Quantum chaos on hyperbolic surfaces

Etienne Le Masson

Contents

CHAPTER 0

Introduction

To explain the general problem of quantum chaos, let us consider M a compact Riemannian manifold (for example a compact surface) which we see as the physical space where a particle evolves. The classical and quantum description can be roughly summarised in the following table.

	Classical mechanics	Quantum mechanics
State	$(x, \xi) \in TM$, a pair in the tangent bundle (phase space) representing the position x and the momentum ξ of a particle.	$\varphi \in L^2(M)$, $\|\varphi\|_2 = 1$, a complex wave function. The quantity $\|\varphi\|^2$ gives the probability density for the particle's position
Observable	$a : TM \to \mathbb{C}$, a function on the phase space. For example the energy of a free particle is given by $a(x, \xi) = \|\xi\|^2$.	$A : L_2(M) \to L_2(M)$, a self-adjoint operator on $L^2(M)$. The spectrum gives the possible values for the observable and the eigenfunctions the corresponding quantum state. For ex. The Laplacian Δ is the energy operator for a free particle.
Dynamics	$\Phi_t : TM \to TM$ a flow on phase space. In particular the geodesic flow: the map $t \mapsto \Phi_t(x, \xi)$ is the unique geodesic passing through x and tangent to ξ.	$U_t : L_2(M) \to L_2(M)$ a unitary group giving the quantum evolution. For example $U_t = e^{it\Delta}$ is the quantization of the geodesic flow
Chaos	A manifestation of chaos is the ergodicity of the geodesic flow $\frac{1}{T}\int_0^T a \circ \Phi_t(x_0, \xi_0)dt \to \int_{\|\xi\|=\|\xi_0\|} a(x, \xi)\, dx d\xi$ when $T \to \infty$. It says that the classical trajectories equidistribute.	Quantum chaos: what are the consequences of classical chaos on the quantum side? Concretely on the spectrum and eigenfunctions of the Laplacian.

Quantum chaos thus tries to make a link between the dynamical properties of the space and the spectral properties of the Laplacian. More generally this connection between the geometry and the spectral theory of the Laplacian is the main object of *spectral geometry*. It arises from the fact that the Laplacian in the simplest differential operator that respects the symmetries (it commutes with the isometries) and this explains its presence in most of the fundamental equations of physics.

We will be interested in the geodesic flow and the Laplacian on hyperbolic surfaces. Hyperbolic surfaces provide a concrete model for chaotic dynamics, and allow to define a lot of objects explicitly. Yet this setting is sufficiently rich and still not completely understood. It is also a bridge between quantum chaos and number theory, via the theory of automorphic forms and arithmetic quantum chaos. We will introduce the latter in the last part of the course.

CHAPTER 1

Hyperbolic geometry

In this chapter we introduce some elements of hyperbolic geometry in dimension 2. A more complete reference is the book of Katok [**Kat92**].

The hyperbolic plane is a model of two dimensional non-Euclidean geometry, where the *parallel postulate*

> Given a line and a point not on it, there is only one line parallel to the given line that can be drawn through the point.

is replaced by the following

> Given a line and a point not on it, **at least two** lines parallel to the given line can be drawn through the point.

Some consequences of this change are that two parallel lines (or *geodesics*) do not stay equidistant but tend to separate from each other, and the sum of angles of a triangle is strictly less than π.

In the same way that flat tori can be seen as quotients of the Euclidean plane $\mathbb{R}^2$ by a lattice such as $\mathbb{Z}^2$, we will define hyperbolic surfaces as a quotient of the hyperbolic plane by a hyperbolic lattice (or Fuchsian group). The geodesic flow on these surfaces constitutes a model of chaotic dynamical system.

1.1. Hyperbolic plane

The hyperbolic plane can be defined abstractly as the unique simply connected Riemannian manifold of dimension 2 of constant curvature -1. We will work concretely with the upper half plane model:

$$\mathbb{H} = \{z = x + iy \,|\, y > 0\}.$$

The tangent bundle is given by

$$T\mathbb{H} = \{(z, v) : z \in \mathbb{H}, v \in \mathbb{C}\},$$

and the Riemannian metric, for $v, w \in T_z\mathbb{H}$ by

$$\langle v, w\rangle_z = \frac{v\bar{w}}{\mathrm{Im}\,(z)^2}.$$

We will write

$$\|v\|_z = \sqrt{\langle v, v\rangle_z} = \frac{|v|}{\mathrm{Im}\,(z)}.$$

DEFINITION 1.1. A *curve* is a piecewise differentiable function $\gamma : [a, b] \to \mathbb{H}$. The *length* of this curve is given by

$$L(\gamma) = \int_a^b \|\gamma'(t)\|_{\gamma(t)}\, dt.$$

We define a distance $d : \mathbb{H} \times \mathbb{H} \to [0, +\infty)$ on $\mathbb{H}$ such that for any $z, z' \in \mathbb{H}$,

$$d(z, z') = \inf_{\gamma} L(\gamma),$$

where the infimum runs over all curves $\gamma : [a, b] \to \mathbb{H}$ such that $\gamma(a) = z$ and $\gamma(b) = z'$.

1.2. Isometries

The *isometry group* of $\mathbb{H}$ is the set $\mathrm{Isom}(\mathbb{H})$ of smooth maps $\varphi : \mathbb{H} \to \mathbb{H}$, equipped with the composition operation and preserving the riemannian metric, i.e. such that

$$\|d_z\varphi(v)\|_{\varphi(z)} = \|v\|_z.$$

We define an action of $\mathsf{SL}(2,\mathbb{R})$ on the extended complex plane $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ via *Möbius transformations.* For any $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathsf{SL}(2,\mathbb{R})$,

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

if $z \in \mathbb{C}$, and

$$g\, \infty = \frac{a}{c}.$$

Note that if $c = 0$, then ∞ is a fixed point of g. The two matrices g and $-g$ define the same Möbius transformation. Therefore the group of Möbius transformations is identified to

$$\mathsf{PSL}(2,\mathbb{R}) = \mathsf{SL}(2,\mathbb{R})/\{\pm 1\}.$$

Lemma 1.2. $\mathsf{SL}(2,\mathbb{R})$ *acts transitively on* $\mathbb{H}$.

$$\mathbb{H} \simeq \mathsf{SL}(2,\mathbb{R})/\,\mathsf{SO}(2),$$

where $\mathsf{SO}(2) = \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} : \theta \in \mathbb{R} \right\}$.

Proof. First note that if $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathsf{SL}(2,\mathbb{R})$, we have

$$\mathrm{Im}\,(gz) = \frac{\mathrm{Im}\,(z)}{|cz+d|^2}. \tag{1.1}$$

So $\mathsf{SL}(2,\mathbb{R})$ acts on $\mathbb{H}$. Now we have

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} i = x + iy,$$

and therefore the action is transitive. It can be checked that the stabilizer of i is

$$\mathrm{Stab}(i) = \mathsf{SO}(2),$$

which gives the isomorphism. □

Remark 1.3. We deduce from the previous proof the Iwasawa decomposition of $\mathsf{SL}(2,\mathbb{R})$. Any element can be written uniquely as a product

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

with $x, y \in \mathbb{R}, y > 0, \theta \in [0, 2\pi)$.

Lemma 1.4. $\mathsf{PSL}(2,\mathbb{R}) \subset \mathrm{Isom}(\mathbb{H})$

Proof. Let $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathsf{PSL}(2,\mathbb{R})$. The differential of the map $z \mapsto gz = \frac{az+b}{cz+d}$ is given by

$$d_z g = \frac{a(cz+d) - c(az+b)}{(cz+d)^2} = \frac{1}{(cz+d)^2},$$

so we have

$$\|d_z g(v)\|_{gz} = \frac{|v|}{|cz+d|^2} \frac{1}{\mathrm{Im}\,(gz)} = \frac{|v|}{\mathrm{Im}\,(z)} = \|v\|_z,$$

using (1.1). □

Remark 1.5. The full group of isometries is $\mathrm{PS^*L}(2,\mathbb{R}) = \mathrm{S^*L}(2,\mathbb{R})/\{\pm 1\}$, where $\mathrm{S^*L}(2,\mathbb{R})$ is the group of real matrices g with $\det(g) = \pm 1$. See for example Katok. It is generated by $\mathsf{PSL}(2,\mathbb{R})$ together with the reflection $z \mapsto -\bar{z}$.

1.3. Geodesics

Definition 1.6. Let $\gamma : \mathbb{R} \to \mathbb{H}$ be a curve such that for any $t, t' \in \mathbb{R}$,

$$L(\gamma_{|[t,t']}) = d(\gamma(t), \gamma(t')). \tag{1.2}$$

Then the image of γ in $\mathbb{H}$ is called a *geodesic*.

There is a one-to-one correspondence between geodesics and curves γ satisfying (1.2) and parametrised by the length, i.e. such that

$$L(\gamma_{|[t,t']}) = |t - t'|.$$

We will thus identify geodesics to the corresponding curve parametrised by the length.

Lemma 1.7. *The geodesics are the vertical lines and the semi-circles centred on the real axis.*

Proof. Let us fix two points z_1 and z_2 and find the geodesic passing through these two points. Assume first that $z_1 = ia$ and $z_2 = ib$ are pure imaginary, with $b > a > 0$. The length of a curve $\gamma : I \to \mathbb{H}$, $\gamma(t) = x(t) + iy(t)$ between these two points satisfies

$$\begin{aligned} L(\gamma) &= \int_I \frac{\sqrt{x'(t)^2 + y'(t)^2}}{y(t)}\, dt \\ &\geq \int_I \frac{|y'(t)|}{y(t)}\, dt \\ &= \int_a^b \frac{dy}{y} = \log(b/a). \end{aligned}$$

And $\log(b/a)$ is the hyperbolic length of the segment of the y-axis joining ia and ib. As this is valid for any point of the imaginary axis, we deduce that this axis is a geodesic.

Now for two arbitrary points $z_1, z_2 \in \mathbb{H}$, we denote by $\mathcal{L}$ the unique vertical line or circle centered on the real axis passing through these two points. There exists a transformation in $\mathsf{PSL}(2,\mathbb{R})$ mapping $\mathcal{L}$ to the imaginary axis. Indeed, if $\mathcal{L}$ is a vertical line crossing the real axis at a point a, we can take $z \mapsto z - a$. If $\mathcal{L}$ is a circle and b is one of its end points on the real axis, the transformation $z \mapsto -\frac{1}{z-b} \in \mathsf{PSL}(2,\mathbb{R})$ maps it to a vertical line and we are back to the first case. As these transformations are isometries, we deduce that $\mathcal{L}$ minimizes the distance between any two of its point, and thus is a geodesic. ☐

Lemma 1.8. *For $z, z' \in \mathbb{H}$, we have*

$$\cosh d(z, z') = 1 + \frac{|z - z'|^2}{2\mathrm{Im}\,(z)\mathrm{Im}\,(z')}. \tag{1.3}$$

Proof. We already know that for any $g \in \mathsf{PSL}(2,\mathbb{R})$, $d(gz, gz') = d(z, z')$, so the left-hand side is invariant by the action of $\mathsf{PSL}(2,\mathbb{R})$. It is a simple exercise to check that the right-hand side is also invariant. As in the previous proof we can thus choose a $g \in \mathsf{PSL}(2,\mathbb{R})$ that brings the geodesic passing through z and z' to the imaginary axis, and assume that $z = ia, z' = ib$, $b > a > 0$. In this case, we have seen also in the previous proof that $d(z, z') = \log(b/a)$ and it is easy to verify that the equality (1.3) holds. ☐

1.4. Geodesic flow

We define the *unit tangent bundle* as

$$T^1\mathbb{H} = \{(z,v) : z \in \mathbb{H}, v \in T_z\mathbb{H}, \|v\|_z = 1\}.$$

The geodesic flow is a one parameter family of maps

$$a_t : T^1\mathbb{H} \to T^1\mathbb{H},$$

for $t \in \mathbb{R}$, such that for any $(z,v) \in T^1\mathbb{H}$, $(\gamma(t), \gamma'(t)) = a_t(z,v)$ is the unique geodesic parametrised by the arc length satisfying

$$\gamma(0) = z, \gamma'(0) = v.$$

Lemma 1.9. $T^1\mathbb{H} \simeq \mathsf{PSL}(2,\mathbb{R})$.

