A Short and Dirty Introduction to Hyperbolic Surfaces

January 9, 2012

Chapter 1

Foreword

The aim of this small series of note is to give a concise and elementary introduction to hyperbolic surfaces. Starting from a synthetic point of view, we shall give the classification of compact oriented connected hyperbolic surfaces, define arithmetic surfaces introduce some elementary dynamical ideas and finally state some of the major conjectures in the subject.

The text is meant to be accessible to a student who knows about the cross ratio and the projective line, what is a covering space and a simply connected object.

This set of notes has been very preliminary for years and the result may or may not evolve to a more correct version.

Many excellent textbooks or notes on the web exist and I would like in particular suggest to consult N. Bergeron remarkable course on a broader subject.

This text contains many errors, wrong statements, no bibliography and its spelling and syntax are pathetic. So please use it at your own risks.

Contents

Chapter 2

Hyperbolic plane

2.1 Synthetic geometry

The complete geometry of the hyperbolic plane can be recovered synthetically from several features, namely *lines* and *boundary at infinity.* Let us define

Definition 2.1.1 [Hyperbolic plane] *A set A together with a*

1. *a subset B called the* boundary at infinity *with a cyclic order,*

2. *a family of* lines *which are subsets of A,*

is a weak (completed) hyperbolic plane *if*

1. *given any two distinct points in A there exists a unique line passing through two points of A,*

2. *Any line intersects the boundary at infinity in exactly two points.*

3. *Two lines with four intertwined ends intersects in a point in $A \setminus B$.*

4. *Two lines with four non intertwined ends do not intersect.*

5. *If x is a point in B, and ι_x is the involution of B exchanging the end points of the lines through y, if x is distinct than y then the* hyperbolic translation $\iota_x \circ \circ \iota_y$ *has exactly two fixed points in B: the end points of the line passing through x and y.*

The set A is completed hyperbolic plane *if moreover the set B – together with its cyclic order – is equipped with a cross ratio that identifies it with $\mathbb{P}(P)$ for some real two plane P, and if furthermore this cross ratio is invariant under the involution above.*

In this case, we use the following notation, we denoted the completed hyperbilic plane by $\overline{\mathsf{H}^2}$, its boundary at infinity by $\partial_\infty \mathsf{H}^2$, and the hyperbolic plane itself is $\mathsf{H}^2 = \overline{\mathsf{H}^2} \setminus \partial_\infty \mathsf{H}^2$.

Exercise : All hyperbolic planes are geometrically equivalent, by which we mean that there exists a bijection sending lines to lines, boundary at infinity to boundary at infinity and preserving the cross ratios. The idea is that any point in $A \setminus B$ defines an involution of the boundary at infinity which preserves the cross ratio by exchanging the end points and which is different than the identity. Conversely any such involution can be produced that way. Therefore we have identified $A \setminus B$ with the set of involutions of the projective line preserving the cross ratio and lines can be now defined accordingly using the hyperbolic translations.

Then almost by definition the group $\mathsf{PSL}(P)$ acts transitively on the hyperbolic plane preserving the geometry.

It remains to construct hyperbolic planes and we shall that they exists thanks to the existence of complex numbers. If Euclid would have known complex numbers, as well as understood Desargues, he would have construct hyperbolic geometry as in the next section ...

2.1.1 The complex projective line and how to build hyperbolic planes

Let E be a vector space of dimension 2 over a field $\mathbb{K}$. For a geometer the most important feature of the group $\mathsf{PSL}(E, \mathbb{K})$ is its action on the projective line $\mathbb{P}(E)$, action which is characterised by the fact it preserves the cross ratio. In an affine chart, the action is given by homographies

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} x = \frac{ax+b}{cx+d}.$$

The cross ratio itself is

$$[x_1, x_2, x_3, x_4] = \frac{x_1 - x_2}{x_1 - x_4}.\frac{x_3 - x_4}{x_3 - x_2}.$$

For $\mathbb{K} = \mathbb{C}$, the fact that $\mathbb{C}$ is an extension of $\mathbb{R}$ defines an extra geometry on the projective line $\mathbb{P}(E)$.

The starting observation is that every complex line in E is a real two plane in E seen as a real vector space.

Conversely, every real two plane P in E determines

1. a *circle* C_P in $\mathbb{P}^1(\mathbb{C})$ which is the set of complex lines that intersects P, and which is identified with $\mathbb{P}^1(\mathbb{R})$.

2. in the case, P is not a complex line, an *involution* preserving C. This involution σ_P reflect the Galois group of the extension of $\mathbb{R}$ by $\mathbb{C}$ and can be characterised in several ways. The more geometric one is the following. if P is a real two-plane and D is a complex line, $\sigma(D)$ is characterised by

$$\forall D_1, D_2, D_3, \in C_P,\ [D_1, D_2, D_3, D] \quad = \quad \overline{[D_1, D_2, D_3, \sigma(D)]}.$$

When $P = \mathbb{R}^2$ in $E = \mathbb{C}^2$, the corresponding involution is given by complex conjugation. All these quantities are equivariant under the action of $\mathsf{PSL}(E, \mathbb{C})$.

The product of two such involutions preserves the cross ratio and is therefore a projective transformation. We finally define the *angle* between two intersecting and oriented circles C_P and C_Q as the argument of the trace of the product of the two involutions σ_P and σ_Q.

We can now proceed to the construction of the hyperbolic plane.

Proposition 2.1.2 *Let P be a real two-plane in a complex two-vector space – for instance $\mathbb{R}^2$ – itself. The* completed hyperbolic plane *is the set of complex lines L so that*

$$\forall D_1, D_2, D_3, \in C_P, \quad \Im([D_1, D_2, D_3, D) \geq 0.$$

A hyperbolic line *or* geodesic *is the intersection of a circle orthogonal to P. The* boundary at infinity *is C_P.*

2.1.2 The upper half space model

Using a chart, we end up with a more familiar picture: the hyperbolic plane is the upper half plane in $\mathbb{C}$.

- The *completed hyperbolic plane* is
$$\overline{\mathsf{H}^2} = \{z \in \mathbb{C}, \ \ \Im(z) \geq 0\} \cup \{\infty\}.$$

- The *hyperbolic plane* is
$$\mathsf{H}^2 = \{z \in \mathbb{C}, \ \ \Im(z) > 0\}.$$

- The *boundary at infinity* is
$$\partial_\infty \mathsf{H}^2 = \{z \in \mathbb{C}, \ \ \Im(z) = 0\} \cup \{\infty\}.$$

- The *lines* are either half circles orthogonal to the real axis or half lines orthogonal to the real axis completed by ∞.

Then the natural action of $\mathsf{PSL}(2,\mathbb{R})$ is given by homographies
$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} x = \frac{ax+b}{cx+d}.$$

Observe finally that the angles – in the upper half plane model – between two oriented geodesics can now be defined using the cross ratios of their end points. In particular, two geodesics intersect orthogonally if their end points form a *harmonic division*, that is if their cross ratio is -1.

2.2 More geometric features: figures and distances

If two geodesics D and L do not intersect, we define the *distance* between them as the logarithm of the cross ratio of their end points. We define the *distance* between two points as the distance between the geodesics that are orthogonal. It is not yet clear that this is a distance, we shall prove that later in some more economic way. For the moment we admit that this is indeed a distance. *Hyperbolic circles* – later on called simply circles – are the circles in our geometric model that do not intersect the boundary at infinity, *horocycles* are circles that intersect the boundary at infinity at exactly one point

A *half space* is a complementary region to a geodesic, and *wedge* is one of the complementary region of two oriented geodesic intersecting orthogonally.

A *convex polygon* is the intersection of half spaces. Among them are triangles, hexagons etc ... two points in a convex polygons are joined by a geodesic arc inside this polygon

2.2.1 Triangles and ideal triangles

An *ideal triangle* is a triangle with three points at infinity, a *2/3-ideal triangle* has two points at infinity and a *1/3-ideal triangle* has one. All ideal triangles are congruent.

Proposition 2.2.1 *Given any number a, b and c satisfying the triangles inequalities there exist a unique triangle – up to the action of* $\mathsf{PSL}(2,\mathbb{R})$ *– whose length are a, b and c.*

Idea: we fix a point and one segment of length a, we consider the segment at angle θ and length b then the distance between the end points is an increasing function of θ. A continuity argument shows this is a bijection onto $[|a-b|, a+b]$.

2.2.2 Right-angled hexagons

Proposition 2.2.2 *Given any number a, b and c, there exist a unique right angled hexagon – up to the action of* $\mathsf{PSL}(2,\mathbb{R})$ *– whose length of non intersecting edges are a, b and c.*

idea : same idea as above.

2.2.3 Regular polygons

Proposition 2.2.3 *Given any integer $n > 4$, there exist a unique – up to the action of* $\mathsf{PSL}(2,\mathbb{R})$ *– regular right-angled n-gon.*

idea : same idea as above.

2.3 The Riemannian interpretation

2.3.1 The length of a curve

We define the length of a parametrised curve $c = (x, y) : [a, b] \to \mathsf{H}^2$ in the hyperbolic plane in the upper half model as

$$\ell(c) = \int_a^b \frac{\sqrt{\dot{x}^2 + \dot{y}^2}}{y} dt.$$

Then, the following facts are true

- The length of the curves is invariant under the action $\mathsf{PSL}(2,\mathbb{R})$. *Hint: check the invariance under well chosen subgroups if you are lazy*
- Geodesics minimise the length of the curve. *Hint: consider first the case of a curve orthogonal to the real axis.*
- The distance between two points is

  $$d(x, y) = \inf\{\ell(c) \mid c \text{ joins } x \text{ to } y\}.$$

  similarly for distance between geodesics.

- The hyperbolic space is a *length space* in the sense that we can recover the length from the distance:

$$\ell(c) = \sup\{\sum_i d(c(x_i), c(x_{i+1})) \mid 0 = x_0 < x_1 < \ldots < x_n = 1\}.$$

Finally, one recover the boundary at infinity from this picture. We say two oriented geodesics are *asymptotic* if given two arc lengths parametrisation of these geodesics $t \to \gamma_1(t)$ and $t \to \gamma_2(t)$ then

$$\lim \sup_{t \to +\infty} (d(\gamma_1(t), \gamma_2(t)) < \infty.$$

Then two oriented geodesics are asymptotic precisely if they have the same end point at $+\infty$.

2.3.2 Area

The area of a measurable set A is

$$\text{area}(A) = \int_A \frac{1}{y^2} dxdy.$$

2.3.3 The Gauss-Bonnet formula

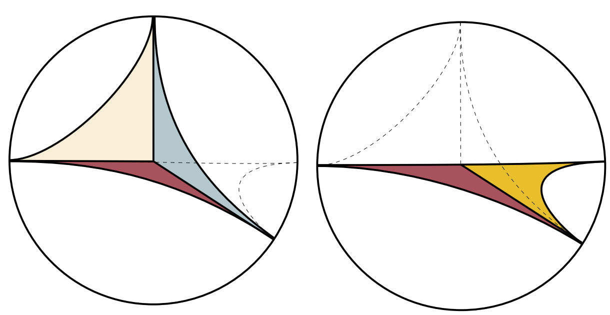

Figure 2.1: Gauss additivity

Theorem 2.3.1 *The area of a triangle is its* angle defect *that is π minus the sum of the interior angles of the triangle.*

Proof : We shall follow Gauss approach. We first check by a direct computation that the area of an ideal triangle is π. Then let $A(\theta)$ the area of a 2/3-ideal triangle with angle $\pi - \theta$. Gauss observation is $A(\theta)$ is an additive function hence a multiple of θ, hence θ by the normalisation. This is done geometrically as follows in Figure 2.1: the area of the union of the pink and blue triangles is the area of yellow triangle, since they are both π minus the area of the purple triangle. One this is done the rest follows.Q.e.d.

