

The Bowen–Series coding and zeros of zeta functions

Mark Pollicott and Polina Vytnova

Department of Mathematics, University of Warwick

Abstract

We give a discussion of the classical Bowen–Series coding and, in particular, its application to the study of zeta functions and their zeros. In the case of compact surfaces of constant negative curvature $\kappa = -1$ the analytic extension of the Selberg zeta function to the entire complex plane is classical, and can be achieved using the Selberg trace formula. However, an alternative dynamical approach is to use the Bowen–Series coding on the boundary at infinity to obtain a piecewise analytic expanding map from which the extension of the zeta function can be obtained using properties of the associated transfer operator. This latter method has the advantage that it also applies in the case of infinite area surfaces provided they don't have cusps. For such examples the location of the zeros is somewhat more mysterious. However, in particularly simple examples there is a striking structure to the zeros when we take appropriate limits. We will try to give some insight into this phenomenon.

The authors are grateful to the organizers of the program at SUSTech in Shenszen (China) and to the head of the mathematics department Jeff Xia for their help and hospitality

Contents

1	Introduction	2
2	Two types of dynamical system	4
3	Hyperbolic Geometry	9
4	Fuchsian groups and the Bowen–Series map	10
4.1	Fuchsian groups	11
4.2	Bowen–Series map	12

5	Dynamical Applications: Compact surfaces	14
5.1	Counting closed geodesics	14
5.2	Taking a leaf out of the number theory book	16
5.3	Error terms in counting closed geodesics	16
5.4	The Banach spaces	19
5.5	Mixing of geodesic flow	20
6	The Bowen–Series map for infinite area surfaces	22
7	Dynamical applications: non-compact surfaces	25
7.1	Counting closed geodesics	25
7.2	The Banach spaces	26
7.3	Measures and mixing	28
8	Location of resonances for infinite area surfaces	28
8.1	Distribution of zeros	29
8.2	Patterns of zeros	29
8.3	Approximating Selberg Zeta function	32

Basic background. Some useful references for the basic material in these lectures are contained in the following books and articles. We will not cover even a fraction of all of the material in these sources, but that does not detract from their interest. The selection reflects the theme we are trying to give to the story we are telling.

1. Basic material: A lot of the relevant material can be found in a very accessible form in the Proceedings of the Trieste Conference on *Ergodic Theory, symbolic dynamics and hyperbolic dynamics* from 1989 [4].
2. Background in ergodic theory: The LMS–LNS volume by Nicholls [19].
3. Trace formulae: The original very readable article of McKean on the use of heat kernels [14]
4. Resonance-free regions: The article of Naud [17].
5. Density of resonances: A second article of Naud [18].

1 Introduction

Dynamical systems and ergodic theory are both subjects rich in concrete examples. To set the scene for these notes we will consider two basic examples of two somewhat different types of dynamical system. The first is an example of a discrete dynamical system (or map) corresponding

to an \mathbb{Z}_+ -action and corresponds to an expanding map on a disjoint union of intervals. This example has the virtue of being both simple and accessible. The second will be an example of a continuous dynamical system (or flow) corresponding to an \mathbb{R} -action and corresponds to the classical example of the geodesic flow for a surface of constant negative curvature. This example has been of paramount importance in the original development of ideas in ergodic theory, particularly in the 1930s.

Both of these examples, the map and the flow, have been the subject of extensive research. However, as we will see later, these two systems are not as unrelated as they might at first appear. The basis for this connection is some elegant ideas in the work of Bowen and Series [6] (and a closely related approach appears in the work of Adler and Flatto [2]).

Unfortunately, the history of this work is tinged with sadness. Robert (Rufus) Bowen was a professor at the University of California at Berkeley and during the collaboration with Caroline Series he died suddenly, at the age of 31. Their joint work was completed by his co-author Caroline Series and published posthumously in a memorial volume of Publications Mathematiques (IHES) dedicated to Bowen. Caroline is currently President of the London Mathematical Society.

There have been other ways to associate to geodesic flows an expanding map. Perhaps the best known is that of Ratner, which she later generalised to Anosov flows, and was further generalised by Bowen to Axiom A flows (which were introduced by Smale his supervisor in Berkeley). This method used two dimensional Markov sections transfer to the flow direction, with “parallelepipeds” playing the role of flow boxes. The associated Markov map corresponds to interval maps induced from the Poincaré map between sections by collapsing them along the stable direction (an observation which, at least, is made explicit in a paper of Ruelle [24]). However, the resulting interval map is not very canonical, as one would anticipate from its potential greater generality.

On the other hand, the Bowen–Series coding is very closely related to the generators of the fundamental group of the surface. Although the choice of generators is not unique, this leads to a very natural approach to understanding orbits. Indeed, part of the appeal of the Bowen–Series coding is its transparency. The work of Bowen and Series pre-dated later work of Cannon and others on the automatic structure of more general Gromov hyperbolic groups, which is in a similar spirit.

The Bowen–Series coding has proved very useful in a number of different applications. This is particularly true when an explicit understanding of the fundamental group plays a role. The Selberg zeta function takes into account precisely one closed geodesic in each conjugacy class of the fundamental group.

In the last section we shall consider an application of one of the simplest forms of such coding (for a three funnelled surface or a “pair of pants”) to the resonances of the associated Selberg zeta function.

To complete the introduction we summarize what we want to do, in a nutshell.

Aim of lectures: To explore which results/methods (particularly in the context of zeta functions) carry over from the case of compact surface V to a surface V having infinite area using the Bowen–Series coding.

2 Two types of dynamical system

We now compare these two types of dynamics.

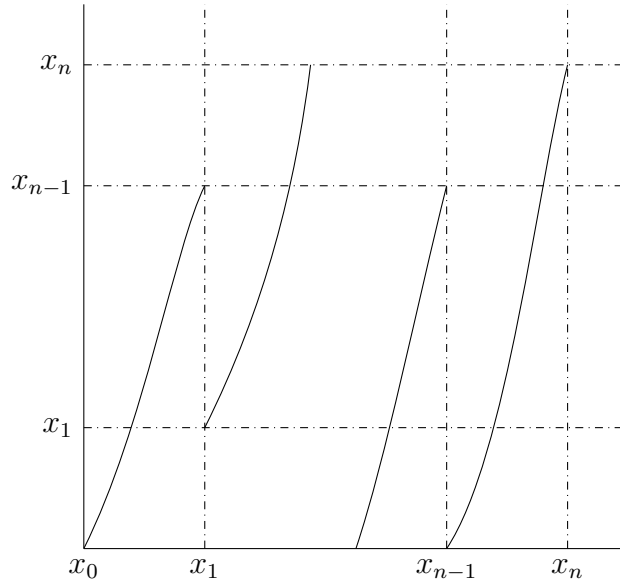


Figure 1: A Markov expanding map of the interval.

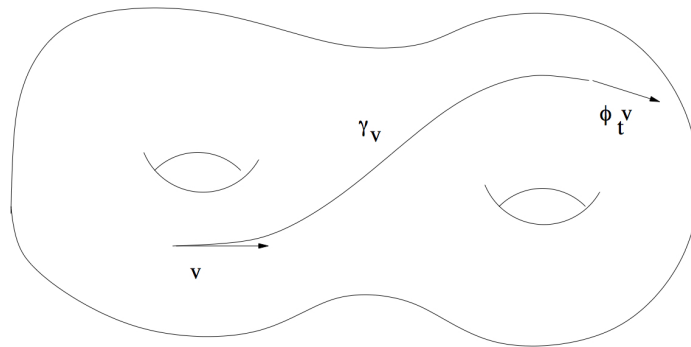


Figure 2: The geodesic flow on the unit tangent bundle of a compact surface of constant negative curvature.

Discrete (Interval maps)	Continuous (Geodesic flows)
Let $X = \bigsqcup_{j=0}^{n-1} I_j$ be a partition of the unit interval where $I_j = [x_j, x_{j+1}]$ and $0 = x_0 < x_1 < \dots < x_n = 1$	Let $Y = \{v \in TV : \ v\ = 1\} =: SV$ be the unit tangent bundle to a surface V with constant curvature $\kappa = -1$
<p>Let $T: X \rightarrow X$ be a map satisfying:</p> <ul style="list-style-type: none"> • T is piecewise-smooth and uniformly expanding, i.e. $T \in C^\infty(I_j)$ and $\exists \beta > 1$ such that $\forall x \in I_i T'(x) \geq \beta$; • T is Markov, i.e., $T(I_j)$ is a union of $\{I_k\}$ • Assume T is transitive, i. e. $\exists x \in X$ with orbit $\{T^n x\} \subseteq X$ is dense; 	<p>Let $\varphi_t: Y \rightarrow Y$ be the geodesic flow:</p> <ul style="list-style-type: none"> • $\forall v \in Y$ choose the unique geodesic $\gamma_v: \mathbb{R} \rightarrow V$ such that $\dot{\gamma}_v(0) = v$ and then define $\varphi_t v = \dot{\gamma}_v(t)$. (This corresponds to parallel transport.) • V is compact, and $\pi_1(V)$ is finitely generated • Assume V is connected, then φ_t is transitive i.e. $\exists v \{\varphi_t v \subset Y \mid t \in \mathbb{R}\}$ is dense;
<p>We can associate a natural T-invariant probability measure:</p> <p>Lemma 1 (Folklore). <i>There exists a unique T-invariant probability measure, which is absolutely continuous with respect to Lebesgue measure.</i></p>	<p>We can associate a natural φ_t-invariant probability measure:</p> <p>Lemma 2. <i>The Liouville measure $d\nu = d(\text{Vol})_V \times d\Theta$ is invariant with respect to the geodesic flow φ_t.</i></p>

Table 1: Correspondence between discrete and continuous dynamics.

The Folklore Lemma 1 has been variously attributed to: Bowen (1979), Adler (1972), Flatto (1969), Weiss (1968), Sinai (1968) and Renyi (1960). For an interesting historical perspective we refer the reader to [1].

Of course there are various classical results we are implicitly using. For example, we will assume that the surface V is orientable and has genus at least 2 (or, equivalently, negative Euler characteristic). This ensures that V can carry a metric of negative curvature. Moreover, there is a number of obvious generalisations that we will not consider. For example, the analogous results in the case of surfaces of variable negative curvature.

Having emphasized the parallels between these two systems, we can look for a more tangible connection. More precisely, we can pose the following question.

Question 1. Can we relate these two simple examples (the Markov map and the geodesic flow)?

One elegant solution to this problem appears in the work of Bowen and Series, as mentioned in the introduction. The basic mantra can be briefly summarized as follows.

Key Idea (Bowen–Series): Given the geodesic flow on a compact surface of constant negative curvature we can associate an interval map (which helps to understand properties of the flow).

If we take a surface of a fixed genus $g \geq 2$ then the associated expanding maps for the geodesic flow for different metrics on V will be topologically conjugate. However, the actual maps themselves will depend fundamentally on the choice of metric on the surface.