Proof. The group $\mathsf{PSL}(2,\mathbb{R})$ acts on $T^1\mathbb{H}$

$$\left(g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, (z,v)\right) \to (g \cdot z, d_z g(v)) = \left(\frac{az+b}{cz+d}, \frac{v}{(cz+d)^2}\right).$$

A convenient way to see the homeomorphism is to use the *Iwasawa decomposition*. Any matrix in $\mathsf{PSL}(2,\mathbb{R})$ can be written as the product

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{1.4}$$

with $x \in \mathbb{R}$, $y \in \mathbb{R}_+$ and $\theta \in [0,\pi)$. We have

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} (i,i) = \left(x+iy, ye^{i2\theta}i\right),$$

which defines the homeomorphism between $\mathsf{PSL}(2,\mathbb{R})$ and $T^1\mathbb{H}$. □

The geodesic $t \mapsto a_t(i,i)$ is the vertical line corresponding to the imaginary axis. As $d(i, i\alpha) = \log\alpha$ for any $\alpha > 1$, we have

$$a_t(i,i) = (e^t i, e^t i) = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} (i,i).$$

And for any $(z,v) \in T^1\mathbb{H}$, if we let $g \in \mathsf{PSL}(2,\mathbb{R})$ such that $g \cdot (i,i) = (z,v)$, we have

$$a_t(z,v) = a_t(g \cdot (i,i)) = g \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} (i,i).$$

So via the identification $T^1\mathbb{H} \simeq \mathsf{PSL}(2,\mathbb{R})$, the geodesic flow is given by the right action of the one parameter subgroup

$$A = \left\{ a_t = \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}, \quad t \in \mathbb{R} \right\}.$$

1.5. Hyperbolic area and Gauss-Bonnet formula

The hyperbolic area corresponding to the metric we defined is given by

$$\frac{dx\,dy}{y^2}.$$

In particular it is invariant under the action of $\mathsf{PSL}(2,\mathbb{R})$, which can also be checked directly.

Theorem 1.10 (Gauss-Bonnet formula). *If T is a hyperbolic triangle with angles α, β, γ, then the area of T is given by*

$$|T| = \pi - \alpha - \beta - \gamma.$$

PROOF. First assume that the triangle has one vertex in $\mathbb{R} \cup \{\infty\}$, or equivalently that one angle is equal to 0. By applying isometries, we can further assume that this vertex is at ∞ and the other two are on the unit circle.

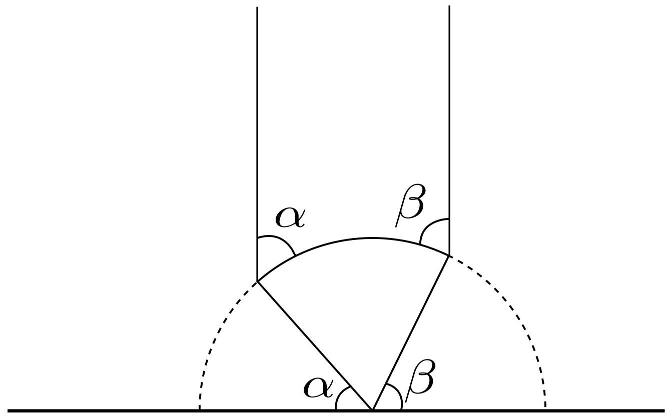


FIGURE 1.

Using the fact that Euclidean angles and hyperbolic angles are the same in the upper-half plane, we have (see Figure 1)

$$\begin{aligned} |T| &= \int_{\cos(\pi-\alpha)}^{\cos\beta} \int_{\sqrt{1-x^2}}^{\infty} \frac{dx\,dy}{y^2} \\ &= \int_{\cos(\pi-\alpha)}^{\cos\beta} \frac{dx}{\sqrt{1-x^2}} \\ &= \pi - \alpha - \beta \end{aligned}$$

We can write a general triangle as the difference of the areas of two triangles with one vertex in $\mathbb{R} \cup \{\infty\}$ as in Figure 2. We find that in this case the area is

$$|T| = \pi - \alpha - (\gamma + \delta) - (\pi - \delta - (\pi - \beta)) = \pi - \alpha - \beta - \gamma.$$

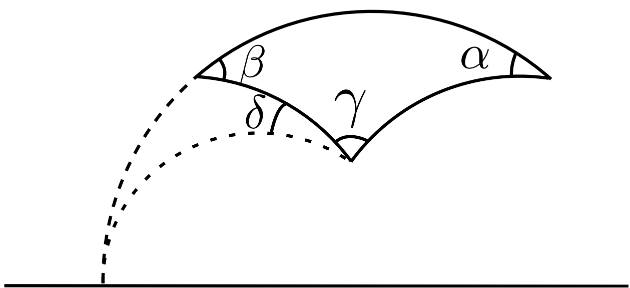


FIGURE 2.

□

COROLLARY 1.11. *The area of a hyperbolic n-gon with angles $\alpha_1, \ldots, \alpha_n$ is*

$$(n-2)\pi - \alpha_1 - \ldots - \alpha_n.$$

1.6. Classification of isometries

The isometries can be classified according to their conjugacy class. The characteristic polynomial of an element $\gamma \in \mathsf{SL}(2,\mathbb{R})$ is given by

$$X^2 - \mathrm{Tr}(\gamma)X + 1,$$

and has a discriminant equal to $\mathrm{Tr}(\gamma)^2 - 4$. Note that the fixed points of an isometry are obtained by solving the equation

$$\frac{az+b}{cz+d} = z \iff cz^2 + (d-a)z - b = 0.$$

Using the fact that $ad - bc = 1$ we find that the discriminant of this equation is given also by

$$(a+d)^2 - 4.$$

The trace of the isometry will thus determine the number and type of fixed points and the conjugacy class for the isometry. Remembering that ∞ is a fixed point iff $c = 0$, we have:

- If $|\operatorname{Tr}\gamma| = 2$, γ has one fixed point in $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\}$ and is called a *parabolic* transformation. There exists $s \in \mathbb{R}$ such that it is conjugate to
$$\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$$
- If $|\operatorname{Tr}\gamma| > 2$, γ has two fixed point in $\widehat{\mathbb{R}}$ and is called *hyperbolic*. There exists $t \in \mathbb{R}$ such that it is conjugate to
$$\begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}$$
- If $|\operatorname{Tr}\gamma| < 2$, γ has one fixed point in the upper half plane (and its conjugate in the lower half plane) and is called *elliptic*. There exists $\theta \in \mathbb{R}$ such that it is conjugate to
$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Definition 1.12. The *length* of an isometry $\gamma \in \mathsf{PSL}(2,\mathbb{R})$ is the non-negative real number l_γ defined by

$$l_\gamma = \inf_{z\in\mathbb{H}} d(z, \gamma z).$$

It can be checked that if γ is elliptic or parabolic, then $l_\gamma = 0$. The only positive length is obtained for hyperbolic transformations, and for a conjugate γ of $\begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix}$ we have $l_\gamma = t$.

1.7. Hyperbolic surfaces

Let Γ be a discrete subgroup of $\mathsf{PSL}(2,\mathbb{R})$ (also called a *Fuchsian group*), a hyperbolic surface is a quotient $M = \Gamma\backslash\mathbb{H}$. Let us just mention that if Γ does not contain any elliptic element, then M is a Riemannian manifold. Otherwise, there are singular points, and M is a more general object called an *orbifold*. In most of the course, we assume Γ does not have any elliptic elements, except in the particular and important case of the modular surface that we introduce at the end of the section.

Definition 1.13. $F \subset \mathbb{H}$ is a *fundamental domain* for Γ if

(1) $\bigcup_{\gamma\in\Gamma} \gamma\overline{F} = \mathbb{H}$
(2) $\forall \gamma \neq \gamma' \in \Gamma, \quad \gamma\mathring{F} \cap \gamma'\mathring{F} = \varnothing$.

EXERCISE 1.14. If F_1 and F_2 are two fundamental domains of Γ and $\mu(F_1) < \infty$, then $\mu(F_1) = \mu(F_2)$.

A special case is given by a *Dirichlet fundamental domain*. For any point $p \in \mathbb{H}$ not fixed by an element of $\Gamma \setminus \{e\}$, it is defined as

$$D = D_p(\Gamma) = \{z \in \mathbb{H} : d(z, p) < d(z, \gamma p) \quad \forall \gamma \in \Gamma \setminus \{e\}\}.$$

We define the *perpendicular bisector* of a geodesic segment $[z_1, z_2]$ as the unique geodesic through z, the mid-point of $[z_1, z_2]$ orthogonal to $[z_1, z_2]$. It is equal to the line

$$\{z \in \mathbb{H} : d(z, z_1) = d(z, z_2)\}.$$

Let us denote by $H_p(\gamma)$ the hyperbolic half plane

$$H_p(\gamma) = \{z \in \mathbb{H} : d(z, p) \leq d(\gamma z, p)\}.$$

The Dirichlet domain can thus be written as

$$D_p(\Gamma) = \bigcap_{\gamma \in \Gamma \setminus \{\mathrm{id}\}} H_p(\gamma).$$

The Dirichlet fundamental domain is therefore a hyperbolically convex set bounded by segments of geodesics and possibly segments of $\mathbb{R} \cup \{\infty\}$. Before proving that this set is indeed a fundamental domain, we need the following lemma.

LEMMA 1.15. *A discrete subgroup Γ of $\mathsf{SL}(2, \mathbb{R})$ acts properly discontinuously. That is for every $z \in \mathbb{H}$ and compact set $K \subset \mathbb{H}$, $\{g \in \Gamma : gz \in K\}$ is finite.*

PROOF. As Γ is discrete, we just need to show that for any $z \in \mathbb{H}$, the map $\varphi_z : g \mapsto gz$ is proper (i.e. the pre-image of a compact set is compact), as we have

$$\{g \in \Gamma : gz \in K\} = \varphi_z^{-1}(K).$$

Without loss of generality, let us show it for $g \mapsto gi$. The Iwasawa decomposition (1.4) of $\mathsf{SL}(2, \mathbb{R})$ gives us the map

$$g = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} y^{1/2} & 0 \\ 0 & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \mapsto gi = x + iy.$$

If $K \subset \mathbb{H}$ is compact, the pre-image by this map is homeomorphic to $K \times \mathsf{SO}(2)$, which is compact. □

PROPOSITION 1.16. *D is a fundamental domain.*

PROOF. We need to show first that $\mathbb{H} \subset \bigcup_{\gamma \in \Gamma} \gamma \overline{D}$. In other words that if $z \in \mathbb{H}$, then there exists $z_0 \in \overline{D}$ such that $z \in \Gamma z_0$. As the action of Γ on $\mathbb{H}$ is properly discontinuous, for any $z \in \mathbb{H}$, the orbit Γz is discrete, and thus there exists $z_0 \in \Gamma z$ such that

$$d(z_0, p) \leq d(\gamma z, p) = d(z, \gamma^{-1} p),$$

for any $\gamma \in \Gamma$. By definition $z_0 \in D$, and it is therefore the element we were looking for.

