 1\}$ on $\bar{\mathbb{D}}$ is bounded independently of s for s in the range described.*

Proof. Since all orbits are necessarily in $\mathbb{D} \subset \bar{\mathbb{D}}$, by definition we have

$$\mu_{x,s}(\bar{\mathbb{D}}) = \frac{\sum_{\gamma\in\Gamma} e^{-sd_{\mathbb{D}}(x,\gamma y)} h\left(e^{d_{\mathbb{D}}(x,\gamma y)}\right)}{\sum_{\gamma\in\Gamma} e^{-sd_{\mathbb{D}}(y,\gamma y)} h\left(e^{d_{\mathbb{D}}(y,\gamma y)}\right)}.$$

As seen in Lemma 4.2.1, the convergence of the modified Poincaré series does not depend on the base point. Thus, this implies that both the numerator and the denominator in the above fraction converge, given $s > \delta$ in the described range. Given Section 4.1, the triangle inequality implies $e^{-sd_{\mathbb{D}}(x,\gamma y)} \leq e^{sd_{\mathbb{D}}(x,y)} e^{-sd_{\mathbb{D}}(y,\gamma y)}$, and so the numerator of the above fraction cannot exceed

$$e^{sd_{\mathbb{D}}(x,y)} \sum_{\gamma\in\Gamma} e^{-sd_{\mathbb{D}}(y,\gamma y)} \frac{h\left(e^{d_{\mathbb{D}}(x,\gamma y)}\right)}{h\left(e^{d_{\mathbb{D}}(y,\gamma y)}\right)} h\left(e^{d_{\mathbb{D}}(y,\gamma y)}\right).$$

Based on a consequence of the second part of Lemma 4.2.1, for all except finitely many terms in the series,

$$\frac{h\left(e^{d_{\mathbb{D}}(x,\gamma y)}\right)}{h\left(e^{d_{\mathbb{D}}(y,\gamma y)}\right)} < 2.$$

Notice that the finitely many terms for which the above bound does not hold can still be bounded independently of s, using the fact $s > \delta$ implies $-s < -\delta$.

Substituting this quantity back into the numerator leads to the bound

$$\mu_{x,s}(\bar{\mathbb{D}}) \leq Ce^{sd_{\mathbb{D}}(x,y)},$$

where C depends only on finite hyperbolic $d_{\mathbb{D}}$-distances and not on s. We similarly obtain a lower bound

$$\mu_{x,s}(\bar{\mathbb{D}}) \geq C'e^{-sd_{\mathbb{D}}(x,y)},$$

and the proof is complete. □

One then applies Prokhorov's Theorem for a decreasing sequence $(s_n)_{n\in\mathbb{N}}$ of $s_n \in (\delta, \delta + 1)$, so that $\mu_{x,s_n} \overset{s_n\to\delta^+}{\longrightarrow} \mu_x$, for $x \in \mathbb{D}$, where μ_x is a measure on $\bar{\mathbb{D}}$ and convergence occurs in the weak topology. We cannot assume that the measure obtained in this way is unique, since different sequences may yield different (weak) limits. The uniqueness of the limit measure depends on the group Γ. In fact, it is closely related to notions of ergodicity. We therefore postpone the discussion of the uniqueness of the Patterson-Sullivan measures to Section 5.5. For the moment, then, we need to take into account the entire family of limit measures: given $x \in \mathbb{D}$ and a monotonically decreasing sequence s_n tending to δ^+, M_x denotes the collection of measures μ_x defined on $\bar{\mathbb{D}}$.

Before showing the relationship between elements of M_x and $M_{x'}$, for $x' \in \mathbb{D}$, we show that these (weak) limit measures are indeed orbital measures.

Theorem 4.4.3. *Let Γ be a Fuchsian group with critical exponent δ. For any $x \in \mathbb{D}$ and any $\mu_x \in M_x$, the measure μ_x is concentrated on the limit set of $\Lambda(\Gamma)$.*

Proof. Due to the use of the Dirac mass in the definition of the measure μ_{x,s_j} (and its interpretation in terms of an indicator function), it is sufficient to prove that an arbitrary disc B centered at some γy and containing no other Γ-image of y, has a null μ_x measure. After all, the Dirac density only assigns mass to Borel sets that are intercepted by Γ-orbits. The fact that it is possible to find such a disc is of course a consequence of the discreteness of Γ, assumed a Fuchsian group.

Since B is open (and thus Borel) we have, as a consequence of Prokhorov's Theorem, that

$$\liminf_{n\to\infty} \mu_{x,s_n}(B) \geq \mu_x(B)$$

where $\{s_n\}$ is the monotonically decreasing sequence to δ^+, such that $\mu_{x,s_n} \to \mu_x$.

The disc has no other Γ-image of y, so

$$\mu_{x,s_n}(B) = \frac{e^{-s_n(x,\gamma y)} h\left(e^{(x,\gamma y)}\right)}{g^*_{s_n}(y,y)}.$$

And, by continuity of the quotient and the numerator, it remains only to show that

$$\lim_{n\to\infty} g^*_{s_n}(y,y) = \infty.$$

Notice that the measures μ_{x,s_n} were defined for $s > \delta$. So, we cannot use continuity to write $\lim_{n\to\infty} g^*_{s_n}(y,y) = g^*_\delta(y,y) = \infty$ because $\mu_{x,s}$ is not defined for $s = \delta$. Still, we can rewrite the limit as a divergent series as follows.

Keeping y fixed, we use the fact that a Fuchsian group is countable to write $\Gamma = \{\gamma_m\}, m = 1, 2, 3, \ldots$, in such a way that the sequence $(\lambda_m)_{m\in\mathbb{N}}$, where $\lambda_m = d_{\mathbb{D}}(y, \gamma_m(y))$, is nondecreasing. Let $a_m = h\left(e^{d_{\mathbb{D}}(y,\gamma_m(y))}\right)$ and recall that $a_m \geq 0, \forall m$ (Lemma 4.2.1). Then we have the Dirichlet series[12]

$$f(z) = \sum_{m=1}^{\infty} a_m e^{-\lambda_m z},$$

which is known to be analytic in the half-plane $\operatorname{Re} z > \delta$.[13] Moreover, the point $s = \delta$ is a singularity for $f(s)$.[14] Thus, if s is real and monotonic decreasing to δ, $f(s)$ is unbounded as required. □

4.5 Relationship between limit measures

For any $x \in \mathbb{D}$, the set M_x (as defined in the previous section) can be understood as a topological space (when equipped with the weak convergence topology). Then, it is possible to deduce a homeomorphism between M_x and $M_{x'}$, for $x' \in \mathbb{D}$ as follows.

[12] See Titchmarsh ([75], p. 289).
[13] See Titchmarsh ([75], p. 290).
[14] See Titchmarsh ([75], p. 294).

Theorem 4.5.1. *Let Γ be a Fuchsian group with critical exponent δ. Choose $x, x' \in \mathbb{D}$ and for $\nu_x \in M_x$ define a new measure $\phi(\nu_x)$ by*

$$\phi(\nu_x)(E) = \int_E \left(\frac{P(x',\xi)}{P(x,\xi)}\right)^\delta d\nu_x(\xi)$$

Then ϕ is a homeomorphism of M_x onto $M_{x'}$. The inverse map is given by

$$\phi^{-1}(\nu_{x'})(E) = \int_E \left(\frac{P(x,\xi)}{P(x',\xi)}\right)^\delta d\nu_{x'}(\xi)$$

Further, if ν_{x,s_j} converges weakly to ν_x as $s_j \to \delta^+$, then ν_{x',s_j} converges weakly to $\phi(\nu_x)$.

Proof. Using Theorem 4.3.3 while swapping the roles of x and x', we have

$$\left[\left(\frac{P(x',\xi)}{P(x,\xi)}\right)^s - \epsilon\right](1-\epsilon)\mu_{x,s}(E(t)) \le \mu_{x',s}(E(t)) \le \left[\left(\frac{P(x',\xi)}{P(x,\xi)}\right)^s + \epsilon\right](1+\epsilon)\mu_{x,s}(E(t)).$$

Take a sequence $\{s_j\}$ of values of s, decreasing to δ, for which μ_{x,s_j} converges weakly to ν_x. Now, take two subsequences $\{s_{j_k}\}, \{s_{j_p}\}$ of the original sequence. Assume that for these subsequences, the measures $\mu_{x',s}$ converge weakly to $\nu_{x'}$ and $\sigma_{x'}$, respectively. From the above inequalities, we get

$$\left[\left(\frac{P(x',\xi)}{P(x,\xi)}\right)^\delta - \epsilon\right](1-\epsilon)\nu_x(E(t)) \le \nu_{x'}(E(t)) \le \left[\left(\frac{P(x',\xi)}{P(x,\xi)}\right)^\delta + \epsilon\right](1+\epsilon)\nu_x(E(t)),$$

with similar ones holding with $\nu_{x'}(E(t))$ replaced by $\sigma_{x'}(E(t))$.

Taking the limit $t \to 0$ shows that $\nu_{x'}, \nu_x$ are absolutely continuous with respect to each other,[15] that is, $\nu_{x'} \ll \nu_x$ and $\nu_{x'} \gg \nu_x$. Similarly, $\sigma_{x'}, \nu_x$ are absolutely continuous with respect to each other. Furthermore, the Radon-Nikodym derivatives are given by

$$\frac{d\nu_{x'}}{d\nu_x}(\xi) = \frac{d\sigma_{x'}}{d\nu_x}(\xi) = \left(\frac{P(x',\xi)}{P(x,\xi)}\right)^\delta.$$

Because the Radon-Nikodym derivatives are unique up to a ν_x-null set, it follows that $\nu_{x'}, \sigma_{x'}$ are the same and so $\phi(\nu_x) = \nu_{x'}$. This proves the last statement of the theorem and shows that ϕ is a map from M_x into $M_{x'}$.

Reversing the roles of x and x', notice that the map ψ on $M_{x'}$ given by

$$\psi(\nu_{x'})(E) = \int_E \left(\frac{P(x,\xi)}{P(x',\xi)}\right)^\delta d\nu_{x'}(\xi)$$

[15] If μ and ν are two measures on the same measurable space (X, Σ), μ is said to be absolutely continuous with respect to ν if $\mu(A) = 0$ for every set $A \in \Sigma$ for which $\nu(A) = 0$. This is usually denoted by $\mu \ll \nu$.

is a map from $M_{x'}$ into M_x. Using the property

$$\frac{d\nu_x}{d\nu_{x'}}(\xi) = \left(\frac{d\nu_{x'}}{d\nu_x}(\xi)\right)^{-1} = \left(\frac{P(x',\xi)}{P(x,\xi)}\right)^{-\delta} = \left(\frac{P(x,\xi)}{P(x',\xi)}\right)^{\delta},$$

we have

$$\psi(\nu_{x'})(E) = \int_E \left(\frac{P(x,\xi)}{P(x',\xi)}\right)^{\delta} d\nu_{x'}(\xi) = \int_E \left(\frac{d\nu_x}{d\nu_{x'}}(\xi)\right) d\nu_{x'}(\xi) = \nu_x(E),$$

which implies that

$$\phi(\psi(\nu_{x'})) = \phi(\nu_x) = \nu_{x'},$$

and similarly $\psi(\phi(\nu_x)) = \nu_x$. Thus, $\psi = \phi^{-1}$.

It remains only to establish the (sequential) continuity of ϕ with respect to the weak topology (the continuity of ϕ^{-1} can be established analogously). Suppose that $\{\nu_x^j\}$, $j = 1, 2, 3, \ldots$, is a sequence of measures in M_x that converges weakly to ν_x in M_x. The goal is to show that $\phi(\nu_x^j)$ converges (weakly) to $\phi(\nu_x)$ for any compactly supported continuous function. Let f be continuous with compact support on $\partial\mathbb{D}$. Then the product $f(\xi)\left(P(x',\xi)/P(x,\xi)\right)^{\delta}$ is also continuous and compactly supported on $\partial\mathbb{D}$ and so, by (weak) convergence $\nu_x^j \to \nu_x$,

$$\int_{\partial\mathbb{D}} f(\xi)\left(\left(\frac{P(x',\xi)}{P(x,\xi)}\right)^{\delta}\right) d\nu_x^j(\xi) \to \int_{\partial\mathbb{D}} f(\xi)\left(\left(\frac{P(x',\xi)}{P(x,\xi)}\right)^{\delta}\right) d\nu_x(\xi).$$

Now, the Radon-Nikodym derivative allows us to rewrite this convergence as

$$\int_{\partial\mathbb{D}} f d\nu_{x'}^j \to \int_{\partial\mathbb{D}} f d\nu_{x'},$$

where $\nu_{x'}^j = \phi\left(\nu_x^j\right)$ and $\nu_{x'} = \phi(\nu_x)$. It follows that $\phi\left(\nu_x^j\right)$ converges weakly to $\phi(\nu_x)$ and so ϕ is continuous in the weak topology, thus a homeomorphism. □

We follow Nicholls ([53], p. 54) and adopt the convention of using the same Greek letter for measures in M_x and $M_{x'}$ if and only if they are equivalent under the homeomorphism ϕ between these spaces. In practice, this means that $\nu_x \in M_x$ and $\nu_{x'} \in M_{x'}$ imply $\nu_{x'} = \phi(\nu_x)$.

4.6 Patterson-Sullivan measure distortion

Now we wish to express the relationship between measures $\nu_x \in M_x, \nu_{x'} \in M_{x'}$, for $x, x' \in \mathbb{D}$, using a Möbius transformation γ that preserves the Poincaré disc and such that $\gamma(x) = x'$. We emphasize that it is not necessary for γ to be an element of Γ, the Fuchsian group that is used to define the measures that converge to the Patterson-Sullivan measures $\nu_x, \nu_{x'}$.

A simple computation shows that for any Möbius transformation γ, and $z, w \in \mathbb{H}$, the relationship $|\gamma(z)-\gamma(w)|^2 = |\gamma'(z)|\,|\gamma'(w)|\,|z-w|^2$ holds. Moreover, this formula is valid in the context of $\mathbb{D}$. Knowing this, it is possible to deduce a transformation rule for the Poisson kernel, starting from

$$P(\gamma(z),\gamma(q))\,|\gamma'(q)| = \frac{1-|\gamma(z)|^2}{|\gamma(z)-\gamma(q)|^2}\,|\gamma'(q)| = \frac{1-|\gamma(z)|^2}{|z-q|^2\,|\gamma'(z)|},$$

for any $z \in \mathbb{D}, q \in \partial\mathbb{D}$.

The term $|\gamma'(z)|$ can be calculated using the transformation coefficients. To obtain a formula in the Poincaré disc setting we use the quotient rule: $|\gamma'(z)| = \left|\frac{\alpha(\bar{\beta}z-\bar{\alpha})-(\alpha z+\beta)\bar{\beta}}{|\bar{\beta}z+\bar{\alpha}|^2}\right| = \frac{1}{|\bar{\beta}z+\bar{\alpha}|^2}$, for $\alpha, \beta \in \mathbb{C}$. Recall that we can assume $|\alpha|^2 - |\beta|^2 = 1$. Substituting this into the expression above and writing $|\gamma(z)|$ in terms of the coefficients, we can simplify the ratio and obtain the Poisson kernel. That is, we find

$$P(\gamma(z), \gamma(q))\,|\gamma'(q)| = P(z, q)$$

Thus, we can use this formula for $\gamma(x) = x' \Leftrightarrow x = \gamma^{-1}(x')$ and $q \in \partial\mathbb{D}$, obtaining

$$|\gamma'(q)| = \frac{P(\gamma^{-1}(x'), q)}{P(x', \gamma(q))}.$$

For simplicity, we take a reference point $x' = 0$, so that $P(x', \gamma(q)) = P(0, \gamma(q)) \overset{\gamma(q)\in\partial\mathbb{D}}{=} 1 = P(0, q)$. Then, the expression

$$|\gamma'(q)| = \frac{P(\gamma^{-1}(0), q)}{P(0, \gamma(q))} = \frac{P(\gamma^{-1}(0), q)}{P(0, q)}$$

combined with Theorem 4.5.1 return

$$\nu_x(E) = \nu_{\gamma^{-1}(0)}(E) = \int_E |\gamma'(\xi)|^\delta \, d\nu_0(\xi),$$

which shows how the critical exponent regulates the relationship between measures $\nu_x \in M_x$ and $\nu_0 \in M_0$. It also shows how any Paterson-Sullivan measure with base point x can be written in terms of the Patterson-Sullivan measure of base point zero given a Möbius transformation γ, not necessarily in Γ, such that $\gamma(x) = 0$ (provided the measures are related by the homeomorphism between the measure spaces M_x and M_0).

In addition, if γ is an element of Γ, we can verify that the critical exponent in this case is equally important in controlling the distortion of a Möbius transformation affecting the measure ν_0. That is, by the definition of $\gamma^*\nu_0$ and since Theorem 4.3.5 has as a consequence $\gamma^*\nu_0 = \nu_{\gamma^{-1}(0)}$, we obtain the expression

$$\nu_0(\gamma(E)) \equiv \gamma^*\nu_0(E) \overset{\gamma\in\Gamma}{=} \nu_{\gamma^{-1}(0)}(E) = \int_E |\gamma'(\xi)|^\delta d\nu_0(\xi).$$

The appearance of δ as such an exponent occurs in analogous relationships that can be deduced for the more general case where x' in $\mu_{x'} \in M_{x'}$ is not necessarily zero. In this situation, for some other homeomorphic measure $\mu_x \in M_x$, and E a Borel set of $\partial\mathbb{D}$, we have

$$\mu_{x'}(\gamma(E)) = \gamma^*\mu_{x'}(E) = \mu_{\gamma^{-1}(x')}(E) = \mu_x(E) = \int_E |\gamma'_{x'}(\xi)|^\delta \, d\mu_{x'}(\xi),$$

and the derivative $|\gamma'_{x'}|$ is not simply the derivative of the Möbius transformation, but a quantity derived from the metric obtained on $\partial\mathbb{D}$ by radial projection.[16]

We will not rely on this more general result in what follows, so we summarize the developments of the section in the lemma hereunder.

[16]The details concerning this more general derivative can be found in Lemma 3.4.2 in Nicholls' book ([53], p. 55). We will revisit this general case in Section 4.9.

Lemma 4.6.1. *Let E be a Borel subset of $\partial\mathbb{D}$ and let $\mu \equiv \mu_0$ denote a Patterson-Sullivan measure constructed from a given Fuchsian group Γ with a critical exponent $\delta \equiv \delta(\Gamma)$. For any $x \in \mathbb{D}$ and any orientation-preserving Möbius transformation γ of the unit disc, with $\gamma(x) = 0$, the transformation rule*

$$\mu_x(E) = \int_E |\gamma'(\xi)|^\delta \, d\mu(\xi),$$

applies. Moreover, if $\gamma \in \Gamma$, then

$$\mu(\gamma(E)) = \gamma^*\mu(E) = \mu_{\gamma^{-1}(0)}(E) = \mu_x(E) = \int_E |\gamma'(\xi)|^\delta \, d\mu(\xi).$$

Under the conditions of the above lemma, it is worthwhile to define the kernel-related function $F : \mathbb{D} \to \mathbb{R}$ given by

$$F(z) \equiv \int_{\partial\mathbb{D}} P(z,\xi)^\delta d\mu(\xi),$$

which can be shown to be invariant under Γ.

Proposition 4.6.2 (Γ-invariance of F)**.** *For $T \in \Gamma$, with Γ a Fuchsian group acting on $\mathbb{D}$, we have $F(Tz) = F(z)$.*

Proof. Take an arbitrary $T \in \Gamma$. By the transformation rule for the Poisson kernel, we have $P(Tz, Tq)\,|T'(q)| = P(z,q)$. Raising both sides to the power of δ and subsequently integrating with respect to the Patterson-Sullivan measure μ, we obtain

$$\begin{aligned}
\int_{\partial\mathbb{D}} P(z,q)^\delta d\mu(q) &= \int_{\partial\mathbb{D}} P(Tz,Tq)^\delta |T'q|^\delta d\mu(q) \\
F(z) &= \int_{\partial\mathbb{D}} P(Tz,Tq)^\delta \left(\frac{P(T^{-1}0,q)}{P(0,q)}\right)^\delta d\mu(q) \\
&= \int_{\partial\mathbb{D}} P(Tz,Tq)^\delta \frac{d\mu_{T^{-1}0}}{d\mu}(q) d\mu(q) \\
&= \int_{\partial\mathbb{D}} P(Tz,Tq)^\delta d\mu(Tq) \\
&= \int_{\partial\mathbb{D}} P(Tz,q')^\delta d\mu(q') \\
&= F(Tz),
\end{aligned}$$

where we have used the expression for $|T'q|$ found above and the Radon-Nikodym derivative (as in Theorem 4.3.3) when passing from the first to the third equality. The fourth equality is due to Lemma 4.6.1. The change of variables at the end is also justified, since for any $q \in \partial\mathbb{D}$, we have $Tq \in \partial\mathbb{D}$. Thus, the Γ-invariance is proved. □

In the remaining sections of the present chapter and all subsequent chapters (except when indicated otherwise), the Patterson-Sullivan measures have zero as a base point. The forthcoming results often focus on non-elementary geometrically finite Fuchsian groups Γ, so due to the alluded divergence of the (absolute) Poincaré series at $\delta(\Gamma)$ in this setting, it is also assumed that the Patterson-Sullivan measures have $h = 1$, where h is the function whose existence was proved in Lemma 4.2.1.

4.7 Usefulness of the Patterson-Sullivan measure

As an illustration of the usefulness of Patterson-Sullivan measures with regard to the critical exponent $\delta(\Gamma)$, we present two basic results that are somewhat different in nature. The first theorem was obtained by Beardon [4, 5] and addresses the aforementioned dependence of the range of $\delta(\Gamma)$ on the Fuchsian group Γ.

Theorem 4.7.1 (Beardon)**.** *For any non-elementary Fuchsian group* Γ *we have* $\delta > 0$*, and if* Γ *contains parabolic elements, then* $\delta > \frac{1}{2}$.

Proof. Assume $\mu \equiv \mu_0 \in M_0$ denotes a Patterson-Sullivan measure. We divide the proof into two parts.

Step 1) $\delta > 0$.

Suppose that $\delta = 0$. Then Lemma 4.6.1 shows that μ is a Γ-invariant measure on $\Lambda(\Gamma)$. If $T \in \Gamma$ is hyperbolic, with fixed points $q_\pm \in \partial\mathbb{D}$, then we can partition $\partial\mathbb{D} - \{q_\pm\}$ into a countably infinite collection $\cup_i U_i$ of disjoint intervals I_i that are mapped to each other by powers of T.That is, by removing points q_-, q_+ from the circle, we end up with two arcs that are homeomorphic to two open intervals on the real line, and these can be partitioned into a countably infinite union of disjoint open intervals. Since Möbius transformations map $\partial\mathbb{D}$ to itself, these intervals are mapped (under positive powers of T) to other intervals that approach q_+ and escape q_-. By invariance and because the total mass is finite, we conclude that $\mu\left(\partial\mathbb{D} - \{q_\pm\}\right) = 0$, otherwise we would have $\mu\left(\partial\mathbb{D} - \{q_\pm\}\right) = \mu\left(\cup_i I_i\right) = \sum_i \mu\left(I_i\right) > \infty$.

Why is this infinite sum diverging? The reason is the following. Notice that if we have $\mu(I_i) > 0$ for some i, then $\sum_i \mu\left(I_i\right) > \infty$, because for $j \neq i$, we have $\mu(I_j) = \mu(T^n I_i) = |(T^n)'|^{\delta=0}\mu(I_i) = \mu(I_i) > 0$ for some n, due to the invariance of μ, and this holds for any j. Therefore, the sum converges only if $\mu(I_i) = 0$, for all i. Now, since $\mu\left(\partial\mathbb{D} - \{q_\pm\}\right) = 0$ implies that $\mu\left(\{q_\pm\}\right) = 1$, we cannot have other limit points, since the whole Patterson-Sullivan probability measure is already assigned to $\{q\pm\}$ (recall that μ is fully concentrated on the boundary, by Theorem 4.4.3). Hence $\{q_\pm\}$ are the only two limit points of Γ and Γ is thus elementary. The same strategy also applies if Γ contains a parabolic element.

Step 2) $\delta > 1/2$ if Γ contains a parabolic element.

Next, assume that Γ is non-elementary and contains a parabolic element $T \in \mathrm{PSU}(2)$. By conjugation, we can assume that T fixes 1 (i.e., $T1 = 1$) and maps i to -1 (i.e., $Ti = -1$). From these equations, we obtain coefficients such that

$$T = \begin{pmatrix} 1 + \frac{i}{2} & -\frac{i}{2} \\ \frac{i}{2} & 1 - \frac{i}{2} \end{pmatrix}.$$

Then, by induction,

$$T^n = \begin{pmatrix} 1 + i\frac{n}{2} & -i\frac{n}{2} \\ i\frac{n}{2} & 1 - i\frac{n}{2} \end{pmatrix}.$$

According to a formula obtained above, the distortion of T^n is given by

$$|(T^n)'(z)| = \frac{1}{\left|(i\frac{n}{2})z + (1 - i\frac{n}{2})\right|^2} = \frac{1}{\left|1 + i\frac{n}{2}(z-1)\right|^2}.$$

Parabolic transformations move points (that are distinct from their sole fixed point) along the boundary $\partial\mathbb{D}$ in such a way that they distance themselves from the fixed point on one of its sides on the boundary and approach it from the other. That is, the movement resembles that of a full circle departing from and returning to the fixed point, except that the fixed point stays where it is. In this sense, applying the parabolic transformation T on $\xi \neq 1, \xi \in \partial\mathbb{D}$ multiple times means that one is moving the point around $\partial\mathbb{D}$ in a counterclockwise direction. Thus, if we let $E = \left\{e^{i\theta} : \frac{\pi}{2} < \theta \leq \pi\right\}$, we get $\{T^n E\}$ as a disjoint cover for $\partial\mathbb{D} - \{1\}$. We claim $\mu(E) > 0$. This holds because if $\mu(E) = 0$, then $\mu(T^n E) = |(T^n)'|^\delta \mu(E) = 0$, which would imply that the measure μ is entirely concentrated at the point 1. This contradicts the assumption of Γ being non-elementary. Thus, using what we have calculated above, we have

$$\begin{aligned}
\mu(\partial\mathbb{D} - \{1\}) &= \sum_n \mu\left(T^n E\right) \\
&= \sum_n \int_E \left|(T^n)'(\xi)\right|^\delta d\mu(\xi) \\
&= \sum_n \int_E \left(\frac{1}{\left|1 + i\frac{n}{2}(\xi - 1)\right|^2}\right)^\delta d\mu(\xi).
\end{aligned}$$

Now we can use the estimate $\left|1 + i\frac{n}{2}(\xi - 1)\right| \leq 1 + \left|\frac{n}{2}(\xi - 1)\right| \leq 1 + \frac{n}{2}(|\xi| + 1) \leq 1 + n$ to rewrite

$$\begin{aligned}
\mu(\partial\mathbb{D} - \{1\}) &= \sum_n \int_E \left(\frac{1}{\left|1 + i\frac{n}{2}(\xi - 1)\right|^2}\right)^\delta d\mu(\xi) \\
&\geq \sum_n \int_E \left(\frac{1}{(1+n)^2}\right)^\delta d\mu(\xi) \\
&= \mu(E) \sum_n \left(\frac{1}{(1+n)^2}\right)^\delta.
\end{aligned}$$

The left-hand side is bounded by 1 and $\mu(E) > 0$ so the series on the right-hand side must converge. This happens when $\delta > \frac{1}{2}$. □

The second result shows a connection between the critical exponent and the spectral theory of the Laplacian operator.

Proposition 4.7.2. *Given a Fuchsian group Γ acting on $\mathbb{D}$ with a critical exponent $\delta(\Gamma) = \delta$, the function F (introduced at the end of Section 4.6) behaves as an eigenfunction of the Laplacian operator defined on $L^2(X)$, where $X \cong \Gamma\backslash\mathbb{D}$.*[17] *That is,*

$$\Delta_X F = \delta(1 - \delta)F.$$

[17] It is an actual eigenfunction for $\delta > \frac{1}{2}$, which we saw holds whenever X is cusped (Theorem 4.7.1). See Sections 3, 4, and 5 in Patterson [55].

Proof sketch. Because hyperbolic surfaces are taken as quotients $X \cong \Gamma\backslash\mathbb{H}$ or $X \cong \Gamma\backslash\mathbb{D}$, functions on such surfaces are usually understood as Γ-invariant functions $\psi : \mathbb{H}$ or $\mathbb{D} \to \mathbb{C}$. It is straightforward to show that the Laplacian on $\mathbb{H}$ (or $\mathbb{D}$) is invariant under local isometries, so it descends as an operator on X, which we denoted above Δ_X. The spectral problem of hyperbolic surfaces is then concerned with the study of such invariant functions ψ that fulfill $\Delta_X\psi = \lambda\psi$, for a real λ, and some integrability condition that depends on the domain of the (unbounded) operator considered. The domain of Δ_X is assumed to be $L^2(X)$.

Now, recall the function F introduced in Section 4.6, given by

$$F(z) \equiv \int_{\partial\mathbb{D}} P(z,\xi)^\delta d\mu(\xi).$$

Proposition 4.6.2 showed that F is Γ-invariant, for F defined via the critical exponent δ associated with Γ. So, F can be thought of as a function defined on the hyperbolic surface X. We will argue that it fulfills $\Delta_X F = \delta(1-\delta)F$.

The Laplacian operator $\Delta_{\mathbb{H}}$ on $\mathbb{H}$ can be written, for $z = x + iy, z \in \mathbb{H}$, as $\Delta_{\mathbb{H}} = -y^2(\partial_x^2 + \partial_y^2)$.[18]

Notice that for any α, we have

$$\begin{aligned}\Delta_{\mathbb{H}} y^\alpha &= (-y^2(\partial_x^2 + \partial_y^2))y^\alpha \\ &= -y^2(\alpha(\alpha-1)y^{\alpha-2}) \\ &= \alpha(1-\alpha)y^\alpha.\end{aligned}$$

Based on the Cayley transform h, the inverse Cayley transform $h^{-1} : \mathbb{D} \to \mathbb{H}$ is defined by $z' \mapsto \frac{z'+1}{iz'-i}$. Now, for $z' = x' + iy', z' \in \mathbb{D}$, since

$$\operatorname{Im} h^{-1}(z') = \frac{1-|z'|^2}{|z'-1|^2} = P(z',1),$$

we see that the Laplacian operator $\Delta_{\mathbb{D}}$ is related to the Poisson kernel.