As an exercise, define an area just using the angle defect and check which properties the angle defect should have and prove them.

As a special we also see that the area of a right angle hexagons is 2π and that of of regular right-angled n-gon is $\frac{\pi}{2}n$.

Chapter 3

Hyperbolic surfaces

3.1 Surfaces

3.1.1 Hyperbolic surfaces

A *hyperbolic surface* is a complete metric space M such that every point in M has a neighbourhood isometric to an open set of the hyperbolic plane.

A *hyperbolic surface with totally geodesic boundary* is a complete metric space M such that every point in M has a neighbourhood isometric to an open set of the hyperbolic plane, or an hyperbolic half plane.

A *hyperbolic surface with totally geodesic boundary and right angles* is a complete metric space M such that every point in M has a neighbourhood isometric to an open set of the hyperbolic plane, or an hyperbolic half plane, or a right-angled wedge.

Given a hyperbolic surface (S, d) with a metric d, we can find a better metric on it. Let $c : [a, b] \to S$ be a curve in S, we can first define its *hyperbolic length* $\ell(c)$ as follows. We first find a subdivision

$$a = t_0 < t_1 < \ldots < t_n = b,$$

so that $c[t_i, t_{i+1}] \subset B_i$, where B_i is a ball for d isometric (by a map ϕ_i) to a hyperbolic ball. Then we define

$$\ell(c) = \sum_{i=0}^{n-1} \ell\left(\phi_i \circ c|_{[t_i, t_{i+1}]}\right).$$

We leave as an exercise the following facts

1. The length $\ell(c)$ does not depend on the subdivision of $[a, b]$.

2. The length $\ell(c)$ does not depend on the parametrisation of c : if ϕ is a diffeormorphism from $[a, b]$ to $[c, d]$ then $\ell(c) = \ell(c \circ \phi)$.

3. If d and d' are two locally isometric metrics on S, both locally isometric to the hyperbolic plane. Then the length for d and the length for d' are equal.

This length allows us to define a new metric on S. For any x and y on S we define the *Riemannian distance* on S by

$$d(x,y)=\inf\{\ell(c)\mid c:[0,1]\to S,\quad c(0)=x,\ c(1)=y\}.$$

One now has the following proposition

Proposition 3.1.1 *The Riemannian distance is a distance on the hyperbolic surface (S,d) which is moreover locally isometric to d. Finally two locally isometric d and d' on S generates the same Riemannian distance.*

From now on, we shall always equip a hyperbolic surface with its Riemannian distance.

It follows from the definitions that hyperbolic surfaces are length spaces (see below). This has an interesting consequence:

Lemma 3.1.2 *Two points in a hyperbolic surface with totally geodesic boundary and right angles can be joined by a geodesic whose length realise the distance between the two points.*

Proof : use the fact that hyperbolic surfaces are locally compact spaces. Then produce a length minimising sequence of curves parametrised by arc length and prove they converge to a geodesic by using Arzela-Ascoli lemma. Q.e.d.

length space

Given a metric space (X,d), one can always define the length of a continuous curve $c:[a,b]\to X$ as

$$\ell(c)=\sup\left\{\sum_{i=0}^{n} d(c(t_i),c(t_{i+1}))\mid (t_0,\ldots,t_{n+1}),\ a=t_0<\ldots<t_{n+1}=b\right\}.$$

This length allows to define a new distance, taking possibly infinite values, the *length distance* on X, by

$$d_\ell(x,y)=\inf\left\{\ell(c)\mid c:[a,b]\to X,\quad c(a)=x,\ c(b)=y\right\}.$$

By definition, a metric space is a *length metric space* if $d_\ell=d$. It turns out that starting from any metric space (X,d) the metric space (X,d_ℓ) is a length metric space.

3.1.2 Local isometries

We begin with the following two lemmas

Lemma 3.1.3 *Every map from a subset U of H^2 containing more than three points to H^2 preserving the distance is the restriction of an isometry. Every orientation preserving map from an open connected subset of H^2 to H^2 preserving the length of curves is the restriction of a homography.*

Proof : Prove the first easy statement then the second one. Q.e.d.

Lemma 3.1.4 *Let ϕ be a local isometry from S to Σ which both are hyperbolic surfaces, then ϕ is a a covering.*

Proof : take a ball in S isometric to a ball in H^2. Then, use Lemma 3.1.2 to show that the preimage of this ball B is a union of disjoint balls all isometric to B. Q.e.d.

3.2 Construction of hyperbolic surfaces by gluing

3.2.1 Gluing two length metric spaces

We can glue two length spaces provided we have a gluing map that preserves the distances, by defining the length of any curves as the sum of the length in each part.

Then one checks that the gluing two hyperbolic half-spaces along their boundary leads to the hyperbolic plane, and gluing two right-angle wedges leads to the hyperbolic half plane.

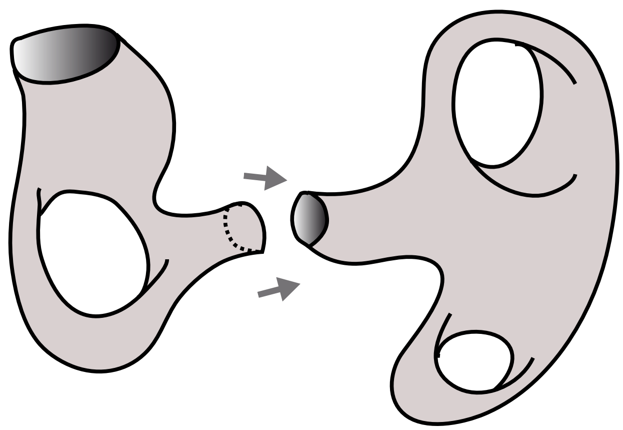

Figure 3.1: Gluing surfaces

3.2.2 Construction of closed surfaces

We can therefore construct hyperbolic surfaces of area $4\pi n$ using $3n$-positive real parameters, and $3n$-angles – i.e elements of $\mathbb{R}/\mathbb{Z}$. Moreover, the surface is given together with an extra topological structure namely an *decomposition into pair of pants*. The construction run as follows. First, we construct pairs of right-angled hexagons, fixing the boundary length. Then, gluing two hexagons together we obtain a hyperbolic pair of of pants whose boundary length are prescribed. Then we glue these pair of pants together using a prescribed identification of the boundaries. We can isometrically identify each boundary component of length a with $\mathbb{R}/a\mathbb{Z}$, in a canonical way: sending a chosen corner of the previous hexagon to zero. Then the gluing between two boundary components is determined by one parameter in $\mathbb{R}/a\mathbb{Z}$.

Two questions remain

1. Are these surfaces isometric ?

2. Are all compact surfaces obtained this way ?

3.3 Cutting and classification

Let us state our main result, which answers partly the question above.

Theorem 3.3.1 *Every compact oriented connected hyperbolic surface is of area $4\pi n$ and can be obtained be decomposed into $2n$ hexagons glued as pair of pants. Moreover, this decomposition, as well as the $3n$ gluing parameters, is fixed as soon as we fix $3n$ homotopy classes of pairwise non intersecting simple curves on S.*

3.3.1 Simply connected surfaces

Theorem 3.3.2 *The universal cover of a hyperbolic surface with geodesic boundary and right-angles is isometric to a convex non compact polygon in the hyperbolic plane.*

We state a useful corollary.

Corollary 3.3.3 *Let S be a hyperbolic surface with right-angles. Then any loop based at a point is homotopic to a geodesic loop. Conversely, any curve homotopic to a geodesic loop, is not homotopic to zero.*

PROOF : We sketch the proof here. By considering the double of the hyperbolic surface, it suffices to prove the result for hyperbolic surface without boundary as we shall we do now. Let $\{U_i\}_{i\in I}$ be a cover of a hyperbolic surface so that every U_i is isometric to a ball in the hyperbolic plane.

Fix an index i_0, a *path* starting from i_0 is a finite sequence of indices $\{i_0, i_1, \ldots, i_n\}$ so that U_{i_j} intersects U_{j+1} and moreover $U_{i_j} \cap U_{j+1}$ is included in a ball isometric to a ball in the hyperbolic space. We now fix once and for all an isometry f_0 of U_{i_0} to a ball in the hyperbolic plane.

The first remark, which follows from Lemma 3.1.3 is the following: given a path P there exists a unique family of isometries $f^P_{i_k}$ of U_{i_k} in H^2 so that $f^P_{i_k}$ coincide with $f^P_{i_{k-1}}$ on the intersection. We call $f^P_{i_n}$ the *final isometry*

Then, we say two paths with fixed extremity i_n are *homotopic* if they can be obtained from each other after a succession of the following elementary operation: *deletion* which is just to remove an index (whenever it is possible) or *insertion* which is the converse operation. Using Theorem 3.3.2, it is easy to see that two homotopic path defines the same final isometries.

Now we can proceed in two ways

1. either use our little knowledge of homotopy theory and notice that in a simply connected hyperbolic surface, each "discrete" path is homotopic to any another path,

2. or construct directly a hyperbolic surface Σ with a covering $\{V_j\}_{j\in J}$ together with a map f from Σ to S so that

   (a) all paths in Σ are homotopic to zero,

   (b) for any $j \in J$, there exists $i \in I$, so that f is an isometry from V_j to U_i.

   This is a fairly easy construction obtained by gluing. Then, by Lemma 3.1.4 Σ is S.

It follows that we can extend uniquely the map f_{i_0} to a map f from S to H^2. Then by Lemma 3.1.4 again, f is a covering, hence an isometry.

Q.E.D.

3.3.2 Curves on surfaces

We summarise here the property of curves on compact hyperbolic surfaces with boundary that we prove.

Theorem 3.3.4 *1. Given two point p and q in S and a path c from p to q there exists a unique geodesic joining p to q and homotopic to c. Moreover this geodesic minimises the length of all curves joining p to q.*

2. Given a closed curve c, there exists a unique closed geodesic γ_c freely homotopic to c, whose length furthermore minimise the length of all curves freely homotopic to c. Moreover

(a) if c is simple so is γ_c,

(b) if c_1 and c_2 do not intersect and if furthermore no non trivial powers of these curves are freely homotopic, γ_{c_1} and γ_{c_2} do not intersect.

Proof : We write $S = \tilde{S}/\Gamma$. The first statement is a consequence of our description of the universal cover.

For the second statement, let $c : [0,1] \to S$ be a closed curve and $\tilde{c} : \mathbb{R} \to \tilde{S}$ a lift of c in the universal cover. If c is not freely homotopic to zero, then there exists an element $\gamma \in \Gamma$ such that $\tilde{c}(x+n) = \gamma^n \tilde{c}(x)$ for all $n \in \mathbb{N}$. Thus we see that c is freely homotopic to the closed geodesic g associated to γ and which join γ^+ to γ^-.