Now let us show that if $z_1, z_2 \in \mathring{D}$, they cannot be on the same orbit. Suppose they are, that is there exists $\gamma \in \Gamma \setminus \{1\}$ such that $z_1 = \gamma z_2$. Then

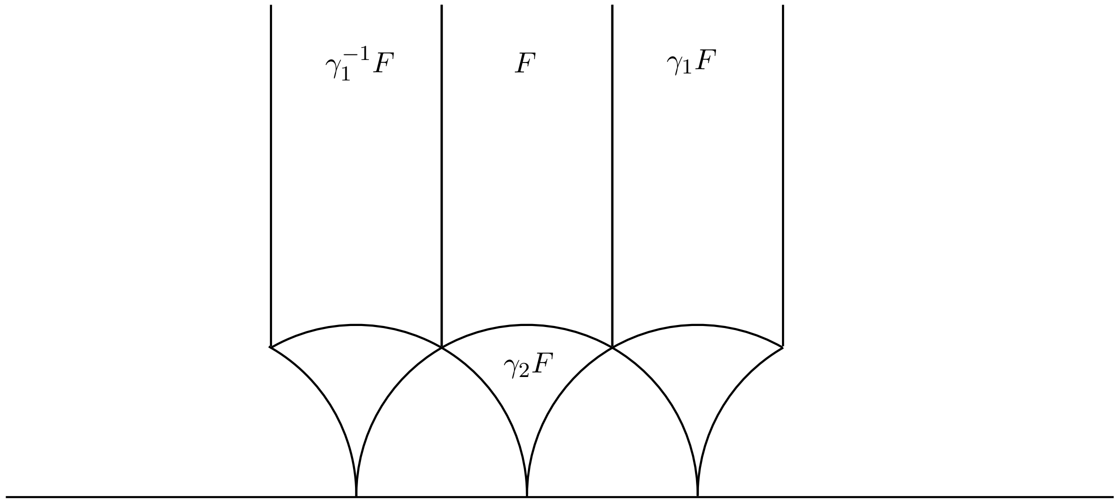
$$d(z_1, p) < d(z_1, \gamma^{-1} p) = d(z_2, p).$$

And similarly $d(z_2, p) < d(z_1, p)$ which is a contradiction. □

EXAMPLE 1.17 (Modular surface). An important example is the modular surface, defined by taking $\Gamma = \mathsf{PSL}(2, \mathbb{Z})$.

Let us determine the Dirichlet domain with respect to $p = 2i$. The group Γ contains the two isometries $\gamma_1 : z \mapsto z + 1$ and $\gamma_2 : z \mapsto -1/z$, so clearly $D_p(\Gamma) \subset F$, where

$$F = \{z \in \mathbb{H} : |\mathrm{Re}\,(z)| \leq 1/2, |z| \geq 1\}.$$



Now assume that $F \neq D$, then there exists $z \in \mathring{F}$ and $\gamma \in \Gamma \setminus \{\mathrm{id}\}$ such that $\gamma z \in \mathring{F}$. We write $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and we have

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

Since $|z| > 1$ and $|\mathrm{Re}\,(z)| < 1/2$,

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

Since $c, d \in \mathbb{Z}$ and $(c,d) \neq (0,0)$, $|cz+d| > 1$. Hence $\mathrm{Im}\,(\gamma z) < \mathrm{Im}\,(z)$. We can apply the same argument to γ^{-1} and obtain $\mathrm{Im}\,(z) < \mathrm{Im}\,(\gamma z)$, which is a contradiction.

To finish this chapter we define the notion of lattice and state two additional results that we are not going to prove (see [**Kat92**]).

Definition 1.18. A discrete subgroup $\Gamma < \mathsf{PSL}(2,\mathbb{R})$ is a *lattice* if it has a fundamental domain F of finite area.

Theorem 1.19 (Siegel). *If Γ is a lattice, then any Dirichlet fundamental domain has finitely many sides.*

Theorem 1.20. *If Γ is co-compact (Γ is also called* uniform*), meaning that the quotient $\Gamma\backslash\mathbb{H}$ is compact, then Γ has no parabolic element.*

CHAPTER 2

Geodesic and horocycle flows

A good reference for this chapter is the book of Einsiedler and Ward [**EW11**].

2.1. Definitions and geometric interpretation

We have seen that the unit tangent bundle

$$T^1\mathbb{H} = \{(z,v) : z = x+iy \in \mathbb{H}, v = ye^{i\theta}\theta \in [0,2\pi)\} \simeq \mathbb{H}\times\mathbb{S}^1$$

can be identified to $\mathsf{PSL}(2,\mathbb{R})$ via the map

$$\begin{aligned}\mathbb{H}\times\mathbb{S}^1 &\to \mathsf{PSL}(2,\mathbb{R})\\ (x+iy,\theta) &\mapsto \begin{pmatrix}1 & x\\ 0 & 1\end{pmatrix}\begin{pmatrix}y^{1/2} & 0\\ 0 & y^{-1/2}\end{pmatrix}\begin{pmatrix}\cos(\theta/2) & \sin(\theta/2)\\ -\sin(\theta/2) & \cos(\theta/2)\end{pmatrix}.\end{aligned}$$

We define the measure

$$d\mu = \frac{dx\,dy}{y^2}d\theta$$

on $T^1\mathbb{H}$. It can be checked that it is invariant with respect to the action of $\mathsf{PSL}(2,\mathbb{R})$. It also gives a left and right-invariant measure on $\mathsf{PSL}(2,\mathbb{R})$ called the *Haar measure.*

The geodesic flow on $\mathsf{PSL}(2,\mathbb{R})$ is given by the action on the right of the one parameter subgroup

$$A = \left\{a_t = \begin{pmatrix}e^{t/2} & 0\\ 0 & e^{-t/2}\end{pmatrix}, \quad t\in\mathbb{R}\right\}.$$

Similarly, we define the *(stable) horocycle flow* as the action of the one parameter subgroup

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

also called the *upper unipotent matrices* group. The *unstable horocycle flow* is associated to the subgroup of lower unipotent matrices

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

As an element $g \in \mathsf{PSL}(2,\mathbb{R})$ is identified to a unit vector $(x,v)\in T^1\mathbb{H}$, on the upper half-plane the geodesic flow is given by $a_t(x,v) = ga_t(i,i)$ and the two horocycle flows by $u_s^\pm(x,v) = gu_s^\pm(i,i)$. The latter move vectors along *horocycles*, that is straight horizontal lines and circles tangent to the real axis in $\mathbb{H}$ (see Figure 2.1).

Now let $\Gamma < \mathsf{PSL}(2,\mathbb{R})$ be a lattice and $M = \Gamma\backslash\mathbb{H}$ the corresponding hyperbolic surface. The unit tangent bundle of the surface $X = T^1M$ is identified with $\Gamma\backslash\mathsf{PSL}(2,\mathbb{R})$. The measure $d\mu$ gives a measure on the quotient, and the flows also pass to the quotient, for example the geodesic flow:

$$\begin{aligned}\Gamma\backslash\mathsf{PSL}(2,\mathbb{R}) &\to \Gamma\backslash\mathsf{PSL}(2,\mathbb{R})\\ \Gamma g &\mapsto \Gamma g\, a_t.\end{aligned}$$

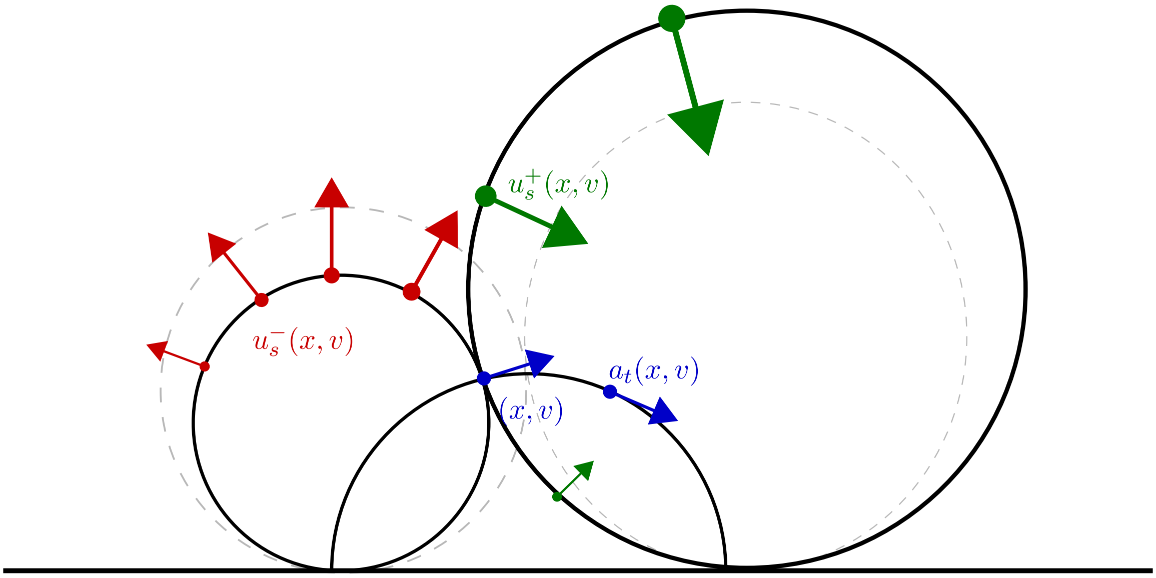


FIGURE 1. The geodesic flow a_t, the stable and unstable horocycle flows $u_s^\pm$ acting on a unit tangent vector (x, v).

REMARK 2.1. These notions generalise to $\mathsf{SL}(n, \mathbb{R})$ and abstract Lie groups. The general study of the action of diagonal and unipotent subgroups of Lie groups constitute the field of *homogeneous dynamics*. We will see an application of one of these results to the problem of Quantum Unique Ergodicity. These tools have also been successfully applied to other problems in mathematical physics and number theory (the resolution of the Oppenheim conjecture by Margulis is another famous example).

2.2. Mixing

DEFINITION 2.2. Let $(X, \mathcal{B}, \mu)$ be a probability space and $g_t : X \to X$ a measure-preserving flow, i.e. such that $g_t^* \mu = \mu$,

(1) The flow g_t is *ergodic* if for every measurable set $A \subset X$ which is g_t invariant, $\mu(A) = 0$ or $\mu(A) = 1$.
(2) The flow g_t is *mixing* if for every measurable sets $A, B \subset X$

$$\mu(A \cap g_t^{-1} B) \to \mu(A)\mu(B),$$

when $t \to \infty$.

These notions are two manifestations of chaotic behaviour of a classical system. Note that mixing implies ergodicity and is thus a stronger form of chaos. A way to understand ergodicity in the physical picture is the following theorem of Birkhoff about the equidistribution of trajectories in phase space.

THEOREM 2.3 (Birkhoff). *The flow g_t is ergodic iff $\forall \varphi \in L^1(X)$,*

$$\frac{1}{T} \int_0^T \varphi \circ g_t(x) dt \to \int_X \varphi \, d\mu,$$

when $T \to +\infty$, for μ-a.e $x \in X$.

The main theorem of this chapter is the following.

THEOREM 2.4. *The geodesic and horocyclic flows are mixing.*

This theorem is a direct consequence of a more general result about matrix coefficients of representations of subgroups of $\mathsf{SL}(2,\mathbb{R})$. A *unitary representation* of a group G on a Hilbert space $\mathcal{H}$ is a group homomorphism

$$\pi : G \to \mathsf{U}(\mathcal{H}),$$

where $\mathsf{U}(\mathcal{H})$ is the group of unitary transformations, so that we have $\|\pi(g)\varphi\| = \|\varphi\|$ for any $\varphi \in \mathcal{H}$. It is *strongly continuous* if for any $\varphi \in \mathcal{H}$, the map $g \mapsto \pi(g)\varphi$ is continuous. For $\varphi, \psi \in \mathcal{H}$, the function $g \mapsto \langle \pi(g)\varphi, \psi\rangle$ is called a *matrix coefficient* of the representation.

In our case, π will be a representation of $G = \mathsf{SL}(2,\mathbb{R})$ on

$$\mathcal{H} = L^2_0(X) = \{\varphi \in L_2(X) : \int_X \varphi \, d\mu = 0\},$$

with $X = \Gamma \backslash \mathsf{SL}(2,\mathbb{R})$, defined by

$$\pi(g)\varphi(h) = \varphi(hg)$$

for any $g \in \mathsf{SL}(2,\mathbb{R})$ and $\varphi \in L^2_0(X)$. A one parameter subgroup of $\mathsf{SL}(2,\mathbb{R})$ defines a flow on X by right multiplication, as we have seen with the geodesic and horocyclic flow. We can translate the definitions of ergodicity and mixing in the new context.