In fact, since by transitivity for any $\xi \in \partial\mathbb{D}$ there exists a transformation γ such that $\gamma(\xi) = 1$ implies $\xi = \gamma^{-1}(1)$, we have

$$P(z',\xi) = P(z',\gamma^{-1}(1)) = P(\gamma(z'),1).$$

Not only $P(z',1)^\alpha$ behaves as an eigenfunction of $\Delta_{\mathbb{D}}$ with the corresponding eingenvalue $\alpha(1-\alpha)$, but a known fact[19] on Möbius transformations also entails that $P(z',\xi)^\alpha$ behaves as an eigenfunction of the Laplacian $\Delta_{\mathbb{D}}$ (with the same corresponding eigenvalue).

Now, since this holds for any α, it is also valid for $\alpha = \delta$, for which we know F to be Γ-invariant. Then, a standard argument for interchanging the Laplacian with the integral in F can be used to conclude $\Delta_X F = \delta(1-\delta)F$.

Finally, it only remains to check that $F \in L^2(X)$, but this can be done if $\delta > \frac{1}{2}$.[20] □

[18] See Borthwick ([9], p. 12).

[19] See Theorem 5.1.1 in Nicholls' book ([53], p. 90).

[20] To avoid a lengthy detour into the topic, in the case of a geometrically finite non-elementary Fuchsian group Γ (and Lebesgue measure), we refer the reader to Theorem 5.3.3. in Nicholls ([53], p. 98) which gives a proof of this result (assuming $\delta > \frac{1}{2}$).

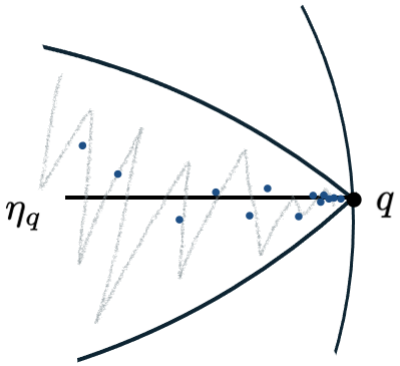


Figure 4.1: Radial limit point.

We barely scratched the surface on the relationship between the constructed Patterson-Sullivan measures, critical exponents, and the spectral theory of hyperbolic surfaces. We refer the reader to Borthwick [8, 9] for a comprehensive overview.

4.8 Patterson-Sullivan measures are atomless

For all practical purposes, an atom can be defined as follows.

Definition 4.8.1. Let $(\bar{\mathbb{D}}, \mathcal{B}(\bar{\mathbb{D}}), \mu)$ denote the measure space where $\mathcal{B}(\bar{\mathbb{D}})$ denotes the Borel σ-algebra of $\bar{\mathbb{D}}$ and $\mu \equiv \mu_0 \in M_0$ is a Patterson-Sullivan measure (as constructed in previous sections). If for some $x \in \bar{\mathbb{D}}$, the singleton $\{x\}$ has measure $\mu(x) \equiv \mu(\{x\}) > 0$, then x is an atom.

As we saw in Theorem 4.4.3, the measure μ is concentrated on the limit set $\Lambda(\Gamma)$. As a Fuchsian group Γ cannot have limit points in $\mathbb{D}$, the atoms of μ can only be found along $\partial\mathbb{D}$. This explains why the focus is on $\partial\mathbb{D}$ and on proving $\mu(q) = 0, q \in \Lambda(\Gamma)$, in what follows.

Let us now introduce some more definitions. We can characterize the points in $\Lambda(\Gamma)$ by the directions from which orbit points accumulate.

Definition 4.8.2. A point $q \in \Lambda(\Gamma)$ is called a radial limit point if there exists a geodesic ray η_q in $\mathbb{D}$ with endpoint q and an orbit Γw such that the set $\{z \in \Gamma w : d_{\mathbb{D}}(z, \eta_q) < r\}$ is infinite for some $r > 0$.

This concept is illustrated in Figure 4.1. Notice that the set that characterizes the radial limit point forms a conical region around η_q. This is why radial limit points are sometimes called conical limit points.

Intuitively, if q is the attractive hyperbolic fixed point of $T \in \Gamma$, then q is radial, since we can take η_q to be a portion of the axis of T. However, parabolic fixed points are not radial. To see why this must be the case, we first introduce an important fact.

Proposition 4.8.3. *Given a Fuchsian group Γ, a fixed point of some hyperbolic element of Γ cannot be fixed simultaneously by a parabolic element of Γ.*

Proof. Assume by conjugation that we have an arbitrary hyperbolic transformation $T \in \Gamma$, of the form $Tz = kz$, for $k \neq 1$, and a nontrivial parabolic transformation $P \in \Gamma$, of the form $Pz = z + b$, for $b \in \mathbb{R}$. Notice that both transformations fix ∞. We can also assume that $k > 1$ (otherwise we just consider the inverse transformation). In this case, we can study the set $A = \{T^{-n}PT^n \in \Gamma : T^{-n}PT^n i \in \overline{B(i, |b|)}\}$, where $\overline{B(i, |b|)}$ is a closed hyperbolic ball. Since $T^{-n}PT^n i = T^{-n}(k^n i + b) = i + \frac{b}{k^n}$, for n_0 sufficiently large such that $d_{\mathbb{H}}\left(i, i + \frac{b}{k^{n_0}}\right) < |b|$, $n > n_0$ implies that A has infinite cardinality. This contradicts the discreteness of Γ (via Theorem 2.8.2), since $\overline{B(i, |b|)}$ is a compact set. □

Having proved this proposition, we now have a clearer picture of why is it the case that parabolic points are not radial. Assuming that a given boundary point q' is a fixed point of a parabolic transformation T' of a Fuchsian group Γ, there cannot be a hyperbolic transformation in Γ whose axis terminates at q'. In other words, all hyperbolic transformations in Γ have axes that must terminate elsewhere in $\partial\mathbb{D}$. Thus, in the limit of infinite iterations, the hyperbolic elements of Γ will send an arbitrary point $w \in \mathbb{D}$ to these distinct fixed points and any cone around $\eta_{q'}$ will include a finite number of images of w under these orbits. On the other hand, the parabolic elements of Γ move points along circles that begin and end at a parabolic fixed point. No matter the choice of geodesic $\eta_{q'}$, the cone around it will contain at most a finite number of iterations of such parabolic transformations applied to w, since these circles may enter the radial cone from one side, but soon necessarily leave it from the other, including the circular trajectories drawn by T'.

Are there other possibilities in addition to radial limit points and parabolic fixed points? In our context, the answer is no, as the next result shows.

Proposition 4.8.4. *For Γ a geometrically finite Fuchsian group, all points in $\Lambda(\Gamma)$ are either parabolic fixed points or radial limit points.*

Proof. Let $\mathcal{F} \subset \mathbb{D}$ be a finite-sided fundamental region for Γ (which exists by geometric finiteness). The proof of geometric finiteness (Theorem 2.11.3) implies that the boundary of $\mathcal{F}$, $\partial\mathcal{F}$, meets $\Lambda(\Gamma)$ only at parabolic fixed points.[21] Let $\widetilde{K} \subset \mathbb{D}$ be the truncated Nielsen region. Recall that the truncated Nielsen region is given by $\widetilde{K} \equiv \widetilde{N} - (\cup O_p)$, where $\widetilde{N} \equiv \mathbb{D} - (\cup H_j)$ is the Nielsen region, with the union $\cup O_p$ taken with respect to the distinct parabolic fixed points $p \in \partial\mathbb{D}$ and the union $\cup H_j$ taken with respect to open half-planes related to each of the countable open intervals I_j (see Section 2.13 for the details). Suppose that γ is a geodesic ray originating in $\mathcal{F} \cap \widetilde{K}$ with endpoint $q \in \Lambda(\Gamma)$. There are two possibilities.

Case A) The geodesic ray γ crosses a finite number of translations of $\mathcal{F} \cap \widetilde{K}$.

Assume that γ crosses only finitely many translations of $\mathcal{F} \cap \widetilde{K}$ under the action of Γ. Then there must be a last face crossed. Say that this face is obtained by $T(\mathcal{F} \cap \widetilde{K})$, for some $T \in \Gamma$. This implies $q \in T(\partial\mathcal{F})$, and since $q \in \Lambda(\Gamma)$, it must therefore be a parabolic fixed point, because $\partial\mathcal{F}$ meets $\Lambda(\Gamma)$ only at

[21] This is most clear in Beardon's book (see Theorem 9.3.8. in [6], p. 224).

parabolic fixed points.[22]

Case B) The geodesic ray γ crosses infinitely many translations of $\mathcal{F} \cap \widetilde{K}$.

Now assume that γ crosses infinitely many images of $\mathcal{F} \cap \widetilde{K}$. In other words, we can find sequences of distinct $w_n \in \mathcal{F} \cap \widetilde{K}$ and $T_n \in \Gamma$ such that $T_n w_n \in \gamma$ for $n \in \mathbb{N}$. Since $\mathcal{F} \cap \widetilde{K}$ is compact (see Theorem 2.13.6), by passing to a subsequence we can assume that $w_n \to w \in \mathcal{F} \cap \widetilde{K}$. Then $d_{\mathbb{D}}(T_n w, \gamma)$ is bounded, since $d_{\mathbb{D}}(T_n w, \gamma) \equiv \inf_{z' \in \gamma} d_{\mathbb{D}}(T_n w, z') \leq d_{\mathbb{D}}(T_n w, T_n w_n)$ for all n, and $d_{\mathbb{D}}(T_n w, T_n w_n) = d_{\mathbb{D}}(w, w_n)$ converges. Since convergence implies boundedness, with some bound r, the set $\{z \in \Gamma w : d_{\mathbb{D}}(z, \gamma) < r\}$ has infinite cardinality, because the $T_n \in \Gamma$ are distinct. Thus, q is a radial limit point. □

We now prove the main result of this section.

Theorem 4.8.5. *For Γ non-elementary and geometrically finite, the measure μ has no atoms.*

Proof. According to Proposition 4.8.4, there are two types of limit point to consider and we show that they have measure zero when taken as singletons.

Case A) The limit point $q \in \Lambda(\Gamma)$ is a radial limit point.

Suppose that $q \in \Lambda(\Gamma)$ is a radial limit point. We can assume that $q = 1$ (as we can conjugate Γ) and we take a geodesic ray γ ending at 1. According to the definition of radial limit point, we can find a sequence of distinct elements $T_n \in \Gamma$, such that $d_{\mathbb{D}}\left(T_n^{-1}0, \gamma\right) < C$, for some constant C.

Observe that the sequence $\{T_n^{-1}0\}$ inside the conical region can only accumulate in the direction of 1, otherwise by truncating the cone, it would be possible to cover the complement of this truncated part containing infinite elements of the sequence with a closed hyperbolic disc, contradicting the discreteness of Γ (Theorem 2.8.2). Thus, the points in the sequence get closer and closer to 1 with respect to the Euclidean distance (i.e., they are approaching the boundary of the unit disc). In other words, $|T_n^{-1}0 - 1| \to 0$.

According to the transformation rule of the Poisson kernel, we have

$$|T_n' 1| = \frac{P(T_n^{-1}0, 1)}{P(0, T_n 1)} = P\left(T_n^{-1}0, 1\right) = \frac{1 - \left|T_n^{-1}0\right|^2}{\left|T_n^{-1}0 - 1\right|^2}.$$

Recall that $T_n 1$ is on the boundary of $\mathbb{D}$, since Γ can be understood as a discrete subgroup of PSU(2), thus preserving the unit disc. This justifies $P(0, T_n 1) = 1$. Now, the fact that $T_n^{-1}0$ approaches 1 within the sector implies that $|T_n' 1| \overset{n \to \infty}{\longrightarrow} \infty$. By Lemma 4.6.1,

$$\mu(\{T_n 1\}) = \mu(T_n(\{1\})) = T_n^* \mu(\{1\}) = |T_n'(\{1\})|^{\delta} \mu(\{1\}).$$

[22] Another way of deducing this result is to note that the geodesic ray cannot reach $\partial\mathbb{D}$ by splitting any H_j, or their translations under Γ, because the intervals I_j that define them (or their translations) contain no limit point, so $\{q\}$ cannot be contained in any of them. Thus, if the geodesic ray γ crosses a finite number of translations, the last boundary it crosses must be the boundary that determines a horocyclic region. In this case, the geodesic ray can only reach the boundary $\partial\mathbb{D}$ at the parabolic point that defines the horocyclic region.

As we saw that $\delta > 0$ for a non-elementary Fuchsian group, and μ is a finite measure, the divergence of $|T_n'(\{1\})|$ implies $\mu(\{1\}) = 0$. Indeed,

$$1 \geq \mu(\{T_n 1\}) = |T_n'(\{1\})|^{\delta} \mu(\{1\}) \Rightarrow \mu(\{1\}) \leq \frac{1}{|T_n'(\{1\})|^{\delta}} \overset{n\to\infty}{\longrightarrow} 0.$$

This concludes the first case.

Case B) The limit point $q \in \Lambda(\Gamma)$ is a parabolic fixed point.

For this case, we return to the $\mathbb{H}$ model. Let $A : \mathbb{H} \to \mathbb{D}$ be the isometry $z \mapsto \frac{z-i}{z+i}$. Notice that it maps ∞ to 1.

Taking the pullback by A of the function $F(z)$ defined previously, we obtain

$$F(Az) = \int_{\partial\mathbb{D}} P(Az, q)^{\delta} d\mu(q).$$

Changing variables from q to $x' = A^{-1}q \in \partial\mathbb{H}$ (i.e., boundaries mapped to boundaries), we will rewrite the integral above using the $\mathbb{H}$ version of the Poisson kernel, which we denote by

$$E(z, x') \equiv \begin{cases} \frac{\operatorname{Im} z}{|z-x'|^2}, \text{ if } x' \in \mathbb{R} \\ \operatorname{Im} z, \text{ if } x' = \infty. \end{cases}$$

We can write then

$$P(Az, Ax') = \begin{cases} E(z, x')\left(1 + x'^2\right), \text{ if } x' \in \mathbb{R} \\ E(z, \infty) = \operatorname{Im}(z), \text{ if } x' = \infty. \end{cases}$$

Substituting back into the integral, evaluating it at infinity, and using $A\infty = 1$, we find

$$\begin{aligned} F(Az) = &\int_{\partial\mathbb{H}=\mathbb{R}\cup\{\infty\}} P(Az, Ax')^{\delta} d\mu(Ax') \\ &\qquad + \int_{\partial\mathbb{H}=\mathbb{R}\cup\{\infty\}} P(Az, Ax')^{\delta} 1_{\mathbb{R}}(x') d\mu(Ax') \\ = &P(Az, A\infty)^{\delta} \mu(A\infty) + \int_{\mathbb{R}} P(Az, Ax')^{\delta} d\mu(Ax') \\ = &E(z, \infty)^{\delta} \mu(1) + \int_{\mathbb{R}} E(z, x')^{\delta} \left(1 + x'^2\right)^{\delta} d\mu(Ax'). \end{aligned}$$

On the imaginary axis, $z = iy$, we have

$$F(A(iy)) = \mu(1) y^{\delta} + y^{\delta} \int_{\mathbb{R}} \left(\frac{1 + x'^2}{y^2 + x'^2}\right)^{\delta} d\mu(Ax').$$

If we consider $y \geq 1$, then $\left|\left(\frac{1+x'^2}{y^2+x'^2}\right)^{\delta}\right| \leq \left(\frac{y^2+x'^2}{y^2+x'^2}\right)^{\delta} \leq 1$, for any $x' \in \mathbb{R}$. Since the total mass of $d\mu(Ax')$ is finite (i.e., Patterson-Sullivan μ is a bounded measure), the dominating function (i.e., 1) is integrable, and since

$\lim_{y\to\infty}\left(\frac{1+x'^2}{y^2+x'^2}\right)^\delta = 0$, the Dominated Convergence Theorem (for the Patterson-Sullivan measure μ) guarantees that

$$\lim_{y\to\infty}\int_{\mathbb{R}}\left(\frac{1+x'^2}{y^2+x'^2}\right)^\delta d\mu\left(Ax'\right) = \int_{\mathbb{R}}\lim_{y\to\infty}\left(\frac{1+x'^2}{y^2+x'^2}\right)^\delta d\mu\left(Ax'\right) = 0.$$

Hence, as $y\to\infty$,

$$F(A(iy)) = \mu(1)y^\delta + o\left(y^\delta\right),$$

where we have used the little-o notation.

This was just laying the groundwork for the proof. Let us assume that $\Gamma\subset \mathrm{PSU}(2)$ has an element with a parabolic fixed point at 1 (hence a limit point at 1). If we conjugate the group Γ by A, obtaining $\Gamma' = A^{-1}\Gamma A\subset \mathrm{PSL}(2,\mathbb{R})$, then Γ' has a parabolic fixed point at ∞, because A is an isometry and $A\infty = 1$, so $\Gamma'\infty = A^{-1}\Gamma A\infty = A^{-1}\Gamma 1 = A^{-1}1 = \infty$. Moreover, by further conjugation, suppose that the associated maximal parabolic subgroup is generated by a translation $T: z\mapsto z+1$.

By the definition of $F(z)$, weak convergence $\mu_{s_j}\to\mu$ and Theorem 4.3.3, write

$$\begin{aligned}
F(z) &\equiv \int_{\partial\mathbb{D}} P(z,q)^\delta d\mu(q)\\
&= \lim_{s_j\to\delta^+}\int_{\bar{\mathbb{D}}}\frac{P(z,q)^{s_j}}{P(0,q)^{s_j}}d\mu_{s_j}(q)\\
&= \lim_{s_j\to\delta^+}\left(\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{D}}(0,S0)}\right)^{-1}\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{D}}(0,S0)}\int_{\bar{\mathbb{D}}}\frac{P(z,q)^{s_j}}{P(0,q)^{s_j}}d\delta_{S0}(q)\\
&= \lim_{s_j\to\delta^+}\left(\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{D}}(0,S0)}\right)^{-1}\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{D}}(0,S0)}\frac{P(z,S0)^{s_j}}{P(0,S0)^{s_j}}\\
&= \lim_{s_j\to\delta^+}\left(\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{D}}(0,S0)}\right)^{-1}\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{D}}(z,S0)},
\end{aligned}$$

where the last equality is a consequence of Lemma 4.3.2.

Using the implication $Ai = 0\Rightarrow i = A^{-1}0$, write

$$\begin{aligned}
&\lim_{s_j\to\delta^+}\left(\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{D}}(0,S0)}\right)^{-1}\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{D}}(z,S0)}\\
&\quad= \lim_{s_j\to\delta^+}\left(\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{D}}(0,SAi)}\right)^{-1}\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{D}}(z,SAi)}\\
&\quad= \lim_{s_j\to\delta^+}\left(\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{H}}(A^{-1}0,A^{-1}SAi)}\right)^{-1}\sum_{S\in\Gamma}e^{-s_j d_{\mathbb{H}}(A^{-1}z,A^{-1}SAi)}\\
&\quad= \lim_{s_j\to\delta^+}\left(\sum_{S'\in\Gamma'}e^{-s_j d_{\mathbb{H}}(i,S'i)}\right)^{-1}\sum_{S'\in\Gamma'}e^{-s_j d_{\mathbb{H}}(A^{-1}z,S'i)}.
\end{aligned}$$

Substituting this expression in the last expression obtained for $F(z)$ above, evaluate $F(A(iy))$ as

$$\begin{aligned} F(A(iy)) &= \lim_{s_j \to \delta^+} \left(\sum_{S' \in \Gamma'} e^{-s_j d_{\mathbb{H}}\left(i, S'i\right)} \right)^{-1} \sum_{S' \in \Gamma'} e^{-s_j d_{\mathbb{H}}\left(A^{-1}A(iy), S'i\right)} \\ &= \lim_{s_j \to \delta^+} \left(\sum_{S' \in \Gamma'} e^{-s_j d_{\mathbb{H}}\left(i, S'i\right)} \right)^{-1} \sum_{S' \in \Gamma'} e^{-s_j d_{\mathbb{H}}\left(iy, S'i\right)}. \end{aligned}$$

We will bound a more general version of this expression,

$$L = \lim_{s_j \to \delta^+} \left(\sum_{S' \in \Gamma'} e^{-s_j d_{\mathbb{H}}\left(i, S'w'\right)} \right)^{-1} \sum_{S' \in \Gamma'} e^{-s_j d_{\mathbb{H}}\left(iy, S'w'\right)},$$

where $A^{-1}w = w' \in \mathbb{H}$

To estimate the sums over Γ', we split $\Gamma' = (\Gamma' - \Gamma_\infty) \cup \Gamma_\infty$, where $\Gamma_\infty \equiv \langle T \rangle$ and T is the unit translation, whose existence in Γ' is a consequence of the hypothesis (as described above). Notice that we can use $\mathcal{F}_\infty \equiv \left\{ z \in \mathbb{H} : -\frac{1}{2} < \operatorname{Re} z \leq \frac{1}{2} \right\}$ as a fundamental domain for Γ_∞. Thus, for an arbitrary $S' \in \Gamma'$ we can write it uniquely as $S' = T^n R$ such that $Rw' \in \mathcal{F}_\infty$ and $n \in \mathbb{Z}$. Then the sum can be written as

$$\sum_{S' \in \Gamma'} e^{-s_j d_{\mathbb{H}}\left(iy, S'w'\right)} = \sum_{n \in \mathbb{Z}} \sum_{\substack{R \in \Gamma': \\ Rw' \in \mathcal{F}_\infty}} e^{-s_j d_{\mathbb{H}}\left(iy, Rw' + n\right)}.$$

Using the $\cosh d_{\mathbb{H}}(\cdot, \cdot)$ formula, it is possible to find the following general bounds: for $z, z' \in \mathbb{H}$, $z = x + iy$, $z' = x' + iy'$,

$$\frac{yy'}{(x - x')^2 + y^2 + y'^2} \leq e^{-d_{\mathbb{H}}(z, z')} \leq \frac{2yy'}{(x - x')^2 + y^2 + y'^2}.$$[23]

Observe that there cannot be a hyperbolic transformation in Γ' that fixes ∞, since infinity is already fixed by the parabolic element T (and Proposition 4.8.3 applies). This means that, under iteration, the points in $\mathcal{F}_\infty$ cannot be sent to

[23] This is done noticing that

$$\frac{e^{d_{\mathbb{H}}(z, z')} - e^{-d_{\mathbb{H}}(z, z')}}{2} = \cosh d_{\mathbb{H}}(z, z') = 1 + \frac{|z - z'|^2}{2yy'},$$

and, consequently, using a Taylor expansion,

$$\begin{aligned} \left| e^{d_{\mathbb{H}}(z, z')} \right| &= e^{-d_{\mathbb{H}}(z, z')} + 2 + \frac{(x - x')^2 + (y - y')^2}{yy'} \\ &\sim 1 - d_{\mathbb{H}}(z, z') + 2 + \frac{(x - x')^2 + (y - y')^2}{yy'}, \end{aligned}$$

whereas the other bound uses

$$\begin{aligned} \left| e^{d_{\mathbb{H}}(z, z')} \right| &= e^{-d_{\mathbb{H}}(z, z')} + 2 + \frac{(x - x')^2 + (y - y')^2}{yy'} \\ &\geq 2 + \frac{(x - x')^2 + (y - y')^2}{yy'}. \end{aligned}$$

∞. Moreover, they cannot accumulate in $\mathcal{F}_\infty$ either, due to the discreteness of the group. Thus, within $\mathcal{F}_\infty$, the points Rw' can only accumulate towards the boundary, that is, the real line and $\operatorname{Im} Rw'$ is bounded from above. Furthermore, $|\operatorname{Re} Rw'| \leq \frac{1}{2}$. Thus, for any $Rw' \in \mathcal{F}_\infty$ we find

$$\begin{aligned}
&\sum_{S'\in\Gamma'} e^{-s_j d_{\mathbb{H}}\left(iy,S'w'\right)} \\
&\quad\leq \sum_{n\in\mathbb{Z}} \sum_{\substack{R\in\Gamma':\\ Rw'\in\mathcal{F}_\infty}} \left(\frac{2\operatorname{Im}(iy)\operatorname{Im}(Rw'+n)}{\left(\operatorname{Re}(iy)-\operatorname{Re}(Rw'+n)\right)^2+\operatorname{Im}^2(iy)+\operatorname{Im}^2(Rw'+n)}\right)^{s_j} \\
&\quad= \sum_{n\in\mathbb{Z}} \sum_{\substack{R\in\Gamma':\\ Rw'\in\mathcal{F}_\infty}} \left(\frac{2y\operatorname{Im}(Rw')}{\left(n+\operatorname{Re}(Rw')\right)^2+y^2+\operatorname{Im}^2(Rw')}\right)^{s_j} \\
&\quad\leq C\sum_{n\in\mathbb{Z}} \sum_{\substack{R\in\Gamma':\\ Rw'\in\mathcal{F}_\infty}} \left(\frac{y\operatorname{Im}(Rw')}{n^2+y^2}\right)^{s_j},
\end{aligned}$$

where C is some constant that can be determined by studying the behavior of $\left(\frac{2(n^2+y^2)}{(n+\operatorname{Re}(Rw'))^2+y^2+\operatorname{Im}^2(Rw')}\right)$ when taking the limits $n\to\pm\infty, y\to 0$ and $y\to\infty$. Since this quotient converges in the limits, one can bound (from above) but a finite number of terms, and these in turn can be bounded (from above) by comparing their maxima.

Similarly, using the other bound, we find

$$\begin{aligned}
\sum_{S'\in\Gamma'} e^{-s_j d_{\mathbb{H}}\left(i,S'w'\right)} &= \sum_{n\in\mathbb{Z}} \sum_{\substack{R\in\Gamma':\\ Rw'\in\mathcal{F}\infty}} e^{-s_j d_{\mathbb{H}}(i,Rw'+n)} \\
&\geq \sum_{n\in\mathbb{Z}} \sum_{\substack{R\in\Gamma':\\ Rw'\in\mathcal{F}\infty}} \left(\frac{\operatorname{Im}(Rw')}{(n+\operatorname{Re}(Rw))^2+1+\operatorname{Im}^2(Rw'+n)}\right)^{s_j} \\
&\geq C'\sum_{n\in\mathbb{Z}} \sum_{\substack{R\in\Gamma':\\ Rw'\in\mathcal{F}\infty}} \left(\frac{\operatorname{Im}(Rw')}{n^2+1}\right)^{s_j},
\end{aligned}$$

where C' is some constant that can be similarly obtained.

Now we use these bounds to estimate L. We start by observing that

$$\begin{aligned}
&\left(\sum_{S'\in\Gamma'} e^{-s_j d_{\mathbb{H}}\left(i,S'w'\right)}\right)^{-1} \sum_{S'\in\Gamma'} e^{-s_j d_{\mathbb{H}}\left(iy,S'w'\right)} \\
&\qquad \leq \frac{C\sum_{n\in\mathbb{Z}}\sum_{\substack{R\in\Gamma':\\ Rw'\in\mathcal{F}_\infty}}\left(\frac{y\,\mathrm{Im}(Rw')}{n^2+y^2}\right)^{s_j}}{\left(C'\sum_{n\in\mathbb{Z}}\sum_{\substack{R\in\Gamma':\\ Rw'\in\mathcal{F}_\infty}}\left(\frac{\mathrm{Im}(Rw')}{n^2+1}\right)^{s_j}\right)} \\
&\qquad = \frac{C}{C'}\left(\frac{\left(\sum_{\substack{R\in\Gamma':\\ Rw'\in\mathcal{F}_\infty}}\mathrm{Im}^{s_j}(Rw')\right)\left(\sum_{n\in\mathbb{Z}}\frac{1}{(n^2+y^2)^{s_j}}\right)}{\left(\sum_{n\in\mathbb{Z}}\frac{1}{(n^2+1)^{s_j}}\right)\left(\sum_{\substack{R\in\Gamma':\\ Rw'\in\mathcal{F}_\infty}}\mathrm{Im}^{s_j}(Rw')\right)}\right) y^{s_j} \\
&\qquad = C''\frac{y^{s_j}\left(\sum_{n\in\mathbb{Z}}(n^2+y^2)^{-s_j}\right)}{\left(\sum_{n\in\mathbb{Z}}(n^2+1)^{-s_j}\right)},
\end{aligned}$$

where we have canceled out the sums of imaginary parts.

Then, taking the limit in s, we find

$$\begin{aligned}
L &= \lim_{s_j\to\delta^+}\left(\sum_{S'\in\Gamma'} e^{-s_j d_{\mathbb{H}}\left(i,S'w'\right)}\right)^{-1}\sum_{S'\in\Gamma'} e^{-s_j d_{\mathbb{H}}\left(iy,S'w'\right)} \\
&\leq C''y^{\delta}\lim_{s_j\to\delta^+}\frac{\left(\sum_{n\in\mathbb{Z}}(n^2+y^2)^{-s_j}\right)}{\left(\sum_{n\in\mathbb{Z}}(n^2+1)^{-s_j}\right)}.
\end{aligned}$$

Passing the limit $s_j\to\delta^+$ under the summation sign (this is allowed because all terms are positive, so we can apply the Monotone Convergence Theorem) and using continuity yields

$$F(A(iy)) \leq C''y^{\delta}\frac{\sum_{n\in\mathbb{Z}}\left(n^2+y^2\right)^{-\delta}}{\sum_{n\in\mathbb{Z}}\left(n^2+1\right)^{-\delta}}.$$

Notice that the bound found for L does not depend on w', so it bounds $F(A(iy))$ as above.