The fundamental group for the surface V acts on the universal cover. Underpinning this strategy is the idea of hyperfiniteness, i.e., the way the action

$$\pi_1(V) \times \partial\mathbb{D} \rightarrow \partial\mathbb{D}$$

of the fundamental group $\pi_1(V)$ on the boundary $\partial\mathbb{D}$ of the universal cover can be replaced by a single transformation. The Bowen–Series transformation is a specific realisation of such a phenomenon. It is defined as a map on the limit set, and more precisely is defined on a union of arcs which form a partition of the limit set. However, we can easily think of them as being maps of the interval, or maps on a Cantor set contained in an interval.

This basic approach has the advantage that complicated problems for the geodesic flow can often be reduced to simpler problems for the expanding map of the interval. We can summarise this method (elevated below to a “philosophy”) as follows.

Philosophy

1. Take a problem for the geodesic flow $\varphi_t: Y \rightarrow Y$;
2. Reduce the problem to a problem for the associated expanding map $T: X \rightarrow X$;
3. Solve the problem for T (assuming one can!);
4. Relate it back to a solution for the original problem for $\varphi_t: Y \rightarrow Y$.

Of course, like all approaches its value probably ultimately depends on how useful it is. More precisely, we could ask:

Question 2. What sort of problem can one address?

One would naturally expect any useful technique to have many different applications, but for simplicity we could divide the types of problem we typically consider into two general classes:

Types of problems

- (a) *Topological* (e.g. properties of closed orbits). As a specific example we could consider results on the distribution of closed orbits which reflect information on their free homotopy classes (i.e., conjugacy classes in the fundamental group). The Bowen–Series coding makes it easy to relate the closed orbits for the flow to periodic orbits $\{x, Tx, \dots, T^{m-1}x\}$ (with

$T^n x = x$) for the associated transformation. Furthermore, the usefulness of this particular coding is that the word length of the corresponding geodesic will be n , except in a finite number of exceptional cases.

- (b) *Measure theoretic* (e.g. properties of the measures μ and ν). The two measures are intimately related by the classical work of Bowen–Ruelle. For example, showing ergodicity for the flow automatically implies ergodicity for the discrete map. In the reverse direction, ergodicity for the discrete map (plus some modest hypothesis on the roof function) implies ergodicity for the flow. Stronger properties such as strong mixing, mixing rates, central limit theorems, etc. can be considered in each context, with varying degrees of complexity for the correspondence between them.

Since we are interested in ergodic theory and dynamical systems it is natural to use the model of the geodesic flow using expanding interval maps to try to understand its dynamics. However, in the interests of being open minded, we should also ask:

Question 3. Is this the best approach?

Since the word “best” is a little subjective there is no real definitive answer. However, the somewhat equivocal answer is:

Sometimes... But, of course, this depends on what we are interested in and what we mean by the question. For example if we are interested in closed geodesics on the surface then these can be viewed dynamically as closed orbits for the geodesic flow. However, in some cases their study can be advanced by other (less dynamical) methods. For example:

1. For V a compact surface of constant negative curvature there are already powerful techniques (e.g. trace formulae, representation theory) which can often give more precise results.
2. But for V being a non-compact surface of infinite area these methods may not apply so well, but the dynamical approach often still applies. At a certain level this can be thought of as being because the dynamical method uses only the compact recurrent set of the geodesic flow and $T : X \rightarrow X$ with X retaining its compactness.

Let us recall a classical example of a compact surface.

Example 1. *Let V be an oriented compact surface of genus $g \geq 2$. This surface not only supports a metric of constant curvature $\kappa = -1$, but the space of such metrics is $(6g - 6)$ -dimensional.*

Let us now consider a couple of examples of infinite area surfaces with constant curvature $\kappa = -1$ to illustrate the last point.

Example 2. *A three funneled surface (or “a pair of pants”). This is homeomorphic to a sphere with three disjoint closed disks removed. It carries metrics with constant negative curvature $\kappa = -1$. We can either consider the complete surface which has three funnels (or ends) or alternatively we can cut along three closed geodesics around each of the funnels to get a surface with three boundary components (as in the figure 3). In point of fact, from a dynamical viewpoint the distinction is unimportant. For the geodesic flow on either version the important dynamical component is the recurrent part of the geodesic flow.*

For the surface with three funnels the recurrent part of the geodesic flow on the 3-dimensional unit tangent bundle is a compact set whose intersection with any transverse two-dimensional set is a Cantor set.

However, none of the orbits can cross any of the three geodesics around the funnels and so the corresponding set appears on the pair of pants.

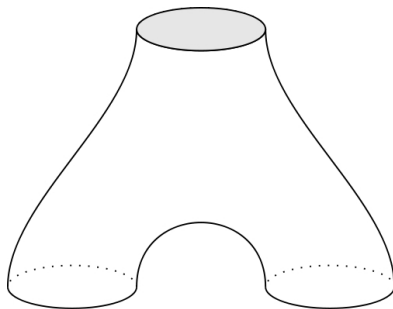


Figure 3: A pair of pants. This is an artistic impression. According to the Hilbert Theorem, no complete surface of constant negative Gaussian curvature can be immersed in \mathbb{R}^3

Example 3. *A punctured torus (or a “torus with a hole”) is a surface homeomorphic to a torus with a closed disk removed. It carries a metric with constant negative curvature $\kappa = -1$. We can either consider the complete surface which has a single funnel (or end) or alternatively we can cut along the closed geodesic around the funnel to get a surface with a single boundary component (as in Figure 4). As in Example 2, from a dynamical viewpoint the distinction is unimportant. For the geodesic flow on either version the important dynamical component is the recurrent part of the geodesic flow. For a one holed torus the recurrent part of the geodesic flow on the three dimensional unit tangent bundle is a compact set whose intersection with any transverse section to two-dimensional set is a Cantor set. However, none of the orbits in the recurrent part can cross the geodesic around the funnel.*

Remark 1. The approach of Bowen and Series is very successful in the case of surfaces. Unfortunately, it does not naturally generalise to higher dimensions.



Figure 4: A one-holed torus. This is an artistic impression. Again there is no immersion in \mathbb{R}^3 by the Hilbert Theorem.

3 Hyperbolic Geometry

In this section we want to introduce the basic setting in which we will be working. This involves recalling some basic definitions and notions in hyperbolic geometry. Hyperbolic geometry was developed in the 19th century by Gauss and Bolyai¹ to show that the 5th postulate of Euclid was not implied by the others.

There are many different models for the hyperbolic space, but we will concentrate on the disk model. We begin with the basic definitions, and refer the reader to the book of Beardon [3] for more details.

Definition 1. Let

$$\mathbb{D}^2 = \{z = x + iy \in \mathbb{C} : |z| < 1\}$$

be the open unit disk in the complex plane. We can equip this with the *Poincaré metric* defined locally by

$$ds^2 = \frac{4(dx^2 + dy^2)}{(1 - x^2 - y^2)^2}.$$

For example, if $z_1, z_2 \in \mathbb{D}$ then

$$d(z_1, z_2) = 2 \tanh^{-1} \left| \frac{z_1 - z_2}{1 - z_1 \bar{z}_2} \right|.$$

This defines a complete Riemannian metric on \mathbb{D}^2 . The factor of 4 in the definition of the metric is ensuring a convenient normalisation for the curvature. The Poincaré metric has a number of useful properties which we briefly summarize as follows.

¹This geometry was independently developed by a Russian mathematician N. I. Lobachevski, and carries his name in post-Soviet countries.

Lemma 3. *Consider the unit disk (\mathbb{D}^2, ds) with the Poincaré metric then:*

1. *the space has Gaussian curvature $\kappa = -1$; and*
2. *the orientation preserving isometries take the special form of linear fractional transformations*

$$\begin{cases} \mathbf{g}: \mathbb{D}^2 \rightarrow \mathbb{D}^2 \\ \mathbf{g}(z) = \frac{az+b}{bz+\bar{a}} \end{cases} \quad \text{with } |a|^2 - |b|^2 = 1, (a, b \in \mathbb{C}).$$

There is some ambiguity with respect to the simultaneous sign change $a, b \longleftrightarrow -a, -b$ which one has to bear in mind.

There are other equivalent models for hyperbolic plane, such as the Poincaré upper half plane model. However the symmetry in the disk model is particularly useful in what follows in later sections.

Given a compact surface V with negative curvature $\kappa = -1$, the Poincaré metric on the unit disk has a particular significance. This can be explained as follows:

Lemma 4. *Let V be a compact surface of constant curvature $\kappa = -1$.*

1. *Then by the Gauss–Bonnet theorem we can deduce that the genus g of V satisfies $g \geq 2$. Equivalently, the Euler characteristic is strictly negative.*
2. *In particular, since $g \geq 2$ we have that the universal cover of V is \mathbb{D}^2 and the lifted metric is the Poincaré metric cf. [3].*
3. *The geodesics on V (locally distance minimizing) lift to geodesics on \mathbb{D}^2 . The latter are semi-circular arcs which meet the unit circle $\partial\mathbb{D}^2$ orthogonally.*

We have identified in Lemma 3 the isometries for the Poincaré metric. We next want to consider discrete subgroups. This brings us onto the next topic.

4 Fuchsian groups and the Bowen–Series map

The next ingredient to consider is a suitable subgroup of the isometries of the Poincaré disk with respect to the Poincaré metric. Given a compact surface of constant curvature $\kappa = -1$, we can write $V = \mathbb{D}^2/\Gamma$ where $\Gamma < \text{Isom}(\mathbb{D}^2, ds^2)$ is a discrete subgroup isomorphic to the Fundamental Group (i.e., $\Gamma = \pi_1(V)$).

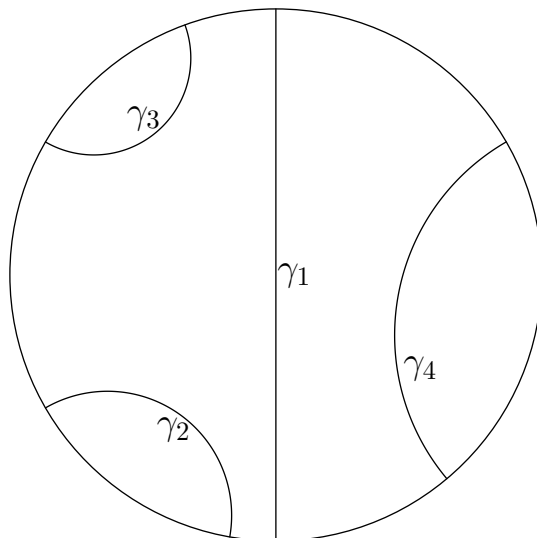


Figure 5: The disk model of the hyperbolic plane. The geodesics are either diameters or semicircular arcs which meet the boundary at right angles.

4.1 Fuchsian groups

We recall the following standard definition of a Fuchsian group of isometries.

Definition 2. We call $\Gamma < \text{Isom}(\mathbb{D}^2, ds^2)$ a *Fuchsian group* if it is a discrete subgroup. In other words, the orbit of the point $0 \in \mathbb{D}^2$ under the Fuchsian group Γ , denoted $\Gamma 0$, doesn't accumulate inside the disk. Moreover, since $\mathbf{g} \in \Gamma$ is a linear fractional transformation, we see that it extends to the circle $\partial\mathbb{D}^2$ (i.e., $\mathbf{g}: \partial\mathbb{D}^2 \rightarrow \partial\mathbb{D}^2$).