Lemma 2.5. *Let U_t be a one parameter subgroup of* $\mathsf{SL}(2,\mathbb{R})$

(1) *The flow associated to U_t is* ergodic *if and only if for any $t \in \mathbb{R}$, $\pi(U_t)$ has no non-trivial invariant vector.*
(2) *The flow associated to U_t is* mixing *if and only if for any $\varphi, \psi \in L^2_0(X)$,*

$$\langle \pi(U_t)\varphi, \psi\rangle \to 0$$

when $t \to \infty$.

The more general theorem from which we will deduce the mixing of the geodesic and horocyclic flow is then.

Theorem 2.6 (Howe-Moore). *Let π be a strongly continuous unitary representation of $\mathsf{SL}(2,\mathbb{R})$ on a Hilbert space $\mathcal{H}$. Assume that π has no non-trivial invariant vector in $\mathcal{H}$. Then if g_n is a diverging*[1] *sequence in $\mathsf{SL}(2,\mathbb{R})$, then*

$$\lim_{n\to+\infty} \langle \pi(g_n)\varphi, \psi\rangle = 0, \tag{2.1}$$

for any $\varphi, \psi \in \mathcal{H}$.

In our case $\mathsf{SL}(2,\mathbb{R})$ acts transitively on X, so any invariant function has to be constant, and is thus equal to 0 in $\mathcal{H} = L^0_2(X)$. So π has no non trivial invariant vector. As $a_t \to \infty$ and $u^\pm_s \to \infty$ when $s, t \to +\infty$, the Howe-Moore theorem implies the mixing of the geodesic and horocycle flows.

We prove the theorem using two lemmas

Lemma 2.7. *Let $\varphi \in \mathcal{H}$ such that the sequence $\{\pi(a_{t_n})\varphi\}_n$ converges weakly to an element φ_0. Then φ_0 is invariant under the group U.*

Lemma 2.8 (Mautner phenomenon). *If $\varphi \in \mathcal{H}$ is invariant under U, then it is invariant under* $\mathsf{SL}(2,\mathbb{R})$.

Let us first show that the two lemmas imply (2.1) for a_t. Suppose there exists φ, ψ and a sequence $t_n \to \infty$ such that $\langle \pi(a_{t_n})\varphi, \psi\rangle$ does not converge to 0.

We have $\|\pi(a_{t_n})\varphi\| = \|\varphi\|$, so passing to a subsequence, we can assume that there exists φ_0 such that $\pi(a_{t_n})\varphi \to \varphi_0$ weakly (by Banach-Alaoglu theorem, any bounded sequence has a subsequence converging weakly). By Lemma 2.7 and 2.8, φ_0 is $\mathsf{SL}(2,\mathbb{R})$-invariant.

[1]That is if for any compact $K \subset \mathbb{H}$, there exists $N \in \mathbb{N}$ such that $\forall n \geq N, g_n \notin K$. We also write $g_n \to \infty$.

As π has no non-trivial invariant vector, $\varphi_0 = 0$. This is a contradiction. Necessarily, $\forall \varphi, \psi \in \mathcal{H}, \langle \pi(a_t)\varphi, \psi\rangle \to 0$ when $t \to \infty$. For a general diverging sequence $g_n \in \mathsf{SL}(2,\mathbb{R})$, we use Cartan decomposition (essentially the polar coordinates in the upper half plane). There exists $k_n, k'_n \in \mathsf{SO}(2)$ and a sequence $t_n \to +\infty$ (because $g_n \to \infty$) such that

$$g_n = k_n a_{t_n} k'_n.$$

Assume again that $\langle \pi(g_n)\varphi, \psi\rangle$ does not tend to 0. Passing to a subsequence, we can assume that $k_n \to k_0$ and $k'_n \to k'_0$. And

$$\langle \pi(g_n)\varphi, \psi\rangle = \langle \pi(a_{t_n})\pi(k'_n)\varphi, \pi(k_n)^{-1}\psi\rangle$$

has the same limit as

$$\langle \pi(a_{t_n})\pi(k'_0)\varphi, \pi(k_0)^{-1}\psi\rangle,$$

which is 0. This is again a contradiction, proving that $\forall \varphi, \psi \in \mathcal{H}, \langle \pi(g_n)\varphi, \psi\rangle \to 0$ when $n \to \infty$.

Let us now prove the two lemmas.

Proof of Lemma 2.7. We first remark that, by simple matrix multiplications,

$$a_{-t}u_s a_t = u_{se^{-t}}.$$

For any $\psi \in \mathcal{H}$,

$$\begin{aligned}
\langle \pi(u_s)\varphi_0 - \varphi_0, \psi\rangle &= \lim_{n\to+\infty} \langle \pi(u_s a_{t_n})\varphi - \pi(a_{t_n})\varphi, \psi\rangle \\
&= \lim_{n\to+\infty} \langle \pi(a_{-t_n} u_s a_{t_n})\varphi - \varphi, \pi(a_{-t_n})\psi\rangle \\
&= \lim_{n\to+\infty} \langle \pi(u_{se^{-t_n}})\varphi - \varphi, \pi(a_{-t_n})\psi\rangle \\
&\leq \lim_{n\to+\infty} \|\pi(u_{se^{-t_n}})\varphi - \varphi\|\|\psi\| = 0.
\end{aligned}$$

We deduce that $\pi(u_s)\varphi_0 = \varphi_0$. □

Proof of Lemma 2.8. We define the function (matrix coefficient)

$$F(g) = \langle \pi(g)\varphi, \varphi\rangle$$

We first remark that φ is bi-U-invariant: for any $u, u' \in U$,

$$F(ugu') = \langle \pi(ugu')\varphi, \varphi\rangle = \langle \pi(g)\varphi, \pi(u^{-1}\varphi\rangle = F(g).$$

We then compute for any $\varepsilon, r, s \in \mathbb{R}$,

$$\begin{pmatrix} 1 & r \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ \varepsilon & 1 \end{pmatrix}\begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 + r\varepsilon & r + s + rs\varepsilon \\ \varepsilon & 1 + s\varepsilon \end{pmatrix}. \tag{2.2}$$

Fix $t \in \mathbb{R}$, $\varepsilon > 0$ and choose $r = \frac{e^t - 1}{\varepsilon}$ and $s = \frac{-r}{1+r\varepsilon}$, so that

$$\begin{pmatrix} 1 + r\varepsilon & r + s + rs\varepsilon \\ \varepsilon & 1 + s\varepsilon \end{pmatrix} = \begin{pmatrix} e^t & 0 \\ \varepsilon & e^{-t} \end{pmatrix}.$$

This means that for any $\varepsilon, t > 0$

$$F\left(\begin{pmatrix} 1 & 0 \\ \varepsilon & 1 \end{pmatrix}\right) = F\left(\begin{pmatrix} e^t & 0 \\ \varepsilon & e^{-t} \end{pmatrix}\right).$$

By continuity of the representation we thus have when $\varepsilon \to 0$

$$\langle \varphi, \varphi\rangle = F\left(\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}\right).$$

So $F(a_t) = \langle \pi(a_t)\varphi, \varphi\rangle = \langle \varphi, \varphi\rangle = \|\varphi\|^2$. This is a case of equality in Cauchy-Schwarz inequality, which happens if and only if $\pi(g)\varphi$ and φ are linearly dependent. By taking $t = 0$ we find that necessarily $\pi(a_t)\varphi = \varphi$. So φ is invariant under the diagonal subgroup A, and by the same argument as for U, we get that F is bi-A-invariant.

We then use

$$\begin{pmatrix} e^{-t/2} & 0 \\ 0 & e^{t/2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ se^{-t} & 1 \end{pmatrix} \begin{pmatrix} e^{t/2} & 0 \\ 0 & e^{-t/2} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix},$$

to get that

$$F\left(\begin{pmatrix} 1 & 0 \\ se^{-t} & 1 \end{pmatrix}\right) = F\left(\begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}\right).$$

And by continuity, when $t \to +\infty$

$$\langle \varphi, \varphi \rangle = F\left(\begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}\right).$$

which gives us that φ is invariant under the group U^-. We deduce that φ is invariant under $\overline{\langle U^+, U^- \rangle} = \mathsf{SL}(2, \mathbb{R})$.[2] $\square$

[2]To see that this last equality occurs, use equality (2.2) to show that any element of $\mathsf{SL}(2,\mathbb{R})$ can be written as a (limit of) the product of the three matrices on the left-hand side with a suitable choice of the parameters r, s, ε.

CHAPTER 3

Laplacian and spectral decomposition

The *Laplace operator* or *Laplacian* is the unique (up to scalar multiplication) second order differential operator that commutes with the action of the isometry group. This symmetry makes it a fundamental operator in physics, where homogeneity and isotropy of space are basic principles. This explains that it appears in most partial differential equations (such as the heat equation, wave equation, Schrödinger equation). It also plays a role in the representation theory of these isometry groups. Eigenfunctions can be seen as fundamental waves in terms of which any function can be decomposed, a process giving birth to harmonic analysis. We are interested in this chapter in the spectral decomposition of the Laplacian on hyperbolic surfaces. For this purpose we introduce notions of harmonic analysis on the hyperbolic plane, and the heat kernel. In the subsequent chapters we will be interested in connecting the spectrum and eigenfunctions of the Laplacian to the geometric and dynamical properties of hyperbolic surfaces. For this chapter and the next ones a reference we use is [**Ber16**].

The Laplacian is defined explicitly on $\mathbb{H}$ by

$$\Delta = -y^2 \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right).$$

Exercise 3.1. Verify that if $g \in \mathsf{PSL}(2, R)$ and $T_g f(z) = f(g^{-1} z)$ is the left translation operator, we have

$$\Delta T_g = T_g \Delta.$$

(Use that any isometry is a product of transformations of the form $z \mapsto az (a > 0), z \mapsto z + b (b \in \mathbb{R}), z \mapsto -1/z, z \mapsto -\bar{z}$.)

The Laplacian can therefore be seen as a differential operator on any hyperbolic surface $M = \Gamma\backslash\mathbb{H}$. In the rest of the chapter we assume M is a compact surface and we fix a fundamental domain $D \subset \mathbb{H}$. A function $f : M \to \mathbb{C}$ is identified to a Γ-invariant function on $\mathbb{H}$ and its integral given by $\int_D f \, d\mu$.

We want to prove the following theorem.

Theorem 3.2. *There exists an orthonormal basis* $\{\varphi_j\}_{j \in \mathbb{N}}$ *of* $L^2(M)$ *and a sequence of real numbers*

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_j \leq \ldots$$

such that $\lambda_j \to +\infty$ *and*

$$\Delta \varphi_j = \lambda_j \varphi_j,$$

for any $j \in \mathbb{N}$.

Although the Laplacian is not a compact operator, the proof relies on the use of the spectral decomposition theorem for compact operators. The idea is to use a family of compact operators (the heat kernel) commuting with the Laplacian and apply the decomposition to these operators. We will then show that they share the same eigenspaces as the Laplacian.

First let us state without proof some important properties of the Laplacian. These properties are true in general for the Riemannian Laplacian and the proof is based on the use of Stoke's theorem.

PROPOSITION 3.3. (1) *Δ is symmetric, i.e. $\langle \Delta f, g\rangle = \langle f, \Delta g\rangle$ for any $f, g \in C^\infty(M)$.*
(2) *If $f \in C^\infty(M)$ is not constant then $\langle \Delta f, f\rangle > 0$.*

3.1. Heat kernel and proof of the spectral decomposition

The heat equation is a partial differential equation used to model diffusion phenomena. For example the evolution of the distribution of heat over time.

$$\begin{cases} \frac{\partial u}{\partial t} + \Delta u &= 0 \\ u_{|t=0} &= f \end{cases}$$

Where f is the initial distribution.