The assumption that Γ is non-elementary and contains a parabolic element guarantees, via Beardon's theorem, that $\delta > \frac{1}{2}$, so both sums converge. We dismiss the denominator because it is some constant. On the other hand, the numerator can be estimated against the integral $\int_{-\infty}^{\infty}\left(t^2+y^2\right)^{-\delta}dt$, which gives

$$\int_{-\infty}^{\infty}\left(t^2+y^2\right)^{-\delta}dt = \frac{\sqrt{\pi}\sqrt{\frac{1}{y^2}}\Gamma\left(\delta-\frac{1}{2}\right)\left(y^2\right)^{1-\delta}}{\Gamma(\delta)} = O\left(y^{1-2\delta}\right).$$

In the above formula, $\Gamma(\cdot)$ indicates the Gamma function, not the Fuchsian group Γ. Therefore, using the big-O notation, after having replaced the constants accordingly, we find

$$F\circ A(iy) \leq C'''y^{\delta}O\left(y^{1-2\delta}\right) = C'''O\left(y^{1-\delta}\right) \Rightarrow F\circ A(iy) = O\left(y^{1-\delta}\right).$$

Equating this result with the expression we previously found, we have for every $\varepsilon > 0$ and constants $M > 0, y_0$, for $y \geq y_0$,

$$\mu(1)y^\delta - \varepsilon y^\delta \leq \mu(1)y^\delta + o\left(y^\delta\right) = F \circ A(iy) = O\left(y^{1-\delta}\right) \leq My^{1-\delta}$$
$$\mu(1) \leq My^{1-2\delta} + \varepsilon \overset{\substack{y\to\infty \\ \varepsilon\to 0}}{\longrightarrow} 0.$$

The implication holds because $\delta > \frac{1}{2}$, proving the assertion. □

4.9 Remarks on conformal densities

Many of the properties of Patterson-Sullivan measures described in the last sections can be inspected from a conformal viewpoint. The main idea is that the boundary $\partial\mathbb{D}$, seen as a 1-manifold, is a conformal manifold in which Riemannian metric tensors $g_x, g_{x'}$ that arise from different base points $x, x' \in \mathbb{D}$ are related by a conformal factor $\left(\frac{P(x',\xi)}{P(x,\xi)}\right)^{\pm 1}$, for $\xi \in \partial\mathbb{D}$.[24] The point here is not to specify such tensor metrics but to address conformal densities.[25]

Definition 4.9.1. Let Γ be a discrete group acting on $\mathbb{D}$. A Γ-invariant conformal density of dimension α is a map σ from the collection of all metrics $\{g_x : x \in \mathbb{D}\}$ into the collection of positive finite measures on $\partial\mathbb{D}$ such that, writing $\sigma_x = \sigma(g_x)$,

1. σ_x is supported on the limit set of Γ;
2. for $x, x' \in \mathbb{D}, \sigma_x$ and $\sigma_{x'}$ are absolutely continuous with respect to each other and the Radon-Nikodym derivative satisfies
$$\left(\frac{d\sigma_{x'}}{d\sigma_x}\right)(\xi) = \left(\frac{P(x',\xi)}{P(x,\xi)}\right)^\alpha;$$
3. $\gamma^*\sigma_x = \sigma_{\gamma^{-1}(x)}$ (or, in the manner of the pushforward, $\gamma_*\sigma_x = \sigma_{\gamma(x)}$) for $\gamma \in \Gamma$.

As a consequence of these three defining properties, we obtain the familiar formulas from Section 4.6, but this time with respect to a conformal density σ of dimension α. That is,

$$\sigma_x(\gamma(E)) = \int_E |\gamma_x'(\xi)|^\alpha \, d\sigma_x(\xi),$$

for E a Borel subset of $\partial\mathbb{D}$ and $\gamma \in \Gamma$. Note the reappearance of the derivative $|\gamma_x'(\xi)|$ from Section 4.6, which is defined in terms of the great circle metric.

In fact, the relationship between Patterson-Sullivan measures and conformal densities is not accidental. If we define a map ν from the collection of tensor metrics to the set of positive finite measures on $\partial\mathbb{D}$ via $\nu(g_x) = \nu_x \in M_x$ (observing

[24] This conformal factor is often described in terms of the Busemann function (see Denker and Stratmann's survey [17]), though this is not explicitly the case in Nicholls [53] nor is it the case in Borthwick [9].

[25] For the interested reader, Chapter 4 in Nicholls' reference work has all the pertinent details ([53], p. 68).

the convention that for $x, x' \in \mathbb{D}, \nu_x, \nu_{x'}$ are the corresponding elements of the classes $M_x, M_{x'}$, as explained in Section 4.5), then if Γ is a Fuchsian group with critical exponent δ, the map ν is a Γ-invariant conformal density of dimension δ.

The language of conformal densities will be particularly helpful in showing some of the next results, especially in questions that touch on the uniqueness (up to scaling) of the Patterson-Sullivan measures and bounds involving the lattice-point counting function π_X.

Chapter 5

Patterson-Sullivan measures and the ergodicity of the geodesic flow

The goal of the chapter is to show that the geodesic flow on hyperbolic surfaces is ergodic. This will be done with respect to a measure m_μ, or more generally m_{μ_z}, to be defined from the Patterson-Sullivan measure μ_z (μ in case $z = 0$), whose development was the topic of the last chapter. We start with a preliminary section on the geodesic flow and the construction of this *ad hoc* measure. Then we prove that the geodesic flow is conserved by this measure, a key step that depends on the non-atomicity of the Patterson-Sullivan measure (for non-elementary Fuchsian groups). Next, we introduce some basic results from ergodic theory which are required in the later sections, specifically Hopf's argument for the ergodicity of the geodesic flow.[1] Although the arguments are presented in the hyperbolic surface setting, in fact most of them carry over to hyperbolic manifolds in higher dimensions.[2] The final section revisits the question of the uniqueness of Patterson-Sullivan measures.

5.1 The geodesic flow and measure m_{μ_z}

As before, we treat hyperbolic surfaces as manifestations of the upper half-plane or the Poincaré disc via the space of orbits generated by the action of a Fuchsian group on these models. Specifically, for this chapter, our model of choice is the unit disc. We will then describe the geodesic flow for this model and later move on to geodesic flows on quotients.

Consider the geodesic flow for the Poincaré disc $\mathbb{D}$. The space where the geodesic flow occurs, denoted by Ω throughout the chapter, can be described in

[1] Hopf's original treatment (see [29] and [28]) was of course not based on Patterson-Sullivan measures, it was based on the measure M' presented in Section 5.1.

[2] One difference lies in the proof of conservation of the geodesic flow. Since the argument put forward depends on the non-atomicity of the Patterson-Sullivan measure (for non-elementary Γ), this has to be proved in higher dimensions. Although still doable, the argument becomes more involved. See Section 3.5 in Nicholls' book ([53], p. 57)

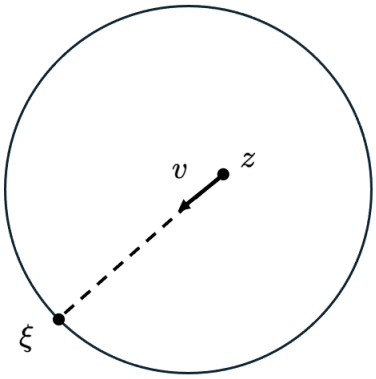


Figure 5.1: Ascribing a direction to every point in $\mathbb{D}$.

slightly different ways. In essence, it is the set $\mathbb{D}$ for which we have assigned a direction at each point in the set.

If we think of $z \in \mathbb{D}$ as representing only a point in a set, frequently called the base or carrier point, then the direction assigned to it can simply be thought of as a point ξ in $\partial\mathbb{H}$. This is because a line can be drawn from z to ξ that would then describe the direction considered (see Figure 5.1). In other words, the geodesic flow takes place in the space of pairs $(x, \xi) \in \mathbb{D} \times \partial\mathbb{D} = \Omega$. This is the approach followed by Nicholls.[3]

However, using the language of differential and Riemannian geometry, it is also possible to describe the directions at the base point $z \in \mathbb{D}$ as the unit norm elements that live in the tangent space identified by this point, that is, $v \in T_z\mathbb{D}, ||v||_z = 1$. Here, $||\cdot||_z$ designates the norm induced by the hyperbolic tensor metric $g_{\mathbb{D}}$ at the point z. In this language, the one employed by Borthwick,[4] we come to identify Ω with $\mathcal{S}\mathbb{D}$, the unit tangent bundle of $\mathbb{D}$. Recall the definition of this construct.

Definition 5.1.1. The unit tangent bundle of a Riemannian manifold (M, g), denoted by $\mathcal{S}M$, is the unit sphere bundle for the tangent bundle $\mathcal{T}M$. In particular, it is a fiber bundle over M whose fiber at each point is the unit sphere in the tangent space.

The elements of $\mathcal{S}\mathbb{D}$ are then pairs (z, v), where z is some point of the manifold $\mathbb{D}$ and v is some tangent direction (of unit length) at z, since

$$\mathcal{S}\mathbb{D} \equiv \bigcup_{z\in\mathbb{D}} \left\{ v \in T_z\mathbb{D} : ||v||_z^2 = (g_{\mathbb{D}})_z(v, v) = 1 \right\},$$

where $(g_{\mathbb{D}})_z$ denotes the hyperbolic tensor metric evaluated at the point z.

It is not difficult to see that both descriptions of Ω are equivalent, since (as alluded to in Figure 5.1) every unit direction v in $(z, v) \in \mathcal{S}\mathbb{D}$ can be identified with a unique boundary point ξ, the point of intersection of $\partial\mathbb{D}$ with the ray

[3] See Chapter 8 in Nicholls [53].
[4] See Section 14.2 in Borthwick [9].

that exits z in the direction of v. Therefore, without loss of generality, we focus on a general space Ω identified with $\mathbb{D} \times \partial\mathbb{D}$.

Having defined the set Ω, we can now talk about a continuous flow that takes place in this space. In particular, we are interested in measure-preserving flows.

Definition 5.1.2. Suppose (Ω, d_Ω) is a metric space equipped with a σ-finite measure μ on its Borel subsets. A measure-preserving flow on Ω is any map $\phi : \Omega \times \mathbb{R} \to \Omega$ with the following properties:

1. $\phi(x, 0) = x, \forall x \in \Omega$;
2. $\phi(\phi(x, s), t) = \phi(x, s + t), \forall x \in \Omega$ and $\forall s, t \in \mathbb{R}$;
3. ϕ is continuous;
4. for any $t \in \mathbb{R}$ and A a measurable subset of Ω, $\phi(A, t)$ is measurable, and $\mu(\phi(A, t)) = \mu(A)$.

It is common to abbreviate $\phi(x, t)$ as $\phi_t(x)$ or $\phi^t(x)$.

What is the distance d_Ω associated with the flow space? A suitable metric on Ω is given by the hyperbolic distance between base points, that is, for $(z_1, \xi), (z_2, \eta) \in \mathbb{D} \times \partial\mathbb{D}$, we have $d_\Omega((z_1, \xi), (z_2, \eta)) \equiv d_\mathbb{D}(z_1, z_2) + |\xi - \eta|$.[5]

Since we will be interested in the action of a Fuchsian group Γ on Ω, one must also ask how orientation-preserving isometries affect the elements of Ω. The action envisaged in this setting is the so-called derivative action.[6] That is, given a Möbius transformation $\gamma \in \Gamma$ (that we can also think of as a matrix, as seen in Section 2.3) and $(z, \xi) \in \Omega$, we define

$$\gamma((z, \xi)) = \left(\gamma(z), \frac{\gamma'(z)}{|\gamma'(z)|}\xi\right).$$

Notice that the choice of metric ensures that d_Ω is invariant under the derivative action of Möbius transformations. It is also assumed that the group Γ is measurable (i.e., if $A \subset \Omega$ is measurable, with respect to some measure, then $\gamma(A)$ is likewise measurable, for $\gamma \in \Gamma$) and non-singular (i.e., if $A \subset \Omega$ has measure zero, then $\gamma(A)$ likewise has measure zero).

The existence of a flow on Ω opens up the possibility of studying its flow-invariant subsets.

Definition 5.1.3. Given a measure-preserving flow $\phi_t, t \in \mathbb{R}$, on Ω, the set $A \subseteq \Omega$ is a flow-invariant set with respect to ϕ_t if $\phi_t(A) = A, \forall t \in \mathbb{R}$.

Crucially, the characterization of ergodicity of a flow depends on such flow-invariant sets.

Definition 5.1.4. A measure-preserving flow $\phi_t, t \in \mathbb{R}$, on Ω is said to be ergodic if, whenever $A \subseteq \Omega$ is a measurable flow-invariant set, either $\mu(A) = 0$ or $\mu(A^c) = 0$.

[5] In fact, for higher dimensions, the metric $d_\Omega : \Omega \times \Omega \to \mathbb{R}$ is given by the sum of the hyperbolic distance between base points and a correction term that can be understood as a difference between angles, if computed for the same base point. The details can be found in Nicholls' book ([53], p. 129), as well as the proof that d_Ω as defined is indeed a metric.

[6] See Einsiedler and Ward's book ([20], p. 280).

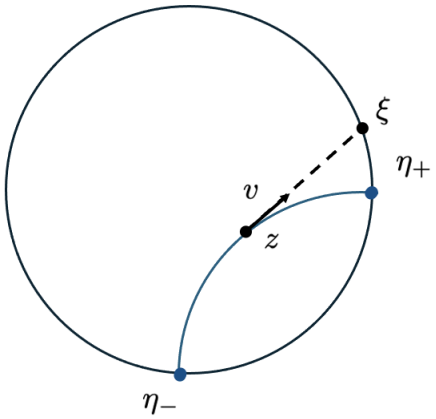


Figure 5.2: The drawing illustrates how every element $(z, \xi) \in \Omega$ determines a geodesic and a point along it which are both describable via Hopf coordinates (η_-, η_+, s).

Later we will rely on the algebraic counterpart of this definition, which can be stated as follows.

Definition 5.1.5. Let Γ be a Fuchsian group acting on $\mathbb{D}$ and assume that σ is a Γ-invariant α-dimensional conformal density. Then, the action of Γ on $\partial\mathbb{D}$ is said to be ergodic with respect to σ if for any Γ-invariant Borel set $A \subseteq \partial\mathbb{D}$, $\sigma_x(A) = 0$ or $\sigma_x(A^c) = 0$, for at least one $x \in \mathbb{D}$.

Although the measure σ_x in this last definition is clear, both the definitions of measure-preserving flows and ergodic flows refer to a measure μ of Ω, not of $\bar{\mathbb{D}}$. But how can this measure μ be defined? The question of assigning a measure to Ω can be approached in different ways. In all cases, we wish for the measure to be invariant under orientation-preserving isometries of $\mathbb{D}$. For example, it is possible to use the hyperbolic area form and define a measure M such that

$$dM(z, \xi) = \frac{dxdy}{y^2} dw(\xi),$$

where $z = x + iy$ and dw denotes the Lebesgue measure on $\partial\mathbb{D}$. The measure thus defined is called the Liouville measure, and it will be of interest only for comparison purposes.

To define a measure derived from a conformal density, specifically a measure derived from the Patterson-Sullivan measures studied in the previous chapter, a change of coordinates is most helpful. In this sense, let us reintroduce the Liouville measure using coordinates known as Hopf coordinates.

Any element $(z, \xi) \in \Omega$ determines a unique directed geodesic that passes through z in the direction of ξ.[7] We call the endpoints of this geodesic η_- and η_+ and, assuming z_0 as the Euclidean midpoint of this geodesic (i.e., its nearest point to the origin of the unit disc), we let s quantify the oriented hyperbolic

[7]For the base point of (x, ξ) in $\mathbb{D}$, there are infinitely many geodesics passing through this point, but only one of them is tangent to the direction attributed to the base point. This is a consequence of the uniqueness of geodesics.

distance from z_0 to z. Figure 5.2 illustrates that the attracting endpoint η_+ of the geodesic identified by (z, ξ) is not necessarily the same as ξ. The coordinates allow a correspondence $\Omega \Leftrightarrow ((\partial\mathbb{D} \times \partial\mathbb{D}) - \Delta) \times \mathbb{R}$, where Δ is the diagonal of $\partial\mathbb{D} \times \partial\mathbb{D}$, so that it becomes feasible to define a measure M' on Ω through

$$dM'(\eta_-, \eta_+, s') \equiv \frac{2dw(\eta_-)dw(\eta_+)ds(s')}{|\eta_+ - \eta_-|^2},$$

where dw is again the Lebesgue measure on $\partial\mathbb{D}$ and ds is the Lebesgue measure on $\mathbb{R}$.

Before we show that this measure is invariant under the action of Möbius transformations, we need to define how the action behaves with respect to the Hopf coordinates. For γ a Möbius transformation of $\mathbb{D}$, define the action of γ on (η_-, η_+, s) as

$$\gamma((\eta_-, \eta_+, s)) = (\gamma(\eta_-), \gamma(\eta_+), s + k),$$

where k is the oriented hyperbolic distance from z_0', the Euclidean midpoint of the geodesic joining $\gamma(\eta_-)$ and $\gamma(\eta_+)$, to $\gamma(z_0)$, with z_0 denoting the Euclidean midpoint of the geodesic joining η_- and η_+.

Then, the proof of invariance follows from the fact that $|\gamma(\eta_+) - \gamma(\eta_-)| \overset{(*)}{=} |\eta_+ - \eta_-|\,|\gamma'(\eta_+)|^{1/2}\,|\gamma'(\eta_-)|^{1/2}$ combined with the identity

$$dw(\gamma(\eta_\pm)) = |\gamma'(\eta_\pm)|dw(\eta_\pm).$$

As indicated previously, M' is the Liouville measure, that is, the measures M and M' on Ω thus defined can be proved equivalent.[8] Since M' is invariant under Möbius transformations, the equivalence implies that M is also invariant.

The geodesic flow on Ω is a particular type of measure-preserving flow on Ω, which can be conveniently described using Hopf coordinates.

Definition 5.1.6. The geodesic flow on Ω is a measure-preserving flow $\phi : \Omega \times \mathbb{R} \to \Omega$ such that $\phi_t((\eta_-, \eta_+, s)) = (\eta_-, \eta_+, s + t)$.

We will denote the geodesic flow by $g_t, t \in \mathbb{R}$, whenever no ambiguity arises from the symbol g in its potential use as a tensor metric.

Notice that the action of g_t on an element $(x, \xi) \in \Omega$ slides the point by t along the oriented geodesic associated with the pair. The map g_t responsible for the geodesic flow on Ω clearly does not affect the measure M', according to the definition of this measure. Since the measure M' is equivalent to M, the latter is also not affected by the action of the geodesic flow. This can also be proved using the derivative action of g_t on $(z, \xi) \in \Omega$. In this case, it can be shown that this action occurs via specific Möbius transformations, which do not affect the measures (as argued above).[9]

Guided by the measure M', we may now construct yet another measure on Ω, this time derived from a conformal density μ of dimension α and a Fuchsian group Γ that preserves $\mathbb{D}$. For $(z, \xi) \in \Omega$, using Hopf coordinates, define

$$dm_{\mu_z}(\eta_-, \eta_+, s') \equiv \frac{d\mu_z(\eta_-)\, d\mu_z(\eta_+)\, ds(s')}{|\eta_+ - \eta_-|^{2\alpha}}.$$

[8] See Theorem 8.1.1 in Nicholls' book ([53], p. 131).

[9] The details of this explicit construction (using matrices) can be found in Einsiedler and Ward's book ([20], p. 286).

Given $\gamma \in \Gamma$ and $\eta_\pm \in \partial\mathbb{D}$, by the properties of conformal densities, we have

$$\frac{d\mu_z(\gamma(\eta))}{d\mu_z(\eta)} = |\gamma'(\eta)|^\alpha ,$$

and using $(*)$ above, we see that the measure m_{μ_z} is also invariant under Möbius transformations, as k in

$$\gamma((\eta_-, \eta_+, s)) = (\gamma(\eta_-), \gamma(\eta_+), s + k),$$

defined as described is independent of s.

The last aspect to be considered in the present section is that of the orbit space $\Gamma\backslash\Omega$. The geodesic flow g_t also induces a geodesic flow on the quotient $\Gamma\backslash\Omega$, since it commutes with Möbius transformations. In Hopf coordinates,

$$\begin{aligned}\gamma(g_t(\eta_-, \eta_+, s)) &= \gamma(\eta_-, \eta_+, s + t)\\ &= (\gamma(\eta_-), \gamma(\eta_+), s + t + k)\\ &= g_t(\gamma(\eta_-), \gamma(\eta_+), s + k)\\ &= g_t(\gamma(\eta_-, \eta_+, s)).\end{aligned}$$

To define a measure M or m_{μ_z} on $\Gamma\backslash\Omega$, we proceed as follows. Given the natural projection $\pi : \Omega \to \Gamma\backslash\Omega$, we say that a subset $E \subseteq \Gamma\backslash\Omega$ is measurable if $\pi^{-1}(E)$ is a Borel subset of Ω. Using a Dirichlet domain $\mathcal{D}_w$ for Γ and defining $E(\mathcal{D}_w)$ as the set of points in $\pi^{-1}(E)$ whose base points are in $\mathcal{D}_w$, let $M(E)$ (or analogously, $m_{\mu_z}(E)$) be defined as the M measure (respectively, m_{μ_z} measure) of the set $E(\mathcal{D}_w)$. The invariance of the measures m_{μ_z}, M under Möbius transformations descends to the measures on the quotient space $\Gamma\backslash\Omega$

The metric d_Ω can be used to define a metric on $\Gamma\backslash\Omega$ in the usual way (i.e., via the infimum). Once $\Gamma\backslash\Omega$ is equipped with a measure and a metric, it can be shown that g_t is a measure-preserving flow, also called geodesic flow. The symbols g_t, M and m_{μ_z} will be used in the next sections for both Ω and $\Gamma\backslash\Omega$ and we distinguish between them when appropriate. We use d_Ω and $d_{\Gamma\backslash\Omega}$ to denote the different metrics.

Having all definitions at hand, we reiterate that the end goal of the chapter is to prove that for hyperbolic surfaces $X \cong \Gamma\backslash\mathbb{D}$, where Γ is a non-elementary geometrically finite Fuchsian group, their measure-preserving geodesic flows g_t are ergodic with respect to the measure m_{μ_z}. In other words, these flows come to sweep the entire surface in the limit in time.

As a final remark, we observe that similar questions can be raised for other flows of interest on geometrically finite hyperbolic surfaces, such as the horocyclic flow, but we do not address these issues here.[10]

5.2 The geodesic flow is conservative

The notion of conservative point for a flow can be stated as follows.

Definition 5.2.1. Suppose $\phi_t, t \in \mathbb{R}$, is a flow on Ω. A point $x \in \Omega$ is a conservative point for the flow ϕ_t if $\phi_t(x)$ remains in some compact set F for a

[10]For results concerning both the geodesic and the horocyclic flow, see Schapira [66] and Roblin [61]. We also briefly discuss some recent results concerning frame flows in Chapter 8.

sequence t tending to infinity. If $\phi_t(x)$ eventually leaves any compact set for t large enough, then x is a dissipative point. The set of conservative points for a flow is denoted by $C \subset \Omega$.

This intuitive geometric notion of conservative set has an analytical counterpart.

Definition 5.2.2. Given a measure space (Ω, Σ, μ) and a flow $\phi_t, t \in \mathbb{R}$, on Ω, ϕ_t is said to be a conservative flow if its conservative set has full measure: $\mu(C) = \mu(\Omega)$.

Although the definition is applicable to any kind of flow, our main concern is with the geodesic flow g_t. At the end of the chapter, we will demonstrate that conservation of the geodesic flow implies its ergodicity. In the current stage, the hypothesis of this implication can already be shown with the tools from the previous chapters.

Theorem 5.2.3. *Given a hyperbolic surface $X = \Gamma\backslash\mathbb{D}$, where Γ is a non-elementary geometrically finite Fuchsian group, let Ω be the flow space described in Section 5.1. Then, the geodesic flow on $\Gamma\backslash\Omega$ is conservative with respect to the measure m_{μ_z} (as indicated in the mentioned section).*

Proof. Let g_t denote a geodesic flow on $\Gamma\backslash\Omega$. Consider the dissipative set $D \subset \Gamma\backslash\Omega$ and take an arbitrary element $x \in D$. As $x \in \Gamma\backslash\Omega$, it can be thought of as a point of the hyperbolic surface along some geodesic. This geodesic must cross the compact core of the surface X (see the proof of Proposition 3.2.3), denoted by $K = \Gamma\backslash\widetilde{K}$, where $\widetilde{K}$ is the truncated Nielsen region. Moreover, since x is dissipative, for any compact set F, $g_t(x)$ never returns to F after some instant t'. Thus, letting $F = K$, it suffices to compute the measure m_{μ_z} (in the quotient sense) of D, whose elements can be assumed to leave K indefinitely, and to show that this measure is zero. This will imply that the conservative set for the flow has full m_{μ_z} measure.

By the definition of the measure m_{μ_z} for the quotient $\Gamma\backslash\Omega$, the measure of the dissipative set D described above is given by the measure m_{μ_z} (now in the non-quotient sense) of the set A of elements in $\pi^{-1}(D) \subset \Omega$ whose base points lie in $\mathcal{D}_w$, a Dirichlet domain for Γ. These elements identify points along specific geodesics in $\mathbb{D}$ by definition. Thus, if we show that the set of all geodesics that leave $\widetilde{K}$ indefinitely has zero measure, by monotonicity this will imply that A has null measure (in the non-quotient sense), which in turn implies that D has zero measure (in the quotient sense). This is the strategy in what follows.

Consider a geodesic in $\mathbb{D}$ that starts inside $\widetilde{K}$. Notice that this geodesic can either remain in $\widetilde{K}$, or leave it, and if it does leave, then it can go away to the boundary at infinity $\partial\mathbb{D}$ by leaving $\widetilde{N}$, the (untruncated) Nielsen region, or it can leave $\widetilde{K}$ by staying in $\widetilde{N}$. We consider each case separately.

Case A) The geodesic leaves $\widetilde{N}$.

Recall the construction of the (untruncated) Nielsen region $\widetilde{N}$. Observe that the geodesic leaves $\widetilde{N}$ only if it has an endpoint in one of the arcs I_j that make up $\partial\mathbb{H} - \Lambda(\Gamma)$. But in this case $m_{\mu_z}(\partial\mathbb{H} - \Lambda(\Gamma)) = 0$, a consequence of the fact that μ_z (the conformal density used to define m_{μ_z}) is unsupported on the

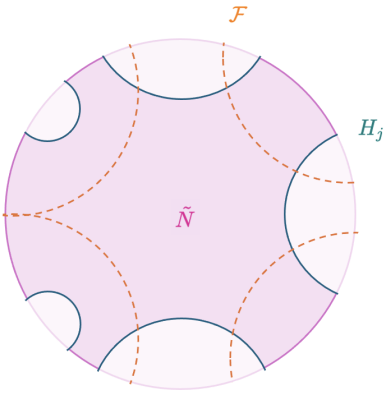


Figure 5.3: Illustration of (untruncated) Nielsen region.

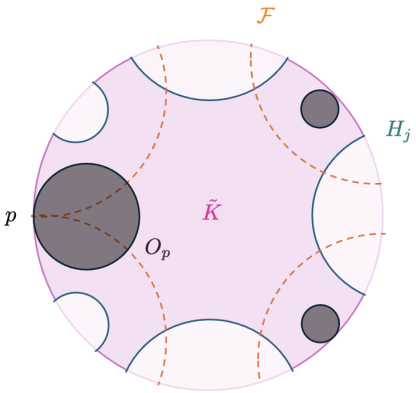


Figure 5.4: Illustration of truncated Nielsen region.

complement of $\Lambda(\Gamma)$ (see Theorem 4.4.3). Thus, the set of geodesics that leave $\widetilde{N}$ has measure zero. In other words, almost every geodesic stays within $\widetilde{N}$.

Case B) The geodesic leaves $\widetilde{K}$, but not $\widetilde{N}$.

A geodesic that stays in $\widetilde{N}$ can only leave $\widetilde{K}$ by crossing one of the horocycles σ_p associated with a parabolic fixed point p. But if the geodesic remains within σ_p after that crossing (as it must, since it leaves $\widetilde{K}$), its endpoint must be p. Because the measure μ_z has no atoms in our context (Theorem 4.8.5, which requires geometric finiteness), the set of geodesics ending in parabolic fixed points has m_{μ_z}-measure zero.

Thus, almost every geodesic on X returns to K at arbitrarily large times. □

Although there is no space for an alternative proof, we remark here that the chain of implications

$$\begin{aligned}\text{Geometric finiteness of } X \cong \Gamma\backslash\mathbb{D} &\Rightarrow m_{\mu_z}(\Gamma\backslash\Omega) < \infty \\ &\Rightarrow \text{ Geodesic flow on } \Gamma\backslash\Omega \text{ is conservative}\end{aligned}$$

is also valid.[11]

5.3 Rudiments of ergodic theory

In this section, we briefly introduce some fundamental results from ergodic theory. The treatment is brief and covers only the bare minimum to proceed. We present a version of Birkhoff's Ergodic Theorem, one of the classical results in the field. This theorem, which is demonstrated for discrete dynamical systems, is then

[11] Patterson proved the first implication (starting from above) for Fuchsian groups essentially by obtaining a lower bound in regard to the Poincaré series (see [55], p. 266). This relies on Theorem 9.1.3 in Nicholls' book ([53], p. 151). Sullivan then generalized the result for any geometrically finite discrete group ([73], p. 270). For the second implication, refer to Theorem 7.2.9 in Nicholls' book ([53], p. 123).

adapted into a continuous version, which is subsequently used in the last part of the chapter.

It is assumed that (Ω, Σ, μ) is a measure space of finite Borel measure μ on which a measure-preserving discrete-time transform T (or a continuous-time flow) acts. The continuous version of the result also requires that (Ω, d_Ω) be a separable, complete, and locally compact metric space, and that the measure of any compact subset of Ω be finite.

It is worth highlighting that these requirements are not too stringent once we adopt Ω and m_{μ_z} as described in Section 5.1. The fact that (Ω, d_Ω) is a metric space is clear from the previous section, with the completeness of d_Ω following that of the metric $d_{\mathbb{D}}$. Since Ω can be thought of as a unit tangent bundle, $\mathcal{S}\mathbb{D}$, and since tangent bundles are manifolds in themselves, this takes care of separability and local compactness. The measure m_{μ_z} is defined via μ_z and the compact sets under the latter have finite measure.

5.3.1 Ergodic theory: discrete-time results

The main goal of this subsection is to prove a version of Birkhoff's Ergodic Theorem, one of the classical results in ergodic theory. Its proof relies on the so-called Maximal Ergodic Theorem.