Moreover since $\tilde{c}$ is at finite distance to g, more precisely there exists λ and K such that for all t

$$d(\tilde{c}(t), \gamma(\lambda.t)) \leq K.$$

It follows that

$$\lim_{t\to\infty} c(t) = \gamma^+, \quad \lim_{t\to-\infty} c(t) = \gamma^-$$

Therefore, if the geodesics loops associated to the two curves intersects then we can find lifts of these geodesics that intersects, then their endpoint are intertwined, hence the corresponding lifts of the curves intersect too by Jordan theorem, and hence the two initial curves also intersect..

The statement about simple curves is an elaboration of this idea. Q.E.D.

3.3.3 Characterisation of pair of pants

We start with the following result that follows from the classification of surfaces.

Theorem 3.3.5 *Let S be a surface that is not topologically a pair of pants, then it contains a simple curve not freely homotopic to a boundary component.*

This result follows from the classification of surfaces but we sketch an independent proof. Let us start with a definition. Let $H^1(S) = \pi_1(S)/[\pi_1(S), \pi_1(S)]$. We say two closed curves are *homologous* if they coincide in $H^1(S)$. We observe that if curves are not cohomologous then, they are not freely homotopic.

Step 1

As a corollary of the description of the universal cover of a hyperbolic surface in Theorem 3.3.2, we have

Proposition 3.3.6 *The fundamental group of a surface with at least one boundary component is free with at least two generators. Moreover if there are n boundary components, then any $n-1$ of them are independent.*

Proof : consider a polygonal fundamental domain, it is a convex set, the image of its boundary in the quotient is a graph, whose complementary region is simply connected and isometric to a convex polygon. This convex polygon as at least one side which lies in the boundary component, we can therefore retract this side to the other sides. The result follows. Q.e.d.

Step 2

Proposition 3.3.7 *Let S be a compact hyperbolic surface with at most two totally geodesic boundary components. Then there exists a non trivial embedded closed geodesic different than the boundary.*

Proof : We first need to master a technique: how to produce a simple curve not homologous to a power of boundary components from a curve not homologous to boundary components. The idea is as follows : you start with a curve not homologous to boundary components with possibly transverse self intersections – using broken geodesic arcs for instance – then you open up intersections as in the Figure 3.2. Then at least one of the simple curves obtained this way is not homologous to a power the boundary component, hence it is not freely homotopic to a power of boundary component.

Thanks to the first step, one easily produce on a surface with at most two boundary components, a curve not homologous to a boundary component.

Q.e.d.

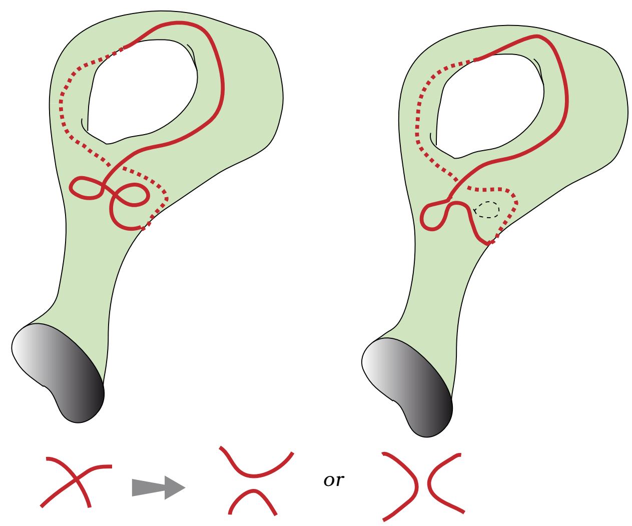

Figure 3.2: Opening up intersections

Step 3

Proposition 3.3.8 *Let S be a compact hyperbolic surface with at least four totally geodesic boundary components. Then there exists a simple closed curve not freely homotopic to a power of a boundary.*

PROOF : Take an arc joining two boundary component and consider thicken it to obtain a simple curve that is homologous to the sum of the two boundary components. This curve is not homologous to any power of a given boundary component, hence not freely homotopic either. Q.E.D.

Step 4

This last step will useful later on.

Proposition 3.3.9 *Let S be a surface with three totally geodesic boundaries. Assume that S does not contain a simple curve not homotopic to any power of a boundary component. Then S can be decomposed uniquely into a pair of right-angled hexagons*

PROOF : We first build simple arcs joining the three components of the boundary. To do so, you find three arcs – for instance using broken geodesic – joining the thee boundaries components with transverse intersection. You open up all intersections wisely, that is proceeding recursively from one boundary component, and always choosing one of the opening up that does not go back to your original boundary. You delete loops and you end up with three arcs. Then you straighten them up using geodesics, that is you find the shortest arc homotopic to them (Hint: use a doubling of the surface). You observe that, by symmetry the geodesics intersect the boundary orthogonally. Then the complementary region is not connected.

We now prove that the remaining two connected components are simply connected. Let c be a loop in one of this region not homotopic to zero. We can assume after opening up intersections, that it is a simple loop, and after straightening up to an embedded closed geodesic, which is different than any of the boundary components since it stay in that complementary region. This contradict the hypotheses. Q.E.D.

3.3.4 Final statement

As an exercise, prove Theorem 3.3.1 by cutting successively using simple closed curves. The process stops at a finite time and the remaining pieces are hyperbolic pair of pants. The stop occurs thanks to the following lemma

Lemma 3.3.10 *A surface with totally geodesic boundary has an area which is a multiple of π.*

PROOF : we use the description of the universal cover and cut the surface into hyperbolic triangles. Then the sum of angle defects is the area, summing these angle defects of vertices show that the sum is a multiple of π. Q.E.D.

As another consequence, we obtain

Theorem 3.3.11 [GAUSS-BONNET FORMULA (BIS)] *Let S be a compact surface admitting a hyperbolic structure. Then two pair of pants decomposition have the same number of pair of pants and this number is even. If $\chi(S)$ is this opposite of this number called the* Euler characteristic of S*, then any hyperbolic structure on the surface has area $-2\pi\chi(S)$.*

Chapter 4

Dynamics

In this chapter, Γ will be a discrete subgroup of $\mathsf{PSL}(2,\mathbb{R})$ so that H^2/Γ is a compact hyperbolic surface.

4.1 The action on the boundary at infinity

We begin by studying the action of Γ on the boundary at infinity $\partial_\infty \mathsf{H}^2$ of H^2. Every element of γ corresponds to a closed geodesic and will therefore preserves exactly two points at infinity $\{\gamma^-, \gamma^+\}$.

Moreover, every γ in Γ has *north-south dynamics* meaning that the sequence of iterates of any point x in $\partial_\infty \mathsf{H}^2$ different than γ^- converges to γ^+, as in Figure 4.1.

$$\lim_{n\to\infty}(\gamma^n(x)) = \gamma^+.$$

We now use the north-south dynamics to show three crucial properties of the action of Γ on the boundary at infinity.

Lemma 4.1.1 [Minimality] *Every orbit of* Γ *is dense on* $\partial_\infty \mathsf{H}^2$.

Proof : Let F be a closed Γ invariant set in $\partial_\infty \mathsf{H}^2$ and E be the *convex envelope* of F in H^2, that is the intersection of all hyperbolic half spaces containing F. The set E is a closed convex set which is Γ invariant. Let d be the function on H^2 defined as a distance to E. Then d is Γ invariant. However d is unbounded as we see from taking a geodesic orthogonal to one of the boundary component of the polygon. This contradicts the compactness of H^2/Γ.Q.e.d.

Lemma 4.1.2 [Density of end points] *The set of end points of geodesics* $\{(\gamma^+,\gamma^-) \mid \gamma \in \Gamma\}$, *is dense in* $\partial_\infty \mathsf{H}^2 \times \partial_\infty \mathsf{H}^2$.

Proof : By the previous lemma the set $\{\gamma^+ \mid \gamma \in \Gamma\}$ is dense in $\partial_\infty \mathsf{H}^2$. Let now (x,y) be pair of points in $\partial_\infty \mathsf{H}^2 \times \partial_\infty \mathsf{H}^2$, we can therefore find a pair of distinct points (η^-,γ^+) associated to

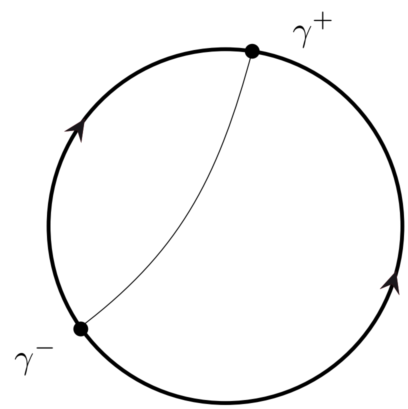

Figure 4.1: North-south dynamics

elements η and γ. We remark that if two elements α and β of the group are such that $\alpha^+ = \beta^+$ then $\alpha^- = \beta^-$ (*Hint:* use the compactness of H^2/Γ).

We therefore assume that all points $\eta^\pm, \gamma^\pm$ are distinct. The final remark is that

$$\lim_{n\to\infty} (\gamma^n \eta^n)^+ = \gamma^+,$$

and symmetrically

$$\lim_{n\to\infty} (\gamma^n \eta^n)^- = \eta^-,$$

The process is described in Figure 4.2 Let U be a small neighbourhood of γ^+. Since γ^+ is different than η^- a high power of η will send U to a very small neighbourhood V of η^+. Since η^+ is different than γ^- a high power of γ will send V to a even smaller neighbourhood of γ^+. It follows that $\xi_n = \gamma^n \eta^n$ maps U into itself. Therefore it has a fixed point in U. This point is necessarily the attractive fixed point of ξ_n. This is what we wanted to prove.

Q.E.D.

Here is another important consequence of this North-South dynamics.

Lemma 4.1.3 [Baby hyperbolic stability] *Let S be a compact surface. Let ρ_1 and ρ_2 be two representations of $\pi_1(S)$ in $\mathsf{PSL}(2,\mathbb{R})$ which are monodromies of hyperbolic structures on S. Then the two corresponding actions on $\partial_\infty \mathsf{H}^2$ are conjugate. More precisely there exists a unique – usually non smooth – homeomorphism Φ of $\partial_\infty \mathsf{H}^2$ so that*

$$\forall x \in \partial_\infty \mathsf{H}^2,\ \forall \gamma \in \pi_1(S),\ \Phi \circ \rho_1(\gamma)(x) = \rho_1 \circ \Phi(\gamma)(x).$$

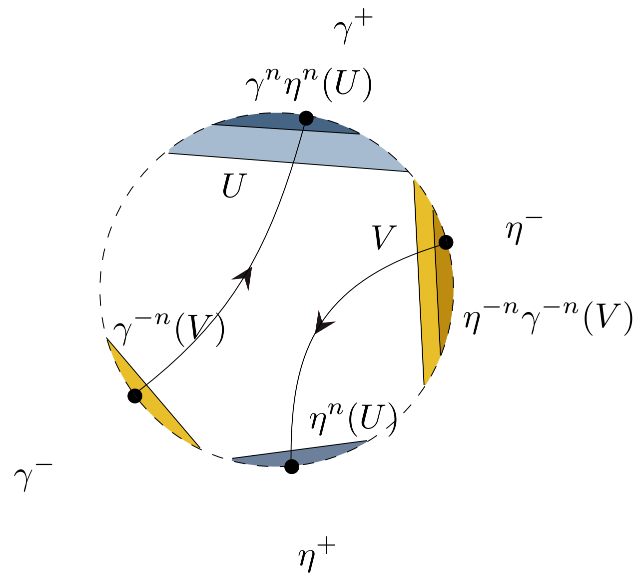

Figure 4.2: Density of pair of fixed points

For the memento we just prove the uniqueness of the conjugation and post-pone the proof of the existence.