DEFINITION 3.4. A *fundamental solution of the heat equation*, or *heat kernel* on $M = \Gamma\backslash\mathbb{H}$ is a family of functions $p_t : M \times M \to \mathbb{R}$, such that $t \mapsto p_t$ is in $C^1([0, +\infty); C^\infty)$.

(1) $\frac{\partial p_t}{\partial t}(z, w) + \Delta_z p_t(z, w) = 0$
(2) $p_t(z, w) = p_t(w, z)$ for any $t \in [0, +\infty)$.
(3) $\lim_{t\to 0^+} \int_M p_t(z, w) f(w) d\mu(w) = f(z)$ for any compactly supported function f and locally uniformly in z.

LEMMA 3.5. *If M is compact, then the solution of the heat equation is unique. In particular the heat kernel is unique and*

$$\int_M p_t(z, w) d\mu(w) = 1. \tag{3.1}$$

PROOF. If u_1, u_2 are solutions for an initial condition f, then $v = u_1 - u_2$ is a solution with initial condition 0. We have

$$\frac{d}{dt}\int_M v^2 d\mu = 2\langle v, \frac{\partial v}{\partial t}\rangle = -2\langle v, \Delta v\rangle \leq 0.$$

Since $v_{|t=0} = 0$, we have $\forall t \geq 0, \int_M v^2 \leq 0$ and we deduce that $v = 0$. □

We denote by P_t the operator associated to the invariant kernel p_t. The function $P_t f$ is solution of the heat equation with initial condition f. A consequence of the unicity is that

$$P_{t+t'} = P_t P_{t'},$$

which implies the positivity of P_t by

$$\langle P_t f, f\rangle = \langle P_{t/2} f, P_{t/2} f\rangle \geq 0.$$

We will now prove the spectral decomposition theorem assuming first the existence of the heat kernel.

PROOF OF THEOREM 3.2. As the heat kernel is a continuous function on a compact space, it is clear that it is in $L^2(M \times M)$, and by the Hilbert-Schmidt theorem, it is then a compact operator. It is also self adjoint as $p_t(z, w) = p_t(w, z)$. We can thus apply the spectral decomposition theorem for compact self-adjoint operators, and we fix an eigenbasis $\varphi_0, \varphi_1, \ldots, \varphi_n, \ldots$ for P_1, with corresponding eigenvalues

$$\eta_0 \geq \eta_1 \geq \ldots \geq \eta_n \geq \ldots \geq 0, \quad \eta_n \to 0.$$

We will show that $\{\varphi_n\}_{n\in\mathbb{N}}$ is also an eigenbasis for Δ. Let us first show that for any $t > 0$ and $j \in \mathbb{N}$,

$$P_t \varphi_j = \eta_j^t \varphi_j$$

If φ is an eigenfunction of $P_{1/k}$ of eigenvalue μ, for some $k \in \mathbb{N}^*$, then it is also an eigenfunction of P_1 of eigenvalue μ^k, which means that the eigenspaces of P_1 and $P_{1/k}$ coincide, and that $P_{1/k}\varphi_j = \eta_j^{1/k}\varphi_j$. More generally we have for any $l, k \in \mathbb{N}^*$

$$P_{l/k}\varphi_j = \eta_j^{l/k}\varphi_j,$$

which by continuity of the heat kernel in time gives $P_t\varphi_j = \eta_j^t\varphi_j$ for any $t > 0$.

Now as $P_t\varphi_j$ is a solution of the Heat equation, we have at least formally

$$0 = \Delta P_t\varphi_j + \frac{\partial}{\partial t}P_t\varphi_j = \eta_j^t(\Delta\varphi_j + \log\eta_j\varphi_j)$$

So if we show that $\eta_0 = 1$, $0 < \eta_j < 1$ for all $j \in \mathbb{N}^*$, and that the φ_j are at least twice differentiable, we obtain that $\{\varphi_j\}_{j\in\mathbb{N}}$ is an eigenbasis for Δ with eigenvalues $\lambda_j = -\log\eta_j$ and

$$0 = \lambda_0 < \lambda_1 \leq \ldots \leq \lambda_n \leq \ldots, \lambda_n \to +\infty.$$

First, by the properties of the heat kernel

$$P_t\varphi_j = \eta_j^t\varphi_j \to \varphi_j$$

when $t \to 0$, so $\eta_j > 0$ for all $j \in \mathbb{N}$. In particular $\varphi_j = \eta_j^{-1}P_1\varphi_j$, and by the fact that the heat kernel is smooth, this gives that φ_j is also smooth.

If φ is a constant function, it is an eigenfunction of P_t with eigenvalue 1. If now φ is a non-constant eigenfunction of P_1 of eigenvalue η.

$$\begin{aligned}\frac{d}{dt}\|P_t\varphi\|^2 &= 2\langle\frac{d}{dt}P_t\varphi, P_t\varphi\rangle \\ &= -2\langle\Delta P_t\varphi, P_t\varphi\rangle \\ &= -2\eta^{2t}\langle\Delta\varphi, \varphi\rangle < 0.\end{aligned}$$

We deduce that $\|P_t\varphi\| = \eta^t\|\varphi\|$ is decreasing as a function of t, so $\eta < 1$. □

In order to construct the heat kernel, we will introduce some notions of harmonic analysis on the hyperbolic plane that will be also useful in the proof of the Selberg trace formula in the next chapter.

3.2. Harmonic analysis on the hyperbolic plane

Definition 3.6. An *invariant kernel* or *point-pair invariant* is a function $k : \mathbb{H}\times\mathbb{H} \to \mathbb{C}$ such that

$$\begin{aligned}&k(\gamma z, \gamma w) = k(z, w) \quad \forall\,\gamma \in \mathrm{Isom}(\mathbb{H}), \forall\,(z, w) \in \mathbb{H}\times\mathbb{H};\\ &k(z, w) = k(w, z) \quad \forall\,(z, w) \in \mathbb{H}\times\mathbb{H}.\end{aligned}$$

Such a kernel defines an *invariant integral operator* A, where for any $f : \mathbb{H} \to \mathbb{C}$, we have at least formally

$$Af(z) = \int_{\mathbb{H}} k(z, w)f(w)\,d\mu(w).$$

This also gives an operator on any surface $M = \Gamma\backslash\mathbb{H}$. Indeed if f is Γ-invariant, then

$$Af(z) = \int_D \left(\sum_{\gamma\in\Gamma} k(z, \gamma w)\right) f(w)\,d\mu(w), \quad \forall z \in D.$$

The invariant kernels we will consider are *radial kernels*. Let $k : (-\infty, +\infty) \to \mathbb{C}$ be an even function, then

$$(z, w) \mapsto k(d(z, w)), \quad (z, w) \in \mathbb{H}\times\mathbb{H}$$

is a point-pair invariant.

PROPOSITION 3.7. *If f is an eigenfunction of the Laplacian of eigenvalue λ, then it is an eigenfunction of the invariant integral operator A associated to k. There exists a function $h : \mathbb{R} \to \mathbb{C}$ such that*

$$Af(z) = \int k(d(z,w))f(w)d\mu(w) = h(\lambda)f(z).$$

We first need the following lemma on radial eigenfunctions.

LEMMA 3.8. *For any $\lambda \in \mathbb{C}$ and $z \in \mathbb{H}$, there exists a unique function $w \mapsto \omega_\lambda(z,w)$, radial around z and such that*

(1) $\omega_\lambda(z,z) = 1$,
(2) $\Delta_w \omega_\lambda(z,w) = \lambda\, \omega_\lambda(z,w)$.

PROOF. In polar coordinates,

$$\Delta = -\frac{\partial^2}{\partial r^2} - \frac{1}{\tanh r}\frac{\partial}{\partial r} - \frac{1}{(2\sinh r)^2}\frac{\partial^2}{\partial \theta^2}.$$

We use these coordinates around z, we are thus looking for a function $F(r)$ satisfying

$$F''(r) + \frac{1}{\tanh r}F'(r) + \lambda\, F(r) = 0.$$

This is the *Legendre differential equation.* It can be shown that the space of solutions to this equation is of dimension 2 and that there is a unique solution $F_s(r)$, with $s \in \mathbb{C}$ such that $\lambda = s(1-s)$, called *Legendre function of the first kind* that is continuous at 0 and such that $F_s(0) = 1$. A formula for this function is

$$F_s(r) = \frac{1}{\pi}\int_0^\pi (\cosh r + \sinh r \cos 2\theta)^{-s}\, d\theta.$$

We have

$$\omega_\lambda(z,w) = F_s(d(z,w)).$$

□

PROOF OF PROPOSITION 3.7. Let f be an eigenfunction of the Laplacian of eigenvalue λ. We define the corresponding radial function

$$\tilde{f}_z(w) = \int_{S_z} f(Tw)\, dT,$$

where S_z is the stabilizer of z in $\mathsf{PSL}(2,\mathbb{R})$ and dT the normalised Haar measure on S_z. It is a radial eigenfunction and by Lemma 3.8 we know that

$$\tilde{f}_z(w) = \omega_\lambda(z,w)f(z).$$

We have

$$\begin{aligned}\int_{\mathbb{H}} k(z,w)f(w)\, d\mu(w) &= \int_{S_z}\int_{\mathbb{H}} k(z,T^{-1}w)f(w)\, d\mu(w)\, dT\\ &= \int_{S_z}\int_{\mathbb{H}} k(z,w)f(Tw)\, d\mu(w)\, dT\\ &= \int_{\mathbb{H}} k(z,w)\tilde{f}_z(w)\, d\mu(w)\end{aligned}$$

so

$$\int_{\mathbb{H}} k(z,w)f(w)\, d\mu(w) = h(\lambda)f(z),$$

where

$$h(\lambda) = \int_{\mathbb{H}} k(z,w)\omega_\lambda(z,w)\, d\mu(w),$$

which is independent of z as for any $g \in \mathsf{PSL}(2,\mathbb{R})$,

$$\int_{\mathbb{H}} k(gz,w)\omega_\lambda(gz,w)\,d\mu(w) = \int_{\mathbb{H}} k(z,g^{-1}w)\omega_\lambda(z,g^{-1}w)\,d\mu(w) = \int_{\mathbb{H}} k(z,w)\omega_\lambda(z,w)\,d\mu(w).$$

□

We define the *Selberg transform* $\mathcal{S}(k)$ of a radial kernel k as

$$\mathcal{S}(k)(\lambda) = h(\lambda) = \int_{\mathbb{H}} k(d(i,w))\omega_\lambda(i,w)\,d\mu(w).$$

We will use the parametrisation $\lambda = s(1-s) = 1/4 + r^2$ with $r \in \mathbb{C}$. By abuse of notation we will write $\mathcal{S}(k)(r) = \mathcal{S}(k)(\lambda)$.