Theorem 5.3.1 (Maximal Ergodic Theorem)**.** *If $f \in L^1(\Omega)$, let E be the set of points $x \in \Omega$ for which at least one of the sums*

$$f(x) + f(T(x)) + f(T^2(x)) + \cdots + f(T^n(x))$$

is non-negative. Then, $\int_E f d\mu \geq 0$.

Proof. See Section §1.3 and Theorem 1.16 in Walters' book for some of the missing details ([78], p. 24 and p. 37).

Let $U : L^1(\Omega) \to L^1(\Omega)$ be a positive linear operator associated with the measure-preserving transform T, sometimes called the Koopman or Composition Operator, such that $Uf \equiv f \circ T$.

Due to the fact that the transform T is measure-preserving, U is an isometry, and so U is unitary.

Define $f_0 = 0, f_n = f + Uf + U^2 f + \cdots + U^{n-1} f$ for $n \geq 1$, and $F_N = \max_{0 \leq n \leq N} f_n \geq 0$, for some integer $N > 0$. Consider the set

$$A(N) \equiv \{x \in \Omega : F_N(x) > 0\}.$$

We will show that $\int_{A(N)} f d\mu \geq 0$, from which the theorem follows.

Observe that $F_N \in L^1(\Omega)$ and that for $0 \leq n \leq N$ we have $F_N \geq f_n$. Hence, $UF_N \geq Uf_n$ by positivity, and so $UF_N + f \geq f_{n+1}$. Therefore,

$$\begin{aligned} UF_N(x) + f(x) &\geq \max_{1 \leq n \leq N} f_n(x) \\ &= \max_{0 \leq n \leq N} f_n(x) \quad \text{when } F_N(x) > 0 \\ &= F_N(x), \end{aligned}$$

which implies that $f \geq F_N - UF_N$ on $A(N)$.

Now,

$$\begin{aligned}\int_{A(N)} f d\mu &\geq \int_{A(N)} F_N d\mu - \int_{A(N)} UF_N d\mu \\ &= \int_{\Omega} F_N d\mu - \int_{A(N)} UF_N d\mu \text{ (since } F_N = 0 \text{ on } \Omega - A(N)) \\ &\geq \int_{\Omega} F_N d\mu - \int_{\Omega} UF_N d\mu \text{ (since F}_N \geq 0 \text{ and hence } UF_N \geq 0) \\ &= 0 \text{ (since } T \text{ is measure-preserving),}\end{aligned}$$

and the theorem is proved. □

The previous result is essentially all that is needed to prove Birkhoff's Ergodic Theorem as envisioned by Hopf. Note that by taking $g \equiv h \equiv 1$ one can recover Birkhoff's theorem in its classical formulation.

Theorem 5.3.2 (Hopf's version of Birkhoff's Ergodic Theorem)**.** *Suppose $f, g \in L^1(\Omega)$ with $g > 0$ and that for almost every $x \in \Omega$,*

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

Then the sequence

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

converges almost everywhere to a measurable function $\phi(x)$. The function ϕ is invariant under T, the function $g\phi$ is integrable, and for every bounded measurable function h invariant under T,

$$\int_{\Omega} g\phi h d\mu = \int_{\Omega} f h d\mu.$$

Proof. We divide the proof of the theorem into two parts.

Step 1) Existence of ϕ.

Take $a, b \in \mathbb{R}$ with $a < b$ and denote by $Y(a, b)$ the set

$$\left\{ x \in \Omega : \liminf_{n\to\infty} \frac{\sum_{j=0}^{n-1} f\left(T^j(x)\right)}{\sum_{j=0}^{n-1} g\left(T^j(x)\right)} < a < b < \limsup_{n\to\infty} \frac{\sum_{j=0}^{n-1} f\left(T^j(x)\right)}{\sum_{j=0}^{n-1} g\left(T^j(x)\right)} \right\}.$$

The set $Y(a, b)$ can be written as the intersection of the pre-images of the above lim inf and lim sup applied to open subsets $(-\infty, a)$ and (b, ∞). Thus, since the transform T is assumed measure-preserving, since f and g are assumed to be measurable, and since lim inf and lim sup of measurable functions are

themselves measurable, $Y(a,b)$ is measurable. Observe that one can write

$$\begin{aligned}\frac{\sum_{j=0}^{n-1} f\left(T^j(T(x))\right)}{\sum_{j=0}^{n-1} g\left(T^j(T(x))\right)} &= \frac{\sum_{j=1}^{n} f\left(T^j(x)\right)}{\sum_{j=1}^{n} g\left(T^j(x)\right)} \\ &= \frac{\frac{\sum_{j=0}^{n-1} f\left(T^j(x)\right)}{\sum_{j=0}^{n-1} g(T^j(x))} + \frac{f(T^n(x))}{\sum_{j=0}^{n-1} g(T^j(x))} - \frac{f(x)}{\sum_{j=0}^{n-1} g(T^j(x))}}{1 + \frac{g(T^n(x))}{\sum_{j=0}^{n-1} g(T^j(x))} - \frac{g(x)}{\sum_{j=0}^{n-1} g(T^j(x))}}.\end{aligned}$$

Taking into account the sequence in n of the last expression, the set $Y(a,b)$ is T-invariant (i.e., $x \in Y(a,b)$ implies $T(x) \in Y(a,b)$) for almost every $x \in \Omega$, since lim inf and lim sup respect the upper and lower bounds a and b, respectively. The result holds almost everywhere, because

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

holds for almost every $x \in \Omega$.

Moreover,

$$\sum_{j=0}^{n-1} \left[f\left(T^j(x)\right) - bg\left(T^j(x)\right)\right] \geq 0 \Leftrightarrow \frac{\sum_{j=0}^{n-1} f\left(T^j(x)\right)}{\sum_{j=0}^{n-1} g\left(T^j(x)\right)} \geq b$$

which is true almost everywhere in $Y(a,b)$ for some n. Now, the Maximal Ergodic Theorem (Theorem 5.3.1), with Ω replaced by $Y(a,b)$ and f replaced by $f - bg$, states that

$$\int_{Y(a,b)} f - bg d\mu \geq 0.$$

Using the other bound,

$$\int_{Y(a,b)} ag - f d\mu \geq 0,$$

and we can combine the results to get

$$(a-b)\int_{Y(a,b)} g d\mu \geq 0.$$

Since $a - b < 0$ and g is positive, the last inequality implies

$$0 = \int_{Y(a,b)} g d\mu \Rightarrow \mu(Y(a,b)) = 0,$$

where the null measure is a consequence of Chebyshev's inequality and continuity of the measure from below.

Applying this to all pairs of rationals (a,b) with $a < b$, the above argument shows that the limit

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

exists almost everywhere in Ω. The limit function $\phi(x)$ is measurable, since defined in terms of measurable functions, and is T-invariant for the same reason

that $Y(a,b)$ is T-invariant.

Step 2) Deriving the equation.

We now know that we can talk about $\phi(x)$ almost everywhere in Ω. Take $L_1 \in \mathbb{R}$. Observe that if $\phi(x) \geq L_1$ then by the properties of the lim inf (and the fact that the limit is equal to lim inf since it exists), for any $\epsilon > 0$ we can find some positive integer n such that

$$\frac{\sum_{j=0}^{n-1} f(T^j(x))}{\sum_{j=0}^{n-1} g(T^j(x))} \geq L_1 - \epsilon \Rightarrow \sum_{j=0}^{n-1} (f - (L_1 - \epsilon)g)(T^j(x)) \geq 0.$$

Then, applying the Maximal Ergodic Theorem to $f - (L_1 - \epsilon)g$, we obtain

$$\begin{aligned} \int_\Omega (f - g(L_1 - \epsilon)) d\mu \geq & 0 \\ \int_\Omega f d\mu \geq & L_1 \int_\Omega g d\mu, \end{aligned}$$

which holds almost everywhere.

Similarly, for $L_2 \in \mathbb{R}$, if $\phi(x) \leq L_2$, then

$$\begin{aligned} \int_\Omega ((L_2 + \epsilon)g - f) d\mu \geq & 0 \\ L_2 \int_\Omega g d\mu \geq & \int_\Omega f d\mu. \end{aligned}$$

Let us now define the set $X(k,n)$ for integers k, n, where

$$X(k,n) \equiv \left\{ x : \frac{k}{2^n} \leq \phi(x) \leq \frac{(k+1)}{2^n} \right\},$$

which is a set defined for almost all Ω.

Notice that this subset of Ω is measurable (i.e., ϕ is a measurable function and closed intervals are measurable with respect to the Borel σ-algebra) and T-invariant (i.e., ϕ is T-invariant). From what was found above, replacing Ω by $X(k,n)$, L_1 by $\frac{k}{2^n}$, and L_2 by $\frac{k+1}{2^n}$, we get

$$\frac{k+1}{2^n} \int_{X(k,n)} g d\mu \geq \int_{X(k,n)} f d\mu \geq \frac{k}{2^n} \int_{X(k,n)} g d\mu.$$

Multiplying the inequality that defines $X(k,n)$ by g (which we recall is positive), we get $\frac{k}{2^n} g(x) \leq \phi(x) g(x) \leq \frac{(k+1)}{2^n} g(x)$, so that we can use the same type of argument to obtain

$$\frac{k+1}{2^n} \int_{X(k,n)} g d\mu \geq \int_{X(k,n)} g\phi d\mu \geq \frac{k}{2^n} \int_{X(k,n)} g d\mu.$$

Subtracting the last inequalities from the first, we get

$$\frac{-1}{2^n} \int_{X(k,n)} g d\mu \leq \int_{X(k,n)} f d\mu - \int_{X(k,n)} g\phi d\mu \leq \frac{1}{2^n} \int_{X(k,n)} g d\mu,$$

which we can sum over k to obtain

$$\left|\int_\Omega f d\mu - \int_\Omega g\phi d\mu\right| \leq \frac{1}{2^n}\int_\Omega g d\mu.$$

Taking the limit $n \to \infty$ we find

$$\int_\Omega f d\mu = \int_\Omega g\phi d\mu.$$

The conclusion regarding h as described in the statement of the theorem is a consequence of using $f_1 = hf$ in the above equation, replacing f, and noting that by T-invariance it is possible to write $\sum_{j=0}^{n} f_1\left(T^j(x)\right) = \sum_{j=0}^{n-1} h(T^j(x)) f\left(T^j(x)\right) = \sum_{j=0}^{n-1} h(x) f\left(T^j(x)\right) = h(x)\sum_{j=0}^{n-1} f\left(T^j(x)\right)$, so that

$$\int_\Omega f h d\mu = \int_\Omega g\phi h d\mu.$$

□

5.3.2 Ergodic theory: continuous-time results

This subsection is devoted to adapting discrete-time results to a continuous-time environment. Since integrals of type $\int_0^u f(g_s(x))ds$ for $f \in L^1(\Omega)$ are important (i.e., they are related to time averages $\lim_{u\to\infty} \frac{1}{u}\int_0^u f(g_s(x))ds$), we refer the reader to a proof of their existence.

Theorem 5.3.3. *If $f \in L^1(\Omega)$ and g_s is a flow on Ω, then $f(g_s(x))$ is measurable as a function on $\Omega \times \mathbb{R}$ with respect to the product measure.*

Proof. For a proof, see Theorem 7.2.4 in Nicholl's book ([53], pp. 119-120). □

Having remarked on this technicality, the continuous-time version of Birkhoff's Ergodic Theorem reads as follows.

Theorem 5.3.4. *If g_s is a flow on Ω, if $f, h \in L^1(\Omega), h > 0$ and*

$$\int_0^u h\left(g_s(x)\right) ds \overset{u\to\infty}{\longrightarrow} \infty$$

almost everywhere on Ω, then the limit

$$\phi(x) = \lim_{u\to\infty} \frac{\int_0^u f\left(g_s(x)\right) ds}{\int_0^u h\left(g_s(x)\right) ds}$$

exists almost everywhere. The function ϕ is measurable and flow invariant. The function $h\phi$ is integrable over Ω and, for any bounded measurable g_s-invariant function ψ,

$$\int_\Omega h\phi\psi d\mu = \int_\Omega f\psi d\mu$$

Proof. Let $T = g_1$ and define

$$F(x) \equiv \int_0^1 f\left(g_s(x)\right) ds, \quad H(x) \equiv \int_0^1 h\left(g_s(x)\right) ds.$$

Using Fubini-Tonelli,

$$\int_\Omega |F(x)|\, d\mu(x) = \int_\Omega \left| \int_0^1 f(g_s(x))ds \right| d\mu(x) \leq \int_\Omega \int_0^1 |f(g_s(x))| ds d\mu(x)$$
$$= \int_0^1 \int_\Omega |f(g_s(x))| d\mu(x) ds,$$

and since g_s is measure preserving (with respect to μ) and $f \in L^1(\Omega)$, we have that $F \in L^1(\Omega)$. This holds similarly for H as defined above.

Observe that by the properties of the flow and $j \in \mathbb{Z}_{\geq 0}$,

$$\begin{aligned}\int_j^{j+1} f\left(g_s(x)\right) ds &= \int_0^1 f\left(g_{u+j}(x)\right) du \\ &= \int_0^1 f\left(g_{j+u}(x)\right) du \\ &= \int_0^1 f\left(g_u\left(g_j(x)\right)\right) du \\ &= F\left(T^j(x)\right),\end{aligned}$$

which implies

$$\sum_{j=0}^{n-1} F\left(T^j(x)\right) = \int_0^n f\left(g_s(x)\right) ds$$

The result now follows from the discrete version of the theorem, Theorem 5.3.2, using the transform T and the functions F, H. □

The next theorem gives a characterization of the ergodicity of the geodesic flow in terms of a relationship between time means and space means.

Theorem 5.3.5. *For a flow g_t on a space Ω, the following are equivalent:*

1. *The flow g_t is ergodic.*
2. *If $f, h \in L^1(\Omega)$ and provided $\int_\Omega h d\mu \neq 0$, then*

$$\lim_{u\to\infty} \frac{\int_0^u f\left(g_t(x)\right) dt}{\int_0^u h\left(g_t(x)\right) dt} = \frac{\int_\Omega f d\mu}{\int_\Omega h d\mu}$$

almost everywhere in Ω.

Proof. We prove the implications separately.

(1) ⇒ (2)

If g_t is ergodic, then the conservative set for the flow comprises all of Ω apart from null measure sets.[12] As a consequence, we may find a function h_0 satisfying

[12]The reason is as follows. Suppose that the dissipative set $D \subset \Omega$ associated with the flow g_t has positive μ measure. Consider the union $V = \bigcup_{t\in\mathbb{R}} g_t(D)$. This set is invariant under the flow, since any application of g_t only shifts the individual sets but not the union. Thus, by ergodicity, we have that $\mu(V) = 0$ or $\mu(V^c) = 0$. By monotonicity, we have $\mu(V) \geq \mu(D) > 0$, which implies that $\mu(V) = \mu(V) + \mu(V^c) = \mu(\Omega) < \infty$. But now the fact that D is dissipative leads to a contradiction, since we can find a disjoint subsequence of translations of D that, by monotonicity, imply $\infty < \mu(V) = \mu(\Omega) < \infty$.

the hypotheses of Theorem 5.3.4,[13] and so

$$\begin{aligned}\lim_{u\to\infty}\frac{\int_0^u f\left(g_t(x)\right)dt}{\int_0^u h\left(g_t(x)\right)dt} &= \lim_{u\to\infty}\frac{\int_0^u f\left(g_t(x)\right)dt}{\int_0^u h_0\left(g_t(x)\right)dt}\cdot\frac{\int_0^u h_0\left(g_t(x)\right)dt}{\int_0^u h\left(g_t(x)\right)dt}\\ &= \lim_{u\to\infty}\frac{\int_0^u f\left(g_t(x)\right)dt}{\int_0^u h_0\left(g_t(x)\right)dt}\cdot\frac{1}{\left(\int_0^u h\left(g_t(x)\right)dt\right)/\left(\int_0^u h_0\left(g_t(x)\right)dt\right)}\\ &= \frac{\int_\Omega f d\mu}{\int_\Omega h d\mu}.\end{aligned}$$

The last equality is a consequence of Theorem 5.3.4, because the function ϕ in that theorem is measurable and flow invariant, so by the ergodicity of the flow it must be constant almost everywhere.

(2) $\Rightarrow$ (1)

We will prove the implication by first proving that if M, M^* are two measurable sets of positive measures, then the hypothesis implies that there exists $t > 0$ such that $g_t(M) \cap M^* \neq \emptyset$.

Proof by contradiction. Let h be a positive integrable function and assume that M, M^* are sets of positive measures such that $g_t(M) \cap M^* = \emptyset$ for all t. If $f = 1_{M^*}$, then we have a contradiction with respect to the hypothesis if $x \in M$, since

$$\lim_{u\to\infty}\frac{\int_0^u f\left(g_t(x)\right)dt}{\int_0^u h\left(g_t(x)\right)dt} = 0 \neq \frac{\mu(M^*)}{\int_\Omega h d\mu}.$$

To show that this auxiliary result implies ergodicity, we again use proof by contradiction. Suppose that A is a measurable flow-invariant set with $\mu(A) > 0$ and $\mu(A^c) > 0$. Then, according to what was found above, there exists a t, such that

$$g_t(A) \cap A^c \neq \emptyset \Rightarrow A \cap A^c \neq \emptyset.$$

The conclusion, $A \cap A^c \neq \emptyset$, is a contradiction. Thus, because we are dealing with finite/probability measures, we must have either $\mu(A) = 0$ or $\mu(A^c) = 0$, verifying the definition of an ergodic flow. □

As seen, the key characterization proposed in the last theorem depends on an appropriately defined function h_0. The next theorem shows that this function exists and behaves as expected for the elements in the conservative set of the flow considered.[14]

Theorem 5.3.6. *If g_t is a flow on the space Ω then there exists a function $h \in L^1(\Omega), h > 0$, such that*

$$\int_0^u h\left(g_t(x)\right)dt \overset{u\to\infty}{\longrightarrow} \infty$$

for all $x \in C$, where C denotes the conservative set associated with the flow (i.e., $x \in C \Leftrightarrow g_t(x) \in F$, some compact set of Ω, for arbitrarily large times t).

[13] This is proved below, in Theorem 5.3.6, for the conservative set C. But here $C = \Omega$ apart from null measure sets, so this is why the hypothesis in Theorem 5.3.4 is validated.

[14] Notice that we are slightly abusing the notation, since h in Theorem 5.3.6 refers to h_0 and not h in Theorem 5.3.5.

Proof. Every separable metric space is second countable and therefore has a countable base.[15] By assumption, (Ω, d_Ω) is separable and locally compact. Thus, it has a countable base, and we can construct a sequence of compact sets

$$F_0 \subset F_1 \subset F_2 \subset \ldots, \quad \text{with } \lim_{n\to\infty} F_n = \Omega.^{16}$$

We assume that for each n there exists ϵ_n such that

$$\{x \in \Omega : d_\Omega(x, F_n) \leq \epsilon_n\} \subset F_{n+1}.^{17}$$

The existence of such ϵ_n ensures that the sets F_i in the above sequence, if not the same, are separated by some positive gap.

Let $\mu(F_0) = m_0 < \infty, \mu(F_n - F_{n-1}) = m_n < \infty, n = 1, 2, \ldots$. In the first case, the measure is finite by assumption, whereas in all the differences of sets both monotonicity and the assumption justify finiteness of the measure. We can write Ω as

$$\Omega = \left(\bigcup_{n=1}^{\infty} (F_n - F_{n-1}) \right) \bigcup F_0,$$

a disjoint union of sets.

Now we define the function $h : \Omega \to \mathbb{R}_{\geq 0}$ by

$$\begin{cases} h(x) = (2^n m_n)^{-1} & \text{if} \quad x \in F_n - F_{n-1} \\ h(x) = m_0^{-1} & \text{if} \quad x \in F_0. \end{cases}$$

The function h is measurable and $\int_\Omega h d\mu = \sum_{n=1}^{\infty} \int_{F_n - F_{n-1}} h(x) d\mu(x) + \int_{F_0} h(x) d\mu(x) = \left(\sum_{n=1}^{\infty} \frac{1}{2^n}\right) + 1 = 2$.

Given $x \in C$, we have that for some compact set F and sequence $\{t_n\}$ tending to infinity, $x_n = g_{t_n}(x) \in F$. For some N, we must have $F \subset F_N$, so we can define $\rho = \left(2^{N+1}\mu(F_{N+1})\right)^{-1} > 0$ such that

$$w \in F_{N+1} \Leftrightarrow w \in F_0 \text{ or } w \in F_n - F_{n-1}, n = 1, \ldots, N+1,$$

and this in turn implies $h(w) \geq \rho$, because

$$h(w) = m_0^{-1} \geq (2^{N+1}\mu(F_{N+1}))^{-1}, \text{ if } w \in F_0,$$

or

$$h(w) = (2^n m_n)^{-1} \geq (2^{N+1}\mu(F_{N+1}))^{-1}, \text{ if } w \in F_n - F_{n-1}, n = 1, \ldots, N+1.$$

Now, for every n we try to find a real number s_n such that $t_n < s_n < t_{n+1}$ and $d_\Omega(g_{t_n}(x), g_{s_n}(x)) > \epsilon_N$. Notice that if s_n does not exist, then the segment $\{g_t(x) : t_n \leq t \leq t_{n+1}\}$ of the trajectory of x does not escape F_{N+1}. Thus, we can consider the only two possible cases: whether the sequence $\{x_n\}$ escapes F_{N+1} infinitely often or not. Notice that it cannot escape F_{N+1} in a definitive fashion, because x is a conservative point for the flow, so it must always return

[15] See Folland ([24], p. 116).

[16] We could also have argued that Ω can be understood as a unit tangent bundle (see Section 5.1), thus a topological manifold, being second countable by definition.

[17] Details regarding this assumption can be found in Nemytskii and Stepanov's book on differential equations ([51], p. 477).

to $F \subset F_N \subset F_{N+1}$.

Case A) The sequence $\{x_n\}$ does not escape F_{N+1} infinitely often.

If the number s_n described above exists only for finitely many n, then for some n_0 we have $g_t(x) \in F_{N+1}$ for all $t \geq t_{n_0}$. Then,

$$\begin{aligned}\int_0^\infty h\left(g_t(x)\right) dt &= \int_0^{t_{n_0}} h\left(g_t(x)\right) dt + \int_{t_{n_0}}^\infty h\left(g_t(x)\right) dt \\ &> \sum_{n=n_0}^{\infty} \int_{t_n}^{t_{n+1}} h(g_t(x)) dt \\ &\geq \sum_{n_0}^{\infty} \rho \\ &= \infty.\end{aligned}$$

Case B) The sequence $\{x_n\}$ escapes F_{N+1} infinitely often.

If infinitely many s_n exist such that $g_{s_n}(x) \notin F_{N+1}$, then

$$\inf_n \left(s_n - t_n\right) = \epsilon > 0.$$

If this were not the case, that is, if $(s_n - t_n) \to 0$, observe that by continuity of the flow, the sequence $\{x_n\}$ would not escape F_{N+1} from some time t onward, contradicting the assumption. Therefore, the above infimum must be positive and

$$\int_0^\infty h\left(g_t(x)\right) dt \geq \sum_{n=1}^{\infty} \int_{t_n}^{s_n} h\left(g_t(x)\right) dt \geq \sum_{n=1}^{\infty} \int_{t_n}^{s_n} \rho dt \geq \sum_{n=1}^{\infty} \rho\epsilon = \infty,$$

concluding the proof. □

5.4 Conservation and ergodicity

The fact that conservation of the geodesic flow implies its ergodicity depends on the construction of an appropriate function λ for which the proof of Theorem 5.3.6 can be applied. The next lemma accomplishes just that.

Lemma 5.4.1. *Let Γ be a Fuchsian group preserving $\mathbb{D}$ and let m_μ be a measure defined on $\Gamma\backslash\Omega$ (as previously described in Section 5.1, with μ a Γ-invariant δ-dimensional conformal density). There exists a positive, m_μ integrable function λ on $\Gamma\backslash\Omega$ such that, for some real constant c and all $x, y \in \Gamma\backslash\Omega$ with $d_{\Gamma\backslash\Omega}(x, y) \leq 1$,*

$$\frac{|\lambda(x) - \lambda(y)|}{\lambda(y)} < c d_{\Gamma\backslash\Omega}(x, y)$$

holds.

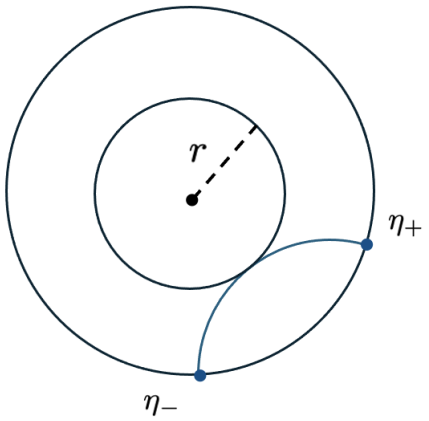


Figure 5.5: Argument used in Lemma 5.4.1.

Proof. Let the function $f : \Gamma\backslash\Omega \to \mathbb{R}_{\geq 0}$ be defined as the distance from the base point of the element of $\Gamma\backslash\Omega$ to the origin: given $\Gamma(x,\xi) \in \Gamma\backslash\Omega$, let $f(\Gamma(x,\xi)) \equiv d_{\mathbb{D}}(\Gamma x, 0)$. Define, for $r > 0$, the set $B_r = f^{-1}([0,r])$. There are three steps to complete the proof of the lemma.

Step 1) Proof that the sets B_r are bounded from above in m_μ-measure.

If we think of $\Gamma(x,\xi)$ in terms of Hopf coordinates, then this is equivalent to studying $(\gamma(\eta_-), \gamma(\eta_+), s + k(\gamma)), \gamma \in \Gamma$, where $k(\gamma)$ is a quantity that depends on γ (see Section 5.1). Each of these identifies a point along a given geodesic. If this point lies within hyperbolic distance r from the origin, then the triad is in B_r, by definition. Let us then ignore the point identified along the geodesic and only consider whether the geodesic itself intersects a hyperbolic disc of radius r. By computing or bounding the m_μ measure of $\Gamma(x,\xi)$ in terms of this criterion, we will then bound $m_\mu(B_r)$ from above, since even though certain points $(\gamma(\eta_-), \gamma(\eta_+), s + k(\gamma))$ may not lie within hyperbolic distance r from the origin, the geodesic identified by $\gamma(\eta_-)$ and $\gamma(\eta_+)$ may still contribute positively to the m_μ measure if it intersects said disc.

With this strategy in mind, let us bound the m_μ measure of the set of all geodesics that intersect a hyperbolic disc of radius r.

Given a point $z \in \mathbb{D}$ and endpoints q_-, q_+ of a geodesic, with $q_+ \neq q_-$, the hyperbolic distance s from z to the geodesic joining q_- and q_+ is given by

$$\cosh s = \frac{2|z - q_-||z - q_+|}{|q_- - q_+|(1 - |z|^2)}.$$ [18]

Consider a hyperbolic circle around the origin having radius r. Consider a geodesic with endpoints q_-, q_+, with $q_- \neq q_+$, tangent to this circle (see Figure 5.5). The hyperbolic distance from the origin to the tangent point z lying on the geodesic is r (i.e., the hemisphere of $\mathbb{D}$ is the unique geodesic passing through

[18] For a proof, see Theorem 1.2.1 in Nicholls'book ([53], pp. 3, 14-15).

the origin and z). Therefore, applying the above formula, we have

$$|q_+ - q_-| = \frac{2|q_-||q_+|}{\cosh r} = \frac{2}{\cosh r}.$$

Using the definition of $\cosh r$ in the expression above, we get the inequality

$$|q_+ - q_-| = \frac{2}{\cosh r} = 2\left(\frac{2}{e^r + e^{-r}}\right) < 4e^{-r}.$$

Thus, if we want to consider all geodesics in $\mathbb{D}$ that pass within a hyperbolic distance r of the origin, we must take Hopf coordinate triads such that the first two coordinates fulfill

$$|q_+ - q_-| \geq 4e^{-r} \Leftrightarrow \frac{1}{|q_+ - q_-|^{2\delta}} \leq 4^{-2\delta} e^{2\delta r}.$$

Moreover, the Euclidean distance between these intersecting geodesics and the origin is $\leq r$. Then, if we call the set of such geodesics A, by definition of the measure m_μ and monotonicity, we have the estimate:

$$\begin{aligned} m_\mu(B_r) \leq m_\mu(A) &= \int_A dm_\mu(q_-, q_+, s) \\ &= \int_A \frac{d\mu(q_-)d\mu(q_+)ds(s')}{|q_+ - q_-|^{2\delta}} \\ &= 4^{-2\delta} e^{2\delta r} \mu(A)^2 \int_A ds(s') \\ &\leq Cre^{2\delta r}, \end{aligned}$$

where C is some constant (due to the finiteness of μ) and $\int_A ds(s')$ should be understood as the arc length of the geodesics in A (and so, since the closure of A is compact, this value is a multiple of r).

Step 2) Proof that a particular choice of λ validates the inequality in the statement of the lemma.

Take $\epsilon > 0$ and define the function λ by

$$\lambda(x') \equiv e^{-(2\delta+\epsilon)f(x')}.$$

Naturally, $\lambda > 0$ for all $x' = \Gamma(x, \xi) \in \Gamma\backslash\Omega$.

By the Mean Value Theorem, for some c' between (but different from) 0 and t, we have $|e^t - 1| = e^{c'}|t| \leq e^{|t|}|t|$. Therefore, the inequality in the lemma statement holds because for $x' = \Gamma(x, \xi) \in \Gamma\backslash\Omega$ and $y' = \Gamma(y, \eta) \in \Gamma\backslash\Omega$,

$$\begin{aligned} \left|\frac{\lambda(x') - \lambda(y')}{\lambda(y')}\right| &= \left|e^{(2\delta+\epsilon)(f(y') - f(x'))} - 1\right| \\ &= \left|e^{(2\delta+\epsilon)(d_{\mathbb{D}}(\Gamma y, 0) - d_{\mathbb{D}}(\Gamma x, 0))} - 1\right| \\ &\leq e^{(2\delta+\epsilon)} e^{|d_{\mathbb{D}}(\Gamma y, 0) - d_{\mathbb{D}}(\Gamma x, 0)|} |2\delta + \epsilon| \, |d_{\mathbb{D}}(\Gamma y, 0) - d_{\mathbb{D}}(\Gamma x, 0)| \\ &\leq ce^{d_{\mathbb{D}}(\Gamma x, \Gamma y)} d_{\mathbb{D}}(\Gamma x, \Gamma y) \\ &\leq ced_{\mathbb{D}}(\Gamma x, \Gamma y), \end{aligned}$$

for some constant c and where the last inequality is a consequence of $1 \geq d_{\Gamma\backslash\Omega}(x', y') = d_\Omega(\Gamma(x,\xi), \Gamma(y,\eta)) \geq d_{\mathbb{D}}(\Gamma x, \Gamma y)$.