Proof : Let E_1 be E_2 be the set of end points of closed geodesics in ∂_∞ of respectively $\rho_1(\pi_1(S))$ and $\rho(\pi_1(S))$.

Our first remark is that Φ satisfies from $\rho_1(\gamma)^+$ to $\rho_2(\gamma)^+$. Indeed, since Φ conjugate the action it sends attractive fixed points to attractive fixed points.

Then the uniqueness follows from the density of E_1. Q.E.D.

One can actually prove that the conjugacy is Hölder, this is the grown up version of Hyperbolic Stability.

This last lemma leads the an abstract definition of the boundary at infinity of a surface group.

Definition 4.1.4 *Let S be a closed connected oriented surface of genus greater than 2. The boundary at infinity $\partial_\infty\pi_1(S)$ of a surface group is a topological circle on which $\pi_1(S)$ in a way which is conjugate to the action of $\rho(\pi_1(S)$ on $\partial_\infty\mathsf{H}^2$, where ρ is the monodromy of a hyperbolic structure.*

There is a beautiful theorem by Matsumoto which characterizes the action of $\pi_1(S)$ on $\partial_\infty\pi_1(S)$.

Theorem 4.1.5 *Let S be a closed surface. Let T be a topological space homeomorphic to the circle $x \in \mathsf{U}S$. Assume that $\pi_1(S)$ acts on T, with the following properties*

- *each non trivial element has exactly one attractive and one repulsive fixed point,*

- *every orbit is dense*

then there is a homeomorphism conjugating the action of $\pi_1(S)$ *between* T *and* $\partial_\infty \pi_1(S)$.

4.2 The unit tangent bundle and flows

The group $\mathsf{PSL}(2,\mathbb{R})$ is the group of orientation preserving isometries of the hyperbolic plane. Moreover the group of orientation preserving isometries acts faithfully and transitively on the unit tangent bundle $\mathsf{U}S$ which we define equivalently as the set U of *pairs* (x, L), where x is a point on the oriented geodesic L, or the set $\mathcal{T}$ of oriented distinct triples (x, y, z) of points in the boundary at infinity.

The *geodesic flow* is the one parameter subgroup $\{\phi_t\}$ of homeomorphisms of $\mathsf{U}S$, that sends (x, L) to (y, L) so that $[L^-, x, L^+, y] = t$. The *stable horocycle flow* is the one parameter subgroup that sends (x, L) to (u, M) so that L and M has a common positive end point z, and x and u are joined by the horocycle centred at z of length one. The following commutation rules, see Figure 4.3 holds and identify these flows with 1-parameter groups of matrices.

$$\phi_t \circ H_s \circ \phi_{-t} \quad = \quad H_{e^t s}, \tag{4.1}$$
$$\tag{4.2}$$

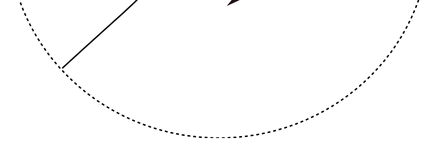

Figure 4.3: Commutation rule

Hence, the geodesic flows and horocycle flows are just the actions of the one-parameter diagonal and upper triangle groups of matrices on $\mathsf{PSL}(2,\mathbb{R})$ respectively.

Closed geodesics can – and will – now be interpreted as closed orbits of the geodesic flow.
Exercise Write the action of these flows on $\mathcal{T}$ using only the cross ratio.

4.2.1 The Anosov property

Let $\mathsf{U}S$ be the unitary tangent bundle of the surface S, which from the discussion above is a quotient of $\mathsf{PSL}(2,\mathbb{R})$ by a discrete group Γ. We therefore have three flows on $\mathsf{U}S$ and the corresponding foliations

1. The geodesic flow ϕ_t
2. The stable horocycle flow whose orbits we call *stable leaves*.
3. The unstable horocycle flow whose whose orbits we call *unstale stable leaves*, obtained by interchanging the role of end points.
4. The *central stable leaf* is the 2-dimensional leaf which is obtained as the orbit under the geodesic flow of the stable leaf.

Then the commutation rules (4.1) translate into the *Anosov property* of the geodesic flow, which we try to depict in Figure 4.4

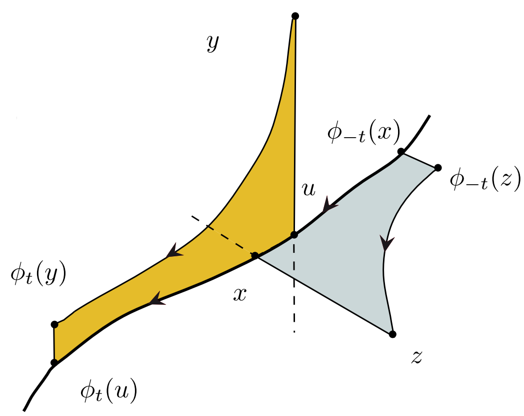

Figure 4.4: Anosov property

1. Two points on the same stable leaf get closer under a *positive* action of the geodesic flow.
2. Two points on the same unstable leaf get closer under a *negative* action of the geodesic flow.

This property is the translation for the geodesic flow of the north-south dynamics of the action of the fundamental group.

The Closing Lemma

The Anosov property has a crucial consequence.

Lemma 4.2.1 [Closing lemma] *For any α and T, there exists $\epsilon > 0$, so that if*

$$d(x, \phi_T(x)) < \epsilon$$

Then there exists y, with $d(x,y) < \alpha$, a positive number s with $|s - T| \leq \alpha$ so that $\phi_s(y) = y$.

Proof : We choose a neighborhood U of x and a parametrisation of this neighborhood by $]-\epsilon, \epsilon[^3$, given by

$$\psi \ : \ (u, v, w) \mapsto \phi_u \circ H^+_v \circ H^-_w(x).$$

Our first step is to prove the following fact:

There exists z such that $d(x,z) < \alpha$, a positive number s with $|s - T| \leq \alpha$ so that there exists u, with $|u| \leq \epsilon$ and

$$\phi_s(z) = H^+_u(z). \tag{4.3}$$

We can first assume, by slightly changing T,

$$\phi_T(x) = H^+_{u_0} \circ H^-_{v_0}(x),$$

for some small u and v.

It follows that for s, $|u| \leq \epsilon$, we have

$$\phi_T(H^+_s(x)) = H^+_{e^{-T}s + u_0} H^-_{v_0}(x).$$

Observe that the contracting map

$$s \mapsto e^{-T}s + u_0,$$

has a fixed point. This proves our first assertion.

Then one obtain the closing Lemma using the same argument but working backward in time. Q.e.d.

The closing lemma implies the density of the reunion of all closed orbits which is also a consequence of Lemma 4.1.2.

The Shadowing Lemma

We say a sequence of points $\{(x_n, T_n)\}_{0 \leq n \leq N}$ in $\mathsf{U}S \times \mathbb{R}$ is an (ϵ, T)-pseudo-orbit if

1. for all n, the distance $d(x_n \phi_T(x_{n-1})) \leq \epsilon$,
2. for all n, the distance $T_n \leq T$,

The following follows from a refinement of the arguments used in the Closing Lemma

Lemma 4.2.2 [Shadowing Lemma] *For every α, there exists some ϵ such that every ϵ-pseudo orbit is α close to an orbit.*

We only sketch the proof. Assume $T > 1$ to avoid having to take too much care of the constants.

We start by a preliminary lemma

Lemma 4.2.3 *Let x_1, x_2, x_3 be three points and T_1, T_2 two points such that $\phi_{T_i}(x_i)$ is ϵ close to x_{i+1}.*

Then there exists y and S_2, with $|S_1 - T_1| \leq \epsilon$

$$\begin{aligned}
\forall s, \quad 0 \leq s \leq \frac{T_1}{2}, \qquad & d(\phi_s(y), \phi_s(x_1)) \leq e^{-T/2}\epsilon, \\
\forall s, \quad \frac{T_1}{2} \leq s \leq T_1, \qquad & d(\phi_s(y), \phi_s(x_1)) \leq \epsilon, \\
\forall s, \quad S_1 \leq s \leq S_1 + \frac{T_2}{2}, \qquad & d(\phi_s(y), \phi_{T+s}(x_2)) \leq \epsilon, \\
\forall s, \quad S_1 + \frac{T_2}{2} \leq s \leq S_1 + T_2, \qquad & d(\phi_s(y), \phi_{T+s}(x_2)) \leq e^{-T/2}\epsilon.
\end{aligned}$$

Proof : The proof of this assertion now follows from similar ideas to the proof of the closing lemma:

We can write

$$x_2 = H_u^- \circ H_v^+ \circ \phi_{T+w}(x_1),$$

with $|u|$, $|v|$, $|w|$ smaller than ϵ.

Then we take $y = \phi_{-T}(z)$ where $z = H_v^+ \circ \phi_{T+w}(x_1)$. The assertion follows from the contraction property. Q.E.D.

We can now proceed to the proof of the Shadowing Lemma.

Proof : We give the rough idea. Assume now for simplicity that $N = 2^p$ and let $\{x_n\}_{0 \leq n \leq N}$ in $\mathsf{U}S$ be an (ϵ, T) pseudo orbit, with all $T_n = T$. The assertion above tells us that we can produce a is a $(\epsilon(1+e^{-T}), 2T)$ pseudo orbit $\{y_2 n\}_{0 \leq n \leq N/2}$. Furthermore the orbit arc $\phi_{[0,T]}(x_i)$ is $(\epsilon(1+e^{-T})$-close to $\phi_{[0,2T]}(y_{[i/2]})$.

We just continue the induction for one more step:

We produce an $(\epsilon(1+e^{-T}+e^{-2T}+e^{-3T}), 4T)$ pseudo orbit $\{z_{4n}\}_{0 \leq n \leq N/4}$ where furthermore $\phi_{[0,T]}(x_i)$ is $(\epsilon(1+e^{-T}+e^{-2T}+e^{-3T})$-close to $\phi_{[0,4T]}(z_{[i/4]}$.

Continuing the induction, we end up with X_0 and X_N which a (α, NT) pseudo-orbit. We ca so that all x_i are α close to the orbit of X_0, where

$$\alpha = \frac{\epsilon}{1 - e^{-T}}.$$

Q.E.D.

Hyperbolic stability at last

Now we can come back to the proof of the Hyperbolic Stability using the Shadow Lemma.

We are going to prove this in the case the corresponding hyperbolic metrics g and g' are close enough. The result would then follow using the fact that the space of hyperbolic metrics is connected.

The proof follows from the following lemma.