PROPOSITION 3.9. *The Selberg transform $\mathcal{S}(k)$ of a radial kernel k is obtained as the Fourier transform*

$$\mathcal{S}(k)(r) = \int_{-\infty}^{+\infty} e^{iru} g(u)\,du$$

of the function

$$g(u) = \sqrt{2}\int_{|u|}^{+\infty} \frac{k(\varrho)\sinh\varrho}{\sqrt{\cosh\varrho - \cosh u}}\,d\varrho.$$

PROOF. We note that the function $f(z) = \mathrm{Im}\,(z)^{1/2+ir}$ is an eigenfunction of the Laplacian of eigenvalue $\lambda = 1/4 + r^2 = s(1-s)$. By abuse of notation we write $h(r) = h(\lambda)$. By Proposition 3.7, we have

$$h(r) = \int k(d(i,z))\,\mathrm{Im}\,(z)^{1/2+ir} d\mu(z).$$

Let $U(\cosh\varrho) = k(\varrho)$ so that

$$U\left(1 + \frac{|z-w|^2}{2\mathrm{Im}\,z\mathrm{Im}\,w}\right) = k(d(z,w)).$$

We have

$$h(r) = \int_{-\infty}^{+\infty}\int_0^{+\infty} U\left(\frac{1+x^2+y^2}{2y}\right) y^{1/2+ir}\frac{dy}{y^2}\,dx.$$

Successive changes of variable $t = t(x) = (1+x^2+y^2)/2y$, $y = e^u$ and $t = \cosh\varrho$ give the expression we want. □

PROPOSITION 3.10. *For a function $h : \mathbb{R} \to \mathbb{C}$, the Selberg transform is inverted using the inverse Fourier transform*

$$g(u) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-iru} h(r)\,dr$$

and the formula

$$k(\varrho) = -\frac{1}{\pi\sqrt{2}}\int_{\varrho}^{+\infty} \frac{g'(u)}{\sqrt{\cosh u - \cosh\varrho}}\,du.$$

PROOF. It is clear that h is obtained from g by inverse Fourier transform. Now we have

$$g(u) = \sqrt{2}\int_{|u|}^{+\infty} \frac{k(\varrho)\sinh\varrho}{\sqrt{\cosh\varrho - \cosh u}}\,d\varrho,$$

and if we do a change of variables $x = \cosh u$ and $t = \cosh\varrho$, we obtain

$$g(\operatorname{arcosh} x) = \sqrt{2}\int_x^{+\infty} \frac{k(\operatorname{arcosh} t)}{\sqrt{t-x}}\,dt.$$

The function $G(x) = g(\operatorname{arcosh} x)$ is called the *Abel transform* of $\Phi(t) = k(\operatorname{arcosh} t)$. We have

$$G(x) = \sqrt{2} \int_x^\infty \frac{\Phi(t)}{\sqrt{t-x}} dt = 2\sqrt{2} \int_0^\infty \Phi(x+\xi^2)\, d\xi.$$

with the change of variable $\xi = \sqrt{t-x}$. Note that

$$G'(x) = 2\sqrt{2} \int_0^\infty \Phi'(x+\xi^2)\, d\xi.$$

Now, assuming Φ is compactly supported, which is the case if k is itself compactly supported,

$$\begin{aligned}
\Phi(t) &= -\int_0^\infty (\Phi(t+\xi^2))'_\xi \\
&= -2\int_0^\infty \Phi'(t+\xi^2)\xi d\xi = -\frac{4}{\pi}\int_0^{\pi/2}\int_0^\infty \Phi'(t+\xi^2)\xi d\xi d\theta \\
&= -\frac{4}{\pi}\int_0^\infty\int_0^\infty \Phi'(t+u^2+v^2)\, dudv \\
&= -\frac{2}{\pi\sqrt{2}}\int_0^\infty G'(t+v^2)\, dv \\
&= -\frac{1}{\pi\sqrt{2}}\int_t^\infty \frac{G'(x)}{\sqrt{x-t}} dx
\end{aligned}$$

Taking $t = \cosh\varrho$ and $x = \cosh u$, which gives in particular $G'(x)dx = g'(u)du$, we obtain

$$k(\varrho) = -\frac{1}{\pi\sqrt{2}} \int_\varrho^{+\infty} \frac{g'(u)}{\sqrt{\cosh u - \cosh\varrho}}\, du.$$

□

3.3. Existence of the heat kernel

3.3.1. Construction on $\mathbb{H}$. We first build the heat kernel on $\mathbb{H}$ before showing that it goes to the quotient $M = \Gamma\backslash\mathbb{H}$. We will look for a fundamental solution of the heat equation in the form of an invariant kernel k_t. According to (1) of Definition 3.4, the Selberg transform h_t of k_t must satisfy the equation[1]

$$\frac{\partial h_t}{\partial t}(r) = -\left(\frac{1}{4}+r^2\right) h_t(r).$$

The property (3) of the heat kernel implies in addition that $h_t(r) \to 1$ when $t \to 0^+$. We obtain the solution

$$h_t(r) = e^{-\left(\frac{1}{4}+r^2\right)t}.$$

Let us compute the inverse Selberg transform of h_t, which should give us the heat kernel we are looking for. We have

$$\begin{aligned}
g(u) &= \frac{e^{-t/4}}{2\pi} \int_{\mathbb{R}} e^{-iru} e^{-r^2 t}\, dr \\
&= \frac{e^{-t/4}}{2\sqrt{\pi t}} e^{-\frac{u^2}{4t}},
\end{aligned}$$

[1]To see that this is true, multiply both sides of the heat equation (1) by an eigenfunction $f(w)$ and integrate over w.

Using the formula for the Fourier transform of a Gaussian function. Then

$$
\begin{aligned}
p_t(z,w) &= -\frac{1}{\sqrt{2}\pi}\int_{d(z,w)}^{+\infty}\frac{g'(u)du}{\sqrt{\cosh u - \cosh d(z,w)}}\\
&= \frac{\sqrt{2}e^{-t/4}}{(4\pi t)^{3/2}}\int_{d(z,w)}^{+\infty}\frac{ue^{-u^2/4t}du}{\sqrt{\cosh u - \cosh d(z,w)}}
\end{aligned}
$$

It now remains to check that this expression indeed gives a fundamental solution of the heat equation on $\mathbb{H}$. The points (1) and (2) are satisfied by construction. For (3) we need a decay estimate for the heat kernel.

LEMMA 3.11. *There exists a constant $C > 0$ such that*

$$p_t(z,w) \le Ct^{-1}e^{-d(z,w)^2/8t}.$$

This lemma implies that $p_t \to 0$ when $t \to 0$ uniformly on any compact that does not contain 0. Together with the fact that by construction

$$\int_{\mathbb{H}} p_t(z,w)d\mu(w) = 1,$$

we obtain the property (3).

3.3.2. On a compact hyperbolic surface. To obtain the heat kernel on a compact hyperbolic surface $M = \Gamma\backslash\mathbb{H}$ we define formally

$$\bar{p}_t(z,w) = \sum_{\gamma\in\Gamma} p_t(z,\gamma w).$$

To show that the sum converges, we use the following lemma

LEMMA 3.12.

$$\#\{\gamma\in\Gamma : d(z,\gamma w) < T\} \le Ce^T.$$

PROOF. Since $d(z,\gamma w) \ge d(w,\gamma w) - d(z,w)$ we will assume that $z = w$. Let D be the compact Dirichlet domain at w. We have

$$
\begin{aligned}
\#\{\gamma\in\Gamma : d(w,\gamma w) < T\} &= \frac{1}{|D|}\left|\bigcup_{\gamma: d(w,\gamma w)<T}\gamma D\right|\\
&\le \frac{1}{|D|}\left|B_{T+\mathrm{diam}(D)}(w)\right|\\
&\le Ce^T,
\end{aligned}
$$

where $B_{T+\mathrm{diam}(D)}(w)$ is the disc of radius $T + \mathrm{diam}(D)$ centred at w. □

This lemma together with Lemma 3.11 allows us to show the convergence of the series

$$\bar{p}_t(z,w) \le C\sum_{n=0}^{\infty}\#\{\gamma\in\Gamma : n \le d(w,\gamma w) < n+1\}t^{-1}e^{-n^2/8t} \le C\sum_{n=0}^{\infty}t^{-1}e^{n}e^{-n^2/8t} < \infty.$$

Similar decay estimates as in Lemma 3.11 can be established for the derivatives of the heat kernel and the corresponding derivatives of $\bar{p}_t(z,w)$ can be shown to converge too.

Let us now check that $\bar{p}_t$ is a function on $M \times M$. This is the case as for any $\gamma_1, \gamma_2 \in \Gamma$,

$$\begin{aligned}\bar{p}_t(\gamma_1 z, \gamma_2 w) &= \sum_{\gamma\in\Gamma} p_t(\gamma_1 z, \gamma\gamma_2 w)\\ &= \sum_{\gamma\in\Gamma} p_t(z, \gamma_1^{-1}\gamma\gamma_2 w)\\ &= \sum_{\gamma\in\Gamma} p_t(z, \gamma w) = \bar{p}_t(z, w).\end{aligned}$$

The fact that $\bar{p}_t$ satisfies the conditions for a heat kernel are clear, for condition (3) it comes from the fact that for any Γ-invariant function f

$$\int_D \bar{p}_t(z, w) f(w) d\mu(w) = \int_D \sum_{\gamma\in\Gamma} p_t(z, \gamma w) f(\gamma w) d\mu(w) = \int_{\mathbb{H}} p_t(z, w) f(w) d\mu(w),$$

which tends to $f(z)$ when $t \to 0$.

CHAPTER 4

Selberg trace formula

Good references for this chapter are [**Mar12**] and [**Ber16**].

The trace formula connects the spectrum of the Laplacian to the set of lengths of closed geodesics. Let $M = \Gamma\backslash\mathbb{H}$ be a compact hyperbolic surface. As M is compact, Γ does not contain any parabolic element. We assume that Γ does not contain any elliptic elements either, so that it is strictly hyperbolic. The spectrum of the Laplacian is denoted by $\lambda_j = \frac{1}{4} + r_j^2$ for $j \in \mathbb{N}$ and as it is real we have $r_j \in \{r \in \mathbb{C} : |\mathrm{Im}\,(r)| \leq 1/2\}$.

Definition 4.1. A *periodic* or *closed geodesic* of M is a geodesic $\gamma : \mathbb{R} \to M$ such that

$$\exists \alpha \in \mathbb{R}, \forall t \in \mathbb{R} \quad (\gamma(t+\alpha), \gamma'(t+\alpha)) = (\gamma(t), \gamma'(t)). \tag{4.1}$$

We denote by $\mathcal{G}(M)$ the set of periodic geodesics of M. The smallest α satisfying (4.1) is called the *period* or *length* of the geodesic denoted by ℓ_γ.

The Selberg trace formula is then given by the following theorem that we are proving in this chapter.

Theorem 4.2. *For any admissible function h (see next remark),*

$$\sum_{j=0}^{\infty} h(r_j) = \frac{\mathrm{Area}(M)}{4\pi} \int_{-\infty}^{+\infty} h(r)\, r \tanh(\pi r)\, dr + \sum_{\gamma \in \mathcal{G}(M)} \sum_{n=1}^{+\infty} \frac{\ell_\gamma g(n\,\ell_\gamma)}{2\sinh(n\,\ell_\gamma/2)},$$

where g is the Fourier transform of h.

Remark 4.3. An admissible function $h : \mathbb{C} \to \mathbb{C}$ will be a function such that

(1) h is analytic on the strip $\{r \in \mathbb{C} : |\mathrm{Im}\,(r)| < 1/2 + \delta\}$ for some $\delta > 0$
(2) h is even, i.e. $h(-r) = h(r)$
(3) $|h(r)| \leq C(1 + |\mathrm{Re}\,(r)|)^{-N}$ for any $N > 1$ uniformly on the strip $|\mathrm{Im}\,(r)| < 1/2 + \delta$.

We will accept without proof the following proposition.

Proposition 4.4. *The function h is admissible if and only if it is the Selberg transform of an even function $k : \mathbb{R} \to \mathbb{C}$ satisfying the bound on the n-th derivative*

$$|k^{(n)}(\varrho)| \leq Ce^{-(1+\delta-\varepsilon)\varrho},$$

for any $\varepsilon > 0$ and any $n \in \mathbb{N}$.

In all this chapter we will assume that the function h is admissible.

4.1. Periodic geodesics and primitive hyperbolic elements

For any hyperbolic element $\gamma \in \mathsf{PSL}(2,\mathbb{R})$, there exists a unique geodesic a_γ in $\mathbb{H}$ invariant with respect to γ, called the *axis* of γ. It is the geodesic passing through the two fixed points of γ in $\widehat{\mathbb{R}}$.

Definition 4.5. An element $\gamma \in \Gamma$ is *primitive* if it cannot be written as a power δ^k, with $k \geq 2$ of another element $\delta \in \Gamma$.