Step 3) Proof that the choice of λ is integrable with respect to m_μ.

Notice that for every $x \in B_{n+1} - B_n$, $f(x) \in (n, n+1]$, so $\lambda(x) = \frac{1}{e^{(2\delta+\epsilon)f(x)}} < \frac{1}{e^{(2\delta+\epsilon)n}}$. Moreover, by monotonicity, $m_\mu(B_{n+1} - B_n) \leq m_\mu(B_{n+1})$, while, based on Step 1), $m_\mu(B_{n+1})$ is at most a constant times $ne^{2\delta n}$. Therefore,

$$\int_{B_{n+1}-B_n} \lambda dm_\mu < e^{-(2\delta+\epsilon)n} \int_{B_{n+1}-B_n} dm_\mu \leq cne^{-\epsilon n},$$

where c is some constant, from which we get

$$\begin{aligned} \int_{\Gamma\backslash\Omega} \lambda dm_\mu &= \sum_{n=0}^{\infty} \int_{B_{n+1}-B_n} \lambda dm_\mu \\ &< \sum_{n=0}^{\infty} cne^{-\epsilon n} < \infty. \end{aligned}$$

This concludes the proof of the lemma. □

It is time to combine all the tools at our disposal and prove the main result of the chapter.

Theorem 5.4.2. *Let Γ be a Fuchsian group preserving $\mathbb{D}$ and consider an α-dimensional conformal density $\mu \equiv \mu_0$ that is invariant under Γ. Let m_μ be the induced measure on $\Gamma\backslash\Omega$. If the geodesic flow on $\Gamma\backslash\Omega$ is conservative with respect to the measure m_μ then it is ergodic.*

Proof. Take a function f defined on $\Gamma\backslash\Omega$ that is integrable with respect to m_μ. Suppose that λ is the function obtained in Lemma 5.4.1. By assumption, g_t is conservative, so we can use the proof of Theorem 5.3.6 to deduce

$$\lim_{T\to\infty} \int_0^T \lambda\left(g_t(x)\right) dt = +\infty$$

almost everywhere, relative to the m_μ measure on $\Gamma\backslash\Omega$.

There are two steps to conclude the proof.

Step 1) Simplifying the problem using a key property of L^1.

Define the function

$$F(T,x) \equiv \frac{\int_0^T f\left(g_t(x)\right) dt}{\int_0^T \lambda\left(g_t(x)\right) dt}.$$

By Theorem 5.3.4, it converges almost everywhere as $T \to \infty$ to a measurable function that is invariant under g_t. Denote

$$F(x) \equiv \lim_{T\to\infty} F(T,x),$$

and using the second part of said theorem, with g replaced by λ, ϕ replaced by F, and h replaced by $\text{sign}(F)$ (a bounded, m_μ-measurable, since F is measurable, and flow-invariant function) we can study the signs of F and f to deduce

$$\int_{\Omega/\Gamma} \lambda |F| dm_\mu \leq \int_{\Omega/\Gamma} |f| dm_\mu.$$

Take a sequence of L^1-functions f_n converging in L^1 (in the m_μ-measure) to f and form the corresponding F_n sequence, that is,

$$F_n(x) \equiv \lim_{T\to\infty} \frac{\int_0^T f_n\left(g_t(x)\right) dt}{\int_0^T \lambda\left(g_t(x)\right) dt}.$$

In the same way that the above inequality was obtained, we get the inequality below

$$\int_{\Omega/\Gamma} \lambda |F - F_n| dm_\mu \leq \int_{\Omega/\Gamma} |f - f_n| dm_\mu \overset{n\to\infty}{\longrightarrow} 0,$$

implying $F_n \to F$.

The end goal of the proof is to show that F is constant almost everywhere. Since λ is m_μ-integrable on $\Gamma\backslash\Omega$ (Lemma 5.4.1) and f is assumed to be m_μ-integrable on $\Omega\backslash\Gamma$, the quotient $\frac{\int_{\Gamma\backslash\Omega} f dm_\mu}{\int_{\Gamma\backslash\Omega} \lambda dm_\mu}$ is also constant. Thus, in the case of constant F almost everywhere, we would have

$$F(x) = \frac{\int_{\Gamma\backslash\Omega} f dm_\mu}{\int_{\Gamma\backslash\Omega} \lambda dm_\mu},$$

which would allow Theorem 5.3.5 to be applied to conclude that the flow g_t on $\Gamma\backslash\Omega$ is ergodic.

Now, recall that continuous functions of compact support are dense in $L^1(m_\mu)$.[19] This fact implies that our arbitrary f can be approximated through a sequence of continuous functions f_n of compact support. Hence, using the last inequality stated above, the constancy of F almost everywhere can be proved by showing that the functions F_n, which are now defined in terms of continuous functions f_n of compact support, are constant almost everywhere.

In this sense, with some abuse of notation, assume that f is continuous with compact support. The goal is now to show that its corresponding function F is constant almost everywhere.

Step 2) The Hopf argument.

Keep in mind that F is always invariant under the flow, including when F is defined by a continuous f of compact support, since we obtained F using Theorem 5.3.4.

Recall as well the Hopf coordinates identification

$$\Omega \Leftrightarrow \mathbb{D} \times \partial\mathbb{D} \overset{\text{Hopf coordinates}}{\Longleftrightarrow} ((\partial\mathbb{D} \times \partial\mathbb{D}) - \Delta) \times \mathbb{R},$$

[19] See Proposition 7.9 in Folland's book ([24], p. 217).

where Δ denotes the diagonal in $\partial\mathbb{D} \times \partial\mathbb{D}$.

The first objective is to prove that F is constant for elements $x, x' \in \Gamma\backslash\Omega$ that are asymptotically the same, in the sense that they share either the first (coset) Hopf coordinate or the second (coset) Hopf coordinate.

Without loss of generality, suppose that x and x' share the second (coset) Hopf coordinate. In practice, this means that x and x' have the property that for a fixed $a \in \mathbb{R}$,

$$(*) \quad d_{\Gamma\backslash\Omega}\left(g_t(x), g_{t+a}\left(x'\right)\right) \to 0 \text{ as } t \to +\infty.$$

Now, observe that

$$\begin{aligned} F\left(x'\right) = F(g_a(x')) &\equiv \lim_{T\to\infty} \frac{\int_0^T f\left(g_t\left(g_a(x')\right)\right) dt}{\int_0^T \lambda\left(g_t\left(g_a(x')\right)\right) dt} \\ &= \lim_{T\to\infty} \frac{\int_0^T f\left(g_{a+t}\left(x'\right)\right) dt}{\int_0^T \lambda\left(g_{a+t}\left(x'\right)\right) dt}, \end{aligned}$$

where we have used the fact that $\int_0^T \lambda\left(g_t\left(x'\right)\right) dt$ is unbounded and F is invariant under the flow.

The first objective entails showing that

$$\lim_{T\to\infty} \left(\frac{\int_0^T f\left(g_t(x)\right) dt}{\int_0^T \lambda\left(g_t(x)\right) dt} - \frac{\int_0^T f\left(g_{t+a}\left(x'\right)\right) dt}{\int_0^T \lambda\left(g_{t+a}\left(x'\right)\right) dt} \right) = 0,$$

since by what was found above this implies $F(x) - F(x') = 0$.

To see that the limit holds, notice that the subtraction

$$\frac{\int_0^T f\left(g_t(x)\right) dt}{\int_0^T \lambda\left(g_t(x)\right) dt} - \frac{\int_0^T f\left(g_{t+a}\left(x'\right)\right) dt}{\int_0^T \lambda\left(g_{t+a}\left(x'\right)\right) dt}$$

can be written as the difference between

$$\text{(I)} \quad \frac{\int_0^T \left[\left(f\left(g_t(x)\right) - f\left(g_{t+a}\left(x'\right)\right)\right) / \lambda\left(g_t(x)\right)\right] \lambda\left(g_t(x)\right) dt}{\int_0^T \lambda\left(g_t(x)\right) dt}$$

and

$$\text{(II)} \quad \frac{\int_0^T f\left(g_{t+a}\left(x'\right)\right) dt}{\int_0^T \lambda\left(g_{t+a}\left(x'\right)\right) dt} \cdot \frac{\int_0^T \left[\left(\lambda\left(g_t(x)\right) - \lambda\left(g_{t+a}\left(x'\right)\right)\right) / \lambda\left(g_t(x)\right)\right] \lambda\left(g_t(x)\right) dt}{f_0^T \lambda\left(g_t(x)\right) dt}.$$

Now, recall that continuous functions attain a maximum value in compact sets. Then, since f is assumed to be continuous of compact support in $\Gamma\backslash\Omega$, $f(g_t(x))$ attains a maximum value where it is supported (a subset of $\Gamma\backslash\Omega$). Similarly for $f(g_{t+a}(x'))$. This means that the difference

$$f\left(g_t(x)\right) - f\left(g_{t+a}\left(x'\right)\right)$$

in (I) is bounded. However, the denominator is unbounded (Theorem 5.3.6), so (I) goes to zero in the limit $T \to \infty$. On the other hand, due to the property

of λ stated in Lemma 5.4.1 and $(*)$ above, (II) also goes to zero as $T \to \infty$. Therefore, the limit is indeed true, and we have $F(x) = F(x')$.

The same fact concerning F holds for $x, x' \in \Gamma\backslash\Omega$ that have a same first (coset) Hopf coordinate, that is, elements for which

$$d_{\Gamma\backslash\Omega}\left(g_{-t}(x), g_{-t+b}(x')\right) \to 0 \text{ as } t \to +\infty,$$

for some $b \in \mathbb{R}$.

This concludes the proof that F has the same value for points that are asymptotically the same, in the sense described.

Now, consider the lift of F from $\Gamma\backslash\Omega$ to Ω, denoted $\widetilde{F} \equiv F \circ \pi$, where $\pi : \Omega \to \Gamma\backslash\Omega$ is the canonical projection. Notice that $\widetilde{F}$ is Γ-invariant, because for any $v \in \Omega$ and $\gamma \in \Gamma$, we have

$$\widetilde{F}(\gamma(v)) = F(\pi(\gamma(v))) = F(\Gamma\gamma(v)) = F(\Gamma v) = F(\pi(v)) = \widetilde{F}(v).$$

Recall that we can describe Ω in terms of Hopf coordinates. This implies that $\widetilde{F}$ is defined on $((\partial\mathbb{D} \times \partial\mathbb{D}) - \Delta) \times \mathbb{R}$. However, because F is flow invariant with respect to the geodesic flow, we can take $\widetilde{F}$ to be a $\partial\mathbb{D} \times \partial\mathbb{D}$ (minus diagonal) function, such that the value of $\widetilde{F}$ depends only on the geodesic considered.

Furthermore, the asymptotic property proved previously for F implies that $\widetilde{F}$ is constant on $\xi \times \partial\mathbb{D}$ and $\partial\mathbb{D} \times \xi$, for almost all $\xi \in \partial\mathbb{D}$, since, as seen, $\widetilde{F}$ is Γ-invariant. The "almost every" qualification comes from the definition of F, whose limit holds everywhere except on null measure sets.

To conclude, observe that F is measurable in $\Gamma\backslash\Omega$ with respect to the m_μ measure (in the quotient sense). Since m_μ is invariant under Γ (because μ is assumed to be invariant under the action of this group), the lift $\widetilde{F}$ is measurable in Ω with respect to m_μ (in the non-quotient sense). Thus, it is possible to use Fubini's Theorem, showing that $\widetilde{F}$ is constant almost everywhere on $\partial\mathbb{D} \times \partial\mathbb{D}$ (minus diagonal).[20] As a consequence, F is constant almost everywhere on $\Gamma\backslash\Omega$. □

Therefore, Theorem 5.4.2 combined with Theorem 5.2.3 implies that the geodesic flow is ergodic. The ergodicity of the geodesic flow in turn has the ergodicity of the action of Γ on $\partial\mathbb{D}$ as its logical consequence, a result we will rely on in the following section.

[20] Suppose $\widetilde{F}$ is the lift of F from $\Gamma\backslash\Omega$ to Ω. Then, by redefining $\widetilde{F}$ to be zero on the diagonal of $\partial\mathbb{D} \times \partial\mathbb{D}$, the integrability of $\widetilde{F}$ implies (for some $s \in \mathbb{R}$) that

$$\begin{aligned}
\infty > & \int_{\partial\mathbb{D}\times\partial\mathbb{D}} \widetilde{F}(\eta_-, \eta_+, s) \frac{d\mu(\eta_-)d\mu(\eta_+)}{|\eta_- - \eta_+|^{2\alpha}} \\
& \overset{\text{Fubini}}{=} \int_{\partial\mathbb{D}} \left(\int_{\partial\mathbb{D}} \widetilde{F}(\eta_-, \eta_+, s) \frac{d\mu(\eta_+)}{|\eta_- - \eta_+|^{2\alpha}} \right) d\mu(\eta_-) \\
& \overset{\text{Constant a.e.}}{=} \int_{\partial\mathbb{D}} \widetilde{F}(\eta_-, \eta_+, s) \left(\int_{\partial\mathbb{D}} \frac{d\mu(\eta_+)}{|\eta_- - \eta_+|^{2\alpha}} \right) d\mu(\eta_-) \\
& \overset{\text{Constant a.e.}}{=} \widetilde{F}(\eta_-, \eta_+, s) \int_{\partial\mathbb{D}} \left(\int_{\partial\mathbb{D}} \frac{d\mu(\eta_+)}{|\eta_- - \eta_+|^{2\alpha}} \right) d\mu(\eta_-) \\
& \overset{\text{Tonelli-Fubini}}{=} \widetilde{F}(\eta_-, \eta_+, s) \int_{\partial\mathbb{D}\times\partial\mathbb{D}} \frac{d\mu(\eta_-)d\mu(\eta_+)}{|\eta_- - \eta_+|^{2\alpha}},
\end{aligned}$$

showing that $\widetilde{F}$ is almost everywhere constant on $\partial\mathbb{D} \times \partial\mathbb{D}$.

Corollary 5.4.3. *The action of* Γ *on* $\partial\mathbb{D}$ *is ergodic with respect to* μ *(defined as in the preceding theorem).*

Proof. Suppose A is a Γ-invariant subset of $(\partial\mathbb{D} \times \partial\mathbb{D}) - \Delta$, where Δ denotes the diagonal. Then $A \times \mathbb{R} \subset \Omega$ is invariant with respect to both Γ and the geodesic flow (i.e., the flow only changes the last Hopf coordinate, but $\mathbb{R}$ contains all possible changes already). Now, since $\Gamma\backslash(A \times \mathbb{R})$ is a flow-invariant subset of $\Gamma\backslash\Omega$ (i.e., the flow commutes with Γ), the ergodicity of the geodesic flow derived above implies that either $m_\mu(A \times \mathbb{R}) = 0$ or $m_\mu(A^c \times \mathbb{R}) = 0$. But these alternatives imply $\tilde{\mu}(A) = 0$ or $\tilde{\mu}(A^c) = 0$,[21] respectively, since $\mathbb{R}$ cannot have a null Lebesgue measure.

Having proven that the product action of Γ on $(\partial\mathbb{D} \times \partial\mathbb{D}) - \Delta$ is ergodic, we use this fact to prove the desired result. Let $E \subset \partial\mathbb{D}$ be invariant with respect to Γ. Consequently, $E' \equiv (E \times E) - \Delta$ is invariant under the product action, so either $\tilde{\mu}(E') = 0$ or $\tilde{\mu}((E')^c) = 0$. In the first case, $(\mu \times \mu)(E') = 0$. Using Fubini's Theorem and the non-atomicity of μ (Theorem 4.8.5), $\mu(E) = 0$. In the second case, since $E^c \times E^c \subset (E \times E)^c$, the same reasoning and monotonicity imply $\mu(E^c) = 0$. □

5.5 Uniqueness of Patterson-Sullivan measures and ergodicity

As touched on in Chapter 4, the question of whether a Patterson-Sullivan measure (when viewed as a conformal density of some dimension) is unique or not depends on the Fuchsian group Γ. In particular, if there exists an invariant conformal density of dimension α, the next theorem establishes a sufficient condition for it to be the unique conformal density of said dimension (up to a multiple by a constant). Interestingly, the outcome hinges on the ergodicity of the group.

Theorem 5.5.1. *Let* Γ *be a Fuchsian group acting on* $\mathbb{D}$ *and, given* $\alpha > 0$*, suppose that* σ *is an* α*-dimensional* Γ*-invariant conformal density. Then, the collection of all* Γ*-invariant conformal densities of dimension* α *is equal to the set of nonzero constant multiples of* σ *if* Γ *is ergodic on* $\partial\mathbb{D}$ *with respect to the class of measures defined by* σ*.*

Proof. Assume that Γ is ergodic on $\partial\mathbb{D}$ with respect to the class of measures σ. If we take another conformal density π of dimension α, define $\nu \equiv (\sigma + \pi)/2$. It is clear ν checks the first and third conditions of Definition 4.9.1. To verify the second condition, since $\sigma_x \ll \sigma_{x'} \ll \nu_{x'}$ and $\pi_x \ll \pi_{x'} \ll \nu_{x'}$ (i.e., if a set has zero measure with respect to $\nu_{x'}$, for some $x' \in \mathbb{D}$, then the set has null measure $\sigma_{x'}$ and $\pi_{x'}$, since these measures are nonnegative finite measures), by

[21] The measure $\tilde{\mu}$ is defined as the non-Lebesgue part of the measure m_μ.

the properties of the Radon-Nikodym derivative we have

$$\begin{aligned}\frac{d(\sigma_x+\pi_x)}{d(\sigma_{x'}+\pi_{x'})}(\xi) &= \frac{d\sigma_x}{d(\sigma_{x'}+\pi_{x'})}(\xi)+\frac{d\pi_x}{d(\sigma_{x'}+\pi_{x'})}(\xi)\\ &= \frac{d\sigma_x}{d\sigma_{x'}}(\xi)\frac{d\sigma_{x'}}{d(\sigma_{x'}+\pi_{x'})}(\xi)+\frac{d\pi_x}{d\pi_{x'}}(\xi)\frac{d\pi_{x'}}{d(\sigma_{x'}+\pi_{x'})}(\xi)\\ &= \left(\frac{P(x,\xi)}{P(x',\xi)}\right)^{\alpha}\left(\frac{d\sigma_{x'}}{d(\sigma_{x'}+\pi_{x'})}(\xi)+\frac{d\pi_{x'}}{d(\sigma_{x'}+\pi_{x'})}(\xi)\right)\\ &= \left(\frac{P(x,\xi)}{P(x',\xi)}\right)^{\alpha}\end{aligned}$$

almost everywhere, for $x, x' \in \mathbb{D}$ and $\xi \in \partial\mathbb{D}$. So, ν is a conformal density of dimension α as well.

As argued above, $\sigma_x \ll \nu_x$ and $\pi_x \ll \nu_x$, so the Radon-Nikodym derivatives $\frac{d\sigma_x}{d\nu_x}$ and $\frac{d\pi_x}{d\nu_x}$ exist and are measurable with respect to ν_x. But both are also measurable with respect to the measure σ_x, since the latter derivative is measurable with respect to π_x and ν_x, and by the definition of ν_x, this implies that it is σ_x-measurable.

To show that as functions on $\partial\mathbb{D}$, these derivatives are Γ-invariant, we proceed as follows. Take $\xi \in \partial\mathbb{D}$ and, for $t > 0$, let $I(\xi, t)$ denote an open interval in $\partial\mathbb{D}$ centered at ξ and having Euclidean radius t. Given the characteristic property of conformal densities discussed in Sections 4.6 and 4.9, we write

$$\frac{\sigma_x[\gamma(I(\xi,t))]}{\nu_x[\gamma(I(\xi,t))]} = \frac{\int_{I(\xi,t)} |\gamma_x'(\eta)|^{\alpha}\, d\sigma_x(\eta)}{\int_{I(\xi,t)} |\gamma_x'(\eta)|^{\alpha}\, d\nu_x(\eta)}.$$

By differentiation of measures and continuity of $|\gamma_x'(\eta)|$,[22] taking the limit $t \to 0$ leads to

$$\left(\frac{d\sigma_x}{d\nu_x}\right)(\gamma(\xi)) = \left(\frac{d\sigma_x}{d\nu_x}\right)(\xi).$$

This shows that $\frac{d\sigma_x}{d\nu_x}$ is not only measurable (in the sense of σ_x) but also Γ-invariant on $\partial\mathbb{D}$. By assumption, the action of Γ is ergodic on $\partial\mathbb{D}$ with respect to σ_x, and so any Γ-invariant measurable function is constant almost everywhere,[23] implying that $\frac{d\sigma_x}{d\nu_x}$ is constant almost everywhere on $\partial\mathbb{D}$.

To see that this constant cannot be zero, consider a Γ-invariant Borel set $A \subset \partial\mathbb{D}$. Then, either

$$0 = \sigma_x(A) = \int_A \frac{d\sigma_x}{d\nu_x}(\xi) d\nu_x(\xi) \text{ or } 0 = \sigma_x(A^c) = \int_{A^c} \frac{d\sigma_x}{d\nu_x}(\xi) d\nu_x(\xi).$$

Notice that if $\frac{d\sigma_x}{d\nu_x}(\xi) = 0$, then $0 = \sigma_x(A) + \sigma_x(A^c) = \sigma_x(\partial\mathbb{D})$, and this is not allowed, since σ_x is a positive finite measure.

[22] This derivative can be expressed as $\frac{P(\gamma^{-1}x,\eta)}{P(x,\eta)}$ (see Lemma 3.4.2 in Nicholls [53], p. 55). It is therefore continuous.

[23] This can be proved by contradiction. Suppose that there exists a Γ-invariant function f such that, for real numbers $a_1 < a_2$, the sets $A_1 = \{z \in \partial\mathbb{D} : f(z) \le a_1\}$ and $A_2 = \{z \in \partial\mathbb{D} : f(z) \ge a_2\}$ have positive σ_x-measure. Then, take some $a \in (a_1, a_2)$ and observe that the set $A = \{z \in \partial\mathbb{D} : f(z) > a\}$ is Γ-invariant. By ergodicity, $\sigma_x(A) = \sigma_x(\partial\mathbb{D})$, and so $\sigma_x(A_1) \le \sigma_x(A^c) = 0$, a contradiction.

Then, by the properties of the Radon-Nikodym derivative and the definition of the measure ν_x, $\frac{d\pi_x}{d\nu_x}$ is also a positive constant almost everywhere. Thus, $\frac{d\pi_x}{d\nu_x} = \frac{d\pi_x}{d\sigma_x}\frac{d\sigma_x}{d\nu_x}$ implies that $\frac{d\pi_x}{d\sigma_x}$ is a positive constant almost everywhere, proving the result. □

Although not needed in what follows, the converse of the theorem's implication is also true.[24]

In the last section, it was proved (under the usual assumptions) that the geodesic flow is ergodic with respect to the m_μ measure and that this ergodicity implies the ergodicity of Γ on $\partial\mathbb{D}$ with respect to a measure μ used to define m_μ (Corollary 5.4.3). By the just proved theorem, this implies that μ is the unique conformal density of dimension δ (up to a constant multiple). That is, the Patterson-Sullivan measure μ is unique.

It is worth mentioning that the ergodicity of a (non-elementary, geometrically finite) group Γ on $\partial\mathbb{D}$ (with respect to some existing conformal density) can also be demonstrated without taking into account the geodesic flow on a hyperbolic surface. Using the converse of the last theorem, this yields another route toward proving the uniqueness of a conformal density of given dimension (up to scaling). The proof involves the set of radial/conical limit points, the non-atomicity of class measures, and arguments that make an appearance in the proof of Sullivan's Shadow Lemma.[25] An investigation of this lemma is the main topic of the next chapter.

[24] See Theorem 4.2.1 in Nicholls' book ([53], p. 70).
[25] Refer to Theorem 4.4.4 in Nicholls' book ([53], p. 78).

Chapter 6

Hausdorff Measure of the Limit Set

As another illustration of the power of Patterson-Sullivan measures, this chapter addresses the relationship between conformal densities and the Hausdorff measure of a limit set $\Lambda(\Gamma)$. Recall that the Hausdorff measure is a classical fractional measure. The focus is on showing that the Hausdorff dimension of the limit set of a Fuchsian group is precisely the critical exponent $\delta(\Gamma)$ controlling the convergence of the (absolute) Poincaré series, that is, the dimension associated with Patterson-Sullivan measures. This is accomplished for convex cocompact groups, a simpler setting in which cusps are avoided, using shadows (to be defined below), although the result is valid more generally for geometrically finite hyperbolic surfaces.[1]

6.1 Hausdorff measure

Recall that s-dimensional Hausdorff measures are essentially fractal analogues of the Lebesgue measure. The Hausdorff measure of a set is found by shrinking a set covering to arbitrarily small scales and computing a type of average of the contributions of the pieces of this compressed cover. The dimension s regulates by how much each piece of the cover contributes to the average, a role similar to that of s in the Poincaré series (which justifies the same notation). Since we are interested in limit sets $\Lambda(\Gamma) \subseteq \partial\mathbb{D}$, the Hausdorff measure is defined for sets lying on the boundary of the unit disc.

Definition 6.1.1. For $A \subset \partial\mathbb{D}$, $s \in [0, \infty)$, define the s-dimensional Hausdorff measure $H^s(A)$ as

$$H^s(A) \equiv \lim_{\varepsilon \to 0} \inf \left\{ \sum_{j=1}^{\infty} |I_j|^s : A \subset \bigcup_{j=1}^{\infty} I_j, |I_j| < \varepsilon \right\},$$

where the infimum is taken over all countable covers of A and $|\cdot|$ denotes the Euclidean arc length in $\partial\mathbb{D}$.

[1] See Sullivan's original result ([73], pp. 260-261) or Section 9.3 in Nicholls' book ([53], p. 154).

The definition of Hausdorff dimension concerns the exponent s which is part of the definition of the measure.

Definition 6.1.2. The Hausdorff dimension $\dim_H(A)$ of $A \subset \partial\mathbb{D}$ is defined by

$$\dim_H(A) \equiv \inf \{s \geq 0 : H^s(A) = 0\},$$

or, equivalently,

$$\dim_H(A) \equiv \sup \{s \geq 0 : H^s(A) = \infty\},$$

except that, when this last set of numbers s is empty, the Hausdorff dimension is zero.

The Hausdorff dimension is then the exponent s for which the Hausdorff measure falls from infinity to zero.[2]

Proposition 6.1.3. *The Hausdorff dimension of $A \subset \partial\mathbb{D}$ is such that*

$$H^s(A) = \begin{cases} \infty, & 0 \leq s < \dim_H(A) \\ 0, & \dim_H(A) < s. \end{cases}$$

Proof. From the definitions. □

We wish to find a relationship between Patterson-Sullivan measures μ_z and the Hausdorff measures H^s. A key development in this endeavor is a result due to Sullivan, namely Sullivan's Shadow Lemma. Before its statement and proof in the next section, we introduce the geometric notion of a shadow and obtain a bound for its arc length measure.

Take $w \in \mathbb{D}$ and $r > 0$ such that $0 \notin B(w, r)$. Imagine a flashlight fixed at the origin, emitting light against the hyperbolic disc $B(w, r)$, such that $B(w, r)$ casts a shadow on $\partial\mathbb{D}$, which can be described by the set

$$I(w; r) \equiv \{q \in \partial\mathbb{D} : d_{\mathbb{D}}([0, q], w) < r\},$$

where $[0, q]$ denotes the geodesic ray from 0 to q.

As illustrated in Figure 6.1, we can consider a geodesic ray from 0 that is tangential to $B(w, r)$ at some point z. We can then form a hyperbolic triangle between geodesic arcs $[0, w]$, $[w, z]$, and $[0, z]$. This triangle has some angle θ at the origin vertex, and we can therefore deduce the arc length size of the shadow set in this case: $|I(w; r)| = 2\theta$. Given the hyperbolic version of the sine rule,[3] it can be used to extract a relationship between θ and the particular parameters of the disc $B(w, r)$. That is,

$$\sin\theta = \frac{\sinh r}{\sinh d_{\mathbb{D}}(0, w)}.$$

We can use the known[4] disc formulas

$$\sinh^2\left(\frac{1}{2} d_{\mathbb{D}}(z, w)\right) = \frac{|z - w|^2}{(1 - |z|^2)(1 - |w|^2)},$$

[2]The reason for this behavior is explained in the beginning of Section 2.2 in Falconer's book ([22], p. 31).

[3]See §7.12 in Beardon's book ([6], p. 148).

[4]See Theorem 7.2.1. in Beardon's book ([6], p. 130).

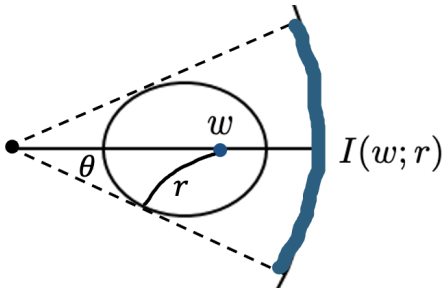


Figure 6.1: Illustration of shadow of $B(w,r)$ on $\partial\mathbb{D}$.

$$\cosh^2\left(\frac{1}{2}d_{\mathbb{D}}(z,w)\right) = \frac{|1-z\overline{w}|^2}{(1-|z|^2)\,(1-|w|^2)},$$

and the relation $\sinh 2x = 2\sinh x \cosh x$ to rewrite $\sinh d_{\mathbb{D}}(0,w)$ above and obtain

$$\sin\theta = \frac{1-|w|^2}{2|w|}\sinh r.$$

Thus, with r fixed and $|w|$ bounded away from zero, we can use the bounds $x < \arcsin x < \frac{1}{\sqrt{1-x^2}}$ to find the estimate

$$|I(w;r)| \asymp 1-|w|.^5$$

Bear in mind that the symbols $|\cdot|$ have different meanings in the above equation, although both denote Euclidean measures. On the left-hand side it means arc length distance, whereas on the right-hand side it denotes Euclidean distance.