Lemma 4.2.4 *Let g and g' two close hyperbolic metrics on S. Let γ' be a geodesic of g'. Then there exists a unique geodetic γ for g which is at bounded distance of γ'.*

The uniqueness is obvious: two geodesics at a bounded distance coincide up to reparametrisation.

The conjugacy (check the details) is given by

$$\Psi(\gamma'(+\infty)) = \gamma(+\infty),$$

We leave the reader check the details which are easy:

- Ψ is well defined,
- Ψ is continous.

We now prove the lemma

PROOF : We denote by ϕ_t the geodesic flow of the first metric and by ϕ'_t the geodesic flow of the second metric. We denote by $\mathsf{U}S$ and $\mathsf{U}S$ the unit tangent bundle for g and g'. Observe that we have a natural map F – linear fiber by fiber – sending U_2S to U_1S.

Using F, we now consider ϕ^2_t as a flow on $U_1(S)$. Our hypothesis implies that ϕ'_1 is ϵ close to ϕ_1.

Then every geodesic γ' for g' defines a ϵ-pseudo orbit of ϕ_t, which is defined by

$$\{\phi'_n(\dot{\gamma}'(0)\}_{n\in\mathbb{N}}.$$

By the Shadow Lemma, this ϵ-pseudo orbit is close to a geodesic γ. Q.E.D.

4.3 Measures and Ergodic Theory

We give here as baby course on measure theoretical properties of the geodesic flow on surfaces. We refer the avid reader to Martine Babillot for a more thorough introduction to the subject.

4.3.1 Invariant measures

Let $\{\phi_t\}_{t\in\mathbb{R}}$ be a flow acting on a measure space. We say that a measure μ is invariant under the flow if for all real t and measurable set $A \subset X$, we have

$$\mu(A) = \mu(\phi_t(A)).$$

Equivalently, for a a compact topological space and a flow of homeomorphisms, a Radon measure is invariant if for all continuous function f and real t, we have

$$\int_X f \, \mathrm{d}\mu = \int_X f \circ \phi_t \, \mathrm{d}\mu.$$

We state now two important elementary results

Theorem 4.3.1 *Let X be a compact measure space and $\{\phi_t\}_{t\in\mathbb{R}}$ a flow of homeomorphisms. Then there exists a $\{\phi_t\}_{t\in\mathbb{R}}$ invariant measure on X.*

PROOF : Let ν be any probability measure on X. Let

$$\nu_t = \frac{1}{t}\int_X \phi_s^* \nu \ \mathrm{d}s.$$

Since X is compact, the set $\mathcal{M}(X)$ of Radon probability measures on X is weakly compact. In other words, there exists a probability measure μ on X, a sequence or real numbers $\{t_n\}_{n\in\mathbb{N}}$ converging to infinity such that for any continuous function f

$$\int_X f \ \mathrm{d}\mu = \lim_{n\to\infty}\int_X f \ \mathrm{d}\nu_{t_n} = \lim_{n\to\infty}\frac{1}{t_n}\int_X\int_0^{t_n} f\circ\phi_s \ \mathrm{d}\mu \ \mathrm{d}s. \tag{4.4}$$

Thus in particular for any real number u,

$$\begin{aligned}\int_X f\circ\phi_u \ \mathrm{d}\mu &= \lim_{n\to\infty}\frac{1}{t_n}\int_X\int_0^{t_n} f\circ\phi_{s+u} \ \mathrm{d}\mu \ \mathrm{d}s \\ &= \lim_{n\to\infty}\frac{1}{t_n}\int_X\int_u^{t_n+u} f\circ\phi_s \ \mathrm{d}\mu \ \mathrm{d}s. \end{aligned}\tag{4.5}$$

It follows that

$$\int_X f\circ\phi_u \ \mathrm{d}\mu - \int_X f \ \mathrm{d}\mu = \lim_{n\to\infty}\frac{1}{t_n}\int_X\left(\int_0^u f\circ\phi_s \ \mathrm{d}s - \int_{t_n}^{t_n+u} f\circ\phi_s \ \mathrm{d}s\right) \ \mathrm{d}\mu \tag{4.6}$$

since,

$$\left|\int_0^u\left(\int_X f\circ\phi_s \mathrm{d}\mu\right)\mathrm{d}s - \int_{t_n}^{t_n+u}\left(\int_X f\circ\phi_s\mathrm{d}\mu\right)\mathrm{d}s\right| \leq 2u\|f\|_\infty,$$

it follows that

$$\int_X f\circ\phi_u \ \mathrm{d}\mu - \int_X f \ \mathrm{d}\mu \quad = \quad 0. \tag{4.7}$$

The result follows. Q.E.D.

Let μ be a measure. Then the *support* of μ is the closed set $\mathrm{Supp}(\mu)$ whose complementary of the set

$$\bigcap_{U\in V} U,$$

where $V := \{U \mid U \text{ open}, \ \mu(U) = 0\}$. The second result is

Theorem 4.3.2 [POINCARÉ RECURRENCE THEOREM] *Let X be a space and $\{\phi_t\}_{t\in\mathbb{R}}$ a flow of homeomorphisms preserving a Radon measure probability measure μ. Let $x\in\mathrm{Supp}(\mu)$, then for any neighborhood U of x and positive T, there exists $t > T$ and $y\in U$ such that*

$$\phi_t(y)\in U.$$

PROOF : Let x and U such as in the theorem. We now that $\mu(U) \neq 0$. Moreover, since μ is invariant by ϕ_t, fro all t $\mu(\phi_t(U)) \neq 0$. The key observation is that there exists $n \neq p$ such that

$$\phi_{T^n}(U) \cap \phi_{T^p}(U) \neq \emptyset.$$

Indeed otherwise,

$$1 = \mu(X) \geq \mu\left(\bigcup_{n\in\mathbb{N}} \phi_{T^n}(U))\right) = \sum_{n\in\mathbb{N}} \mu\left(\phi_{T^n}(U)\right) = \infty.$$

It follows that there exists $p < 0$, such that

$$U \cap \phi_{T^{-p}}(U) \neq \emptyset.$$

Let thus $y \in U \cap \phi_{T^{-p}}(U)$, then by definition $\phi_{T^p}(y) \in U$ and $y \in U$. The theorem follows. Q.E.D.

4.3.2 Invariant measures by the geodesic flow

The unit tangent bundle of the hyperbolic space is identified with the group $\mathsf{PSL}(2,\mathbb{R})$. The group $\mathsf{PSL}(2,\mathbb{R})$ is *unimodular*: in other words the following is true

Proposition 4.3.3 *The group* $\mathsf{PSL}(2,\mathbb{R})$ *possesses a bi-invariant measure.*

PROOF : Let $\mathfrak{G}$ be the Lie algebra of $\mathsf{PSL}(2,\mathbb{R})$. Let $\mathrm{Det}(\mathfrak{G}) = \Lambda^3\mathfrak{G}^*$.

Since $\mathrm{Det}(\mathfrak{G})$ has dimension 1, and $\mathsf{PSL}(2,\mathbb{R})$ does not have any non trivial homomorphisms in $\mathbb{R}$, it follows that $\mathsf{PSL}(2,\mathbb{R})$ acts trivially on $\mathrm{Det}(\mathfrak{G})$. There exist thus an invariant volume form –hence a measure– bi-invariant on $\mathsf{PSL}(2,\mathbb{R})$. Q.E.D.

We just gave a rather general existence proof. In our case, $\mathfrak{G}$ is the set of trace free (2×2)-matrices on $\mathbb{R}^2$ a natural invariant volume form on $\mathcal{G}$ is

$$\Omega(A, B, C) := \mathrm{Trace}(A.B.C)$$

This corresponding measure is called the *Liouville measure*.

Otherwise, any closed geodesic γ defines an invariant measure μ_γ. This measure is the unique invariant probability measure whose support is that closed geodesic and is sometimes called the *Dirac measure* supported on the closed geodesic. This measure is defined as follows: for any continuous function f for any x in γ, we define

$$\int_X f \,\mathrm{d}\mu_\gamma = \frac{1}{\ell(\gamma)} \int_0^{\ell(\gamma)} f \circ \phi_s(x)\,\mathrm{d}s.$$

4.3.3 Ergodicity

The unit tangent bundle has a probability measure μ_0 – that we call the *Lebesgue* measure – which comes from the Haar measure of $\mathsf{PSL}(2,\mathbb{R})$ and which is invariant under the geodesic flow.

We say a probability measure μ is *ergodic* under the flow $\{\phi_t\}_{t\in\mathbb{R}}$ if for all invariant set A either $\mu(A)=0$ or $\mu(A)=1$. We state in these notes without proof two important results, the first one is relatively easy to prove.

Theorem 4.3.4 [Ergodic Decomposition Theorem]
Let $\{\phi_t\}_{t\in\mathbb{R}}$ be a flow acting on a compact space X. We denote by $\mathcal{M}(X)$ the convex set of probability Radon measures on X, and $\mathcal{M}_0(X)$ the subset of $\{\phi_t\}_{t\in\mathbb{R}}$ invariant measures. Let μ_0 be an element of $\mathcal{M}_0(X)$, then there exists a probability measure ν_0 on $\mathcal{M}_0(X)$ supported on ergodic measures so that

$$\mu_0 = \int_{\mathcal{M}_0(X)} \mu \, \mathrm{d}\nu_0(\mu).$$

The second one is a deeper result.

Theorem 4.3.5 [Birkhoff Ergodic Theorem] *Let $\{\phi_t\}_{t\in\mathbb{R}}$ be a flow acting on a compact space X. Let μ_0 be an ergodic probability Radon measures on X. Let f be a measurable function. Then there exists a set A of full measure in X so that for all $x\in A$, we have*

$$\lim_{t\to\infty}\left(\frac{1}{t}\int_0^t f(\phi_s(x))\,\mathrm{d}s\right) = \int_X f(x)\,\mathrm{d}\mu(x).$$

The quantity

$$\mathrm{M}f(x,t) := \frac{1}{t}\int_0^t f(\phi_s(x))\,\mathrm{d}s,$$

is called a *Birkhoff sum.*

As an exercise we shall prove *Von Neumann ergodic* L^2*-Theorem* where we further assume that f is in $\mathrm{L}^2(X,\mu_0)$.

We also leave as an exercise the following proposition which follows from Birkhoff ergodic theorem and the ergodic decomposition theorem

Proposition 4.3.6 *Let μ be a measure invariant by a flow $\{\phi_t\}_{t\in\mathbb{R}}$, Let f be a continuous function. Then there exists a set of μ full measure A an invariant function M_f on A, such that for all x in A,*

$$\lim_{t\to+\infty}\frac{1}{t}\int_0^t f\circ\phi_s(x)\ \mathrm{d}s = \lim_{t\to+\infty}\frac{1}{t}\int_0^t f\circ\phi_{-s}(x)\ \mathrm{d}s = M_f(x). \tag{4.8}$$

Moreover, if for every f there exists a subset of full measure $B\subset A$ on which M_f is constant, then μ is ergodic.

The rest of this paragraph is devoted to the proof of the following result.

Theorem 4.3.7 *Let S be a finite volume surface. Then the Lebesque measure is ergodic with respect to the geodesic flow.*

In Paragraph **??**, we shall give a proof of this result using considerations on unitary representations of $\mathsf{SL}(2,\mathbb{R})$ on $\mathrm{L}^2(\mathsf{U}S,\mu)$ and the fact that the geodesic flow comes from an action of $\mathsf{PSL}(2,\mathbb{R})$.