We can also associate to the Fuchsian group Γ an important closed subset of the boundary circle $\partial\mathbb{D}$ called the limit set.

Definition 3. The limit set $\Lambda \subset \partial\mathbb{D}^2$ consists of Euclidean accumulation points of the orbit of the centre 0 , i.e,

$$\Lambda = \overline{\Gamma 0} - \Gamma 0.$$

It is not essential to take the orbit of 0 ; any point in the interior of \mathbb{D}^2 would do. The limit set is always compact. Moreover we know the following:

Lemma 5. *There are two possibilities: Either*

1. *the limit set is the entire boundary circle (i.e., $\Lambda = \partial\mathbb{D}^2$); or*
2. *the limit set is a Cantor set of Hausdorff dimension $0 < \delta < 1$.*

These are usually referred to as limit sets of type I and type II, respectively.

For the moment, we will concentrate on the case where V is a compact surface and the associated Fuchsian group Γ is called co-compact. In this case the limit set is the entire circle. The elements of the Fuchsian group Γ have a number of useful and important properties.

Lemma 6. *Let V be a compact surface of constant negative curvature.*

1. *For all $\mathfrak{g} \in \Gamma \setminus \{e\}$ there are 2 fixed points $x_{\mathfrak{g}}^-, x_{\mathfrak{g}}^+ \in \partial\mathbb{D}^2$. We let $x_{\mathfrak{g}}^+$ denote the attracting fixed point, i.e., $|\mathfrak{g}'(x_{\mathfrak{g}}^+)| < 1$. Similarly, we let $x_{\mathfrak{g}}^-$ denote the repelling fixed point, i.e., $|\mathfrak{g}'(x_{\mathfrak{g}}^-)| > 1$.*
2. *Γ is finitely generated and finitely presented:*

$$\Gamma = \langle a_1, b_1, \dots, a_{2g}, b_{2g}: (a_1 b_1 a_2^{-1} b_2^{-1}) \dots (a_{2g-1} b_{2g-1} a_{2g}^{-1} b_{2g}^{-1}) = 1 \rangle.$$

(This is the standard presentation.)

4.2 Bowen–Series map

Given a Fuchsian group we want to associate a piecewise continuous expanding Markov map called the *Bowen–Series map* defined on the boundary $\partial\mathbb{D}^2$.

To construct the map $T: \partial\mathbb{D}^2 \rightarrow \partial\mathbb{D}^2$ we take the natural generators a_i, b_i , ($i = 1, \dots, 2g$) and their inverses in Γ . We first describe a partition of the circle into subarcs and then we want to define a map which takes the form of the action of one of these $2g$ maps (or their inverses) on the unit circle. We briefly describe how this construction works.

- We associate to each generator \mathfrak{g} the *isometric circle*:

$$C(\mathfrak{g}) = \{z \in \mathbb{D}^2 \mid |\mathfrak{g}'(z)| = 1\} \quad (\mathfrak{g} = a_i, b_i, i = 1, \dots, 2g),$$

The name isometric here doesn't refer to the Poincaré metric, of course, since \mathfrak{g} is an isometry on \mathbb{D}^2 . However, each of the curves $C(\mathfrak{g})$ is actually a geodesic in the Poincaré metric. Observe that if $\mathfrak{g}(z) = \frac{az+b}{bz+\bar{a}}$ then $\mathfrak{g}'(z) = (\bar{b}z + \bar{a})^{-2}$. Thus the $C(\mathfrak{g})$ are parts of circular arcs and are actually geodesics. These have a simple property: $\mathfrak{g}C(\mathfrak{g}) = C(\mathfrak{g}^{-1})$. The region in the disk \mathbb{D}^2 “exterior” to all of the isometric circles is a Fundamental Domain for the Fuchsian group.

- We divide up $\partial\mathbb{D}^2$ into arcs (intervals) using the $8g$ endpoints of the $4g$ isometric circles.

In point of fact it is convenient to make an additional assumption on the way the surface group acts on the hyperbolic plane:

Assumption. The tessellation associated to the union of the isometric circles $C(\mathfrak{g})$, with $\mathfrak{g} \in \Gamma$, is preserved by the the action of $\Gamma = \pi_1(V)$ on \mathbb{D}^2 .

This ensures that the boundary map is Markov.

Remark 2. In many cases this assumption may not hold, but then another choice of generators can be made to ensure that this property holds. The boundary map is then Markov (on $(8g-4)$ intervals). This alternative choice was written down independently by both Series and Adler–Flatto [2]. However, in the interests of clarity and historical reverence we will use the original construction of Bowen and Series [6].

The configuration of the isometric circles takes the form illustrated in Figure 6 below (taken from an article of Series).

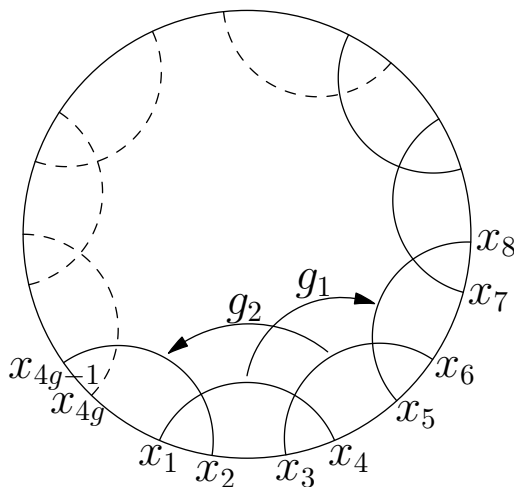


Figure 6: The boundary of the fundamental domain is given by isometric circles. Each of these may intersect with its direct neighbours, but no more.

We can now use the $4g$ points on the circle to divide it into the same number of arcs, and associate a piecewise analytic map to each of these arcs. This gives the required map of the interval (where the arcs are thought of as intervals, in the usual way). More precisely, we introduce a numbering on the generators of the group Γ as follows $\mathfrak{g}_{2k+1} = a_k^+$, $\mathfrak{g}_{2k+2} = a_k^-$, $\mathfrak{g}_{2k+2g+1} = b_k^+$, and $\mathfrak{g}_{2k+2g+2} = b_k^-$; for $k = 1, \dots, g$. Then we define

$$T(x) = \mathfrak{g}_k(x), \text{ if } x \in [x_{2k}, x_{2k+1 \bmod 4g}] \text{ or } x \in [x_{2k-1}, x_{2k}], \text{ for } k = 1, 2, \dots, 2g. \quad (1)$$

We summarise the main properties of the interval map T in the following lemma.

Lemma 7. *For the interval map T we have the following properties.*

1. T is C^ω (real analytic) on each arc;
2. T is expanding (i.e. $\exists \beta > 1 \ |T'(z)| \geq \beta \ \forall z \in \partial\mathbb{D}^2$);
3. T is Markov (i.e. image of each arc in the partition is a union of arcs in the partition).

These properties are very classical and lead to a clear understanding of the dynamics of the map. Moreover, these can be used to understand the action of the Fuchsian group Γ on the circle $\partial\mathbb{D}^2$. In particular, we would like to replace the action

$$\Gamma \times \partial\mathbb{D}^2 \rightarrow \partial\mathbb{D}^2 \quad (g, x) \mapsto g(x)$$

by a single transformation $T: \partial\mathbb{D}^2 \rightarrow \partial\mathbb{D}^2$ (up to finitely many points).

Lemma 8. *We can denote by Γx the orbit of x under the action of the group, i.e., $\Gamma x = \{gx : g \in \Gamma\}$. Then for a typical x the orbit Γx can be written as an equivalence class*

$$\Gamma x = \{y \mid \exists n, m \geq 0 : T^n x = T^m y\},$$

where T is defined by (1).

Example 4. *Consider a regular right-angled hyperbolic octagon. We can label each of the edges in sequence as $a^+, b^+, a^-, b^-, c^+, d^+, c^-, d^-$. Then we may consider the orientation preserving isometries Γ_0 of the hyperbolic plane which identify sides which are labelled by symbols and their inverses. Let Γ be the group generated by Γ_0 . The symmetry of the fundamental domain ensures that there are 16 arcs (of continuity) in the Bowen–Series map, moreover eight of them have one length and another eight have a second length.*

5 Dynamical Applications: Compact surfaces

We now return to the study of the dynamical properties of the geodesic flows, particularly those which can be approached via the perspective of the associated Bowen–Series map T on the unit circle (or more accurately, on a disjoint union of intervals). We would like to highlight two particularly well known properties.

5.1 Counting closed geodesics

We begin with a topological result. Let V be a compact surface with constant curvature. It is easy to see that there are infinitely many closed geodesics since each free homotopy class (i.e., conjugacy classes) in $\pi_1(V)$ contains a closed geodesic. Furthermore, there are countably many closed geodesics since each free homotopy class for a negatively curved surface contains precisely one closed geodesic. By a closed geodesic we mean the directed closed geodesics, i.e., each curve considered as a set actually corresponds to two directed geodesics, which differ by their orientation.

Let γ denote a (primitive) closed geodesic of length $\ell(\gamma)$. These are in bijection with the closed orbits of the geodesics flow of the same length.

Definition 4. For $t > 0$ we denote by

$$N(t) = \#\{\gamma : \ell(\gamma) \leq t\}$$

the number of closed geodesics whose length is at most t . It is well known that $N(t)$ is finite.

Now we recall the following well known properties.

Lemma 9. *For a compact negatively curved surface we have:*

1. $N(t)$ is monotone increasing; and

$$2. \lim_{t \rightarrow \infty} \frac{1}{t} \log N(t) = 1.$$

This is one of many ways of establishing the topological entropy for the geodesic flow, i. e. the growth rate of $N(t)$ equals 1.

In fact the second part follows from the definition of topological entropy. A much stronger result than Part (2) of Lemma 9 is the following asymptotic.

Theorem 1 (Huber).

$$N(t) \sim \frac{e^t}{t} \text{ as } t \rightarrow \infty, \text{ i.e. } \lim_{t \rightarrow \infty} \frac{N(t)}{e^t} \cdot t = 1.$$

The natural proof of Huber's theorem uses the Selberg zeta function $Z_V(s)$ and the location of its zeros. We recall the definition.

Definition 5. We formally define the Selberg zeta function as an infinite product

$$Z_V(s) = \prod_{n=0}^{\infty} \prod_{\gamma} (1 - e^{-(s+n)\ell(\gamma)}), \quad s \in \mathbb{C}, \quad (2)$$

where γ denotes a closed geodesic on V of length $\ell(\gamma)$. The function is well defined, as the product converges to an analytic function for $\Re(s) > 1$.

The convergence of the infinite product from (2) on $\Re(s) > 1$ follows from the estimate in Lemma 9. The asymptotic estimates for the counting function $N(t)$ follows completely by analogy with the Prime Number Theorem for primes and the zeta function $Z_V(s)$ plays an analogous to the Riemann ζ -function.

The zeta function $Z_V(s)$ formally defined on $\Re(s) > 1$ can be extended to the entire complex plane using the famous Selberg trace formula [14]. This is explained in the book of Hejhal [12] in considerable detail. We recall the basic properties:

Lemma 10. *Let V be compact surface of constant negative curvature.*

1. $Z_V(s)$ analytic and non-zero for $\Re(s) > 1$;
2. $Z_V(s)$ has a simple zero at $s = 1$;
3. $Z_V(s)$ has an analytic extension to \mathbb{C} .