6.2 Sullivan's Shadow Lemma

We can now state and prove Sullivan's Shadow Lemma, explicitly expressing the connection between Patterson-Sullivan and Hausdorff measures.

Lemma 6.2.1 (Sullivan's Shadow Lemma)**.** *Let μ be a Patterson-Sullivan measure of dimension δ and Γ a non-elementary geometrically finite Fuchsian group acting on $\mathbb{D}$. For fixed r sufficiently large and all but finitely many $T \in \Gamma$, we have*

$$\mu(I(T0;r)) \asymp |I(T0;r)|^{\delta},$$

with constants independent of T.

Proof. We divide the proof into four steps.

Step 1) Setting up the strategy.

Let $w = T0$ for some $T \in \Gamma$. Assuming that $d_{\mathbb{D}}(w,0) > r$, we apply T^{-1} to the shadow $I(w;r)$. As is known, T^{-1} maintains the boundary, so $T^{-1}I(w;r) \subset \partial\mathbb{D}$.

[5]The symbol $\asymp$ here denotes comparability. Specifically, when $x \asymp y$, this means that there exist constants c_1, c_2 such that $c_1 y \le x \le c_2 y$.

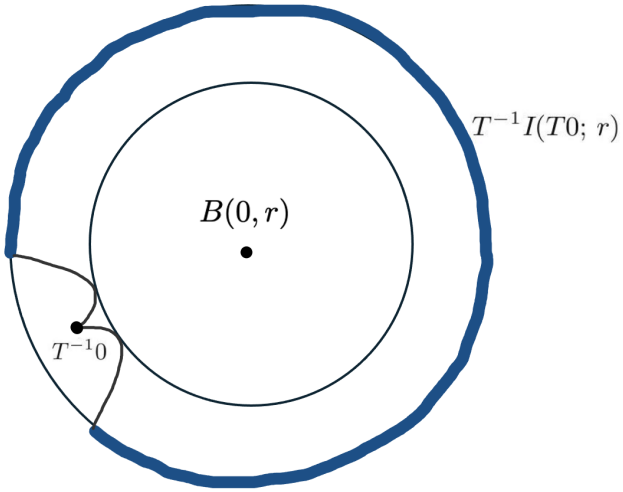


Figure 6.2: Illustration of a pullback of a shadow.

We can then consider it as the shadow of a hyperbolic disc in $\mathbb{D}$. In fact, as illustrated in Figure 6.2, $T^{-1}I(w;r)$ is the shadow cast by $B(0,r)$ using a light source at $T^{-1}0$, since we can use the fact that the elements of Γ are isometries to find

$$\begin{aligned} I(0;r) &\equiv \{q \in \partial\mathbb{D} : d_{\mathbb{D}}([T^{-1}0, q], 0) < r\} \\ &= \{q \in \partial\mathbb{D} : d_{\mathbb{D}}([T^{-1}0, q], T^{-1}w) < r\} \\ &= \{q \in \partial\mathbb{D} : d_{\mathbb{D}}([0, Tq], w) < r\} \\ &= \{T^{-1}q' \in \partial\mathbb{D} : d_{\mathbb{D}}([0, q'], w) < r\} \\ &= T^{-1}I(w;r). \end{aligned}$$

In the next steps, we will show that $\mu\left(T^{-1}I(w;r)\right)$ can be appropriately bounded, and this in turn will imply the estimate by the distortion rule $T^*\mu = |T'|^\delta \mu$. We emphasize here that it must be assumed that $T^{-1}0 \notin B(0,r)$ for $I(w;r)$ to be well defined in the first place, since (per definition) the light source cannot be located inside the disc that originates the shadow. This is the reason why the present lemma holds for all but possibly finitely many elements of Γ (i.e., there cannot be an infinite number of elements $T \in \Gamma$ fulfilling $T^{-1}0 \in B(0,r)$, because this would contradict the discreteness of Γ).

Before we address the bounds, an auxiliary result is proved.

Step 2) Proof of the claim: $\exists\, \varepsilon > 0, \forall I \subset \partial\mathbb{D}, |I| < \varepsilon$ (in arc length) $\Rightarrow \mu(I) < \frac{1}{2}$.

Toward a contradiction, suppose that this does not hold. In other words, suppose that there exists a sequence of closed $I_j \subset \partial\mathbb{D}$ with $|I_j| \to 0$ but $\mu(I_j) \geq \frac{1}{2}$. Since for probability/finite measures it must be true that $\mu\left(\bigcup I_j\right) \leq 1$, infinitely many of the intervals I_j must overlap; otherwise, we would be able to use countable additivity to make the measure of the union diverge. Thus, passing to a subsequence, we can assume that $\bigcap I_j \neq \emptyset$. Because the arc lengths of

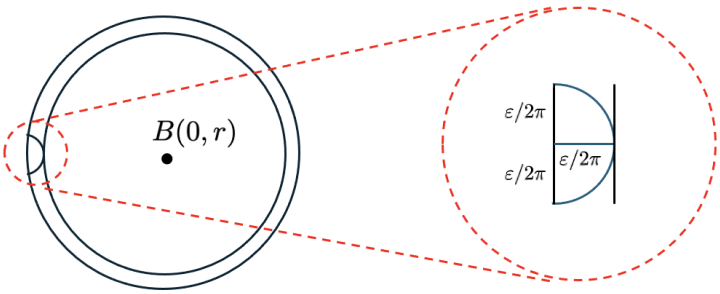


Figure 6.3: Approximation used in the argument of the upper bound.

the elements of the cover tend to zero, $\bigcap I_j$ consists of a single point. However, by continuity from above, we have $\mu(\bigcap I_j) = \lim_{j\to\infty} \mu(I_j) \geq \frac{1}{2}$, and this contradicts the non-atomicity of the Patterson-Sullivan measures (when Γ is non-elementary and geometrically finite), Theorem 4.8.5. Therefore, the claim holds.

Step 3) Upper bound.

Fix ε and choose r sufficiently large so that not only $B(0,r)$ is within Euclidean distance $\varepsilon/2\pi$ of $\partial\mathbb{D}$, but also that the inequality $\left|\left(T^{-1}I(w;r)\right)^{\mathrm{c}}\right| \approx 2\frac{\varepsilon}{2\pi} < \varepsilon$ holds in arc length (the complement is taken with respect to $\partial\mathbb{D}$). The approximation is illustrated in Figure 6.3. Using the claim proved above, we can obtain (by complementarity of the probability measure) the following bounds

$$\begin{aligned}\mu\left(\left(T^{-1}I(w;r)\right)^{\mathrm{c}}\right) < \frac{1}{2} &\Rightarrow 1-\mu\left(T^{-1}I(w;r)\right) < \frac{1}{2}\\ &\Rightarrow \frac{1}{2} < \mu\left(T^{-1}I(w;r)\right) \leq 1,\end{aligned}$$

an inequality that holds for any $T \in \Gamma$ such that $w = T0$ lies outside of $B(0,r)$ (which excludes at most finitely many T, again by discreteness). The upper bound is a consequence of the fact that the measure is a finite/probability measure.

By the distortion formula (Lemma 4.6.1), we have

$$(*)\quad \mu\left(T^{-1}I(w;r)\right) = \left(T^{-1}\right)^{*}\mu\left(I(w;r)\right) = \int_{I(w;r)} \left|(T^{-1})'\right|^{\delta} d\mu.$$

To control the size of $\left|T^{-1\prime}\right|$, we write

$$\left|(T^{-1})'(q)\right| = \frac{P(T0,q)}{P(0,T^{-1}q)} = \frac{P(w,q)}{P(0,T^{-1}q)} \overset{T^{-1}q\in\partial\mathbb{D}}{=} P(w,q) = \frac{1-|w|^2}{|w-q|^2}.$$

Since $|w-q| \geq 1-|w|$, we have $\frac{1}{|w-q|} \leq \frac{1}{1-|w|}$. Hence, $\frac{1}{|w-q|^2} \leq \frac{1}{(1-|w|)^2}$ (here $|\cdot|$ denotes the Euclidean distance). We can then find an upper bound

$$\left|(T^{-1})'(q)\right| = \frac{1-|w|^2}{|w-q|^2} \leq \frac{1-|w|^2}{(1-|w|)^2} = \frac{1+|w|}{1-|w|} < \frac{2}{1-|w|}.$$

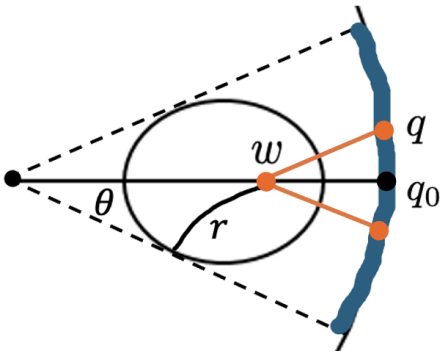


Figure 6.4: Illustration of the lower bound argument.

Step 4) Lower bound.

For the lower bound, suppose that $q \in I(w;r)$, that q_0 is the midpoint of the shadow, and let $\theta = |I(w;r)|/2$ be the angle of the vertex as in Figure 6.4.

Using the Euclidean triangle inequality, since $|q_0 - w| = 1 - |w|$ and $|q - q_0| \leq \widehat{q_0 q} \leq \frac{|I(w;r)|}{2} = \theta$,[6] we have

$$|w - q| \leq |q_0 - w| + |q - q_0| \leq 1 - |w| + \theta.$$

As shown in Section 6.1, $2\theta = |I(w;r)| \asymp (1 - |w|)$, for r held constant and $|w|$ bounded away from zero, so we can write $\theta \leq C'(1 - |w|)$, for some constant C'. Then, using this in the inequality above, for any $q \in I(w;r)$ we can get the bound

$$|w - q| \leq C(1 - |w|),$$

with C independent of T. This implies $\frac{1}{|w-q|^2} \geq \frac{1}{C^2(1-|w|)^2}$. Hence,

$$\frac{1 - |w|^2}{|w - q|^2} \geq \frac{1 - |w|^2}{C^2(1 - |w|)^2} = \frac{1 + |w|}{C^2(1 - |w|)} > \frac{1}{C^2(1 - |w|)} = \frac{c}{1 - |w|},$$

for some constant c independent of T. In view of $\left|T^{-1\prime}(q)\right| = \frac{1-|w|^2}{|w-q|^2}$, we can substitute the previous results and get uniform bounds

$$\frac{c}{1 - |w|} < \left|(T^{-1})'(q)\right| < \frac{2}{1 - |w|}.$$

Now we raise all terms to the power of $\delta > 0$ and integrate with respect to the Patterson-Sullivan measure μ on $I(w;r)$. By $(*)$ above, we get

$$\left(\frac{c}{1 - |w|}\right)^{\delta} \mu(I(w;r)) < \mu(T^{-1}I(w;r)) < \left(\frac{2}{1 - |w|}\right)^{\delta} \mu(I(w;r)).$$

[6] We have denoted the arc length between q_0 and q along the unit circle using $\widehat{q_0 q}$ in order to differentiate it from the Euclidean distance between both points, $|q - q_0|$.

From these inequalities (and finiteness of the measure) we find the desired bounds, obtaining

$$\mu(I(w;r)) \asymp (1-|w|)^{\delta} \asymp |I(w;r)|^{\delta},$$

where the last comparison on the right-hand side follows from the last part of Section 6.1. □

6.3 Hausdorff dimension and convex cocompact groups

To use Sullivan's Shadow Lemma (Lemma 6.2.1) to estimate Hausdorff measures, one must approximate small intervals in $\partial\mathbb{D}$ by shadows of the form $I(T0;r)$. So far, the assumptions on the Fuchsian group Γ have been mild (i.e., only that it ought to be non-elementary and geometrically finite). However, for the next result, a restriction is imposed on Γ: the assumption that Γ is convex cocompact makes for a substantial simplification. Lemma 2.13.3 is the reason why neighborhoods of parabolic fixed points are not susceptible to approximations by shadows of the type proposed.

Lemma 6.3.1. *Assume that Γ is a convex cocompact Fuchsian group. For $q \in \Lambda(\Gamma)$, let I_q denote an interval in $\partial\mathbb{D}$ centered on q. There exists $\varepsilon > 0$, such that for any $|I_q| < \varepsilon$,*

$$\mu\left(I_q\right) \asymp |I_q|^{\delta},$$

uniformly in q.

Proof. Recall that a convex cocompact Fuchsian group by definition has a convex core that is compact. Thus, $N = \Gamma\backslash\widetilde{N}$ is compact, where $\widetilde{N}$ denotes the Nielsen region. Since we intend to apply Sullivan's Shadow Lemma, we assume Γ is geometrically finite.[7] Let $\mathcal{F}$ be a finite-sided fundamental region for Γ. Moreover, use conjugation (if needed) to assume $0 \in \mathcal{F}_c \equiv \mathcal{F} \cap \widetilde{N}$ and denote $d_0 = \operatorname{diam} \mathcal{F}_c$ (in the hyperbolic sense).

For an arbitrary I_q and $r > 0$, Figure 6.5 illustrates that one can choose a unique point $w \in \widetilde{N}$ along the (geodesic) segment joining 0 and q, so that the set I_q is expressed as a shadow $I(w;r)$. Observe that $w \in \widetilde{N}$ because $0 \in \widetilde{N}$ and $q \in \Lambda(\Gamma)$, so the geodesic segment $[0,q]$ would only escape $\widetilde{N}$ if q was not a limit point of Γ. Moreover, the choice of positive r is free, because we can always find a hyperbolic disc closer to the boundary, where the distances increase arbitrarily.

Now, since fundamental domains tessellate $\mathbb{D}$ (see Section 2.9), it is clear that $w \in T\mathcal{F}$ for some $T \in \Gamma$. Furthermore, since according to Lemma 2.10.2 $\Lambda(\Gamma)$ is Γ-invariant, and since $\widetilde{N}$ is defined as the convex hull of the union of the geodesics whose defining endpoints lie in the limit set, $\widetilde{N}$ is also Γ-invariant. Hence, $w \in T\widetilde{N}$, and so $w \in T\mathcal{F} \cap T\widetilde{N} = T\left(\mathcal{F} \cap \widetilde{N}\right) = T\mathcal{F}_c$, for some $T \in \Gamma$.

[7] One could also observe that convex cocompactness implies geometric finiteness. See §15.1 and §15.3 in Fenchel and Nielsen's book ([23], p. 115 and p. 119).

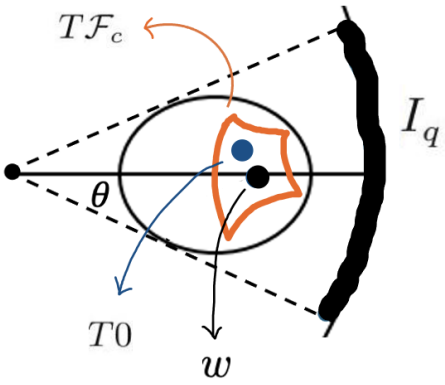


Figure 6.5: Illustration of approximation by shadows.

In this case, we have $d_{\mathbb{D}}(T0, w) = d_{\mathbb{D}}(0, T^{-1}w) < d_0$, and by the triangle inequality,

$$\begin{aligned} d_{\mathbb{D}}(w, z_1) &\leq d_{\mathbb{D}}(w, T0) + d_{\mathbb{D}}(T0, z_1) < d_0 + (r - d_0) = r \\ d_{\mathbb{D}}(T0, z_2) &\leq d_{\mathbb{D}}(T0, w) + d_{\mathbb{D}}(w, z_2) < d_0 + r, \end{aligned}$$

which implies

$$B(T0, r - d_0) \subset B(w, r) \subset B(T0, r + d_0),$$

if we assume $r > d_0$ (recall that there is freedom to choose positive r).

Now, the shadows of the above discs fulfill

$$I(T0; r - d_0) \subset I_q = I(w; r) \subset I(T0; r + d_0).$$

By monotonicity,

$$\mu(I(T0; r - d_0)) \leq \mu(I_q) \leq \mu(I(T0; r + d_0)),$$

and Sullivan's Shadow Lemma can be applied to the bounds of $\mu(I_q)$. In particular, we can choose $\varepsilon > 0$ sufficiently small, so that $|I_q| < |I(T0; r+d_0)| < \varepsilon$ verifies Step 2) in the proof of Lemma 6.2.1. Then, through $\mu(I(T0; r - d_0)) \asymp |I(T0; r - d_0)|^\delta$ and $\mu(I(T0; r + d_0)) \asymp |I(T0; r + d_0)|^\delta$, we get the estimates

$$c_1 |I(T0; r - d_0)|^\delta \leq \mu\left(I_q\right) \leq c_2 |I(T0; r + d_0)|^\delta,$$

for some constants c_1, c_2.

Furthermore, because $|T0|$ is bounded away from zero, we can use the relation $|I(z; r)| \asymp 1 - |z|$ from Section 6.1 and deduce

$$c_1'(1 - |T0|)^\delta \leq c_1 |I(T0; r - d_0)|^\delta \leq \mu\left(I_q\right) \leq c_2 |I(T0; r + d_0)|^\delta \leq c_2'(1 - |T0|)^\delta$$

and

$$c_3(1 - |T0|) \leq |I(T0; r - d_0)| \leq |I_q| \leq |I(T0; r + d_0)| \leq c_4(1 - |T0|),$$

where c_1', c_2', c_3, and c_4 are all constants.

Now, raising the last inequalities to the power of δ and multiplying by the appropriate constants, the result $\mu(I_q) \asymp |I_q|^\delta$ is proved. □

With the relationship above, the next lemma is almost an immediate consequence.

Lemma 6.3.2. *For Γ a convex cocompact Fuchsian group, there exists a constant $b > 0$ such that if A is a Borel subset of the limit set of Γ, then*

$$b\mu(A) \leq H^{\delta}(A).$$

Proof. Consider a countable cover $\bigcup_{j=1}^{\infty} I_j$ of A consisting of intervals I_j centered at some limit point in $\Lambda(\Gamma)$. Then, by countable subadditivity, we have

$$\mu(A) \leq \sum_{j=1}^{\infty} \mu(I_j).$$

Applying Lemma 6.3.1, swap the bound on the right by

$$\mu(A) \leq C \sum_{j=1}^{\infty} |I_j|^{\delta},$$

where $C > 0$ is a constant. This holds, provided that all intervals have an arc length $\epsilon < \epsilon_0$, for some fixed ϵ_0. Since what matters is the behavior in the limit of covers composed of shadows of decreasing arc length, we can assume that this requirement is true.

Thus, the inequality holds for any such cover; so it holds also for the infimum (with respect to the sums $\sum_{j=1}^{\infty} |I_j|^{\delta}$) of all such covers. Then, taking the limit of intervals of decreasing arc length, the result is verified. □

Now, using Lemma 6.3.2 for $A = \Lambda(\Gamma)$, the inequalities

$$0 < b\mu(\Lambda(\Gamma)) \leq H^{\delta}(\Lambda(\Gamma))$$

demonstrate that the Hausdorff dimension of the limit set is at least δ.

Finding the other bound is not as straightforward, but it can be done. Recall the notion of a conical or radial limit point presented in Section 4.8. Denote the set of radial or conical limit points in $\Lambda(\Gamma)$ by $C(\Gamma) \subseteq \Lambda(\Gamma)$, or simply C if Γ is clear from context.

Lemma 6.3.3. *Let Γ be a Fuchsian group acting on $\mathbb{D}$ and assume that μ is an α-dimensional Γ-invariant conformal density. There exists a constant b such that if $A \subseteq C(\Gamma)$ is Borel with $\mu_x(A) > 0$, then $H^{\alpha}(A) \leq b\mu_x(A)$.*

Proof. By a variant of Lebesgue's Density Theorem,[8] $\mu_x(A) > 0$ implies that almost every point (with respect to μ_x, for some $x \in \mathbb{D}$) of A is a density point of the measure. That is, for almost every $a \in A$, the limit

$$\lim_{t \to 0} \frac{\mu_x\left[\tilde{I}(a,t) \cap A\right]}{\mu_x[\tilde{I}(a,t)]} = 1$$

holds, where $\tilde{I}(a,t)$ denotes an open interval centered at a in $\partial\mathbb{D}$ of arc length t with respect to both sides of its center.

[8] See Theorem 3.22 in Folland's book ([24], p. 99).

Thus, given $\delta > 0$, there is a subset A' of A with $\mu_x(A - A') < \delta$ and an instant $t_0 > 0$ such that for all $a \in A'$ and $t < t_0$

$$\frac{\mu_x\left[\tilde{I}(a,t) \cap A\right]}{\mu_x[\tilde{I}(a,t)]} \overset{(*)}{\geq} 1 - \delta.$$

Since every conical limit point can be approximated by shadows (i.e., one can find a sequence of discs in the cone of a radial limit point, as there are infinite Γ-images to choose from, and consider their shadows), using a light source at the origin it is possible to cover each radial limit point in A' by a shadow of the form $I(\gamma_i(0), c)$,[9] with $i = 1, 2, \ldots$ and $\gamma_i \in \Gamma$, for some fixed positive radius c, such that their arc length distances respect

$$|I(\gamma_i(0); c)| > |I(\gamma_{i+1}(0); c)|,$$

with $|I(\gamma_1(0); c)| < \epsilon < t_0$, and such that for each i, the center of the shadow $I(\gamma_{i+1}(0); c)$ is outside the union

$$\bigcup_{k=1}^{i} I(\gamma_k(0); c).$$

Moreover, under the aforementioned conditions, by considering shadows of discs having the same centers but with radii kc, where k is some appropriate constant, a disjoint collection of shadows can be obtained. Thus, without loss of generality, assume that the collection Ω of shadows of the type described above is disjoint. By Lemma 6.2.1, we have

$$\sum_i |I(\gamma_i(0); c)|^\alpha \leq b \sum_i \mu_x(I(\gamma_i(0); c)) = b\mu_x(\Omega),$$

for some constant b (recall that by the mentioned lemma, the constant is the same irrespective of the γ_i considered).

On the other hand, by the definition of Hausdorff measure,

$$H^\alpha_\epsilon(A') \leq \sum_i |I(\gamma_i(0); c)|^\alpha$$

and so $H^\alpha_\epsilon(A') \leq b\mu_x(\Omega)$.

[9] This uses the fact that all limit points of Γ can be seen as limit points of a single Γ-orbit, in this case the Γ-orbit of zero.

Using $(*)$ obtained above (since $\epsilon < t_0$), we have

$$\begin{aligned}
\mu_x(\Omega) = \mu_x\left(\bigcup_{I(\gamma_i(0);c)\in\Omega} I(\gamma_i(0);c)\right) &= \sum_{I(\gamma_i(0);c)\in\Omega} \mu_x\left[I(\gamma_i(0);c)\right] \\
&\overset{(*)}{\leq} \sum_{I(\gamma_i(0);c)\in\Omega} \frac{\mu_x\left[I(\gamma_i(0);c)\cap A'\right]}{1-\delta} \\
&\leq \frac{\mu_x\left[\bigcup_{I(\gamma_i(0);c))\in\Omega} I(\gamma_i(0);c)\cap A'\right]}{1-\delta} \\
&\leq \frac{\mu_x\left(A'\right)}{1-\delta} \\
&= \frac{\mu_x(A)-\mu_x(A-A')}{1-\delta} \\
&\leq \frac{\mu_x(A)-\delta}{1-\delta}.
\end{aligned}$$

Therefore,

$$H^\alpha_\epsilon\left(A'\right) \leq b\frac{\mu_x(A)-\delta}{1-\delta}.$$

Taking the limit $\epsilon \to 0$ and letting $\delta \to 0$, we find in sequence

$$H^\alpha\left(A'\right) \leq b\frac{\mu_x(A)-\delta}{1-\delta}$$

and

$$H^\alpha\left(A\right) \leq b\mu_x(A).$$

□

Corollary 6.3.4. *Given a Γ-invariant conformal density of dimension α, if $\dim(C(\Gamma))$ is the Hausdorff dimension of the conical limit set, then*

$$\dim_H(C(\Gamma)) \leq \alpha.$$

In particular, $\dim_H(C(\Gamma)) \leq \delta(\Gamma)$ in the case of a Fuchsian group Γ with a critical exponent $\delta(\Gamma)$.

Proof. We follow the proof of Lemma 6.3.3 using $A = C(\Gamma)$ as before until $H^\alpha_\epsilon(C(\Gamma)) \leq b\mu_x(\Omega)$ is reached. Then, by monotonicity, the inequality implies that $H^\alpha_\epsilon(\Omega) \leq b\mu_x(\partial\mathbb{D}) \overset{\epsilon\to 0}{\longrightarrow} H^\alpha(\Omega) \leq b\mu_x(\partial\mathbb{D})$, and so $\dim_H(C(\Gamma)) \leq \alpha$. In case Γ is a Fuchsian group having a critical exponent $\delta(\Gamma)$, a Γ-invariant conformal density exists and its dimension is the critical exponent itself (see Section 4.9). □

Proposition 4.8.4 showed that a limit point in $\Lambda(\Gamma)$ for a geometrically finite Fuchsian group Γ can only be parabolic or conical/radial. In the convex cocompact setting, no parabolic limit points are allowed. Thus, if $C(\Gamma) \subseteq \Lambda(\Gamma)$ is the conical limit set, then $C(\Gamma) = \Lambda(\Gamma)$. Therefore, $\dim_H(\Lambda(\Gamma)) \leq \alpha$. Combining this upper bound on the Hausdorff dimension with the lower bound previously obtained, we have the following theorem.

Theorem 6.3.5. *For a geometrically finite Fuchsian group* Γ *with critical exponent* $\delta(\Gamma)$ *and limit set* $\Lambda(\Gamma)$*,*

$$\dim_H(\Lambda(\Gamma)) = \delta(\Gamma).$$

Under the same assumptions, it is possible to go even further and demonstrate that the Patterson-Sullivan measure μ is a constant multiple of the Hausdorff measure H^δ restricted to the limit set $\Lambda(\Gamma)$.[10] The only difficulty in the proof is related to proving that the action of Γ is ergodic with respect to the Patterson-Sullivan measure μ. However, as is now known, the ergodicity of the group action is deducible from the ergodicity of the geodesic flow (Corollary 5.4.3), so even this deeper conclusion is within reach in addition to being linked to ergodic notions.

[10]See, for instance, Theorem 4.6.4 in Nicholls' book ([53], p. 84).

Chapter 7

The Prime Geodesic Theorem

In this chapter, we revisit the problem of counting the lengths of closed prime geodesics on a geometrically finite hyperbolic surface $X \cong \Gamma\backslash\mathbb{H}$. In the first section, we obtain a similar bound to that of Proposition 3.2.3, but this time for the lattice-point counting function and with an explicit dependence on the critical exponent $\delta(\Gamma)$ associated with the Fuchsian group Γ. This result illustrates the relationship between conformal densities and the problem of counting the number of Γ-images of a point that can be found on a given hyperbolic disc. The later sections showcase deeper relationships between Patterson-Sullivan measures and hyperbolic geometry, establishing the PGT presented in the introduction using Newman's Tauberian argument and the Selberg zeta function. Although the asymptotic behavior of π_X is seen to depend on $\delta(\Gamma)$, which was observed to be a key parameter when studying Patterson-Sullivan measures, the proof of the PGT is at least superficially quite independent of almost everything presented in the last chapters. In fact, most of the interconnections arise in the background, in the context of spectral theory, when studying the poles and zeros of the Selberg zeta function $Z_X(s)$. To give an idea of how important the Patterson-Sullivan theory developed in the last chapters is to this matter, we sketched an intuitive proof of why no resonances occur along the critical line, albeit this is done for the (absolute) Poincaré series associated with a non-elementary geometrically finite Fuchsian group. The canonical presentation of this and related statements that refer to the adopted formulation of the PGT deserves its own comprehensive treatment.[1] Thus, the first section of the chapter is mostly self-contained, while the subsequent sections avoid some of the technicalities of the subject.

7.1 A second bound on the lattice-point counting function

In Section 4.1, we saw that the bound obtained in Proposition 3.2.3 for the length counting function π_X has a counterpart affecting the lattice-point counting function of X, for $X \cong \Gamma\backslash\mathbb{H}$ a geometrically finite hyperbolic surface. Recall that the lattice-point counting function of X is defined as

$$\pi_X(t, x, y) \equiv \#\{T \in \Gamma : d_{\mathbb{D}}(x, Ty) < t\},$$

[1] We relied on Borthwick [9], which is an obvious source on the topic.

for $x, y \in \mathbb{D}$ and $t > 0$.

Now we show that it is possible to find a bound on the lattice-point counting function that depends explicitly on the dimension of a conformal density associated with the surface. In fact, most of the work has already been done, since the next theorem is basically an application of Sullivan's Shadow Lemma (Lemma 6.2.1). Recall that this lemma holds in a general setting. Only when it was used to investigate the relationship between Hausdorff measures and Patterson-Sullivan measures did we assume convex cocompactness.

Theorem 7.1.1. *Let Γ be a Fuchsian group acting on $\mathbb{D}$ and μ a Γ-invariant α-dimensional conformal density. Then, for $x, y \in \mathbb{D}$, there exists a constant C that depends on Γ, on x, and on y such that for $t > t_0$,*

$$\pi_X(t, x, y) < Ce^{t\alpha}.$$

Proof. Abusing notation, we consider a measure $\mu \equiv \mu_0$ arising from a conformal density μ.

Take $x = y = 0$. For a positive integer k and a light source fixed at the origin, consider the collection of shadows $I(\gamma(0); c)$ for which $k - \frac{1}{2} \leq d_{\mathbb{D}}(0, \gamma(0)) < k + \frac{1}{2}$.

Notice that a point $\xi \in \partial\mathbb{D}$ is an element of the shadow $I(\gamma(0); c)$ if $\gamma(0)$ is within hyperbolic distance c of the geodesic segment $\{\lambda\xi : d_{\mathbb{D}}(0, \lambda\xi) \in [k - \frac{1}{2}, k + \frac{1}{2})\}$ contained in a geodesic that passes through the origin. This is a consequence of adopting zero as the light source that generates such shadows. Then, for $\gamma_i(0) \neq \gamma_j(0)$ (in the annular region identified) whose closest points along the geodesic segment described are z_i, z_j, respectively, the remark made above and the triangle inequality imply

$$d_{\mathbb{D}}(\gamma_i(0), \gamma_j(0)) \leq d_{\mathbb{D}}(\gamma_i(0), z_i) + d_{\mathbb{D}}(\gamma_j(0), z_j) + d_{\mathbb{D}}(z_i, z_j) < c + c + 1.$$

The inequality $d_{\mathbb{D}}(z_i, z_j) < 1$ follows from the size of the annular region (with respect to zero).