Hopf argument

We are now giving a proof of Theorem which can be extended (with some extra work) to general Anosov flows preserving a volume form.

Proof : We shall use a weak consequence of the Anosov property. We have these three foliations $\mathcal{L}^+$ the stable foliation, $\mathcal{L}^-$ the unstable foliation, and $\mathcal{L}^0$ the foliation by the orbit of the geodesic flow $\{\phi_t\}_{t\in\mathbb{R}}$. We denote by $\mathcal{L}^*_x$ the leaf of $\mathcal{L}^*$ passing through x. These foliations are locally a product, meaning that we can find for every x a neighborhood U of x so that we have the identification

$$U = (\mathcal{L}^+_x \cap U) \times (\mathcal{L}^-_x \cap U) \times (\mathcal{L}^0_x \cap U)$$

and that in this identification the three foliations come from the product structure.

Moroever (this is an important feature) these foliations are absolutely continuous with respect to the Liouville measure. This means that at least locally we can decompose the Liouville measure λ can be written in the coordinates that gives the product structure as

$$\lambda = \lambda^+ \otimes \lambda^- \otimes \lambda_0$$

This property, called the absolute continuity of the stable and unstable foliations is obvious in our case. In the general case of Anosov flows, this is a difficult theorem by Anosov. Assuming this theorem, ergodicity follows from the same scheme of ideas.

We now use Proposition 4.3.6 and consider the function M_f for a continuous function f. By assumption M_f is constant along the leaves of $\mathcal{L}_0$. We now prove that M_f is constant along the leaves of $\mathcal{L}^+$.

Since $\mathsf{U}S$ is compact, f is uniformly continuous. Thus f is bounded by K, and that for every ϵ there exists α such that

$$d(u,v) \leq \alpha \implies |f(u) - f(v)| \leq \epsilon.$$

Now let x and y belong to A and the same leaf of $\mathcal{L}^+$. In particular by definition, when $t > t_0$

$$d(\phi_t(x), \phi_t(y)) \leq \alpha.$$

It thus follows that considering the Birkhoff sums for $t > t_0$, we get

$$\begin{aligned} |Mf(x,t) - Mf(y,t)| &\leq \frac{1}{t}\int_0^{t_0} |f\circ\phi_s(x) - f\circ\phi_s(y)| \,\mathrm{d}s + \frac{1}{t}\int_{t_0}^{t} |f\circ\phi_s(x) - f\circ\phi_s(y)| \,\mathrm{d}s \\ &\leq \frac{K.t_0}{t} + \epsilon. \end{aligned}$$

It follows that for all ϵ

$$|M_f(x) - M_f(y)| \leq \epsilon.$$

And thus for all x, y in the same leaf of $\mathcal{L}^+$ then $M_f(x) = M_f(y)$. A similar argument works for $\mathcal{L}^-$.

Now we leave as an exercise the proof of the following fact: since A is of full measure and the foliations are absolutely continuous with respect to the Liouville measure, locally there exist three sets of full λ^*-measure B^* in $\mathcal{L}^*_x$ so that

$$B = B^+ \times B^- \times B^0 \subset A.$$

Then M_f is constant on B. Since B has full measure, this means by Proposition 4.3.6 that λ is ergodic. Q.E.D.

Actually, you can check that it suffices to use Von Neumann mean ergodic L^2 Theorem.

4.3.4 Ergodicity and mixing

Let $\{\phi_t\}_{t\in\mathbb{R}}$ be a flow acting on a compact space X, preserving a measure μ, we say that $\{\phi_t\}_{t\in\mathbb{R}}$ is *mixing* if for any positive functions f and g we have

$$\lim_{t\to\infty}\int_X (f\circ\phi_t).g \, \mathrm{d}\mu = \int_X f \, \mathrm{d}\mu \int_X g \, \mathrm{d}\mu.$$

Observe the following

Proposition 4.3.8 *Every mixing flow is ergodic.*

Proof : Let f be an invariant L^2-function. Then the integral $\int_X f\circ\phi_s.f\mathrm{d}\mu$ is constant in s. Assuming mixing we obtain that

$$\int |f|^2\mathrm{d}\mu = \left|\int f\mathrm{d}\mu\right|^2.$$

Thus f is constant. Q.E.D.

We now prove the following resut

Theorem 4.3.9 *The geodesic and horcyclic flows are mixing and ergodic.*

4.3.5 The spectral approach

Let X be a space equipped with a probability measure μ. Let $\{\phi_t\}_{t\in\mathbb{R}}$ be a flow preserving μ acting on X. Let $\mathrm{L}^2_0(X,\mu)$ be the vector subspace of $\mathrm{L}^2(X,\mu)$ consisting of functions whose integral is zero. Observe that any measure preserving mapping f from X to X define a unitary operator A_f on $\mathrm{L}^2_0(X,\mu)$ by $A_f : g \to g\circ f$. Let $U_t = A_{\phi_t}$.

We now observe the following

Proposition 4.3.10 [Spectral interpretation]
The flow $\{\phi_t\}_{t\in\mathbb{R}}$ is ergodic if and only if the one parameter group $\{U_t\}_{t\in\mathbb{R}}$ has no non trivial invariant vectors.
The flow $\{\phi_t\}_{t\in\mathbb{R}}$ is mixing if and only for any function f and g in $\mathrm{L}^2_0(X,\mu)$ we have

$$\lim_{t\to\infty}\langle U_t.f,g\rangle = 0.$$

Thus Theorem 4.3.9 (as well as the ergodicity of any non compact group!) follows at once from the following result

Theorem 4.3.11 [Moore] *Assume that we have a strongly continuous unitary representation π of $\mathsf{SL}(2,\mathbb{R})$ on a Hilbert space H. Assume that π has no non trivial invariant vector, then if g_n is a diverging sequence in $\mathsf{SL}(2,\mathbb{R})$ then for any f and g in H,*

$$\lim_{t\to\infty}\langle \pi(g_n)f,h\rangle = 0.$$

We recall that a representation π of a topological group G on a Hilbert space is *unitary and strongly continuous*, if for every g in G, $\pi(g)$ is a unitary operator ion H, and moreover for every $f\in H$ the map $g\mapsto \pi(g).f$ is continuous.

Corollary 4.3.12 *If the group $\mathsf{SL}(2,\mathbb{R})$ acts ergodically on a space X preserving probability measure, then every non compact subgroup acts ergodically and is mixing.*

In particular since $\mathsf{SL}(2,\mathbb{R})$ acts transitively on $\mathsf{U}S$, it acts ergodically and Theorem 4.3.9 follows.

4.3.6 Proof of Moore's Theorem

Our first Lemma is the following

Lemma 4.3.13 [Mautner phenomenon] *Let f be an element in $\mathcal{H}$ so that $\{a_{t_n}f\}_{n\in\mathbb{R}}$ weakly converges to f_0. Then f_0 is invariant under the one parameter group $\{n_t\}_{t\in\mathbb{R}}$.*

Recall that u_n weakly converges to u, if for all z,

$$\lim_{n\to\infty}\langle u_n,z\rangle = \langle u,z\rangle$$

Proof : We have for any g

$$\begin{aligned}
|\langle\pi(n_s)f_0,g\rangle - \langle f_0,g\rangle| &= \lim_{k\to\infty}\left(|\langle\pi(n_s a_{t_k})f,g\rangle - \langle\pi(a_{t_k})f,g\rangle|\right)\\
&= \lim_{k\to\infty}\left(|\langle\pi(a_{-t_k}n_s a_{t_k})f,\pi(a_{-t_k})g\rangle - \langle f,\pi(a_{-t_k})g\rangle|\right)\\
&\leq \lim_{k\to\infty}\left(\|\pi(a_{-t_k}n_s a_{t_k})f - f\|\right).\|g\| = 0,
\end{aligned} \tag{4.9}$$

Since

$$\lim_{k\to\infty} a_{-t_k}n_s a_{t_k} = \lim_{k\to\infty} n_{e^{-t_k}s} = 1,$$

the result follows by the definition of weak continuity. Q.e.d.

Our second Lemma is the following

Lemma 4.3.14 *Let f be an element in $\mathcal{H}$ invariant under the one parameter group $\{n_t\}_{t\in\mathbb{R}}$. Then f is invariant by* $\mathsf{SL}(2,\mathbb{R})$.

Proof : Let f be any vector. Let ϕ be the continuous function on $\mathsf{SL}(2,\mathbb{R})$ given by

$$\phi_f(g) = \langle \pi(g)f, f\rangle.$$

Then the following are equivalent for any closed subgroup Q of $\mathsf{SL}(2,\mathbb{R})$

1. $\phi_f|_Q = 1$
2. ϕ_f is Q-binvariant: for all $h \in H$ and $g \in G$, $\phi_f(hg) = \phi_f(gh) = \phi_f(g)$
3. f is $\pi(Q)$-invariant.

It is obvious that (3) $\implies$ (2) $\implies$ (1). Then (1) implies (3) since for a unitary operator u

$$\|u(f) - f\|^2 = 2\left(\|u\|^2 - \langle u(f), f\rangle\right).$$

We can now proceed to the proof and let f be a N-invariant vector. Let P be the group generated by N and $\{a_t\}_{t\in\mathbb{R}}$.

Since f is invariant by $N = \{n_t\}_{t\in\mathbb{R}}$. It follows that f is a left and right N invariant function on $\mathsf{SL}(2,\mathbb{R})$. Now an easy exercise show that any left and right invariant N function on $\mathsf{SL}(2,\mathbb{R})$ is constant on P (Hint: $\mathsf{SL}(2,\mathbb{R})/N = \mathbb{R}^2 \setminus \{0\}$). It follows that f is invariant by P, and thus ϕ_f is biinvraiant by P. Again an easy exercise show that ϕ_f is constant (Hint: $\mathsf{SL}(2,\mathbb{R})/P = \mathbb{RP}^1$). Q.e.d.

Now we can proceed to the proof by contradiction. Assume that there exists F and G in $\mathcal{H}$, a diverging subsequence $\{u_n\}_{n\in\mathbb{N}}$ in G so that $\langle \pi(u_n)f, g\rangle$ does not converge to zero. We can write

$$u_n = k_n.a_{t_n}.\bar{k}_n,$$

where k_n and $\bar{k}_n$ belongs to S^1. Thus we can as well assume that $\bar{k}_n$ and k_n converges to $\bar{k}_0$ and k_0 respectively. Thus, after taking $f = \pi(\bar{k}_0)F$ and $g = \pi(k_0^{-1})G$, we deduce that $\langle \pi(a_{t_n})f, g\rangle$ does not converge to zero. By the weak compactness theorem, after extracting a subsequence, we can as well assume that $\pi(a_{t_n})f$ converges weakly to f_0.

By the Mautner phenomenon, f_0 is invariant by $\{n_t\}_{t\in\mathbb{R}}$. By the second Lemma f_0 is invariant by $\mathsf{SL}(2,\mathbb{R})$. Thus $f_0 = 0$.