Remark 3. Margulis gave a generalisation of Huber's theorem to the case of surfaces of variable negative curvature. Margulis' approach to Huber's theorem (for $\kappa < 0$) uses strong mixing of a suitable invariant measure (nowadays called the Bowen–Margulis measure) which maximises the entropy. The approach introduced by Margulis has now been adapted to a number of interesting generalizations.

We want to consider error terms in the basic asymptotic stated in Theorem 1. To put this into perspective we need to recall what might happen in the corresponding setting in number theory.

5.2 Taking a leaf out of the number theory book

Let us begin by recalling some classical ideas from number theory. These are useful in understanding the role of the dynamical analogues.

Riemann ζ -function. We can recall the definition of the Riemann ζ -function for primes:

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p (1 - p^{-s})^{-1}$$

where the Euler product is over the prime numbers. Below are some of the properties of this complex function.

Lemma 11. *Let $\zeta(s)$ be the Riemann zeta function. Then*

1. $\zeta(s)$ is analytic and non-zero for $\Re(s) > 1$;
2. $\zeta(s)$ has a simple pole at $s = 1$;
3. $\zeta(s)$ has analytic extension to $\mathbb{C} \setminus \{1\}$.

Of course there is additionally another property which has (famously) not yet been established.

Riemann hypothesis (1859): The zeros of $\zeta(s)$ in $\Re(s) > 0$ only lie on a line $\Re(s) = \frac{1}{2}$.

This question is also known as Hilbert's 8'th Problem and a Clay Institute Millennium Problem. There have been many attempts to prove this result, but so far the conjecture has resisted attempts. One particularly popular approach is the following:

Hilbert–Polya approach: Try to relate the zeros of $\zeta(s)$ to eigenvalues of some self-adjoint operator.

The motivation for this is because the spectrum of a self-adjoint operator is necessarily contained in \mathbb{R} . Whereas this has proved elusive for the Riemann zeta function, the analogue for the Selberg zeta function has been more successful.

5.3 Error terms in counting closed geodesics

There is an improvement to Huber's basic Theorem 1 at least as stated. In fact, the original result of Huber included error terms, so now we are actually stating more of his original theorem.

Theorem 2 (Huber). *Let V be a compact surface with $\kappa = -1$. There exists $\varepsilon > 0$ such that*

$$N(t) = \int_2^{e^t} \frac{1}{\log u} du (1 + O(e^{-\varepsilon t})) = \text{li}(e^t) (1 + o(e^{-\varepsilon t})) \quad \text{as } t \rightarrow +\infty.$$

In other words, there exist $\varepsilon > 0$ and $C > 0$ such that $\left| \frac{N(t)}{\text{li}(e^t)} - 1 \right| \leq C e^{-\varepsilon t}$.

Note that ε here is responsible for the rate of convergence, rather than “just” for the error term itself.

As is well known,

$$\operatorname{li}(x) := \int_2^x \frac{du}{\log u} \sim \frac{x}{\log x} \text{ as } x \rightarrow +\infty.$$

Therefore this statement is consistent with the original statement of Theorem 1, but now has an additional error term.

The proof uses properties of Z_V and the rate $\varepsilon > 0$ depends on location of zeros. This is completely analogous to the situation in number theory where one studies a counting problem for prime numbers by using the Riemann zeta function, except in the present context we have stronger results on the zeta function.

For compact surfaces there is a natural self-adjoint operator. The extension and zeros of Selberg zeta function are related to *the Laplacian* $\Delta: \mathcal{L}^2(V) \rightarrow \mathcal{L}^2(V)$. This can be defined as an operator on real analytic functions $\Delta: C^\infty(V) \rightarrow C^\infty(V)$ and then extended to square integrable functions. Let

$$\Delta\psi_n + \lambda_n\psi_n = 0$$

be the eigenvalue equation. There are infinitely many eigenvalues

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots \nearrow +\infty$$

for the operator $-\Delta$. Moreover, we have the following results:

1. $\#\{\lambda_n \leq t\} \sim \frac{t}{4\pi} \operatorname{Area}(V)$ as $t \rightarrow \infty$ (Weyl, 1911);
2. The zeros of $Z_V(s)$ in $0 < \Re(s) < 1$ are of form $s_n = \sqrt{\frac{1}{4} - \lambda_n}$, where λ_n are eigenvalues of the Laplacian [26].

The second statement implies that $\{s_n\}$ lie on $[0, 1] \cup \{\Re(s) = \frac{1}{2}\}$. We can formulate this as follows:

Corollary 1. *Let V be a compact surface and let s_n be a zero $Z_V(s_n) = 0$. Then either $s_n \in [0, 1]$ or $\Re(s_n) = \frac{1}{2}$.*

Warning: This approach fails if V has infinite area. Thus there is no reason for zeros to lie on such lines — and indeed this is not the case as we shall see later in examples.

The value $\varepsilon > 0$ for the error term for counting closed orbits in Theorem 2 and in exponential mixing can be expressed in terms of the eigenvalues of the Laplacian.

The proof uses trace formulae to relate the zeros of Z_V to the eigenvalues of the Laplacian. It is the self-adjointness of the Laplacian which ultimately leads to this property; The value $\varepsilon > 0$ corresponds to the distance of the zeros of $Z_V(s)$ from the line $\Re(s) = \frac{1}{2}$ (although one may need to take ε slightly smaller).

This can be thought of as an analogue of the “Riemann Hypothesis”.

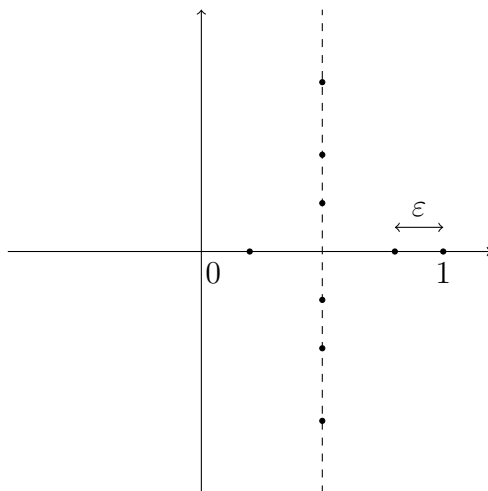


Figure 7: The zeros of the Selberg zeta function lie on $[0, 1] \cup \{s : \Re(s) = \frac{1}{2}\}$

Remark 4. The value of $\varepsilon > 0$ can be arbitrary small. This follows from its spectral interpretation (in terms of the Laplacian) and corresponds to having eigenvalues close to 0. Using classical results of Cheeger [7], Schoen–Wolpert–Yau [25] one can show that $\varepsilon \asymp \ell(\gamma_0)$ where $\ell(\gamma_0)$ is the length of the shortest closed dividing geodesic.

In the context of non-compact surfaces we need an alternative approach using the Bowen–Series map. In particular, the Bowen–Series map can be used to give a correspondence between closed geodesics and closed orbits. More precisely,

Lemma 12. *There is a bijection between*

1. *Closed (prime) geodesics γ ; and*
2. *Closed (prime) orbits $\{x, Tx, \dots, T^{n-1}x \mid T^n x = x\}$, where T is defined by (1),*

except for a finite number of geodesics corresponding to the endpoints.

A geodesic is called prime if it traces its path exactly once. A prime periodic orbit is one for which $T^k x \neq x$ for any $1 < k < n$.

An advantage of the Bowen–Series construction is that the map T defined by (1) can be used to recover the lengths of closed geodesics. More precisely,

Lemma 13. *If γ is a closed geodesic of length $\ell(\gamma)$ with associated periodic T -orbit*

$$\{x, Tx, \dots, T^{n-1}x \mid T^n x = x\},$$

then the following holds $\ell(\gamma) = \log |(T^n)'(x)|$.

In particular, we can interpret the length of γ as the Lyapunov exponent of the periodic orbit, which gives some clue to how to prove the Lemma.

To take advantage of this correspondence to study zeta function Z_V , we introduce the family of complex Ruelle–Perron–Frobenius transfer operators (indexed by $s \in \mathbb{C}$). This might be referred to as the “Thermodynamic Viewpoint” [20]. We begin by recalling the definition.

Definition 6. Given an infinite disjoint union of arcs X , we can define the *transfer operators* $\mathcal{L}_s: C^0(X) \rightarrow C^0(X)$ on a space of continuous functions by

$$\mathcal{L}_s f(x) = \sum_{Ty=x} e^{-s \log |T'(y)|} f(y),$$

for each $s \in \mathbb{C}$.

Remark 5. When $s = 1$, this reduces to the “usual” Ruelle–Perron–Frobenius operator for an expanding interval map.

However, in order to proceed we need to replace the large space $C^0(X)$ by a smaller Banach space upon which the transfer operator has good spectral estimates.

5.4 The Banach spaces

We now want to be more specific about the spaces involved. Let $\{I_j\}$ denote the arcs in the Bowen–Series coding.

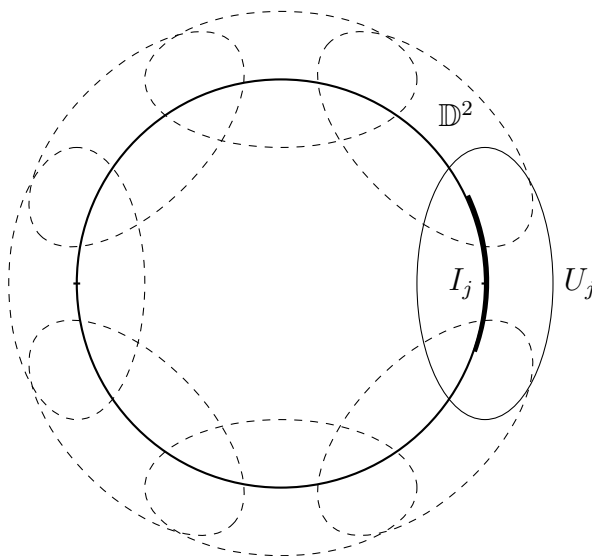


Figure 8: Each arc I_j in the boundary $\partial\mathbb{D}^2$ has a neighbourhood U_j .

1. Choose (small) open neighbourhoods $I_j \subset U_j \subset \mathbb{C}$ such that for any $I_j \subset T(I_k)$ we have that $\bar{U}_j \subset T(U_k)$ (i.e., the closure of the images of the neighbourhoods under the contractions are contained in the corresponding open sets);
2. Let $B \subset C^0(X)$ be the Banach space of bounded analytic functions on the (disjoint) union $\sqcup_k U_k$ with the supremum norm

$$\|f\| = \sup_{z \in \sqcup_k U_k} |f(z)|;$$

3. By construction we see that the Banach space is preserved by any transfer operator, i.e., $\mathcal{L}_s: B \rightarrow B$ for all $s \in \mathbb{C}$;
4. On the Banach space B we have that \mathcal{L}_s is trace class; in other words, there exists only countable many non-zero eigenvalues and their sum is finite²;
5. In particular, we can now observe that the zeta function defined by (2) satisfies

$$Z_V(s) = \det(I - \mathcal{L}_s) = \exp\left(-\sum_{n=1}^{\infty} \frac{1}{n} \text{trace}(\mathcal{L}_s^n)\right)$$

and the latter converges to an analytic function for all $s \in \mathbb{C}$; and finally

6. We have the classical observation: $Z_V(s) = 0$ if and only if there exists a non-zero function $f \in C^\omega(V)$ such that $\mathcal{L}_s f = f$.