Thus, all centers $\gamma(0)$ whose discs $B(\gamma(0), c)$ possibly touch geodesic segments terminating in ξ of the form described are contained in a hyperbolic disc of radius $1 + 2c$. Due to the discreteness of Γ, any such disc contains at most M Γ-images of zero, where M depends only on Γ and c. Thus, the collection $I(\gamma(0); c)$ with $k - \frac{1}{2} \leq d_{\mathbb{D}}(0, \gamma(0)) < k + \frac{1}{2}$ covers any point $\xi \in \partial\mathbb{D}$ with multiplicity bounded by M. In other words, by monotonicity,

$$\sum_{\substack{\gamma \in \Gamma \\ k - \frac{1}{2} \leq d_{\mathbb{D}}(0, \gamma(0)) < k + \frac{1}{2}}} \mu[I(\gamma(0); c)] \leq M\mu(\partial\mathbb{D}).$$

Denote

$$n(k) \equiv \#\{\gamma \in \Gamma : k - \frac{1}{2} \leq d_{\mathbb{D}}(0, \gamma(0)) < k + \frac{1}{2}\}.$$

Using Sullivan's Shadow Lemma (Lemma 6.2.1) and the relationship found in Section 6.1, we get

$$n(k)b'(1 - |\gamma(0)|)^{\alpha} \leq n(k)b|I(\gamma(0); c)|^{\alpha} \leq M\mu(\partial\mathbb{D}),$$

for b, b' constants. Recall that this lemma holds for all but finitely many transformations in Γ. Thus, all relations deduced in what follows hold with respect to $k > k_0$.

But for $\gamma \in \Gamma$ such that $k - \frac{1}{2} \leq d_{\mathbb{D}}(0, \gamma(0)) < k + \frac{1}{2}$, multiplying this formula by -1 and obtaining the corresponding inequalities for the exponential function, we can use the hyperbolic formula for the natural logarithm and $d_{\mathbb{D}}(0, \gamma(0))$ (see Lemma 4.3.2) to find an absolute constant A such that $(1 - |\gamma(0)|)^{\alpha}$ is greater than $Ae^{-k\alpha}$. So,

$$n(k)Ae^{-k\alpha}b' \leq M\mu(\partial\mathbb{D}) < \infty$$

Grouping the constants into a new constant C, we have

$$n(k) \leq Ce^{k\alpha},$$

where C depends on Γ.

Now, by the definition of $\pi_X(t) \equiv \pi_X(t, 0, 0)$ and the inequality above, we can write

$$\begin{aligned}
\pi_X(t) \leq & n(1) + n(2) + \cdots + n(\lceil t \rceil) \\
\leq & C\left(1 + e^{\alpha} + e^{2\alpha} + \cdots + e^{\lceil t \rceil \alpha}\right) \\
= & C\frac{(e^{\alpha})^{\lceil t \rceil + 1} - 1}{e^{\alpha} - 1} \\
\leq & C\frac{e^{(t+1)\alpha}}{e^{\alpha} - 1} \\
\leq & C'e^{t\alpha},
\end{aligned}$$

where the fixed terms were subsumed into a new constant C'. We still need to address the cases $x \neq 0$ or $y \neq 0$, but this is simple: use conjugation to send x and y to zero. This will affect the constants, but otherwise the bound remains the same. □

Since a conformal density is known to exist in relation to a Fuchsian group Γ with critical exponent $\delta(\Gamma)$, and since this conformal density has dimension $\delta(\Gamma)$ (see Section 4.9), the above result implies that

$$\pi_X(t, x, y) \leq Ce^{t\delta},$$

for $t > t_0$ and a constant C that depends on Γ, x, and y.

In the convex cocompact scenario, not only a lower bound on $\pi_X(t, x, y)$ is feasible, but even the nature of the constants involved changes. That is, one can show that the constants in this case depend only on Γ, and not on x or y.[2]

7.2 Selberg zeta functions

Just as Newman's proof of the PNT relies on the Riemann zeta function, the forthcoming proof of the PGT also depends on a zeta function, known as Selberg's zeta function. The definition of this construct is as follows.

Definition 7.2.1. For a hyperbolic surface X associated with a Fuchsian group Γ of critical exponent δ, the Selberg zeta function is defined for $\operatorname{Re} s > \delta$ by the product

$$Z_X(s) \equiv \prod_{k=0}^{\infty} \prod_{\ell \in \mathcal{L}_X} \left(1 - e^{-(s+k)\ell}\right),$$

[2]See Theorem 4.6.5 and Lemma 4.6.6 in Nicholls' book ([53], pp. 85-86).

where $\mathcal{L}_X$ denotes the (primitive) length spectrum of X.

The range of convergence of the Selberg zeta function can be deduced from

$$\sum_{\ell \in \mathcal{L}_X} e^{-s\ell} < \infty, \quad \text{for } \operatorname{Re} s > \delta.^{3}$$

Observe that

$$\begin{aligned} Z_X(s) &\equiv \prod_{k=0}^{\infty} \prod_{\ell \in \mathcal{L}_X} \left(1 - e^{-(s+k)\ell}\right) < \infty \text{ absolutely} \\ &\Leftrightarrow \sum_{k=0}^{\infty} \ln \left(\prod_{\ell \in \mathcal{L}_X} \left(1 - e^{-(s+k)\ell}\right) \right) < \infty \text{ absolutely.} \end{aligned}$$

Each length in $\mathcal{L}_X$ is associated with a primitive oriented closed geodesic on X having a bounded multiplicity (at most two, one count for each orientation). Since these geodesics can be understood as axes of particular transformations in the Fuchsian group associated with the hyperbolic surface, a countable set, then $\mathcal{L}_X$ is also countable. This makes intuitive sense because infinite products are defined in terms of limits of partial products. As a consequence, we can use the continuity of the natural logarithm in order to obtain

$$Z_X(s) < \infty \text{ absolutely} \overset{(*)}{\iff} \sum_{k=0}^{\infty} \sum_{\ell \in \mathcal{L}_X} \ln \left(1 - e^{-(s+k)\ell}\right) < \infty \text{ absolutely.}$$

The convergence can then be proved as follows. Using the Mean Value Theorem for $f(t) \equiv \ln(1-t)$, there exists a number $c \in (0, x)$ for $0 < x < 1$ such that

$$\ln(1-x) = f(x) - f(0) = f'(c)(x-0) = \left(\frac{-1}{1-c}\right) x.$$

Taking the absolute value , we find

$$|\ln(1-x)| = \frac{|x|}{1-c}.$$

We can then write

$$\sum_{k=0}^{\infty} \sum_{\ell \in \mathcal{L}_X} \left| \ln \left(1 - e^{-(s+k)\ell}\right) \right| = \sum_{k=0}^{\infty} \sum_{\ell \in \mathcal{L}_X} \frac{e^{-(s+k)\ell}}{1-c}.$$

Now, consider $\ell_0 \equiv \inf\{\ell : \ell \in \mathcal{L}_X\}$. Notice that we must have $\ell_0 > 0$, due to

[3] This is in itself a consequence of the fact that the critical exponent does not change if we let $w = z$ in $\sum_{T \in \Gamma} e^{-s d_{\mathbb{D}}(z, Tw)}$ (see Section 4.1) combined with the inequality (I) in the proof of Proposition 3.2.3.

the discreteness of the Fuchsian group that characterizes X. We can then write

$$\begin{aligned}\sum_{k=0}^{\infty}\sum_{\ell\in\mathcal{L}_X}\frac{e^{-(s+k)\ell}}{1-c} &= \frac{1}{1-c}\sum_{k=0}^{\infty}\sum_{\ell\in\mathcal{L}_X} e^{-(s+k)\ell} \\ &= \frac{1}{1-c}\sum_{\ell\in\mathcal{L}_X} e^{-s\ell}\sum_{k=0}^{\infty}\left(e^{-\ell}\right)^k \\ &= \frac{1}{1-c}\sum_{\ell\in\mathcal{L}_X} e^{-s\ell}\frac{1}{1-e^{-\ell}} \\ &\leq \frac{1}{(1-c)\left(1-e^{-\ell_0}\right)}\sum_{\ell\in\mathcal{L}_X} e^{-s\ell} < \infty,\end{aligned}$$

where we have used the range of convergence identified above and the fact that we can swap the order of sums of positive terms.

In other words, one can manipulate $Z_X(s)$ analytically, such as performing differentiation in terms of s or rewriting factors. The proof of the PGT will be seen to depend on such operations.

7.3 The zeros and poles of Selberg zeta functions

Recall the Laplacian operator, which can be globally defined for a Riemannian manifold (and thus for a hyperbolic surface) as the divergence of the gradient.

Technically, the notion of resonance is defined in terms of the poles of the resolvent of the Laplacian operator.[4] These poles can then be shown to determine the zeros and poles of the Selberg zeta function.[5]

However, in this section, we proceed naively and avoid discussing the resolvent of the Laplacian operator and the extensive technicalities that the topic engenders. We therefore define resonances as follows.

Definition 7.3.1. For a geometrically finite hyperbolic surface $X \cong \Gamma\backslash\mathbb{D}$, the poles of the meromorphic extension of the Poincaré series (associated with the Fuchsian group Γ) are called the resonances of X.

We can then assume (also naively) that the poles of the Poincaré series determine the zeros and poles of the Selberg zeta function.[6] The intuitive reason for this fact is that, as seen in Section 7.2, the convergence of the Selberg zeta function associated with a given hyperbolic surface $X \cong \Gamma\backslash\mathbb{D}$ depends on the convergence of the Poincaré series defined in terms of the Fuchsian group Γ.[7]

The main theorem of this section illustrates the importance of the Patterson-Sullivan measures in describing the resonances of the Poincaré series, and consequently, by the above argument, the poles and zeros of the Selberg zeta function.

[4] See Definition 8.1. in Borthwick ([9], p. 143).

[5] In essence, this is a consequence of the next theorem (Theorem 7.3.2) combined with Proposition 14.17 and Proposition 14.18 in Borthwick [9], in addition to the characterization of the Selberg zeta function put forward in Borthwick, Judge, and Perry [10].

[6] Rigorously speaking, this hinges on the characterization of the lift of the kernel of the Laplacian resolvent, as it can be written in terms of the Poincaré series. See Proposition 14.17 in Borthwick ([9], p. 341).

[7] Recall that we can rewrite the Poincaré series, which is a sum of transformations $T \in \Gamma$, as a sum of lengths in the spectrum $\mathcal{L}_X$.

Theorem 7.3.2. *For a non-elementary geometrically finite hyperbolic surface $X \cong \Gamma\backslash\mathbb{H}$, there are no resonances on the line $\operatorname{Re} s = \delta$ with $s \neq \delta$.*

Proof sketch. Recall the notation

$$g_s(z,w) \equiv \sum_{T\in\Gamma} e^{-s d_{\mathbb{D}}(z,Tw)}.$$

By contradiction, suppose that $g_s(z,w)$ has a pole at $\zeta \neq \delta$, with $\operatorname{Re}\zeta = \delta$.

For $\operatorname{Re} s > \delta$, we have $g_s(z,w) \leq g_{\operatorname{Re} s}(z,w)$, with the order of the pole at ζ being limited by the order of the pole at δ (both equal to one).

There are three steps to the proof.

Step 1) Constructing a Patterson-Sullivan analogue using the pole.

The proof consists in constructing a complex measure σ analogous to the Patterson-Sullivan measure, but using the existence of the pole to counterbalance the divergence of the (absolute) Poincaré series at the critical exponent. In detail, for some sequence $s_j \to \zeta$, consider the complex measure

$$\sigma = \lim_{s_j\to\zeta} (s_j - \zeta) \sum_{T\in\Gamma} e^{-s_j d(0,T0)} \delta_{T0}$$

obtained in the weak limit.

It can be shown that σ has some of the same properties of the Patterson-Sullivan measure μ proved in Chapter 4 (e.g., σ is a finite nonzero measure supported on $\Lambda(\Gamma)$). The distortion rule also holds (replacing δ for ζ) and one can use it to construct a Γ-invariant measure $\tilde{\sigma}$ on $(\partial\mathbb{D}\times\partial\mathbb{D}) - \Delta$, as was done previously in Section 5.1 (the non-Lebesgue part $\tilde{\mu}$ of m_μ). In fact, there exists a deeper connection, since the weights in front of the Dirac masses in σ are bounded by those in μ, which implies that σ (respectively, $\tilde{\sigma}$) is absolutely continuous with respect to μ (respectively, $\tilde{\mu}$). Hence, for some function ψ defined on $\Lambda(\Gamma)$,

$$d\sigma = \psi d\mu.$$

Now, as seen in the proof of Corollary 5.4.3, the product action of Γ on $(\partial\mathbb{D}\times\partial\mathbb{D}) - \Delta$ is ergodic with respect to $\tilde{\mu}$, so any Γ-invariant function defined on this product is constant almost everywhere. Hence, $\tilde{\sigma} = c\tilde{\mu}$ for some constant c. Using the definitions of the measures, we get

$$(*) \quad \psi(q)\psi(q') = c\,|q-q'|^{2(\zeta-\delta)},$$

for almost every q, q' in $(\Lambda(\Gamma)\times\Lambda(\Gamma)) - \Delta$.

Step 2) Extending $(*)$ to every point in $(\Lambda(\Gamma)\times\Lambda(\Gamma)) - \Delta$.

We will show that any nonempty open subset of $(\Lambda(\Gamma)\times\Lambda(\Gamma)) - \Delta$ has a positive measure with respect to $\tilde{\mu}$. Notice first that the right-hand side of $(*)$ is continuous except on the diagonal. Due to the equality, by altering ψ (if needed) on a set of measure zero, we can assume that this function is continuous on $\Lambda(\Gamma)$. Second, observe that the set where $(*)$ fails to hold is open, because we may define the continuous function $F(q,q') \equiv \psi(q)\psi(q') - c|q-q'|^{2(\zeta-\delta)}$ and

consider the preimage $F^{-1}(\mathbb{R} - \{0\})$. Thus, we may as well take into account only open subsets of $\Lambda(\Gamma)$. Finally, recall that any point in $\Lambda(\Gamma)$ is a limit point of hyperbolic fixed points (for Γ non-elementary).[8] Thus any nonempty open set $U \subset \Lambda(\Gamma)$ contains a hyperbolic fixed point q. As a consequence, $\cup_n T^n U = \Lambda(\Gamma) - \{q_-\}$, and so, by non-atomicity, $\mu\left(\cup_n T^n U\right) = 1$, implying $\mu(U) > 0$ by Patterson-Sullivan measure distortion. Therefore, the set where $(*)$ is not valid can only be the empty set.

Step 3) Deriving the contradiction.

Take any $p \in \Lambda(\Gamma)$ and a sequence $\{q_j\}$ in $\Lambda(\Gamma)$ converging to p. Now, since $(*)$ holds everywhere on $\Lambda(\Gamma) \times \Lambda(\Gamma)$ (except on the diagonal), and because ψ is continuous, we get

$$c\,|q_j - p|^{2(\zeta-\delta)} = \psi\left(q_j\right)\psi(p) \overset{q_j \to p}{\longrightarrow} \psi(p)^2 = 0.$$

Hence, σ vanishes on $\Lambda(\Gamma)$, which contradicts the existence of a resonance at ζ. □

7.4 The Prime Geodesic Theorem

As disclosed in the Introduction (Chapter 1), the forthcoming proof of the PGT follows that of the PNT constructed by Newman in the 1980s. It relies on two special components. The first is Newman's Tauberian Theorem, which is presented in full below. The second is a fact about the holomorphicity of an appropriate function that is employed in the validation of Newman's Tauberian Theorem hypothesis. As will be seen, this second auxiliary result depends on the poles and zeros of the Selberg zeta function.

With that said, let us deduce the first auxiliary theorem.

Theorem 7.4.1 (Newman's Tauberian Theorem)**.** *Suppose that $f(t)$ is bounded and locally integrable*[9] *on $[0, \infty)$, and assume that*

$$g(z) \equiv \int_0^\infty f(t)e^{-zt}dt$$

exists for $\operatorname{Re} z > 0$. If g extends to a holomorphic function on a neighborhood of $\{\operatorname{Re} z \geq 0\}$, then $\int_0^\infty f(t)dt$ exists and

$$\int_0^\infty f(t)dt = g(0)$$

Proof. For $T > 0$, let us define the function

$$g_T(z) \equiv \int_0^T f(t)e^{-zt}dt.$$

[8] This is checked in the proof of Proposition 2.13 in Borthwick ([9], pp. 17-19).

[9] Locally integrable functions are functions which are integrable on compact subsets of their domain of definition. Although this formulation of the theorem is standard (see Bak and Newman [3] and Zagier [80]), it is somewhat redundant, because a measurable bounded function is necessarily locally integrable. In this sense, essentially, Newman's Tauberian Theorem requires that f be measurable and bounded. However, even the boundedness requirement can be dropped (see the remark made right after the proof of the present theorem).

This function fulfills all conditions for complex differentiation under the integral sign. So, g_T is entire.

Our goal is to show that $\lim_{T\to\infty} g_T(0) = g(0)$.

Using the assumption that $g(z)$ has an analytic extension, for any $R > 0$ we can choose $\varepsilon > 0$ such that g is analytic on $\Omega = \{|z| \le R, \operatorname{Re} z \ge -\varepsilon\}$. Then, by Cauchy's Integral Formula applied to

$$m(z) \equiv [g(z) - g_T(z)]\, e^{zT} \left(1 + \frac{z^2}{R^2}\right),$$

which is analytic on Ω due to g being analytic there, we obtain

$$\begin{aligned} m(0) &= \frac{1}{2\pi i} \int_{\partial\Omega} m(z) \frac{dz}{z} \\ g(0) - g_T(0) &= \frac{1}{2\pi i} \int_{\partial\Omega} [g(z) - g_T(z)]\, e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z}, \end{aligned}$$

with $\partial\Omega$ oriented counterclockwise.

Now, separate the boundary $\partial\Omega$ into components $\eta_+ \equiv \{z \in \partial\Omega : \operatorname{Re} z > 0\}$ and $\eta_- \equiv \{z \in \partial\Omega : \operatorname{Re} z < 0\}$. Then we denote the corresponding pieces of the previous integral by $I_\pm(T)$. Our goal is to estimate these integrals individually. Let $M = \sup |f|$ (i.e., f is assumed to be bounded).

Case A) Estimating $I_+(T)$.

On η_+ we use the fact that f is bounded to estimate

$$\begin{aligned} |g(z) - g_T(z)| &= \left| \int_T^\infty f(t) e^{-zt} dt \right| \\ &\le \int_T^\infty |f(t)|\, |e^{-zt}| dt \\ &\le M \int_T^\infty |e^{-\operatorname{Re}(z)t}| |e^{-i \operatorname{Im}(z)t}| dt = M \frac{e^{-(\operatorname{Re} z)T}}{\operatorname{Re} z}. \end{aligned}$$

Since on the boundary $|z| = R$, we have

$$\begin{aligned} |I_+(T)| &\le \left|\frac{1}{2\pi i}\right| \int_{\eta_+} |g(z) - g_T(z)| \left|e^{\operatorname{Re}(z)t}\right| \left|e^{i \operatorname{Im}(z)t}\right| \left|1 + \frac{z^2}{R^2}\right| \left|\frac{dz}{z}\right| \\ &\le \frac{M}{2\pi} \int_{\eta_+} \frac{1}{\operatorname{Re} z} \left|1 + \frac{z^2}{R^2}\right| \frac{|dz|}{R}. \end{aligned}$$

For z on the boundary, we also have $\operatorname{Re}^2 z + \operatorname{Im}^2 z = R^2$, so we can compute $\left|1 + z^2/R^2\right| = 2 \operatorname{Re} z/R$, and the length of $\eta_+$is half the total length of the circle $|z| = R$, that is, πR. Therefore,

$$|I_+(T)| \le \frac{M}{R}.$$

Case B) Estimating $I_-(T)$.

We subdivide the estimate of I_- according to the contributions of g and g_T.

Case B.1) Contribution of g_T to $I_-(T)$.

Since g_T is entire, $g_T(z)e^{zT}\left(1+\frac{z^2}{R^2}\right)\frac{1}{z}$ is holomorphic anywhere on the complex plane except near the origin. Thus, it is path independent for paths that do not approach the origin. We can then deform η_- into the arc of a circle which is the continuation of η_+, obtaining (since both paths avoid the origin and have the same initial and final points, $\frac{\pi}{2}$ and $\frac{3\pi}{2}$)

$$\int_{\eta_-} g_T(z)e^{zT}\left(1+\frac{z^2}{R^2}\right)\frac{dz}{z} = \int_{\eta_R} g_T(z)e^{zT}\left(1+\frac{z^2}{R^2}\right)\frac{dz}{z},$$

where $\eta_R(\theta) = Re^{i\theta}$ for $\theta \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$.

After obtaining the estimate

$$\begin{aligned}
|g_T(z)| &\leq \int_0^T |f(t)|e^{-\operatorname{Re}(z)t}dt \\
&\leq M\int_0^T e^{-\operatorname{Re}(z)t}dt \\
&= M\frac{e^{-\operatorname{Re}(z)T}-1}{-\operatorname{Re} z} \\
&= M\frac{e^{|\operatorname{Re} z|T}-1}{|\operatorname{Re} z|},
\end{aligned}$$

we can bound the whole integral as follows

$$\begin{aligned}
\left|\frac{1}{2\pi i}\int_{\eta_-} g_T(z)e^{zT}\left(1+\frac{z^2}{R^2}\right)\frac{dz}{z}\right| &= \left|\frac{1}{2\pi i}\int_{\eta_R} g_T(z)e^{zT}\left(1+\frac{z^2}{R^2}\right)\frac{dz}{z}\right| \\
&\leq \frac{1}{2\pi}\int_{\eta_R}\left|g_T(z)e^{zT}\left(1+\frac{z^2}{R^2}\right)\frac{1}{z}dz\right| \\
&\leq \frac{1}{2\pi}\int_{\eta_R} M\frac{e^{|\operatorname{Re} z|T}-1}{|\operatorname{Re} z|}e^{\operatorname{Re}(z)T}\frac{2\operatorname{Re} z}{R^2}|dz| \\
&= \frac{M}{\pi R^2}\int_{\eta_R}(1-e^{\operatorname{Re}(z)T})|dz| \\
&\leq \frac{M}{\pi R^2}\int_{\eta_R}|dz| \\
&= \frac{M}{\pi R^2}\pi R \\
&= \frac{M}{R}.
\end{aligned}$$

Case B.2) Contribution of g to $I_-(T)$.

For the contribution of g to I_-, because $\left|e^{zT}\right| = e^{\operatorname{Re}(z)T} \overset{T\to\infty}{\longrightarrow} 0$ uniformly

on compact subsets of $\{\operatorname{Re} z < 0\}$, we have

$$\lim_{T\to\infty} \left| \frac{1}{2\pi i} \int_{\eta_-} g(z) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z} \right| = 0.^{10}$$

Combining cases B.1) and B.2), we get the bound

$$\begin{aligned} |I_-(T)| &\leq \left| \frac{1}{2\pi i} \int_{\eta_-} g(z) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z} \right| \\ &\quad + \left| -\frac{1}{2\pi i} \int_{\eta_-} g_T(z) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z} \right| \\ &\leq \left| \frac{1}{2\pi i} \int_{\eta_-} g(z) e^{zT} \left(1 + \frac{z^2}{R^2}\right) \frac{dz}{z} \right| + \frac{M}{R}, \end{aligned}$$

which is valid for all T.

Taking lim sup with respect to T on both sides, we find

$$\limsup_{T\to\infty} |I_-(T)| \leq \frac{M}{R},$$

where we have used the zero value of the limit found above.

Now, combining the estimates of $I_+(T)$ and $I_-(T)$ as follows

$$\begin{aligned} |g(0) - g_T(0)| &= |I_-(T) + I_+(T)| \\ &\leq |I_-(T)| + |I_+(T)| \\ &\leq |I_-(T)| + \frac{M}{R}, \end{aligned}$$

and taking again lim sup with respect to T, we get

$$\limsup_{T\to\infty} |g(0) - g_T(0)| \leq \frac{2M}{R}.$$

Since R was arbitrary, this completes the proof. □

We remark that the requirement in Newman's Tauberian Theorem that the function f be bounded can be dropped since it can be shown to follow from the other hypotheses via properties of the Laplace transform.[11]

Applying Newman's Tauberian Theorem with a particular $f(t)$ and $g(z)$ defined essentially by the logarithmic derivative of the Riemann zeta function leads to the PNT. To obtain the PGT, we follow the same strategy, but now make use of the Selberg zeta function.

Proof of the Prime Geodesic Theorem. Recall that $\mathcal{L}_X$ denotes the set of all primitive oriented closed geodesics on a geometrically finite hyperbolic surface X and that

$$Z_X(s) \equiv \prod_{\ell\in\mathcal{L}_X} \prod_{k=0}^{\infty} \left(1 - e^{-(s+k)\ell}\right)$$

[10] For details, see Bak and Newman's proof ([3], pp. 288-289).

[11] Refer to Proposition 5.2. in Korevaar [35].

denotes the Selberg zeta function. Define the function

$$\eta(s) \equiv \sum_{\ell \in \mathcal{L}_X} \ell e^{-s\ell}.$$

The proof has four steps.

Step 1) Proof that $\eta(s) - \frac{1}{s-\delta}$ is holomorphic for $\operatorname{Re} s \geq \delta$.

For $\operatorname{Re} s > \delta$, using differentiation with respect to s, write

$$\begin{aligned}\frac{Z_X'(s)}{Z_X(s)} &= \frac{\sum_{\ell \in \mathcal{L}_X} \sum_{k=0}^{\infty} \prod_{\ell' \neq \ell, \ell' \in \mathcal{L}_X} \prod_{k'=0, k' \neq k}^{\infty} \ell e^{-(s+k)\ell}(1 - e^{-(s+k')\ell'})}{\prod_{\ell \in \mathcal{L}_X} \prod_{k=0}^{\infty}(1 - e^{-(s+k)\ell})} \\ &= \sum_{\ell \in \mathcal{L}_X} \sum_{k=0}^{\infty} \frac{\ell}{e^{(s+k)\ell} - 1}.\end{aligned}$$

Subtracting $\eta(s)$, observe that

$$\begin{aligned}\frac{Z_X'(s)}{Z_X(s)} - \eta(s) &= \sum_{\ell \in \mathcal{L}_X} \sum_{k=0}^{\infty} \frac{\ell}{e^{(s+k)\ell} - 1} - \sum_{\ell \in \mathcal{L}_X} \ell e^{-s\ell} \\ &= \sum_{\ell \in \mathcal{L}_X} \sum_{k=1}^{\infty} \frac{\ell}{e^{(s+k)\ell} - 1} + \sum_{\ell \in \mathcal{L}_X} \left(\frac{\ell}{e^{s\ell} - 1} - \ell e^{-s\ell} \right) \\ &= \sum_{\ell \in \mathcal{L}_X} \frac{\ell}{e^{s\ell}\,(e^{s\ell} - 1)} + \sum_{\ell \in \mathcal{L}_X} \sum_{k=1}^{\infty} \frac{\ell}{e^{(s+k)\ell} - 1}.\end{aligned}$$

It can be shown that the first sum on the right is holomorphic for $\operatorname{Re} s > \frac{\delta}{2}$ and the second is holomorphic for $\operatorname{Re} s > \delta - 1$. This implies that the right-hand side is holomorphic for $\operatorname{Re} s > \frac{\delta}{2}$, since (by Proposition 4.1.3) $\max(\{\frac{\delta}{2}, \delta - 1\}) = \frac{\delta}{2}$. Moreover, since $\frac{Z_X'(s)}{Z_X(s)}$ is meromorphic, which means that it is holomorphic everywhere except at the zeros and poles of $Z_X(s)$, we can isolate $\eta(s)$ in the above equation so that it inherits the poles and residues of $Z_X(s)$ that can be found in the region $\operatorname{Re} s > \frac{\delta}{2}$. In other words, $\eta(s)$ is holomorphic in $\operatorname{Re} s > \frac{\delta}{2}$ except at the zeros and poles of $Z_X(s)$. By Theorem 7.3.2, the only problematic point is a simple pole occurring at $s = \delta$. Thus, we can remove this pole of unit residue, obtaining $\eta(s) - \frac{1}{s-\delta}$ holomorphic for $\operatorname{Re} s \geq \delta$.

Step 2) Applying Newman's Tauberian Theorem.

In the same vein as in Proposition 4.1.3, write $\eta(s)$ as the following integral

$$\eta(s) = \int_0^{\infty} e^{-st} d\theta(t),$$

where we define

$$\theta(t) \equiv \sum_{\ell \in \mathcal{L}_X : \ell \leq t} \ell,$$

and use integration by parts to obtain

$$\begin{aligned}\eta(s) &= \int_0^\infty e^{-st} d\theta(t) \\ &= \left[e^{-st}\theta(t)\right]_0^\infty - \int_0^\infty (-s)e^{-st}\theta(t)dt \\ &= s\int_0^\infty e^{-st}\theta(t)dt.\end{aligned}$$

This holds for $\operatorname{Re} s > 1$, since $\theta(t) \leq t\pi_X(t)$, and for s in this range, the bound obtained in Proposition 3.2.3 gives

$$e^{-st}\theta(t) \leq e^{-st}tCe^t \leq Cte^{(1-s)t} \overset{t\to\infty}{\longrightarrow} 0,$$

for some constant C.