Proposition 4.6. *$\mathcal{G}(M)$ can be identified with the set of conjugacy classes in Γ of primitive hyperbolic elements. That is the quotient of the set of hyperbolic elements in Γ by the equivalence relation $\gamma \sim \gamma' \iff \exists \delta \in \Gamma, \gamma = \delta\gamma'\delta^{-1}$.*

Proof. Let $\varrho \in \mathcal{G}(M)$ and $\{\tilde{\varrho}\}$ be the class of geodesics of $\mathbb{H}$ that project on ϱ. Fix a representative $\tilde{\varrho}$ of this class. Now let $\mathsf{Stab}(\tilde{\varrho})$ be the subgroup of $\mathsf{PSL}(2,\mathbb{R})$ of transformations fixing $\tilde{\varrho}$. It is conjugate to the subgroup fixing the imaginary axis, which is

$$\left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}, \quad a \in \mathbb{R} \right\}.$$

In particular $\mathsf{Stab}(\tilde{\varrho}) \simeq \mathbb{R}$, and as $\mathsf{Stab}(\tilde{\varrho}) \cap \Gamma$ is homomorphic to a discrete subgroup of $\mathbb{R}$, it is cyclic. Let $\gamma \in \Gamma$ such that

$$\mathsf{Stab}(\tilde{\varrho}) \cap \Gamma = \langle \gamma \rangle.$$

Necessarily, γ is primitive. Otherwise, there exists $\delta \in \Gamma, k \geq 2$ such that $\gamma = \delta^k$. Then δ is hyperbolic with same fixed points as γ and therefore fixes $\tilde{\varrho}$. But this means $\delta \in \mathsf{Stab}(\tilde{\varrho}) \cap \Gamma = \langle \gamma \rangle$ which is impossible.

Any other representative $\widehat{\varrho}$ of the class $\{\tilde{\varrho}\}$, is obtained as $\widehat{\varrho} = g\tilde{\varrho}$ with $g \in \Gamma$, and

$$\mathsf{Stab}(\widehat{\varrho}) \cap \Gamma = \langle g\gamma g^{-1} \rangle.$$

Let $\{\gamma\}$ be the conjugacy class of γ in Γ. We thus built a map $\varrho \mapsto \{\tilde{\varrho}\} \mapsto \{\gamma\}$ associating a conjugacy class in Γ of primitive hyperbolic elements to a periodic geodesic, which is clearly one-to-one. □

4.2. Pretrace formula

Let $(\varphi_j)_{j\in\mathbb{N}}$ be an orthonormal basis of (real) eigenfunctions of Δ on $L^2(M)$.

Theorem 4.7 (Pretrace formula). *Let h be the Selberg transform of the radial kernel k, we have*

$$\sum_{j=0}^{\infty} h(r_j)\varphi_j(z)\varphi_j(w) = \sum_{\gamma\in\Gamma} k(z,\gamma w),$$

where the convergence is absolute and uniform. When $z = w$, we get

$$\sum_{j=0}^{\infty} h(r_j)|\varphi_j(z)|^2 = \frac{1}{4\pi}\int_{-\infty}^{\infty} h(\varrho)\tanh(\pi\varrho)\varrho\, d\varrho + \sum_{\gamma\in\Gamma\setminus\{1\}} k(z,\gamma z). \tag{4.2}$$

Proof. If we see the φ_j as Γ-invariant functions on $\mathbb{H}$, these are eigenfunctions of Δ on $\mathbb{H}$ and we have by definition of the Selberg transform of k that

$$\int_{\mathbb{H}} k(z,w)\varphi_j(w)\, d\mu(w) = h(r_j)\varphi_j(z).$$

At the same time

$$\int_{\mathbb{H}} k(z,w)\varphi_j(w)\, d\mu(w) = \int_D \sum_{\gamma\in\Gamma} k(z,\gamma w)\varphi_j(w)\, d\mu(w).$$

Now because the (φ_j) form a basis we have

$$\sum_{\gamma\in\Gamma} k(z,\gamma w) = \sum_{j\in\mathbb{N}} \left(\int_D \sum_{\gamma\in\Gamma} k(z,\gamma z')\varphi_j(z')\, d\mu(z') \right) \varphi_j(w) = \sum_{j\in\mathbb{N}} h(r_j)\varphi_j(z)\varphi_j(w),$$

where the convergence of the series is in L^2. It is possible to show that the convergence is absolute and uniform in z, w (see for example [**Hej76**]).

For the equality (4.2) we need to show that

$$k(z,z) = \frac{1}{4\pi}\int_{-\infty}^{\infty} h(r)\tanh(\pi r) r\, dr.$$

We start by writing the inverse Selberg transform

$$k(z,z) = k(0) = -\frac{1}{\sqrt{2}\pi}\int_0^{+\infty} \frac{g'(u)}{\sqrt{\cosh u - 1}}\,du = -\frac{1}{2\pi}\int_0^{+\infty} \frac{g'(u)}{\sinh(u/2)}\,du$$

where

$$g(u) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} h(r)e^{-iur}dr,$$

so

$$\begin{aligned} k(z,z) &= \frac{1}{4\pi^2}\int_0^{+\infty}\int_{-\infty}^{+\infty} h(r)\frac{\sin(ur)}{\sinh(u/2)}r\,dr\,du \\ &= \frac{1}{4\pi^2}\int_{-\infty}^{+\infty} h(r)\left(\int_0^{+\infty}\frac{\sin(ur)}{\sinh(u/2)}du\right) r\,dr, \end{aligned}$$

where the rapid decay of h allows us to interchange the integrals. We now write

$$\frac{1}{\sinh(u/2)} = \frac{2}{e^{u/2}(1-e^{-u})} = 2e^{-u/2}\sum_{n=0}^{+\infty} e^{-nu},$$

so that we have

$$\begin{aligned} \int_0^{+\infty}\frac{\sin(ur)}{\sinh(u/2)}du &= 2\sum_{n\geq 0}\int_0^{+\infty} e^{-(2n+1)u/2}\sin(ur)du \\ &= 2\sum_{n\geq 0}\frac{4r}{4r^2+(2n+1)^2} \\ &= \sum_{n\in\mathbb{Z}}\frac{4r}{4r^2+(2n+1)^2} \end{aligned}$$

by double integration by parts. Note that we have the Fourier correspondence

$$\int_{-\infty}^{+\infty} e^{-2i\pi x\xi}\frac{2r}{r^2+\xi^2}\,d\xi = 2\pi e^{-2\pi r|x|}$$

and we use the Poisson summation formula[1]

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}}\int_{-\infty}^{+\infty} e^{-2i\pi n\xi}f(\xi)\,d\xi$$

to obtain

$$\sum_{n\in\mathbb{Z}}\frac{2r}{r^2+(n+1/2)^2} = 2\pi\sum_{n\in\mathbb{Z}} e^{i\pi n}e^{-2\pi r|n|} = 2\pi\frac{1-e^{-2\pi r}}{1+e^{-2\pi r}} = 2\pi\tanh(\pi r).$$

We thus have

$$\int_0^{+\infty}\frac{\sin(ur)}{\sinh(u/2)}du = \pi\tanh(\pi r),$$

which gives the result. □

[1]This formula can be obtained simply by expanding the periodic function $g(\varrho) = \sum_{n\in\mathbb{Z}} f(\varrho+n)$ in its Fourier series and setting $\varrho = 0$.

4.3. Decomposition in hyperbolic cylinders

To obtain the trace formula, we need to integrate (4.2) over a fundamental domain D. We get

$$\sum_{j=0}^{\infty} h(r_j) = \frac{\text{Area}(M)}{4\pi} \int_{-\infty}^{\infty} h(\varrho) \tanh(\pi\varrho)\varrho \, d\varrho + \sum_{\gamma \in \Gamma \setminus \{1\}} \int_D k(z, \gamma z) \, d\mu(z).$$

In order to write the last term as a sum over closed geodesics, we will group the terms by conjugacy classes $\{\gamma\}$ with $\gamma \in \Gamma$. If $\gamma_0 \in \{\gamma\}$, then $\gamma_0 = \gamma_1^{-1}\gamma\gamma_1$ for some $\gamma_1 \in \Gamma$, and

$$\begin{aligned}\int_D k(z, \gamma_0 z) \, d\mu(z) &= \int_D k(z, \gamma_1^{-1}\gamma\gamma_1 z) \, d\mu(z) \\ &= \int_D k(\gamma_1 z, \gamma\gamma_1 z) \, d\mu(z) \\ &= \int_{\gamma_1 D} k(z, \gamma z) \, d\mu(z).\end{aligned}$$

So we have

$$\sum_{\gamma \in \Gamma \setminus \{1\}} \int_D k(z, \gamma z) \, d\mu(z) = \sum_{\{\gamma\} \neq 1} \int_{D_\gamma} k(z, \gamma z) \, d\mu(z),$$

with

$$D_\gamma = \bigcup_{\gamma_1 \in \Gamma_\gamma \setminus \Gamma} \gamma_1 D,$$

and Γ_γ the *centralizer* of γ, i.e.

$$\Gamma_\gamma = \{\gamma' \in \Gamma : \gamma\gamma' = \gamma'\gamma\}.$$

LEMMA 4.8. *For any $\gamma \in \Gamma$, there is a unique primitive element δ such that $\delta^k = \gamma$ for some $k \geq 1$, and the centralizer is given by*

$$\Gamma_\gamma = \{\delta^n, n \in \mathbb{Z}\}.$$

PROOF. We have seen in the proof of Proposition 4.6 that the subgroup of Γ fixing the axis a_γ is cyclic and generated by a primitive element δ, so that $\gamma = \delta^k$ for some $k \geq 1$. Any other primitive element δ_0 such that $\gamma = \delta_0^k$ has the same axis as γ, so it is necessarily in the group fixing the axis a_γ, hence $\delta_0 = \delta$. Clearly $\langle\delta\rangle$ is a subgroup of the centralizer Γ_γ. To see it is exactly the centralizer, take an element $\alpha \in \Gamma_\gamma$. Then $\gamma\alpha a_\gamma = \alpha\gamma a_\gamma = \alpha a_\gamma$, and by unicity of the axis of γ, we have $\alpha a_\gamma = a_\gamma$, so $\alpha \in \langle\delta\rangle$. □

Using the previous lemma, we can further decompose the sum over conjugacy classes of primitive elements, that is elements of $\mathcal{G}(M)$ by Proposition 4.6.

$$\sum_{\{\gamma\} \neq 1} \int_{D_\gamma} k(z, \gamma z) \, d\mu(z) = \sum_{\gamma \in \mathcal{G}(M)} \sum_{n=1}^{\infty} \int_{D_\gamma} k(z, \gamma^n z) \, d\mu(z).$$

Notice that D_γ is a fundamental domain for the *hyperbolic cylinder* $\Gamma_\gamma \backslash \mathbb{H}$.