We will return to the use of Banach spaces later in the special case of infinite area surfaces. In this case, the compactness enters via the limit set Λ .

5.5 Mixing of geodesic flow

We next consider a closely related problem in the general area of smooth ergodic theory. We begin with a definition.

Definition 7. Given two smooth functions $F, G: Y \rightarrow \mathbb{R}$ and a flow φ_t , we define a correlation function:

$$\rho(t) = \int_Y F(\varphi_t x) G(x) d\nu(x) - \int_Y F d\nu \int_Y G d\nu, \quad t \geq 0,$$

where ν denotes (normalised) Liouville measure on Y . We say that the flow is mixing if $\rho(t) \rightarrow 0$ as $t \rightarrow +\infty$. The rate of convergence in this case is referred to as *the rate of mixing*.

²This follows directly from the work of Grothendieck [10], [11] and Ruelle [24]. We also refer the reader to [13].

In the case of constant negative curvature the Liouville measure corresponds to the normalised Haar measure.

Mixing is a property in the hierarchy of ergodic properties. It implies ergodicity and is in turn implied by the Bernoulli property. In the case of integrable functions $F, G \in \mathcal{L}^2(\nu)$ there is no reason to expect to be able to say anything about the speed of mixing (i.e., the rate at which $\rho(t)$ tends to zero). But since we are assuming F and G are smooth more can be said.

Of course the definition can also be reformulated for different invariant measures, but for our present purposes we only need to consider the measure ν .

The following result dates back over sixty years and was the forerunner of method that has proved to be very successful over subsequent years.

Theorem 3 (Fomin–Gelfand). *The geodesic flow φ_t is strong mixing, in other words $\rho(t) \rightarrow 0$ as $t \rightarrow +\infty$.*

The proof of Theorem 3 uses unitary representations of $SL(2, \mathbb{R})$ on the space $\mathcal{L}^2(Y)$ given by

$$U_{\mathfrak{g}}f(x) = f(\mathfrak{g}^{-1}x).$$

The representation of $U_{\mathfrak{g}}$ is reducible and therefore the problem of mixing is reduced to the understanding of the explicit operators on each of the component representations. This brings us to a basic question:

Question 4. Can we estimate the speed of convergence $\rho(t) \rightarrow 0$ or improve on these results in any way?

The following results on mixing don't directly depend on the zeta function, but the underlying mechanisms are similar as will hopefully soon become apparent.

Theorem 4. *Let V be a compact surface of constant curvature $\kappa = -1$. Let $F, G \in C^\infty(X)$. There exists $\varepsilon > 0, C > 0$ such that*

$$|\rho(t)| \leq Ce^{-\varepsilon t} \quad \text{for } t \geq 0,$$

where C depends on F and G , but ε depends only on the geodesic flow φ_t .

This follows from closer look at representation theory approach. In fact the correlation functions correspond to “decay of matrix coefficients” in representation theory [16].

We have used the same constant $\varepsilon > 0$ for both results, in the counting geodesics problem Theorem 2, and for the speed of mixing of geodesic flow in Theorem 4. Assuming this wasn't carelessness, one might ask the following question:

Question 5. How are the ε 's related in two problems?

The simple answer is: “*They are the same!*”. To see this, we can associate to $\rho(t)$ the Laplace transform

$$\hat{\rho}(z) = \int_0^\infty e^{-zt} \rho(t) dt, \quad z \in \mathbb{C}.$$

Thus we are introducing another complex function to describe the mixing rate. The definition makes sense for $\Re(z) > 0$ since the integral converges on this half-plane. However, there is much more that one can say. We recall the following properties of $\hat{\rho}(z)$.

Lemma 14. *The Laplace transform $\hat{\rho}(z)$ has the following properties:*

1. *The function $\hat{\rho}$ extends meromorphically to \mathbb{C} ;*
2. *Poles z_n for $\hat{\rho}$ are related to zeros s_n for the zeta function Z_V by $z_n = s_n - 1$; and*
3. *The absence of poles for $\hat{\rho}(z)$ for a large half-plane $\Re(z) > -\varepsilon$ implies exponential decay of ρ .*

The existence of meromorphic extension of $\hat{\rho}$ can be obtained by establishing a connection between the function $\hat{\rho}$ and the resolvent of the transfer operators and using properties of the spectrum of the transfer operator. The second part follows from the observation that Z_V has a similar connection to the spectra of essentially the same transfer operators. The last part is classical harmonic analysis using the Paley–Wiener theorem.

Remark 6. The poles $\{z_n\}$ (or sometimes s_n) are usually called “resonances”.

6 The Bowen–Series map for infinite area surfaces

Henceforth, we now want to consider the case of infinite area surfaces V . However, we will only consider the case that V has no cusps (or equivalently the corresponding Fuchsian group Γ has no parabolic points).

We want to return to the boundary map associated to the action of the group on the limit set. Recall that in this case the limit set Λ is a Cantor set. But the same basic approach applies — except that it is even easier. (This time we have an expanding map $T: \Lambda \rightarrow \Lambda$ on the Cantor set).

This is best illustrated by an example. We recall our main example of the three funnelled surface.

Example 5 (Three funnelled surface). *Let V be the three funnelled surface described in Example 2. We can write $V = \mathbb{D}^2/\Gamma$ where Γ is the associated Fuchsian group of isometries. In this particular case, we can write $\Gamma = \langle a, b \rangle$ as a free group on two isometries $a, b \in \text{Isom}(\mathbb{D}^2)$.*

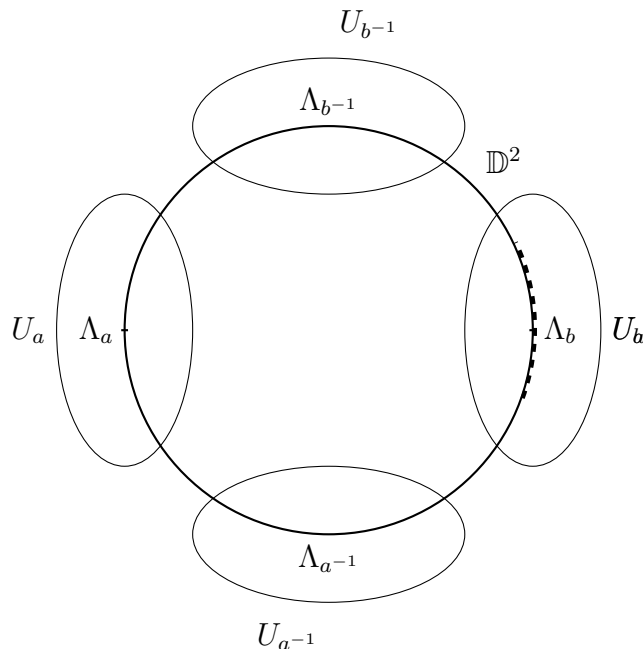


Figure 9: Partition of the Cantor set and the neighbourhoods of the elements.

Since V is an infinite area surface the limit set Λ associated to Γ is a Cantor set (cf. Lemma 5). We can naturally partition it into four disjoint pieces as follows. Let $\Lambda_{\mathfrak{g}}$ be the part of Λ under $C(\mathfrak{g})$ and thus in $U_{\mathfrak{g}}$ for $\mathfrak{g} \in \{a^{\pm 1}, b^{\pm 1}\}$. Then we have a simple partition

$$\Lambda = \Lambda_a \cup \Lambda_{a^{-1}} \cup \Lambda_b \cup \Lambda_{b^{-1}},$$

where $\Lambda_{\mathfrak{g}} = \bigcup_{\mathfrak{g}' \in \{a^{\pm 1}, b^{\pm 1}\} \setminus \mathfrak{g}} \Lambda_{\mathfrak{g}'}$, for $\mathfrak{g} \in \{a^{\pm 1}, b^{\pm 1}\}$.

We can then define a map $T: \Lambda \rightarrow \Lambda$ by

$$T(x) = \begin{cases} ax & \text{if } x \in \Lambda_a \\ a^{-1}x & \text{if } x \in \Lambda_{a^{-1}} \\ bx & \text{if } x \in \Lambda_b \\ b^{-1}x & \text{if } x \in \Lambda_{b^{-1}}. \end{cases} \quad (3)$$

It is easy to see that this is strictly expanding (i.e., $|T'(x)| > 1$ for $x \in \Lambda$) and Markov (in the sense that the image $T(\Lambda_{\mathfrak{g}})$ is the union of three of the elements of the partition, i.e. $\Lambda_{\mathfrak{g}} = \bigcup_{\mathfrak{g}' \in \{a^{\pm 1}, b^{\pm 1}\} \setminus \mathfrak{g}} \Lambda_{\mathfrak{g}'}$).

As in the case of compact surfaces we can associate a function $\tau: \Lambda \rightarrow \mathbb{R}^+$ defined by

$$\tau(x) = \log |T'(x)|. \quad (4)$$

In Lemma 5 we introduced Hausdorff dimension of the limit $\Lambda \subset \partial \mathbb{D}^2$. It turns out we can write δ in terms of the transformation $T: \Lambda \rightarrow \Lambda$ and the function $\tau: \Lambda \rightarrow \mathbb{R}$. In order to formulate this we need the following definition of a pressure functional.

Definition 8. We define the *pressure functional* $P : C(\Lambda) \rightarrow \mathbb{R}$ by

$$P(f) := \lim_{n \rightarrow \infty} \frac{1}{n} \log \left(\sum_{T^n x = x} \exp \left(\sum_{j=0}^{n-1} f(T^j x) \right) \right).$$

Equivalently, we can define the pressure using the variational principle:

$$P(f) := \sup \left\{ h_{\text{top}}(\mu) + \int f d\mu \mid \mu \text{ is a } T\text{-invariant probability measure.} \right\}$$

We now have the following promised characterisation of $\delta = \dim_H \Lambda$ in terms of the functions on the boundary.

Lemma 15. *The Hausdorff dimension $\delta = \dim_H \Lambda$ satisfies $P(-\delta\tau) = 0$, where τ is defined by (4).*

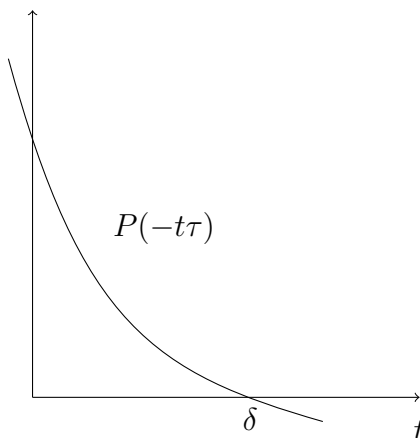


Figure 10: A typical plot of the pressure function.