Now, by defining

$$g(z) \equiv \frac{\eta(z+\delta)}{z+\delta} - \frac{1}{\delta z},$$

write

$$\begin{aligned}g(z) &= \frac{(z+\delta)\int_0^\infty e^{-(z+\delta)t}\theta(t)dt}{(z+\delta)} - \frac{1}{z\delta} \\ &= \int_0^\infty e^{-(z+\delta)t}\theta(t)dt - \int_0^\infty \frac{e^{-zt}}{\delta}dt \\ &= \int_0^\infty e^{-zt}\left[e^{-\delta t}\theta(t) - \frac{1}{\delta}\right]dt.\end{aligned}$$

In Step 1) it was seen that $\eta(s) - \frac{1}{s-\delta}$ is holomorphic for $\operatorname{Re} s \geq \delta$. Consequently,

$$\frac{\eta(s)}{s} - \frac{1}{\delta(s-\delta)}$$

is also holomorphic for $\operatorname{Re} s \geq \delta$. Notice that $s \neq 0$, because $\operatorname{Re} 0 \geq \delta$ implies $\delta = 0$, but we are assuming that δ refers to a non-elementary Fuchsian group, so $\delta > 0$. Thus, letting $s = z + \delta$, we find that

$$\frac{\eta(z+\delta)}{z+\delta} - \frac{1}{\delta z} \equiv g(z) \text{ is holomorphic for } \operatorname{Re}(z+\delta) \geq \delta.$$

But since $\operatorname{Re}(z+\delta) = \operatorname{Re}(z) + \delta$, $g(z)$ is holomorphic for $\operatorname{Re}(z) \geq 0$.

Thus, Newman's Tauberian Theorem (Theorem 7.4.1) applies,[12] and we get

$$\int_0^\infty \left[e^{-\delta t}\theta(t) - \frac{1}{\delta}\right] dt < \infty.$$

Now it is just a matter of deducing the desired asymptotic behavior from this integrability. We first prove an intermediate asymptotic relationship.

Step 3) Proof that $\theta(t) \sim \frac{e^{\delta t}}{\delta}$.

By way of contradiction, take $\varepsilon > 0$ and assume that $\theta(t_j) \geq e^{\delta(t_j+\varepsilon)}/\delta$ for some sequence $t_j \to \infty$. Since $\theta(t)$ increases, we would then have

$$\begin{aligned}
\int_{t_j}^{t_j+\varepsilon} \left[e^{-\delta t}\theta(t) - \frac{1}{\delta}\right] dt &\geq \int_{t_j}^{t_j+\varepsilon} \left[e^{-\delta t}\theta(t_j) - \frac{1}{\delta}\right] dt \\
&\geq \int_{t_j}^{t_j+\varepsilon} \left(\frac{e^{\delta(t_j+\varepsilon-t)}}{\delta} - \frac{1}{\delta}\right) dt \\
&= \frac{e^{\delta\varepsilon}}{\delta^2} - \frac{1}{\delta^2} - \frac{\varepsilon}{\delta} \\
&\geq \frac{\delta\varepsilon}{\delta^2} + \frac{1}{\delta^2} + \frac{\delta^2\varepsilon^2}{2\delta^2} - \frac{1}{\delta^2} - \frac{\delta\varepsilon}{\delta^2} \\
&= \frac{\varepsilon^2}{2}.
\end{aligned}$$

Since $t_j \to \infty$ and ε is arbitrary, this contradicts the finiteness of the integral $\int_0^\infty \left[e^{-\delta t}\theta(t) - \frac{1}{\delta}\right] dt$ as proved in Step 2).

The contradiction allows us to conclude that

$$\begin{aligned}
\theta(t_j) &< \frac{e^{\delta(t_j+\varepsilon)}}{\delta} \\
\delta e^{-\delta t_j}\theta(t_j) &< e^{\delta\varepsilon} \xrightarrow{\varepsilon\to 0} 1 \\
\limsup_{t\to\infty} \delta e^{-\delta t}\theta(t) &\leq 1
\end{aligned}$$

A similar argument can be used to find $1 \leq \liminf_{t\to\infty} \delta e^{-\delta t}\theta(t)$. Thus,

$$1 \leq \liminf_{t\to\infty} \delta e^{-\delta t}\theta(t) \leq \limsup_{t\to\infty} \delta e^{-\delta t}\theta(t) \leq 1 \Rightarrow \lim_{t\to\infty} \delta e^{-\delta t}\theta(t) = 1,$$

[12] It was previously mentioned that one can drop the requirement imposed by Newman's Tauberian Theorem that the function $e^{-\delta t}\theta(t) - \frac{1}{\delta}$ be shown bounded. This is a consequence of Proposition 5.2. in Korevaar ([35], p. 1114). More explicitly, to apply Korevaar's result to the case at hand let $k \equiv \frac{s}{\delta}$. Then $s = k\delta$ and we can define $f(k) \equiv \eta(k\delta)$, so that

$$f(k) = \delta k \int_0^\infty e^{-\delta k t}\theta(t)dt.$$

By change of variables, let $u \equiv \delta t$ and set $S(u) \equiv \theta(\frac{u}{\delta})$. Hence, $du = \delta dt$ and

$$f(k) = \delta k \int_0^\infty e^{-\delta k t}\theta(t)dt = \delta k \int_0^\infty e^{-ku} S(u)\frac{du}{\delta} = k\int_0^\infty e^{-ku}S(u)du.$$

By the holomorphicity found in Step 1) and making the necessary substitutions, we get that $\eta(s) - \frac{1}{s-\delta}$ is holomorphic for $\operatorname{Re} s \geq \delta$ if and only if $f(k) - \frac{1}{\delta(k-1)}$ is holomorphic for $\operatorname{Re} k \geq 1$. Therefore, Proposition 5.2. in Korevaar applies and $e^{-u}S(u) - \frac{1}{\delta} = e^{-\delta t}\theta(t) - \frac{1}{\delta}$ is bounded.

establishing the asymptotic relationship.

Step 4) Rewrite the asymptotic relationship in terms of $\pi_X(t)$.

By definition, note that $\theta(t) \equiv \sum_{\ell \in \mathcal{L}_X : \ell \leq t} \ell \leq t\pi_X(t)$ for all t, since $\pi_X(t) \equiv \#\{\ell \in \mathcal{L}_X : \ell \leq t\}$. So,

$$\frac{t\pi_X(t)}{\theta(t)} \geq 1 \Rightarrow \liminf_{t\to\infty} \frac{t\pi_X(t)}{\theta(t)} \geq 1.$$

On the other hand, for any $\lambda < 1$ we have

$$\begin{aligned}\theta(t) &\geq \sum_{\ell \in \mathcal{L}_X : \lambda t \leq \ell \leq t} \ell \\ &\geq \lambda t \left[\pi_X(t) - \pi_X(\lambda t)\right].\end{aligned}$$

As δ is the critical exponent of the considered Fuchsian group, $\sum_\ell e^{-s\ell}$ converges for any $s > \delta$.

Now, since

$$\sum_{\ell \in \mathcal{L}_X : \ell \leq t} e^{-s\ell} \geq \pi_X(t) e^{-st},$$

if we take some $\varepsilon > 0$, as $\delta + \varepsilon > \delta$, said convergence implies $0 < \sum_{\ell \in \mathcal{L}_X : \ell \leq t} e^{-s\ell} = M < \infty$. Thus,

$$e^{(\delta+\varepsilon)t} M \geq \pi_X(t) \Rightarrow \pi_X(t) = O\left(e^{(\delta+\varepsilon)t}\right).$$

Returning to $\theta(t) \geq \lambda t \left[\pi_X(t) - \pi_X(\lambda t)\right]$ and using the previous estimate, write

$$\begin{aligned}\frac{t\pi_X(t)}{\theta(t)} &\leq \frac{1}{\lambda} + \frac{t\pi_X(t\lambda)}{\theta(t)} = \frac{1}{\lambda} + \frac{\delta t M e^{(\delta+\varepsilon)t\lambda} e^{\delta t}}{e^{\delta t}\delta\theta(t)} \\ \frac{t\pi_X(t)}{\theta(t)} &\leq \lambda^{-1} + M' t e^{(\delta+\varepsilon)\lambda t - \delta t} \frac{e^{\delta t}}{\delta\theta(t)},\end{aligned}$$

where we have used $\delta \leq 1$ (Proposition 4.1.3) and M' is some constant.

We want $(\delta + \varepsilon)\lambda - \delta < 0$, so that the exponential on the numerator of the right-hand side of the last inequality goes to zero as $t \to \infty$. Thus, for the choice $\varepsilon < \delta(1-\lambda)/\lambda$, we have

$$\limsup_{t\to\infty} \frac{t\pi_X(t)}{\theta(t)} \leq \lambda^{-1}$$

for any $\lambda < 1$, where we have used the asymptotic relationship found in Step 3) (i.e., $\limsup_{t\to\infty} \frac{\delta\theta(t)}{e^{\delta t}} = 1$). Taking $\lambda \to 1$, the above established inequalities of $\liminf$ and $\limsup$ combined with said asymptotic relationship imply

$$\lim_{t\to\infty} \frac{t\pi_X(t)}{\theta(t)} = 1 \Rightarrow \theta(t) \sim t\pi_X(t) \Rightarrow \frac{e^{\delta t}}{\delta} \sim t\pi_X(t),$$

and the proof is complete. □

Chapter 8

Some recent developments

In this concluding chapter, we briefly discuss some recent results of interest that can be found in the literature and that are related to the PGT.

We start with two celebrated developments that are closely related to the PGT and that have appeared since the original work by Lalley [36] and Guillopé [26] in the 1980s. The first is a result by Naud [50], who obtained a spectral gap for the Selberg zeta function in the convex cocompact case building on ideas introduced by Dolgopyat [18]. For $\delta \leq \frac{1}{2}$, this gap has deep consequences on resonances and leads to the following error estimate for the PGT:

$$\pi_X(t) = \mathrm{li}(e^{\delta t}) + O\left(e^{(\delta-\alpha)t}\right),$$

for some $\alpha > 0$ and where $\mathrm{li}(x) \equiv \int_2^x \frac{dt}{\log t}$ denotes the log-integral function that also appears in precise formulations of the PNT. The second is an improvement of Bourgain and Dyatlov [11] that removes the dependence of the essential spectral gap constant ε on the surface. In particular, they show that for a positive critical exponent δ associated with a convex cocompact hyperbolic surface, there exists ε depending only on δ, such that Z_X has finitely many zeros s with $\mathrm{Re}\, s > \delta - \varepsilon$.

More recently, Sarkar and Winter [63] addressed the problem of counting closed geodesics beyond the realm of hyperbolic surfaces, thus revisiting the high dimensional version of the problem previously tackled by Perry [56], while establishing other significant and relatively more general results. It is instructive to discuss some of Sarkar and Winter's achievements, since it showcases how the orbit-counting problem is interconnected to different issues at the forefront of modern discoveries in dynamics.

The authors considered convex cocompact manifolds $X \cong \Gamma\backslash\mathbb{H}^n$, where Γ is a discrete subgroup of G, the group of orientation-preserving isometries of $\mathbb{H}^n$, the high dimensional variant of the hyperbolic plane. In this general setting, it is sometimes useful to consider measures that are constructed from Patterson-Sullivan measures but that instead live in the unit tangent bundle of $X \cong \Gamma\backslash\mathbb{H}^n$. These measures $\mathbf{m}$ have interesting properties and have been called Bowen-Margulis-Sullivan (probability) measures in the literature.[1] It is worth

[1]The Bowen-Margulis-Sullivan measures were first introduced independently by Margulis [44] and Bowen [13, 12]. Kaimanovich [33, 32] applied the ideas from Patterson-Sullivan measures, which as seen were constructed for spaces of constant negative curvature, for spaces of nonconstant negative curvature, establishing in the process the link between the Bowen-

noting that these measures have been as actively studied as Patterson-Sullivan measures, although they did not appear in the earlier chapters.

We also did not have space in this thesis to discuss mixing, but this concept can be taken as a property of certain measure-preserving dynamical systems or flows, such as the geodesic flow, a powerful property that implies ergodicity.[2] In an intuitive sense, the mixing property implies that measure-preserving (dynamical) systems are able to evolve past their initial conditions. This property can be further qualified as exponentially mixing, depending on the rate at which this reset process occurs. Moreover, as previously mentioned, other types of flow can be defined, such as horocyclic and frame flows. The latter type of flow, which is roughly similar to a geodesic flow that acts on the orthonormal frame bundle of a given manifold (and not its unit tangent bundle), has been known for some time to be mixing with respect to $\mathbf{m}$. This knowledge is due to independent work by Winter [79] and Babillot [2]. It turns out that the question of whether the frame flow is exponentially mixing with respect to $\mathbf{m}$ is deeply related to orbit-counting problems.

This issue was first resolved for the geodesic flow, which was first shown to be exponentially mixing in a lattice setting by Moore [49] and Ratner [60], using the L^2 spectral gap of the Laplacian. Later, this was extended to a general convex cocompact Γ by Stoyanov [68], who also relied on the methods developed by Dolgopyat [18]. To each discrete subgroup Γ of G (as defined above) one can find a critical exponent $\delta(\Gamma) \in (0, n-1]$ that behaves much the same way as the critical exponent discussed in the previous chapters. After this improvement by Stoyanov focusing on the geodesic flow, there followed proofs of the frame flow as exponentially mixing for particular ranges of $\delta(\Gamma)$, on account of work by Mohammadi and Oh [48] and a paper by Edwards and Oh [19], and in the lattice setting (without the restrictions on ranges), on account of the Selberg trace formula as used in Sarnak and Wakayama [65]. Sarkar and Winter's contribution can be understood as the next logical step in this story, since the authors proved that the frame flow is exponentially mixing without assumptions involving the range of the critical exponent associated with a general convex cocompact Γ. This breakthrough opened the doors for further advances by Chow and Sarkar [16], who extended this result to locally symmetric spaces, and by Li, Pan, and Sarkar [40], who finished writing this story, having settled the matter for geometrically finite hyperbolic manifolds. This last paper uses a symbolic coding approach previously introduced by Li and Pan [39] for geodesic flows, in addition to recasting Dolgopyat's methods in the manner of Tsujii and Zhang [76].

The asymptotics of the length counting function are a consequence of this characterization of frame flows, as it is used (together with an auxiliary result that depends on stable and unstable Burger-Roblin measures) to obtain Theorem 1.3. (in the authors' article and preprint numbering), which establishes the exponential equidistribution of holonomies[3] for closed geodesics on convex cocompact manifolds (Sarkar and Winter [63]) and geometrically finite manifolds

Margulis measures and Patterson-Sullivan measures. For a brief practical appraisal of the relationship between Bowen-Margulis-Sullivan measures and Patterson-Sullivan measures see, for instance, Roblin ([61], p. 12).

[2]For a proper definition, see, for example, Nicholls ([53], p. 125). The fact that the mixing property is stronger than ergodicity is proved in Theorem 7.2.13. in Nicholls [53].

[3]Simply put, holonomies measure how parallel transport around closed geodesics on manifolds fails to conserve the geometric data transported.

(Li, Pan, and Sarkar [40]).

We remark here that these results are seen as the counterpart of an earlier result by Oh and Winter [54], which established that holonomies are exponentially equidistributed as well for hyperbolic rational maps.[4] In this case, there is no set Γ and δ plays the role of the Hausdorff dimension of the Julia set[5] of the rational map studied. This duality illustrates the phenomenon known in the literature as Sullivan's dictionary, in which certain properties of rational maps find a match among properties of (usually Kleinian) groups acting on hyperbolic spaces.[6]

Furthermore, this complex dynamics prime orbit theorem due to Oh and Winter has recently been generalized from hyperbolic (uniformly expanding) rational maps to a non-uniformly expanding setting, on account of work done by Li and Zheng [41, 42, 43]. The lack of uniform expansion of a rational map arises from the presence of critical points in the Julia set. The most general version of Li and Zheng's results deals with expanding Thurston maps, which are indispensable tools in modern complex dynamics. The prime orbit theorem in Oh and Winter's work relies on a geometric potential that is guaranteed Lipschitz continuous for hyperbolic rational maps. This is no longer the case when considering the non-uniformly expanding setting, since the critical points affect the continuity of the geometric potential.

In conclusion, the foregoing concise remarks serve to convey to the reader a sense of the current enthusiasm surrounding some topics that intersect with those encountered in this thesis. The PGT as presented in this thesis is to some extent only a classical result, a consequence of more general spatial properties undergoing active study by professional mathematicians.

[4] Any holomorphic map $f : \hat{\mathbb{C}} \to \hat{\mathbb{C}}$ on the Riemann sphere can be expressed as the ratio of two polynomials (see Milnor [47], p. 41), hence the name. The adjective hyperbolic refers to a rational map that is (dynamically) expanding on its Julia set (see Milnor [47], p. 205).

[5] On the topic of Julia sets and complex dynamics, consult Milnor [47].

[6] On Sullivan's dictionary, see, for example, Fraser and Stuart [25].

Bibliography

[1] Ian Agol. *Tameness of hyperbolic 3-manifolds*. 2004. eprint: `arXiv:math/0405568`.

[2] Martine Babillot. "On the mixing property for hyperbolic systems". In: *Israel Journal of Mathematics* 129 (2002), pp. 61–76.

[3] Joseph Bak and Donald J. Newman. *Complex Analysis*. 3rd Edition. Springer, 2010.

[4] Alan F. Beardon. "Inequalities for certain Fuchsian groups". In: *Acta Mathematica* 127.3 (1971), pp. 221–258.

[5] Alan F. Beardon. "The Exponent of Convergence of Poincaré Series". In: *Proceedings of the London Mathematical Society* s3-18.3 (1968), pp. 461–483.

[6] Alan F. Beardon. *The Geometry of Discrete Groups*. Springer, 1983.

[7] Vladimir I. Bogachev. *Measure Theory*. Vol. 1–2. Springer, 2007.

[8] David Borthwick. *Introduction to Spectral Theory on Hyperbolic Surfaces*. Lecture notes. [Online; accessed 19-June-2025]. 2010. URL: `http://www.mathcs.emory.edu/~davidb/spths.pdf`.

[9] David Borthwick. *Spectral Theory of Infinite-Area Hyperbolic Surfaces*. 2nd Edition. Birkhäuser, 2016.

[10] David Borthwick, Chris Judge, and Peter Perry. "Determinants of Laplacians and Isopolar Metrics on Surfaces of Infinite Area". In: *Duke Mathematical Journal* 118.1 (2003), pp. 61–102.

[11] Jean Bourgain and Semyon Dyatlov. "Fourier dimension and spectral gaps for hyperbolic surfaces". In: *Geometric and Functional Analysis* 27.4 (2017), pp. 744–771.

[12] Rufus Bowen. "Periodic orbits for hyperbolic flows". In: *American Journal of Mathematics* 94 (1972), pp. 1–30.

[13] Rufus Bowen. "Periodic points and measures for Axiom A diffeomorphisms". In: *Transactions of the American Mathematical Society* 154 (1971), pp. 377–397.

[14] Jürg Peter Buser. *Geometry and Spectra of Compact Riemann Surfaces*. Birkhäuser, 2010.

[15] Danny Calegari and David Gabai. "Shrinkwrapping and the Taming of Hyperbolic 3-Manifolds". In: *Journal of the American Mathematical Society* 19.2 (2006), pp. 385–446.

[16] Michael Chow and Pratyush Sarkar. "Local Mixing of One-Parameter Diagonal Flows on Anosov Homogeneous Spaces". In: *International Mathematics Research Notices* 2023.18 (2023), pp. 15834–15895.

[17] Manfred Denker and Bernd O. Stratmann. "The Patterson Measure: Classics, Variations and Applications". In: *Contributions in Analytic and Algebraic Number Theory*. Springer, 2012, pp. 171–195.

[18] Dmitry Dolgopyat. "On decay of correlations in Anosov flows". In: *Annals of Mathematics* 147.2 (1998), pp. 357–390.

[19] Sam Edwards and Hee Oh. "Spectral gap and exponential mixing on geometrically finite hyperbolic manifolds". In: *Duke Mathematical Journal* 170.15 (2021), pp. 3417–3458.

[20] Manfred Einsiedler and Thomas Ward. *Ergodic Theory: with a view towards Number Theory*. Springer, 2010.

[21] Paul Erdős. "On a New Method in Elementary Number Theory Which Leads to An Elementary Proof of the Prime Number Theorem". In: *Proceedings of the National Academy of Sciences* 35.7 (1949), pp. 374–384.

[22] Kenneth Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. 2nd Edition. John Wiley & Sons, 2003.

[23] Werner Fenchel and Jakob Nielsen. *Discontinuous Groups of Isometries in the Hyperbolic Plane*. De Gruyter, 2002.

[24] Gerald B. Folland. *Real Analysis: Modern Techniques and Their Applications*. 2nd Edition. John Wiley & Sons, 1999.

[25] Jonathan M. Fraser and Liam Stuart. "A new perspective on the Sullivan dictionary via Assouad type dimensions and spectra". In: *Bulletin of the American Mathematical Society* 61.1 (2024), pp. 103–118.

[26] Laurent Guillopé. "Sur la distribution des longueurs des geodesiques fermees d'une surface compacte a bord totalement geodesique". In: *Duke Mathematical Journal* 53 (1986), pp. 827–848.

[27] Dennis A. Hejhal. *The Selberg Trace Formula for PSL(2, $\mathbb{R}$), Vol. I.* Vol. 548. Lecture Notes in Mathematics. Springer-Verlag, 1976.

[28] Eberhard Hopf. "Ergodic theory and the geodesic flow on surfaces of constant negative curvature". In: *Bull. Amer. Math. Soc. 77(6)* 6.77 (1971), pp. 863–877.

[29] Eberhard Hopf. "Statistik der geodätischen Linien in Mannigfaltigkeiten negativer Krümmung." In: *Ber. Sächs. Akad. Wiss. Leipzig, math.-phys. Kl. 91* (1939), pp. 261–304.

[30] Heinz Huber. "Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen". In: *Mathematische Annalen* 138.1 (1959), pp. 1–26.

[31] Albert E. Ingham. *The distribution of prime numbers*. Reissued. Cambridge University Press, 1990.

[32] Vadim A. Kaimanovich. "Bowen-Margulis and Patterson measures on negatively curved compact manifolds". In: *Dynamical systems and related topics (Nagoya, 1990)*. Vol. 9. World Scientific, 1991, pp. 223–232.

[33] Vadim A. Kaimanovich. "Invariant measures of the geodesic flow and measures at infinity on negatively curved manifolds". In: *Annales de l'Institut Henri Poincaré - Physique Théorique* 53.4 (1990), pp. 361–393.

[34] Svetlana Katok. *Fuchsian Groups*. University of Chicago Press, 1992.

[35] Jaap Korevaar. "The Wiener-Ikehara Theorem by Complex Analysis". In: *Proceedings of the American Mathematical Society* 134.4 (2006), pp. 1107–1116.

[36] Steven P. Lalley. "Renewal theorems in symbolic dynamics, with applications to geodesic flows, noneuclidean tessellations and their fractal limits". In: *Acta Mathematica* 163 (1989), pp. 1871–2509.

[37] John M. Lee. *Introduction to Riemannian Manifolds*. 2nd Edition. Springer, 2019.

[38] John M. Lee. *Introduction to Topological Manifolds*. 2nd Edition. Springer, 2010.

[39] Jialun Li and Wenyu Pan. "Exponential mixing of geodesic flows for geometrically finite hyperbolic manifolds with cusps". In: *Inventiones Mathematicae* 231 (2023), pp. 931–1021.

[40] Jialun Li, Wenyu Pan, and Pratyush Sarkar. *Exponential mixing of frame flows for geometrically finite hyperbolic manifolds*. 2023. eprint: `arXiv:2308.06203`.

[41] Zhiqiang Li and Tianyi Zheng. "Prime orbit theorems for expanding Thurston maps: Dirichlet series and orbifolds". In: *Advances in Mathematics* 443 (2024), p. 109600.

[42] Zhiqiang Li and Tianyi Zheng. "Prime Orbit Theorems for expanding Thurston maps: Genericity of strong non-integrability condition". In: *Advances in Mathematics* 450 (2024), p. 109765.

[43] Zhiqiang Li and Tianyi Zheng. "Prime orbit theorems for expanding Thurston maps: Lattès maps and split Ruelle operators". In: *Advances in Mathematics* 449 (2024), p. 109723.

[44] Gregory Margulis. "Certain measures that are related to Anosov flows". In: *Funkcional'nyi Analiz i Priloženie* 4.1 (1970), pp. 62–76.

[45] Jens Marklof. "Selberg's Trace Formula: An Introduction". In: *Hyperbolic Geometry and Applications in Quantum Chaos and Cosmology*. Cambridge University Press, 2011, pp. 83–120.

[46] William S. Massey. *Algebraic Topology: An Introduction*. Springer, 1977.

[47] John Milnor. *Dynamics in One Complex Variable*. 3rd Edition. Princeton University Press, 2006.

[48] Amir Mohammadi and Hee Oh. "Matrix coefficients, counting and primes for orbits of geometrically finite groups". In: *J. Eur. Math. Soc. (JEMS)* 17.4 (2015), pp. 837–897.

[49] Calvin C. Moore. "Exponential decay of correlation coefficients for geodesic flows". In: *Group representations, ergodic theory, operator algebras, and mathematical physics (Berkeley, Calif., 1984)*. Vol. 6. Math. Sci. Res. Inst. Publ. Springer, 1987, pp. 163–181.

[50] Frédéric Naud. "Expanding maps on Cantor sets and analytic continuation of zeta functions". In: *Annales scientifiques de l'École Normale Supérieure* 38.4 (2005), pp. 116–153.

[51] Viktor V. Nemytskii and Vyacheslav V. Stepanov. *Qualitative Theory of Differential Equations*. Reissued. Princeton University Press, 2016.

[52] Donald J. Newman. "Simple Analytic Proof of the Prime Number Theorem". In: *The American Mathematical Monthly* 87.9 (1980), pp. 693–696.

[53] Peter J. Nicholls. *The Ergodic Theory of Discrete Groups*. Cambridge University Press, 1989.

[54] Hee Oh and Dale Winter. "Prime number theorems and holonomies for hyperbolic rational maps". In: *Inventiones Mathematicae* 208.2 (2017), pp. 401–440.

[55] Samuel J. Patterson. "The limit set of a Fuchsian group". In: *Acta Mathematica* 136.1 (1976), pp. 241–273.

[56] Peter A. Perry. "Asymptotics of the length spectrum for hyperbolic manifolds of infinite volume". In: *Geometric and Functional Analysis* 11.1 (2001), pp. 132–141.

[57] Jean-François Quint. "An overview of Patterson-Sullivan theory". [Online; accessed 05-November-2025]. 2006. URL: `https://imag.umontpellier.fr/~quint/Publications/courszurich.pdf`.

[58] Burton Randol. "On the Asymptotic Distribution of Closed Geodesics on Compact Riemann Surfaces". In: *Transactions of the American Mathematical Society* 233 (1977), pp. 241–247.

[59] John G. Ratcliffe. *Foundations of Hyperbolic Manifolds*. 3rd Edition. Springer, 2019.

[60] Marina Ratner. "Markov partitions for Anosov flows on n-dimensional manifolds". In: *Israel Journal of Mathematics* 15 (1973), pp. 92–114.

[61] Thomas Roblin. "Ergodicité et équidistribution en courbure négative". In: *Mémoires de la Société Mathématique de France*. [Online; accessed 05-November-2025]. 2003. URL: `https://api.semanticscholar.org/CorpusID:117172616`.

[62] Daniel J. Rudolph. "Ergodic behaviour of Sullivan's geometric measure on a geometrically finite hyperbolic manifold". In: *Ergodic Theory and Dynamical Systems* 2 (1982), pp. 491–512.

[63] Pratyush Sarkar and Dale Winter. "Exponential mixing of frame flows for convex cocompact hyperbolic manifolds". In: *Compositio Mathematica* 157.12 (2021), pp. 2585–2634.

[64] Peter Sarnak. "Prime Geodesic Theorems". PhD Thesis. Stanford University, 1980.

[65] Peter Sarnak and Masato Wakayama. "Equidistribution of holonomy about closed geodesics". In: *Duke Mathematical Journal* 100.1 (1999), pp. 1–57.

[66] Barbara Schapira. "Dynamics of Geodesic and Horocyclic Flows". In: *Ergodic Theory and Negative Curvature*. Vol. 2164. Lecture Notes in Mathematics. Springer, 2017, pp. 129–155.

[67] Atle Selberg. “An Elementary Proof of the Prime-Number Theorem”. In: *Annals of Mathematics* 50.2 (1949), pp. 305–313.

[68] Luchezar Stoyanov. “Spectra of Ruelle transfer operators for Axiom A flows”. In: *Nonlinearity* 24.4 (2011), pp. 1089–1120.

[69] Bernd O. Stratmann. “Fractal geometry on hyperbolic manifolds”. In: *Non-Euclidean Geometries, János Bolyai Memorial Volume*. Springer, 2006, pp. 227–249.

[70] Bernd O. Stratmann and Mariusz Urbański. “The box-counting dimension for geometrically finite Kleinian groups”. In: *Fundamenta Mathematicae* 149 (1996), pp. 83–93.

[71] Bernd O. Stratmann and Mariusz Urbański. “The geometry of conformal measures for parabolic rational maps”. In: *Mathematical Proceedings of the Cambridge Philosophical Society* 128 (2000), pp. 141–156.

[72] Dennis Sullivan. “Conformal dynamical systems”. In: *Geometric Dynamics*. Vol. 1007. Lecture Notes in Mathematics. Springer, 1983, pp. 725–752.

[73] Dennis Sullivan. “Entropy, Hausdorff measures old and new, and limit sets of geometrically finite Kleinian groups”. In: *Acta Mathematica* 153 (1984), pp. 259–277.

[74] Dennis Sullivan. “The density at infinity of a discrete group of hyperbolic motions”. In: *Publications Mathématiques de l’IHÉS* 50 (1979), pp. 171–202.

[75] Edward C. Titchmarsch. *The Theory of Functions*. 2nd Edition. Oxford University Press, 1939.

[76] Masato Tsujii and Zhiyuan Zhang. “Smooth mixing Anosov flows in dimension three are exponentially mixing”. In: *Annals of Mathematics* 197.1 (2023), pp. 65–158.

[77] Charles Walkden. “Hyperbolic geometry”. [Online; accessed 05-November-2025]. 2019. URL: `https://personalpages.manchester.ac.uk/staff/charles.walkden/hyperbolic-geometry/hyperbolic_geometry_1920.pdf`.

[78] Peter Walters. *An introduction to ergodic theory*. Springer, 1982.

[79] Dale Winter. “Mixing of frame flow for rank one locally symmetric spaces and measure classification”. In: *Israel Journal of Mathematics* 210.1 (2015), pp. 467–507.

[80] Don B. Zagier. “Newman’s Short Proof of the Prime Number Theorem”. In: *The American Mathematical Monthly* 104.8 (1997), pp. 705–708.