4.3.7 Equidistribution and growth of geodesics

The unit tangent bundle has a probability measure μ_0 – that we call the *Lebesgue* measure – which comes from the Haar measure of $\mathsf{PSL}(2,\mathbb{R})$ and which is invariant under the geodesic flow. Every closed geodesic γ also defines a geodesic flow invariant probability measure μ_γ on $\mathsf{U}S$ by the formula

$$\int f \, \mathrm{d}\mu_\gamma = \frac{1}{\ell(\gamma)} \int_0^{\ell(\gamma)} f(\phi_t(x)) \, \mathrm{d}t,$$

where $\ell(\gamma)$ is the length of γ and x is any point of γ.

These measures are related by the following deep result

Theorem 4.3.15 [Bowen, Margulis] *The closed geodesics are* equidistributed *with respect to the Lebesgue measure:*

$$\lim_{T\to\infty} \frac{1}{\sharp(\Gamma_T)} \left(\sum_{\gamma\in\Gamma_T} \mu_\gamma \right) = \mu_0,$$

where Γ_T is the set of closed geodesics of length smaller than T.

Moreover, the following theorem counts asymptotically number of closed geodesics.

Theorem 4.3.16 [Margulis] *Let Γ_T be the set of closed geodesics of length smaller than T. Then*

$$\lim_{T\to\infty} 2Te^{-T}\sharp(\Gamma_T) = 1.$$

4.3.8 Unique ergodicity and complements

A flow is said to be *uniquely ergodic* if is possesses a unique invariant measure. By the ergodic decomposition theorem, such a measure is necessarily ergodic. So equivalently a flow is to uniquely ergodic if is possesses a unique ergodic invariant measure. Obviously, the geodesic flow is uniquely ergodic: all closed geodesics define invariant measures. A deeper result says

Theorem 4.3.17 [Furstenberg] *The horocyclic flow is uniquely ergodic for a finite volume surface.*

All the previous results have extensions in higher dimensions and for general Anosov flows.

Chapter 5

Discrete subgroups and closed surfaces

We now assume that S is oriented.

5.1 Monodromies of hyperbolic structures

Every surface gives rise to an embedding of the fundamental group of $\pi_1(S)$ with discrete image, moreover this group has no torsion. Indeed, every torsion element of $\mathsf{PSL}(2,\mathbb{R})$ fixes a point in the hyperbolic plane.

Conversely every torsion free subgroup of $\mathsf{PSL}(2,\mathbb{R})$ acts properly freely on the hyperbolic plane and is the monodromy of a – non necessarily compact – hyperbolic surface. In order to complete the picture, we recall

Lemma 5.1.1 [Selberg] *Every finitely generated linear group possesses a finite index subgroup without torsion.*

It follows that every faithful representation of the fundamental group of a surface with discrete image is the monodromy of a hyperbolic structure. A little extra work shows that if moreover S is compact then $\mathsf{H}^2/\rho(\pi_1(S)$ is homeomorphic to S, so therefore

Proposition 5.1.2 *Every faithful representation of the fundamental group of a compact surface S with discrete image is the monodromy a hyperbolic structure on S.*

Exercise: show that this last statement fails for a non compact surface.

Proposition 5.1.3 [Borel density Theorem] *Every monodromy of a finite volume hyperbolic surface is Zariski dense.*

Proof : We prove it for a compact surface $S = \mathsf{H}^2/\Gamma$. From the description of hyperbolic surfaces, it follows that we can find two elements in Γ which generate different hyperbolic translations. Then, considering the very short list of algebraic subgroups of $\mathsf{PSL}(2,\mathbb{R})$, we see that this cannot happen unless the algebraic group is $\mathsf{PSL}(2,\mathbb{R})$ itself Q.e.d.

5.1.1 Teichmüller space and space of representations

The *Teichmüller space* $\tau(S)$ is the space of all representations of $\pi_1(S)$ with discrete image up to conjugacy. It is diffeomorphic to a ball, we almost proved it ... . Let us introduce an invariant of representation of $\pi_1(S)$. For that, let us choose a presentation of $\pi_1(S)$

$$\pi_1(S) = \langle a_1, b_1, \ldots, a_g, b_g \mid \prod_{i=1} g[a_i, b_i] = 1\rangle,$$

where $[c,d] = cdc^{-1}d^{-1}$.

Observe now that $\mathsf{PSL}(2,\mathbb{R})/\mathrm{SO}(2) = \mathsf{H}^2$, hence that $\mathsf{PSL}(2,\mathbb{R})$ has the homotopy type of S^1. We therefore have an exact sequence

$$\mathbb{Z} \to \widetilde{\mathsf{PSL}(2,\mathbb{R})} \to \mathsf{PSL}(2,\mathbb{R}) \to 0.$$

Let us choose a map σ from $\mathsf{PSL}(2,R)$ to $\widetilde{\mathsf{PSL}(2,\mathbb{R})}$ that splits that sequence, σ will actually never be continuous, nor a group morphism. Then we have

Proposition 5.1.4 [Euler Class] *Let ρ be a representation of $\pi_1(S)$ to $\mathsf{PSL}(2,\mathbb{R})$ The element*

$$e(\rho) = \prod_{i=1} g[\sigma(\rho(a_i)), \sigma(\rho(b_i)],$$

is an element of the centre of $\widetilde{\mathsf{PSL}(2,R)}$ which we identify to $\mathbb{Z}$. This number is independent of the choice of σ, of the presentation of $\pi_1(S)$ and is constant under local deformations of ρ. The number $e(\rho)$ is called the Euler class *of the representation.*

Let now $\chi(S)$ be the Euler characteristics of S. Then

Theorem 5.1.5 [Milnor-Wood inequality] *Let ρ be a representation of $\pi_1(S)$ to $\mathsf{PSL}(2,\mathbb{R})$ Then*

$$|e(\rho)| \leq |\chi(S)|.$$

We can use the Euler class to distinguish connected connected components of space of representations. More precisely

Theorem 5.1.6 [Goldman] *The map from the space of connected components of* $\mathrm{Hom}(\pi_1(S), \mathsf{PSL}(2,\mathbb{R}))$ *to $\{\chi(S), \chi(S)+1, \ldots, -\chi(S)\}$ is a bijection. Moreover, mondromies of hyperbolic structures are exactly representations such that $|e(\rho)| = |\chi(S)|$.*

It follows from this theorem that we can check whether a representation is the monodromy of a hyperbolic structure just from a presentation of the group.

5.2 Arithmetic surfaces

5.2.1 Commensurator group

Let S be a hyperbolic surface $S = \mathsf{H}^2/\Gamma$. The commensurator group of a subgroup of Γ is

$$\mathrm{Comm}(\Gamma) = \{g \in \mathsf{PSL}(2,\mathbb{R}) \mid \sharp(g\Gamma g^{-1} \cap \Gamma)\backslash\Gamma) < \infty\}.$$

We check for instance that

$$\mathrm{Comm}(\mathsf{PSL}(2,\mathbb{Z})) = \mathsf{PSL}(2,\mathbb{Q}).$$

5.2.2 Arithmetic hyperbolic surfaces

By a non standard definition that rests on a deep result by Margulis, we define an *arithmetic hyperbolic surface* to be a finite volume hyperbolic surface whose commensurator of the fundamental group is dense in $\mathsf{PSL}(2,\mathbb{R})$.

Arithmetic hyperbolic surfaces are classified, they are more or less obtained through the following procedure: first we find a representation of $\mathsf{PSL}(2,\mathbb{Q})$ in $\mathsf{PSL}(n,\mathbb{Q})$ then Γ is the intersection of the image of $\mathsf{PSL}(2,\mathbb{Q})$ with $\mathsf{PSL}(n,\mathbb{Z})$. It is not straightforward to construct compact arithmetic surfaces but they exist. Moreover, by a simple cardinality argument, non arithmetic surfaces also exist.

Here is another remark.

Lemma 5.2.1 *If a surface is not arithmetic, then* Γ *has finite index in* $\mathrm{Comm}(\Gamma)$.

PROOF : Let H the closure of $\mathrm{Comm}(\Gamma)$. We first prove that it is discrete: indeed otherwise, every element in $\mathrm{Comm}(\Gamma)$ would fix the Lie algebra of the connected component of H of the origin. But this defines a Zariski closed condition. By Borel density theorem, $H = \mathsf{PSL}(2,\mathbb{R})$ hence S is arithmetic which is a contradiction. It follows that $\mathrm{Comm}(\Gamma)$ is discrete. Then we have a covering map from $\mathsf{PSL}(2,\mathbb{R})/\Gamma$, which is compact, to $\mathsf{PSL}(2,\mathbb{R})/\mathrm{Comm}(\Gamma)$. Hence, the fibres of this map are finite sets which exactly means that Γ has finite index in $\mathrm{Comm}(\Gamma)$. Q.E.D.

5.2.3 Hecke correspondences and arithmetic dynamics

The main feature of arithmetic surfaces are the existence of many correspondences. A *finite correspondence* between two sets X and Y is a subset of Z in the product $X \times Y$ so that the preimage of very point in each of the factor is finite and non empty. In particular, an element g in the commensurator group of a hyperbolic surface S gives rise to such a correspondence which is furthermore a local isometry see Figure 5.1. Indeed, we consider the map of H^2 into $S \times S$ given by $x \to (\pi(x), \pi(gx))$ where π is the covering map. To say g is in the commensurator group, is just to say that this map is a covering map of compact image and that its image is a correspondence Z_g. Actually, the correspondence is just determined by the class of g in $\Gamma\backslash \mathrm{Comm}(\Gamma)/\Gamma$.

We now define *correspondences for hyperbolic surfaces* to be correspondences which are local isometries. The dichotomy between arithmetic surfaces and non arithmetic surfaces is then the dichotomy between finitely many and infinitely many self correspondences.

A self correspondence gives rise to two types of dynamics. First *quantum dynamics* acting on the space of L^2 functions on S. So if p_1 and p_2 are the two projections – of degree q – of the correspondence $Z \subset S \times S$ on each factor, then

$$\mathcal{H}_g(f)(x) = \frac{1}{q} \sum_{z \in p_1(x)} f(p_2(z),$$

is a self adjoint operator called the *Hecke operator* of the correspondence.

Secondly, we can associate *classical dynamics*. There is a classical dynamical way to turn non bijective map or more generally correspondence into a bijective map. So, to settle notation,

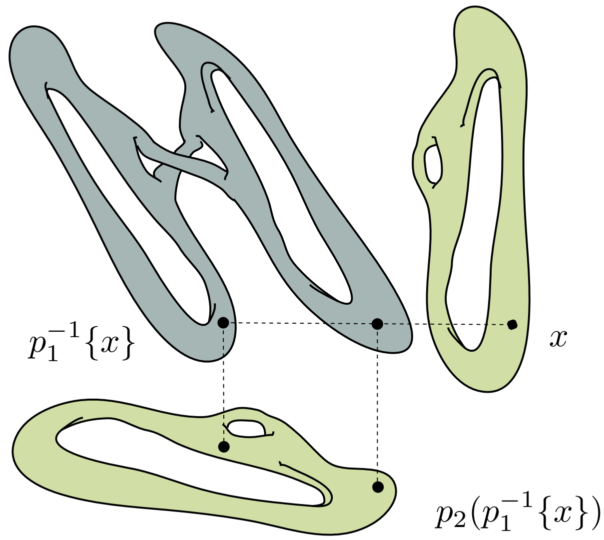

Figure 5.1: A correspondence

let Z be a correspondence and p_1 and p_2 be the two projections, we say $x\mathcal{R}y$ if $p_1^{-1}(x)$ intersects $p_2^{-1}(y)$. Then we consider the set

$$\mathcal{L}_Z = \{f : \mathbb{Z} \to S \mid f(n)\mathcal{R}f(n+1)\} \subset S^{\mathbb{Z}}.$$

The *shift* σ is the map from $\mathcal{L}_Z$ to itself given by $\sigma(f)(n) = f(n+1)$. The shift is now a homeomorphism and its dynamics reflect that of the correspondence.