By analogy with the case of compact surfaces, we can introduce a family of complex Ruelle–Perron–Frobenius transfer operators ($s \in \mathbb{C}$) initially defined on continuous³ functions $C^0(\Lambda)$ on the limit set Λ .

Definition 9. We can define the *transfer operators* $\mathcal{L}_s : C^0(\Lambda) \rightarrow C^0(\Lambda)$ by

$$\mathcal{L}_s f(x) = \sum_{Ty=x} e^{-s \log |T'(y)|} f(y), \quad (5)$$

for each $s \in \mathbb{C}$.

This gives us another interpretation of the value δ .

Lemma 16. *The operator \mathcal{L}_δ has maximal eigenvalue 1.*

To study the zeta function Z_V using the transfer operators we need to replace $C^0(\Lambda)$ by a smaller Banach space upon which the transfer operator has good spectral estimates (see §7.2).

³in the induced topology on $\partial\mathbb{D}^2$.

7 Dynamical applications: non-compact surfaces

We want to find analogues of the counting and mixing results from Theorems 2 and 4, respectively, in the case that we have a surface of infinite area. We begin with the counting problem for closed geodesics.

7.1 Counting closed geodesics

As in the case of compact surfaces, one has in the non-compact case that there are infinitely many closed geodesics (with one in each free homotopy class). The obvious question is the following.

Question 6. What are the analogous results on the asymptotic behaviour of the number of closed geodesics in non-compact case?

Assume that $V = \mathbb{D}^2/\Gamma$ is a surface of infinite area, where Γ denotes the associated Fuchsian group. We first recall that δ denotes the Hausdorff Dimension of the limit set and let us again denote by $\#\{\gamma: \ell(\gamma) \leq t\} =: N(t)$ the number of prime closed geodesics of length at most t . Now we can state the first result, which gives that δ is also the asymptotic growth rate of $N(t)$.

Lemma 17. *It is easy to show that the value δ is the growth rate of the number of closed geodesics:*

$$\delta = \lim_{t \rightarrow \infty} \frac{1}{t} \log \#\{\gamma: \ell(\gamma) \leq t\}.$$

The stronger asymptotic result in the case of closed geodesics for compact surfaces (Theorem 1) can be extended to this case.

Theorem 5. *We can write*

$$N(t) \sim \frac{e^{\delta t}}{\delta t} \text{ as } t \rightarrow +\infty$$

(i.e., $\lim_{t \rightarrow +\infty} N(t)e^{-\delta t}\delta t = 1$).

(Compare with the compact case result in Theorem 1 taking into account that $\delta = 1$.)

However, whereas the proof in the case of compact surfaces used the Selberg trace formulae this is not available to us in the case of general infinite area surfaces. Instead in the present setting the transfer operators can be used to give an extension to the zeta function Z_V and this leads to a proof of Theorem 5.

As in the compact case there is a stronger form of the asymptotic formula with an error term.

Theorem 6 (Naud 2005). *There exists $\varepsilon > 0$ such that*

$$N(t) = \int_2^{e^t} \frac{1}{\log u} du (1 + O(e^{-\varepsilon t})) = \text{li}(e^t) + O(e^{-\varepsilon t}), \text{ as } t \rightarrow +\infty.$$

However, as in the case of Theorem 5 to prove the stronger result in Theorem 6 requires the dynamical approach of transfer operators. This method also works in the case of compact surfaces, but in that case we had the luxury of using the trace formulae, which served us even better in as much as they gave both an analytic extension and additional information on the location of the zeros. Despite its apparent greater generality, or perhaps because of it, the transfer operator approach has the downside that there is less control over the location of zeros. The next lemma is the analogue of Lemma 12 for compact surfaces.

Lemma 18. *There is a correspondence between*

1. *Closed (prime) geodesics γ ; and*
2. *Periodic (prime) orbits $\{x, Tx, \dots, T^{n-1}x\}$ where $T^n x = x$, and the transformation T is given by (3).*

In fact, Lemma 18 has one improvement on Lemma 12 in as much as there is a bijection, without having to concern ourselves with a finite number of exceptional geodesics.

As in the case of compact surfaces we have the following correspondence for the lengths (compare with Lemma 13).

Lemma 19. *If γ is a closed (prime) geodesics which is not a closed bounding curve then we can associate a periodic (prime) orbit $\{x, Tx, \dots, T^{n-1}x\}$, $T^n x = x$ where $\ell(\gamma) = \log |(T^n)'(x)|$ and T is given by (3).*

We again define Selberg Zeta function

$$Z_V(s) = \prod_{n=0}^{\infty} \prod_{\gamma} (1 - e^{-(s+n)\ell(\gamma)}) \quad (6)$$

where γ denotes a prime closed geodesic of length $\ell(\gamma)$, as before in (2). Note that in the case of infinite area surfaces the closed geodesics are contained in the *recurrent* part of the flow. The infinite product on the right hand side of (6) converges for $\Re(s) > \delta$ as is easily seen using Theorem 5 and defines an analytic function.

To proceed we need to replace $C^0(X)$ by a smaller Banach space upon which the transfer operator has good spectral estimates.

7.2 The Banach spaces

The main case we want to concentrate on is the three funnelled surface (or “pair of pants”) from Example 2. This will be a paradigm for a general case.

In order to study the zeta function Z_V using the transfer operator we need to be more restrictive in the type of Banach space upon which the transfer operator acts. We will consider a Banach space of analytic (or holomorphic) functions.

1. Choose four (small) open neighbourhoods $\Lambda_{\mathfrak{g}} \subset U_{\mathfrak{g}} \subset \mathbb{C}$ where $\mathfrak{g} \in \Gamma_0 = \{a, a^{-1}, b, b^{-1}\}$;
2. Let $B \subset C^0(\Lambda)$ be the *Banach space* of bounded analytic functions on the (disjoint) union of the disks $\bigsqcup_{\mathfrak{g} \in \Gamma_0} U_{\mathfrak{g}}$ with the norm

$$\|f\| = \sup_{z \in \bigsqcup_{\mathfrak{g} \in \Gamma_0} U_{\mathfrak{g}}} |f(z)|.$$

Of course, this is equivalent to the direct sum of spaces of analytic functions on each of the four neighbourhoods.

3. By construction we see that the Banach space is preserved by transfer operators, i.e., $\mathcal{L}_s: B \rightarrow B$ for any $s \in \mathbb{C}$ given by (5).
4. On the Banach space B we have that \mathcal{L}_s is trace class (i.e. in particular, there exists only countable set of non-zero eigenvalues which sum is finite)⁴.
5. In particular, we can now observe that the following formal identity holds

$$Z_V(s) = \det(I - \mathcal{L}_s) = \exp\left(-\sum_{n=1}^{\infty} \frac{1}{n} \text{trace}(\mathcal{L}_s^n)\right)$$

and the right hand side converges to a function analytic on \mathbb{C} .

6. Finally, we have the classical observation: $Z_V(s) = 0$ if and only if there exists a function $f \neq 0 \in B$ such that $\mathcal{L}_s f = f$.

In order to show that there is a zero free strip it suffices to show that there exists $\varepsilon_0 < \delta$ such that for any $\varepsilon_0 < \Re(s) < \delta$ the operator L_s doesn't have 1 as an eigenvalue.

However, to get more information on location of zeros we want to restrict to a smaller space of functions (e.g. Hölder, C^1 , C^k , C^ω , the smaller space the better the results).

As we have observed already, closed geodesics in the case of non-compact surfaces lie in the recurrent part of the geodesic flow. We also adopt the convention that they are oriented geodesics (i.e., a pair of geodesics which are identical as sets but have different orientations counted as two geodesics).

We recall basic properties of the zeta function (cf. [21]).

Lemma 20. *Let V be an infinite area surface.*

1. *The infinite product (6) converges for $\Re(s) > \delta$ and therefore Z_V is a well defined analytic function on this half-plane;*
2. *The zeta function Z_V has an analytic extension to \mathbb{C} .*

We recall that there is a natural correspondence between closed geodesics and periodic orbits for the transformation on the Cantor set Λ given by Lemma 18.

⁴This again follows directly from the work of Grothendieck [10], [11] and Ruelle [24].

7.3 Measures and mixing

To move to the case of surfaces of infinite area we first need a new measure replacing the Haar measure. In this case we are concerned with measures supported on the recurrent part of the geodesic flow. The easiest way to construct such measures is by using the classical construction of measures on the limit set and then to convert these into measures invariant by the geodesic flow and supported on the recurrent part of the flow.

Definition 10. We can define a measure on the limit set Λ by

$$\nu_\delta(A) = \lim_{t \searrow \delta} \frac{1}{t - \delta} \sum_{\mathfrak{g} \in \Gamma} \delta_{\mathfrak{g}0}(A) e^{-td(0, \mathfrak{g}0)}, \quad A \subset \Lambda,$$

where $\delta_{\mathfrak{g}0}$ is the Dirac delta measure supported on the point $\mathfrak{g}0$.

This is the standard construction of the Patterson–Sullivan measure on Λ (cf. a book of Nicholls [19], for example). We use the Patterson–Sullivan measure ν_δ to define a flow invariant measure $\widehat{\nu}_\delta$ on the unit tangent bundle to the universal cover, which can be identified with $\partial\mathbb{D} \times \partial\mathbb{D} \setminus \text{diagonal} \times \mathbb{R}$ by:

$$d\widehat{\nu}_\delta(x, y, t) = \frac{d\nu_\delta \times d\nu_\delta \times dt}{|x - y|^{2\delta}}$$

where dt is the Lebesgue measure along the flow direction. This corresponds to an invariant measure on $\partial\mathbb{D} \times \partial\mathbb{D}$ under the diagonal action of Γ , i.e., $\mathfrak{g}(x, y) = (\mathfrak{g}x, \mathfrak{g}y)$. Finally, by considering the quotient by Γ and normalizing we arrive at a φ_t -invariant measure ν on Y .

Given $F, G: X \xrightarrow{C^\infty} \mathbb{R}$ we define the correlation function associated to the flow φ_t by

$$\rho(t) = \int_Y (F \circ \varphi_t) G d\nu - \int_Y F d\nu \int_Y G d\nu$$

In the case of infinite area surfaces the approach of unitary representations does not apply so naturally. However, there are dynamical approaches, mostly rooted in the work of Dolgopyat [8].

Theorem 7 (Naud, 2005). *There exists $C, \varepsilon > 0$ such that $|\rho(t)| \leq Ce^{-\varepsilon t}$ for $t \geq 0$.*

The proof makes use of transfer operators. These are used to analyze the Laplace transform $\widehat{\rho}$ by showing that the transfer operator L_s given by (5) has no poles for $\Re(s) \geq -\varepsilon$.

8 Location of resonances for infinite area surfaces

Finally, we want to describe the behaviour of the distribution of the zeros for the analytic extension of the Selberg Zeta function in the case of infinite area surfaces.

8.1 Distribution of zeros

To demonstrate the method, we will consider the case of a pair of pants. This is the simplest example to study and in this case the results are clearer to interpret. However, the analysis can be adapted to other examples and some of these are discussed in the appendix.

We know that there is a zero at $s = \delta$. We also know that there are no poles in a (possibly very) small strip $\delta - \varepsilon < \Re(s) < \delta$. A natural question to ask is what happens when we first venture into the region $\Re(s) < \delta - \varepsilon$?

Question 7. What about the other zeros not so near $\Re(s) = \delta$?

Since we expect that there will be zeros in a larger region, it is therefore more appropriate to ask about the density of zeros. Naud showed the following:

Theorem 8 (Naud, 2014). *Fix $\sigma \in (\frac{\delta}{2}, \delta)$. Then there exists $0 < \eta < \delta$:*

$$\#\{s_n = \sigma_n + it_n \mid Z_V(s_n) = 0: \sigma \leq \sigma_n \leq \delta, |t_n| \leq t\} = O(t^{1+\eta}).$$

The proof is particularly elegant and uses both the Bowen–Series map and properties of the pressure.

8.2 Patterns of zeros

Our final result describes the strange patterns of zeros that one sees empirically. Consider again the example of a pair of pants and associate the Selberg zeta function (cf. (6))

$$Z_V(s) = \prod_{n=0}^{\infty} \prod_{\gamma} (1 - e^{-(s+n)\ell(\gamma)}), \quad s \in \mathbb{C},$$

where γ is a (prime) closed geodesic or an orbit of the flow of length $\ell(\gamma)$. We have already seen (cf. Lemma 20) that the infinite product converges for $\Re(s) > \delta = \dim_H \Lambda$, the Hausdorff dimension of the limit set of the associated Bowen–Series map and there is an analytic extension to \mathbb{C} (using trace of the associated transfer operator).

In pioneering experimental work, D. Borthwick has studied the location of the zeros for the zeta function in specific examples of infinite area surfaces [5]. The plot in Figure 11 is fairly typical for the distribution of zeros in the critical strip for a symmetric three funnelled surface where each of the three simple closed geodesics corresponding to a funnel has the same length.

This is an infinite area surface whose compact core corresponds to a sphere with three disks removed (called a “pair of pants”). As we have discussed, the closed orbits of geodesic flow belong to the recurrent part, or to the convex core.

A modern laptop allows one to study symmetric three funnelled surface with the length of the three defining closed geodesics at least 8 without much difficulty.

The main object of study in this section is the zero set of the function Z_V :

$$\mathcal{S}_X := \{s \in \mathbb{C} \mid Z_V(s) = 0\}. \tag{7}$$

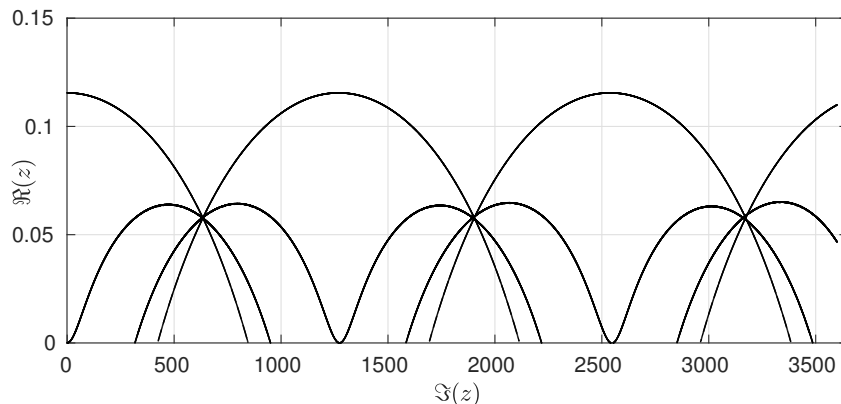


Figure 11: The zeros of the associated zeta function Z_V in the critical strip. The individual zeros are so close in the plot that it creates the illusion that they form well defined smooth curves.

Question 8. Where are the zeros of $Z_V(s)$, in strip $0 < \Re(s) < \delta$ located?

We can make the following simple empirical observations (i.e., statements that experimentally appear to be true, although they may not be exactly accurate).

Informal Qualitative Observations. Let X_b be the three funnelled surface defined by three simple closed geodesics of equal length⁵ $2b$ then for a sufficiently large b .

O1 : The set of zeros \mathcal{S}_{X_b} *appears* to be an almost periodic set with translations

$$\tau = \{i(\pi k e^b + \varepsilon_k) \mid k \in \mathbb{N}\}$$

where $\varepsilon_k = o(e^{-b/2})$ as $b \rightarrow +\infty$.

O2 : The set of zeros \mathcal{S}_{X_b} *appears* to lie on a few distinct curves, which *seem* to have a common intersection point at $\frac{\delta}{2} + i\frac{\pi}{2}e^b$, as $b \rightarrow +\infty$.

In order to understand these observations we need to estimate the approximation of the zeta function by simpler expressions as $b \rightarrow +\infty$. In particular, we obtain an approximating function whose zero set belongs to four explicit curves.

The first issue is that for large b the value of $\delta = \dim_H \Lambda$ will be small. We begin by recalling the following result of McMullen [15]:

Lemma 21. *The largest real zero $\delta \sim \frac{\ln 2}{b}$ as $b \rightarrow +\infty$, i.e., $\lim_{b \rightarrow +\infty} b\delta = \ln 2$.*

This defines the width of the critical strip. Therefore, in the limit $b \rightarrow +\infty$ the zero set converges to imaginary axis. However, a suitable affine rescaling allows one to see the pattern

⁵This normalization makes formulae in subsequent calculations shorter.

of zeros for large values of b . A natural choice for rescaling factors is the approximate period of the pattern (in the imaginary direction) and the approximate reciprocal of the width of the critical strip (in the real direction).

Definition 11. We will be using the following notation.

- (a) A compact part of the critical strip of height $T > 0$ which we denote by

$$\mathcal{R}(T) = \{s \in \mathbb{C} \mid 0 \leq \Re(s) \leq \delta \text{ and } |\Im(s)| \leq T\};$$

and a compact part of the normalized critical strip of height $S > 0$ which we denote by

$$\widehat{\mathcal{R}}(S) = \{s \in \mathbb{C} \mid 0 \leq \Re(s) \leq \ln 2 \text{ and } |\Im(s)| \leq S\}.$$

- (b) We denote the rescaled set of zeros by

$$\widehat{\mathcal{S}}_{X_b} := \left\{ \sigma b + ie^{-bt} \mid \sigma + it \in \mathcal{S}_X \right\}.$$

where evidently, $0 < \Re(\widehat{\mathcal{S}}_{X_b}) \leq \ln 2$.

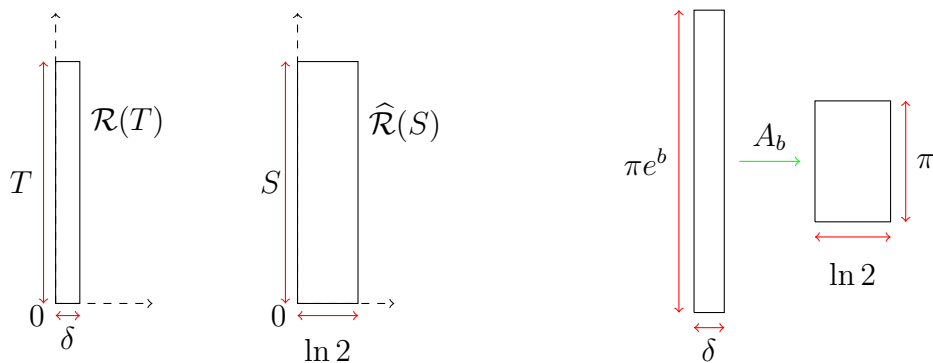


Figure 12: (a) The strips $\mathcal{R}(T)$ and $\widehat{\mathcal{R}}(S)$; (b) By renormalizing the strip $\mathcal{R}(\pi e^b)$ to $\widehat{\mathcal{R}}(\pi)$ we can compare the zeros of zeta functions for different b , as b tends to infinity.

We now introduce a family of four curves approximating $\widehat{\mathcal{S}}_{X_b}$ as $b \rightarrow +\infty$.

- (c) Let $\mathcal{C} = \cup_{j=1}^4 \mathcal{C}_j$, where the \mathcal{C}_i are explicit curves given by

$$\begin{aligned} \mathcal{C}_1 &= \left\{ \frac{1}{2} \ln |2 - 2 \cos(t)| + it \mid t \in \mathbb{R} \right\}; \\ \mathcal{C}_2 &= \left\{ \frac{1}{2} \ln |2 + 2 \cos(t)| + it \mid t \in \mathbb{R} \right\}; \\ \mathcal{C}_3 &= \left\{ \frac{1}{2} \ln \left| 1 - \frac{1}{2} e^{2it} - \frac{1}{2} e^{it} \sqrt{4 - 3e^{2it}} \right| + it \mid t \in \mathbb{R} \right\}; \\ \mathcal{C}_4 &= \left\{ \frac{1}{2} \ln \left| 1 - \frac{1}{2} e^{2it} + \frac{1}{2} e^{it} \sqrt{4 - 3e^{2it}} \right| + it \mid t \in \mathbb{R} \right\}. \end{aligned}$$

The key thing to note is that the curve \mathcal{C} in Figure 13 look very similar to empirical plots in Figure 15.

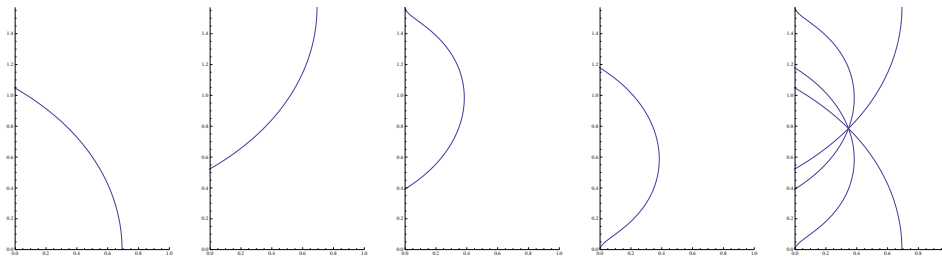


Figure 13: Plots of the curves $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3, \mathcal{C}_4$ and their union \mathcal{C} .

We can now formally state the approximation result, which provides an explanation for the Observations.

Theorem 9. *The sets $\widehat{\mathcal{S}}_{X_b}$ and \mathcal{C} are close in the Hausdorff metric d_H on a large part of the strip $0 < \Re(s) < \log 2$. More precisely, there exists $\varkappa > 0$ such that*

$$d_H(\widehat{\mathcal{S}}_{X_b} \cap \widehat{\mathcal{R}}(e^{\varkappa b}), \mathcal{C} \cap \widehat{\mathcal{R}}(e^{\varkappa b})) \rightarrow 0 \text{ as } b \rightarrow +\infty.$$

The theorem implies that every rescaled zero $s \in \widehat{\mathcal{S}}_{X_b} \cap \widehat{\mathcal{R}}(e^{\varkappa b})$ belongs to a neighbourhood of \mathcal{C} which is shrinking as $b \rightarrow \infty$. On the other hand, the rescaled zeros are so close, that the union of their shrinking neighbourhoods contains \mathcal{C} .

The most significant feature of this result is that the height $e^{\varkappa b}$ of the rescaled strip $\widehat{\mathcal{R}}(e^{\varkappa b})$ is larger than the period of the curves \mathcal{C} , and it corresponds to a part of the original strip of the height $e^{(1+\varkappa)b}$ which allows us to find a set of ε -translations.