This construction is not sufficient for our purpose. We indeed would like to see all correspondences as acting in the same space. Let us first take a look at the space $\mathcal{L}_Z$ in special case.

Let $P_{[n,p]}$ be the map from $\mathcal{L}_Z$ to S^{n-p} given by

$$f \to (f(n), \dots, f(p)).$$

By construction the image of this map is a compact surface, which we call $S_{[n,p]}$, in the product and moreover any restriction map is a covering map. We can therefore describe $\mathcal{L}_Z$ as a limit of some coverings.

We generalise this construction. Let S be a compact hyperbolic surface. The *Sullivan solenoid* $\mathcal{S}(S)$ is the "limit" of all coverings of S. Let us give a definition as a set. The *Sullivan solenoid* is the set of sequences $\{(x_n, S_n, p_n)\}_{n\in\mathbb{N}^*}$ so that x_n is a point in S_n, p_n is a covering from S_n to S_{n-1} – where by convention $S_0 = S$ – such that $p_n(x_n) = x_{n-1}$ up to the following *solenoid equivalence*: two sequences $\{(x_n^0, S_n^0, p_n^0)\}_{n\in\mathbb{N}}$ and $\{(x_n^1, S_n^1, p_n^1)\}_{n\in\mathbb{N}}$ are equivalent if there exists a third one $\{(y_n, \Sigma_n, q_n)\}_{n\in\mathbb{N}}$ together with covering maps q_n^i from Σ_n to S_n^i satisfying the commuting diagram conditions

$$\begin{aligned} q^i(n)(y_n) &= x_n^i, \\ p_n^i \circ q_{n+1}^i &= q_n^i \circ p_{n+1}^i. \end{aligned}$$

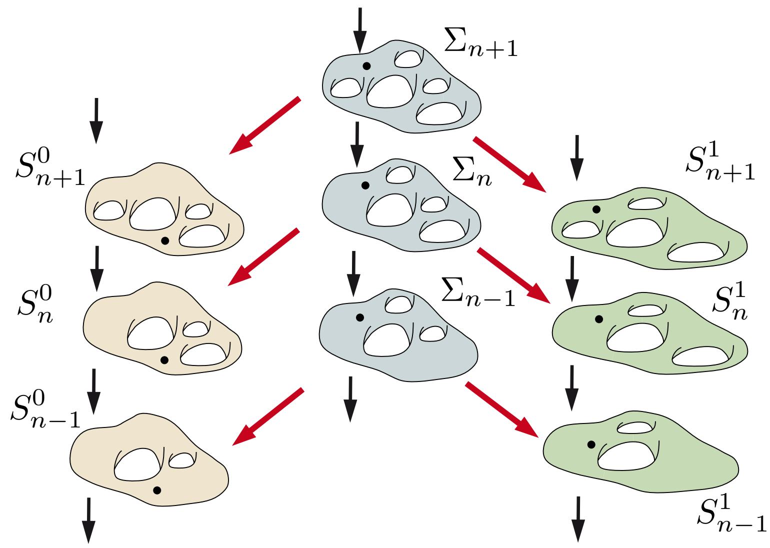

Figure 5.2: Solenoid equivalence

We give two alternate descriptions that describe the topology of this set

1. The combinatorial construction runs as follows. Let W_S be the 2-dimensional complex, whose vertices are surfaces which are finite covers of S, oriented edges correspond to covering between the extremities, and faces correspond to commuting diagrams of coverings. Let Z_S be the universal cover of this complex, and V_S be the set of vertices of this graph, which we consider as surfaces. If e is an edge of V_S from e^- to e^+, then it gives rise to a covering p_e from e^- to e^+ seen as surfaces. Then

$$\mathcal{S}(S) = \{(x_\Sigma)_{i\in V_S} \mid x_\Sigma \in \Sigma, \quad p_e(x_{e^+}) = x_{e^-}\}.$$

2. Alternatively, we can describe the Sullivan solenoid as fibre bundle over S whose structure group is the pro-finite completion of the fundamental group of S. The Sullivan solenoid is the "universal cover for finite covers", in the sense that it solves a universal problem for finite covers of S.

As a topological space, the *Sullivan solenoid* $\mathcal{S}(S)$ is a *hyperbolic laminated space* as in Figure 5.3: every point has a neighbourhood – called chart – which is homeomorphic to a product a ball in H^2 with a topological space, such that moreover the coordinates changes when we change charts are isometries on the hyperbolic factor. A *leafwise isometry* of a hyperbolic laminated space is a homeomorphism that is a local isometry on the hyperbolic factors.

Now, here is the fundamental though obvious remark.

Lemma 5.2.2 *Let p a local isometry from S to Σ. Then the corresponding injection of $\mathcal{S}(S)$ to $\mathcal{S}(\Sigma)$ is a leafwise isometry.*

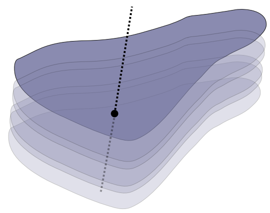

Figure 5.3: A small open set in a laminated space

PROOF : use the fact that you can induce coverings. Q.E.D.

As a consequence, a correspondence on a surface acts as a leafwise isometry on the Sullivan solenoid. Therefore, the Sullivan solenoid of an arithmetic surface has very rich dynamics.

As an exercise, we ask the reader to interpret the universal solenoid for a finite index torsion free subgroup of $\mathsf{PSL}(2,\mathbb{Z})$ in adélic terms.

Chapter 6

Harmonic functions

6.1 Harmonic functions

We finally move to the last topic of these notes. We encourage strongly the reader to have a look at N. Bergeron beautiful set of notes on the Laplacian on hyperbolic surfaces. We say a function f on S is an *eigenfunction* of the Laplacian of eigenvalue λ if $\Delta(f) = \lambda f$. The multiplicity of the eigenvalue λ is the dimension of the space of eigenfunctions. The Laplacian is defined locally as in the hyperbolic plane. Eigenfunctions could also have been defined just using integration on small discs and balls. Since S is assumed to be compact a general theorem asserts that the sum of the multiplicity of eigenvalues less than a given value is finite.

6.1.1 The trace formula

The length of closed geodesics and the eigenvalues of the Laplacian are related by many deep results. *Selberg trace formula* is certainly the most striking. Selberg trace formula is a generalisation of Poisson summation formula, it reads

Theorem 6.1.1 [Selberg-trace version] *Let S be a compact hyperbolic surface and h be an even test function satisfying some restriction. Let $\{\lambda_n\}$ be the set of eigenfunctions of the Laplacian. Let $\mu_n^2 + 1/4 = \lambda_n$, with either the imaginary part or the real part of μ_n is positive. Let $\mathcal{G}$ be the set of closed geodesics and $\ell(\gamma)$ be the length of the closed geodesic γ and $m(\gamma)$ be the multiplicity of γ. Then*

$$\sum_{n=0}^{+\infty} h(\mu_n) = -\frac{\chi(S)}{2}\int_{-\infty}^{+\infty} h(s)s\tanh(\pi s)ds + \sum_{\gamma\in\mathcal{G}} R_\gamma \hat{h}(\ell(\gamma)),$$

where

$$R_\gamma = \frac{\ell(\gamma)}{m(\gamma)2\sinh(\ell(\gamma)/2)}, \quad \hat{h}(s) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} h(u)e^{-ius}du.$$

We now state it using a slightly non standard approach due to Cartier and Voros. We need to introduce two generalised zeta functions.

1. The *generalised ζ function* reflects the analytic side. It is a defined as

$$\zeta(s,a) = \mathrm{Tr}(\Delta_S + a)^{-s} := \sum_{i=0}^{\infty} \frac{1}{(\lambda_n + a)^s},$$

   where (λ_n) is the set of eigenvalues repeated with multiplicities.

2. The *Selberg zeta function* reflects the dynamical side. Let $\mathcal{P}$ be the set of *primitive* closed geodesics. Let us define

$$z_S(s) = \prod_{\gamma \in \mathcal{P}} \prod_{k=0}^{\infty} \left(1 - e^{\ell(\gamma)(k+s)}\right).$$

   where λ_n is the set of eigenvalues of the Laplacian on S.

Then we have after taking the analytic continuation of these functions.

Theorem 6.1.2 [SELBERG-DETERMINANT VERSION] *Let S be a compact hyperbolic surface. Then*

$$\frac{\partial}{\partial s}\Big|_{s=0} \zeta_S\left(s, u^2 - \frac{1}{4}\right) = \psi(u)^{\chi(S)} z_S\left(\frac{1}{2} + u\right),$$

where $\psi(u)$ is an explicit function only depending on u, which can be interpreted as related to a spectral problem on the two-sphere. The left hand side term is usually interpreted as a regularised determinant.

Chapter 7

Conjectures

We can now state two very famous conjectures on hyperbolic surfaces, and a slighty less well known one which is still interesting and has been settled recently. The *quantum unique ergocity conjecture* is the quantum pendent of the equidistribution of the orbits of the geodesic flows. This conjecture due to Sarnak claims the following

Conjecture 7.0.3 [SARNAK] *Let $\{\phi_n\}$ be a sequence of eigenfunctions of the Laplacian on a compact hyperbolic surface S, such that the corresponding eigenvalues go to infinity. Let μ be the hyperbolic measure on S, then for all continous function f, we have*

$$\lim_{n\to\infty} \frac{1}{\int_S |\phi_n|^2 d\mu} \int_S f|\phi_n|^2 d\mu = \int f d\mu.$$

This conjecture is known to be "almost surely true" by Schnirelman, Zelditch and Colin de Verdière as the *quantum ergodicity theorem* in the sense that it converges for a – and then plenty – subsequence of density 1 in the sequence of functions. The following breakthrough has been obtained recently

Theorem 7.0.4 [E. LINDENSTRAUSS] *Let S be an arithmetic surface. Then the quantum ergodicity conjecture holds if we furthermore assume the sequence of functions are Hecke eigenfunctions.*

The approach is ergodic and uses the extra dynamics coming from Hecke correspondences on a space related to the solenoid.

Conjecture 7.0.5 [SELBERG] *The first eigenvalue of the Laplacian on $\mathsf{H}^2/\Gamma_0(N)$ is greater than $1/4$.*

Conjecture 7.0.6 [EHRENPREIS] *For any ϵ, for any pair of compact hyperbolic surfaces, there exists a common covering such that the two induced distances are $1+\epsilon$ bi-Lipschitz equivalent.*

This conjecture has a nice dynamical translation on the Teichmüller space of the solenoid. It has been recently proved by Jeremy Kahn and Vladimir Markovic in a series of beautiful articles.