Because of the natural symmetries of X_b , it is convenient to choose a presentation of the associated Fuchsian group in terms of three reflections (as in [15], for example). More precisely, we can fix a value $0 < \alpha \leq \frac{\pi}{3}$ and consider the Fuchsian group $\Gamma = \Gamma_\alpha := \langle R_1, R_2, R_3 \rangle$ generated by reflections R_1, R_2, R_3 in three disjoint equidistant geodesics $\beta_1, \beta_2, \beta_3$.

In the previous description of the Bowen–Series coding we chose our generators to be orientation preserving. However, to exploit the symmetry of X_b these orientation reversing generators are more convenient, although the same general theory applies as before.

Although the three individual generators are orientation reversing the resulting quotient surface \mathbb{D}^2/Γ is an oriented infinite area surface.

We now explain the pattern of the zeros for Z_{X_b} (for large b) in terms of those of a simpler approximating function.

8.3 Approximating Selberg Zeta function

Finally, we come to a peculiar issue.

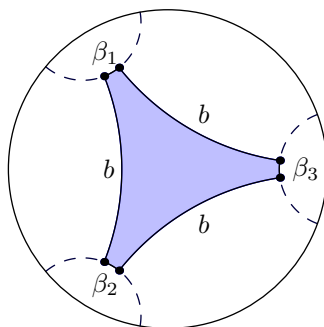


Figure 14: Three circles of reflection in Poincaré disk with pairwise distance b .

Question 9. Where do these curves come from?

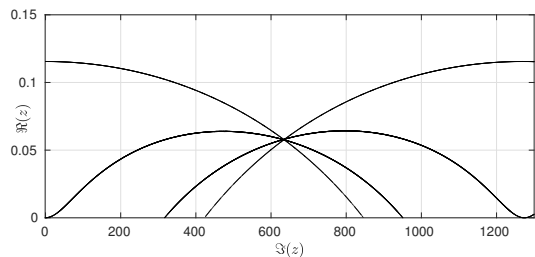
Consider the 6×6 complex matrix function

$$B(z) = \begin{pmatrix} 1 & z & 0 & 0 & z^2 & z \\ z & 1 & z^2 & z & 0 & 0 \\ 0 & 0 & 1 & z & z & z^2 \\ z^2 & z & z & 1 & 0 & 0 \\ 0 & 0 & z & z^2 & 1 & z \\ z & z^2 & 0 & 0 & z & 1 \end{pmatrix}$$

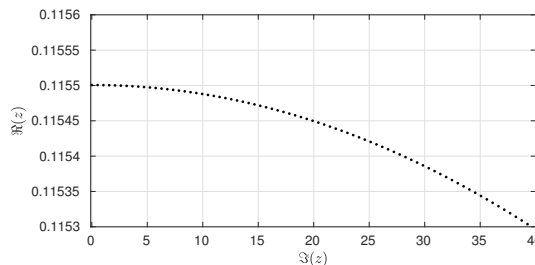
We can use this to give an approximation to Z_{X_b} and its zeros.

Theorem 10 (Approximation Theorem). *Using the notation introduced above, the real analytic function $Z_{X_b}(\frac{\sigma}{b} + ite^b)$ converges uniformly to $\det(I - \exp(-2\sigma - 2itbe^b)B(e^{it}))$, i.e.,*

$$\left| Z_{X_b} \left(\frac{\sigma}{b} + ite^b \right) - \det \left(I - \exp(-2\sigma - 2itbe^b)B(e^{it}) \right) \right| \rightarrow 0 \text{ as } b \rightarrow +\infty.$$



(a)



(b)

Figure 15: (a) The zeros of the determinant $\det(I - e^{-2b\sigma}B(\exp(ie^{-bt})))$ approximating Z_{X_6} ; (b) A zoomed version in a neighbourhood of δ showing individual zeros. The distance between imaginary parts of consecutive zeros is approximately $\frac{\pi}{6}$.

Key observation: The importance of the matrix B is that for $s = \sigma + it$ to lie on the curves C_1, C_2, C_3, C_4 corresponds to eigenvalues λ_t for $B(e^{-2it})$ to satisfy $e^{-\sigma}|\lambda_t| = 1$.

We will briefly show how Theorem 10 implies Theorem 9. We shall show that for all $\varepsilon > 0$ and $T > 0$ there exists $b_0 > 0$ such that for any $b > b_0$ the zeros of the function $Z\left(\frac{\sigma}{b} + ite^b\right)$ with $0 \leq \sigma \leq \ln 2$ and $|t| \leq e^{(2-\varepsilon)b}$ belong to a neighbourhood $\cup_k V(\mathcal{C}_k, \varepsilon)$ of the union of the curves $\cup_k \mathcal{C}_k$.

Given $\varepsilon > 0$ and a point $z_0 = \sigma_0 + it_0$ outside of ε -neighbourhood of $\cup_{j=1}^4 \mathcal{C}_j$ we see that the determinant

$$|\det(I - \exp(-2\sigma_0 - it_0 b e^b) B(\exp(it_0)))| > \exp(-6\varepsilon)(\exp \varepsilon - 1)^6 > 0$$

is bounded away from zero and the bound is independent of b . Thus we see that outside of the neighbourhood $\cup_{j=1}^4 V(\mathcal{C}_j, \varepsilon)$ the determinant has modulus uniformly bounded away from 0. But using Theorem 10 the zeta function $Z_{X_b}\left(\frac{\sigma}{b} + ite^b\right)$ can be approximated arbitrarily closely by the determinant. Therefore for b sufficiently large all zeros of the function $Z_{X_b}\left(\frac{\sigma}{b} + ite^b\right)$ belong to the ε -neighbourhood of $\cup_{j=1}^4 \mathcal{C}_j$.

Remark 7. There is a natural φ -invariant measure of maximal entropy which is supported on the recurrent part. One can consider $F, G: R \rightarrow \mathbb{R}$ and the correlation function

$$\rho(t) := \int F\varphi_t \cdot G d\nu - \int F d\nu \cdot \int G d\nu, \quad t \geq 0.$$

The rate that $\rho(t) \rightarrow 0$ controlled by poles of Laplace transform

$$\hat{\rho}(z) = \int_0^\infty e^{-zt} \rho(t) dt$$

There zeros of this function are closely related to zeros of Z_{X_b} (see [22] and [23] for specific results).

References

- [1] Adler, R. Afterword to “Invariant measures for Markov maps of the interval” by R. Bowen, *Comm. Math. Phys.* 69 (1979), no. 1, 1–17.
- [2] Adler, R. and Flatto, L. Geodesic flows, interval maps, and symbolic dynamics *Bull. Amer. Math. Soc. (N.S.)* Volume 25, Number 2 (1991), 229–334.
- [3] Beardon, A. F. *The geometry of discrete groups*. Graduate Texts in Mathematics, 91. Springer-Verlag, New York, 1995. xii+337 pp.
- [4] Bedford, T., Keane, M. and Series, C., *Ergodic theory, Symbolic Dynamics and Hyperbolic Spaces*, C.U.P., Oxford, 1989.

- [5] Borthwick, D. Distribution of resonances for hyperbolic surfaces. *Exp. Math.* 23 (2014), no. 1, 25–45.
- [6] Bowen, R, and Series, C. *Publications Mathématiques de l’IHÉS*
- [7] Cheeger, J. A lower bound for the smallest eigenvalue of the Laplacian. *Problems in analysis (Papers dedicated to Salomon Bochner, 1969)*, pp. 195–199. Princeton Univ. Press, Princeton, N. J., 1970.
- [8] Dolgopyat, D. On decay of correlations in Anosov flows. *Ann. of Math. (2)* 147 (1998), no. 2, 357–390.
- [9] Falconer, K. *Fractal geometry. Mathematical foundations and applications*. John Wiley & Sons, Ltd., Chichester, 1990. xxii+288
- [10] Grothendieck, A. Produits tensoriels topologiques et espaces nucléaires. (French) *Mem. Amer. Math. Soc.* No. 16 (1955), 140 pp.
- [11] Grothendieck, A. La théorie de Fredholm. *Bull. Soc. Math. France* 84 (1956), 319–384.
- [12] Hejhal, D. *The Selberg Trace Formula for $PSL(2, \mathbb{R})$* , *Lecture Notes in Mathematics* 548, Springer, Berlin, 1976.
- [13] Mayer, D. H. Continued fractions and related transformations. *Ergodic theory, symbolic dynamics, and hyperbolic spaces (Trieste, 1989)*, 175–222, *Oxford Sci. Publ.*, Oxford Univ. Press, New York, 1991.
- [14] McKean, H. Selberg’s trace formula as applied to a compact Riemann surface. *Comm. Pure Appl. Math.* 25 (1972), 225–246
- [15] McMullen, C. T. Hausdorff dimension and conformal dynamics. III. Computation of dimension. *Amer. J. Math.* 120 (1998), no. 4, 691–721.
- [16] Moore, C. C. Exponential decay of correlation coefficients for geodesic flows. *Group representations, ergodic theory, operator algebras, and mathematical physics (Berkeley, Calif., 1984)*, 163–181, *Math. Sci. Res. Inst. Publ.*, 6, Springer, New York, 1987.
- [17] Naud, F. Expanding maps on Cantor sets and analytic continuation of zeta functions, *Ann. Sci. Ecole Norm. Sup.* 38 (2005), 116–153.
- [18] Naud, F. Density and location of resonances for convex co-compact hyperbolic surfaces. *Invent. Math.* 195 (2014), no. 3, 723–750.
- [19] Nicholls, P. J. *The ergodic theory of discrete groups*, LMS–LNS 143, C.U.P., Cambridge 1990;

-
- [20] Parry, W. and Pollicott, M. Zeta functions and the periodic orbit structure of hyperbolic dynamics. *Astérisque* No. 187–188 (1990), 268 pp.
- [21] Pollicott, M. Some applications of thermodynamic formalism to manifolds with constant negative curvature. *Adv. Math.* 85 (1991), no. 2, 161–192.
- [22] Pollicott, M. Meromorphic extensions of generalised zeta functions. *Invent. Math.* 85 (1986), no. 1, 147–164.
- [23] Pollicott, M. On the rate of mixing of Axiom A flows. *Invent. Math.* 81 (1985), no. 3, 413–426.
- [24] Ruelle, D. Zeta-functions for expanding maps and Anosov flows. *Invent. Math.* 34 (1976), no. 3, 231–242.
- [25] Schoen, R.; Wolpert, S.; Yau, S. T. Geometric bounds on the low eigenvalues of a compact surface. *Geometry of the Laplace operator (Proc. Sympos. Pure Math., Univ. Hawaii, Honolulu, Hawaii, 1979)*, pp. 279–285, *Proc. Sympos. Pure Math.*, XXXVI, Amer. Math. Soc., Providence, R.I., 1980.
- [26] Selberg, A. Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. *J. Indian Math. Soc. (N.S.)* 20 (1956), 47–87.
- [27] Series, C. Geometrical Markov coding of geodesics on surfaces of constant negative curvature, *Ergod. Th. and Dynam. Sys.*, 6 (1986) 601–625