4.4. Hyperbolic terms

We now fix a primitive element $\gamma \in \Gamma$. Up to a conjugation by an isometry, which does not change the value of the integral we are computing, we can assume that

$$\gamma = \begin{pmatrix} e^{\ell/2} & 0 \\ 0 & e^{-\ell/2} \end{pmatrix},$$

where $\ell = \ell_\gamma$ is the length of the isometry. A fundamental domain for $\Gamma_\gamma \backslash \mathbb{H}$ is then

$$\{z = x + iy \in \mathbb{H} : 1 \leq y < e^\ell\}.$$

We use the auxiliary function $U(\cosh\varrho) = k(\varrho)$. For $z' = e^{n\ell}z$ we have

$$\cosh d(z, \gamma^n z') = 1 + \frac{|z - z'|^2}{2\mathrm{Im}\, z \mathrm{Im}\, z'} = 1 + 2\frac{|z|^2 \sinh^2(n\ell/2)}{y^2}$$

so that

$$\begin{aligned}
\int_{D_\gamma} k(z, \gamma^n z)\, d\mu(z) &= \int_1^{e^\ell} \int_{-\infty}^{+\infty} U\left(1 + 2\sinh^2(n\ell/2)(1 + x^2/y^2)\right) y^{-2} dx dy \\
&= \int_1^{e^\ell} y^{-1}\, dy \int_{-\infty}^{+\infty} U\left(1 + 2\sinh^2(n\ell/2)(1 + x^2)\right) dx \\
&= \frac{\ell}{\sinh(n\ell/2)} \int_{\sinh^2(n\ell/2)}^{+\infty} \frac{U(1+2u)}{\sqrt{u - \sinh^2(n\ell/2)}} du \\
&= \frac{\ell}{\sinh(n\ell/2)\sqrt{2}} \int_{n\ell}^{+\infty} \frac{k(\varrho)\sinh\varrho}{\sqrt{\cosh\varrho - \cosh(n\ell)}} d\varrho
\end{aligned}$$

where for the last equality we used the change of variable $\cosh\varrho = 1 + 2u$ and the formula $1 + 2\sinh^2(n\ell/2) = \cosh(n\ell)$. We thus have

$$\int_{D_\gamma} k(z, \gamma^n z)\, d\mu(z) = \frac{\ell g(n\ell)}{2\sinh(n\ell/2)}.$$

Putting everything together we have

$$\begin{aligned}
\sum_{j\in\mathbb{N}} h(r_j) &= \frac{\mathrm{Area}(M)}{4\pi} \int_{-\infty}^{\infty} h(\varrho)\tanh(\pi\varrho)\varrho\, d\varrho + \sum_{\gamma\in\mathcal{G}(M)} \sum_{n=1}^{\infty} \int_{D_\gamma} k(z, \gamma^n z)\, d\mu(z) \\
&= \frac{\mathrm{Area}(M)}{4\pi} \int_{-\infty}^{\infty} h(\varrho)\tanh(\pi\varrho)\varrho\, d\varrho + \sum_{\gamma\in\mathcal{G}(M)} \sum_{n=1}^{\infty} \frac{\ell g(n\ell)}{2\sinh(n\ell/2)},
\end{aligned}$$

which is the form we wanted.

4.5. Application: Weyl law

Theorem 4.9. *Let $N(\lambda) = \#\{j \in \mathbb{N} : \lambda_j \le \lambda\}$ be the number of eigenvalues smaller than λ. We have the asymptotic law*

$$N(\lambda) \sim \frac{\mathrm{Area}(M)}{4\pi}\lambda,$$

when $\lambda \to +\infty$.

Proof. We give just the general idea. Apply the trace formula to $h(r) = e^{-tr^2}$ (the Selberg transform of the heat kernel). We have

$$g(u) = \frac{1}{\sqrt{4\pi t}} e^{-\frac{u^2}{2t}}.$$

Using that $\lambda_j = 1/4 + r_j^2$ we get

$$\sum_{j=0}^{\infty} e^{-t\lambda_j} = \frac{\mathrm{Area}(M)}{4\pi t} + O(1),$$

when $t \to 0$. Then we can use Karamata's Tauberian theorem stated just after this proof. □

Let $\widehat{\mu}(t) = \int_0^\infty e^{-tx}\, d\mu(x)$, where μ is a Borel measure on $[0, \infty)$.

Theorem 4.10 (Karamata). *Let $r \ge 0$, $a \in \mathbb{R}$. The following convergences are equivalent*

(1) $\lim_{t\to 0} t^r \widehat{\mu}(t) = a$

(2) $\lim_{\lambda\to+\infty} \lambda^{-r}\mu([0,\lambda]) = \frac{a}{\Gamma(r+1)}$.

where Γ *is the gamma function, in particular* $\Gamma(n) = (n-1)!$ *for* $n \in \mathbb{N}, n \geq 1$.

CHAPTER 5

Microlocal lift and quantum ergodicity

In this chapter, we are starting the study of the eigenfunctions of the Laplacian on hyperbolic surfaces. We will study some of their properties in the large eigenvalue limit. Our first goal will be to lift the eigenfunctions to the unit tangent bundle and to show that asymptotically this lift gives a probability measure invariant under the geodesic flow. The quantum ergodicity theorem will then tell us that in the large eigenvalue limit, most eigenfunctions are close to the uniform measure. For more details on this chapter, see for example the lecture notes [**EW10**].

5.1. Microlocal lift

5.1.1. Lie algebra and differential operators. For any matrix $X \in M_2(\mathbb{R})$ we define the *exponential map* as

$$\exp(X) = \sum_{n=0}^{\infty} \frac{X^n}{n!}.$$

The *Lie algebra* of $\mathsf{SL}(2,\mathbb{R})$ is then given by

$$\mathfrak{g} = \mathfrak{sl}_2(\mathbb{R}) = \{X \in M_2(\mathbb{R}) \;:\; \forall t \in \mathbb{R} \quad \exp(tX) \in G\}.$$

We have

$$\mathfrak{sl}_2(\mathbb{R}) = \{X \in M_2(\mathbb{R}) : \mathrm{Tr}(X) = 0\},$$

which follows from the fact that

$$\det(\exp(X)) = \exp(\mathrm{Tr}(X)).$$

Definition 5.1. For $X \in \mathfrak{g}$, we define the *differential operator* $D_X : C^\infty(G) \to C^\infty(G)$ by

$$D_X f(g) = \frac{d}{dt} f\left(g \cdot \exp(tX)\right)_{|t=0},$$

for any $f \in C^\infty(G)$. We will also use frequently the notation $Xf = D_X f$.

The differentiation is linear, we have $D_{\alpha X + \beta Y} = \alpha D_X + \beta D_Y$. Note that in general $D_X D_Y \neq D_{XY}$. However the commutators are preserved $D_X D_Y - D_Y D_X = D_{[X,Y]}$ where $[X,Y] = XY - YX \in \mathfrak{g}$.

We extend the differentiation to any $Z = X + iY \in \mathfrak{g} \otimes \mathbb{C}$ by defining $D_Z = D_X + iD_Y$. For two smooth compactly supported functions f, f' we define $\langle f, f' \rangle = \int f(g)\overline{f'(g)}\, dg$. Then we have

$$\langle D_Z f, f' \rangle = -\langle f, D_{\overline{Z}} f' \rangle.$$

We can see the geodesic and horocycle flows as $x_t = \exp(tX)$, where X is the *direction* of the flow. Define

$$H = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}, U^+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, U^- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, W = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then H corresponds to the direction of the geodesic flow, U^+ and U^- to the stable and unstable directions of the horocycle flow, and W is the direction of the circular flow $K = \mathsf{SO}(2)$.

We define the *Casimir operator*

$$\Omega = D_H D_H + \frac{1}{2} D_{U^+} D_{U^-} + \frac{1}{2} D_{U^-} D_{U^+},$$

and leave the following important proposition as an exercise.

PROPOSITION 5.2. (1) *The Casimir operator commutes with all differential operators.*
(2) *Recall that $G/K = \mathbb{H}$. The restriction of Ω to K-invariant functions coincides with the Laplacian on $\mathbb{H}$.*

5.1.2. Weight spaces decomposition. We now restrict ourselves to $G = \mathsf{PSL}(2, \mathbb{R})$. Let

$$k_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \in K$$

Define the space of *K-eigenfunctions of weight* $2n$ as

$$A_{2n} = \{ f \in \mathbb{C}^\infty(\Gamma\backslash G) : f(gk_\theta) = e^{2in\theta} f(g) \}$$

Note that the space A_0 is the space of smooth K-invariant function, which can be identified to the space of smooth functions on $\Gamma\backslash\mathbb{H}$.

The following exercise is nothing more than the Fourier decomposition along K, the circle group.

EXERCISE 5.3. Show that A_{2n} and $A_{2n'}$ are orthogonal when $n \neq n'$ and that

$$\overline{\bigoplus_{n\in\mathbb{Z}} A_{2n}} = C^\infty(\Gamma\backslash G).$$

DEFINITION 5.4. A function $f \in C^\infty(\Gamma\backslash G)$ is K-finite if there exists $N \in \mathbb{N}$ such that

$$f \in \bigoplus_{n=-N}^{N} A_{2n}.$$

Note that the weight spaces are eigenspaces for the differential operator D_W, i.e.

$$A_{2n} = \{ f : D_W f = 2inf \}.$$

We also have the property

$$A_{2n} A_{2n'} \subset A_{2(n+n')}.$$

Define now the *raising operator* E^+ and the *lowering operator* E^- by

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

The commutation relation $[W, E^\pm] = \pm 2iE^\pm$ implies the following fundamental property of the raising and lowering operators.

$$E^\pm(A_{2n}) \subset A_{2(n\pm1)}.$$

Indeed if $f \in A_{2n}$, $WE^\pm f = E^\pm W f + [W, E^\pm] f = (2in \pm 2i) E^\pm f$.

Let us finally record two more expressions for the Casimir operator that we will use:

$$\Omega = E^+E^- - \frac{1}{4}W^2 + \frac{i}{2}W = E^-E^+ - \frac{1}{4}W^2 - \frac{i}{2}W.$$

5.1.3. Construction of the microlocal lift. We fix an L^2 normalised eigenfunction φ of the Laplacian, of eigenvalue $\lambda = \frac{1}{4} + r^2$. It can be seen as a K-invariant function on the group G. In order to construct the lift that will satisfy the invariance property under the geodesic flow we define inductively

$$\begin{aligned} \varphi_0(g) &= \varphi(gK) \in A_0 \\ \varphi_{2n+2} &= \frac{1}{ir + \frac{1}{2} + n} E^+ \varphi_{2n}, \text{ if } n \geq 0 \\ \varphi_{2n-2} &= \frac{1}{ir + \frac{1}{2} - n} E^- \varphi_{2n}, \text{ if } n \leq 0 \end{aligned}$$

EXERCISE 5.5. Show that $\Omega \varphi_{2n} = \lambda \varphi_{2n}$ and that $\|\varphi_{2n}\|_2 = 1$ for all $n \in \mathbb{Z}$.

For any K-finite function f, we now define the following linear functional that constitutes the microlocal lift

$$I_\varphi(f) = \langle f \sum_{n \in \mathbb{Z}} \varphi_{2n}, \varphi_0 \rangle.$$

As f is K-finite, only a finite number of terms of the sum are non zero. Although the distribution I_φ is not a measure it projects for functions on $M = \Gamma \backslash \mathbb{H}$ to the probability measure associated to the eigenfunction: if $f \in A_0$,

$$I_\varphi(f) = \int_M f |\varphi|^2 \, d\mu.$$

The definition of this functional is due to Zelditch [**Zel87**] and arises from a quantization procedure of classical observables adapted to the hyperbolic plane. For a function f on the unit tangent bundle $\Gamma \backslash \mathsf{PSL}(2, \mathbb{R})$ seen as a classical observable, Zelditch defines a quantization $\mathrm{Op}(f)$ that associates with f an operator on $L^2(M)$ with standard pseudo-differential operator properties. The microlocal lift is then the quantity $\langle \varphi, \mathrm{Op}(f) \varphi \rangle$ representing the quantum average of the observable $\mathrm{Op}(f)$ in the state φ. It can then be shown ([**Zel87**] proof of Proposition 2.2) that $\langle \varphi, \mathrm{Op}(f) \varphi \rangle = I_\varphi(f)$. Note that the equality for any eigenfunctions φ, ψ of an orthonormal basis

$$\langle \varphi, \mathrm{Op}(f) \psi \rangle = \langle f \sum_{n \in \mathbb{Z}} \varphi_{2n}, \psi_0 \rangle$$

entirely defines $\mathrm{Op}(f)$, although this is not the most natural way to do it.

We will use the notation $\varphi_\infty = \sum_{n \in \mathbb{Z}} \varphi_{2n}$ so that

$$I_\varphi(f) = \langle f \varphi_\infty, \varphi_0 \rangle.$$

We will now show that the distribution I_φ is asymptotically a probability measure that is invariant under the geodesic flow. For this purpose let $N = [\sqrt{r}]$ and define

$$\psi = \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{N} \varphi_{2n}.$$

This is an eigenfunction of the Casimir operator that can be seen as a particular lift of the eigenfunction φ. It gives a probability measure on $\Gamma \backslash G$,

$$\nu = |\psi|^2(g) dg.$$

In a first lemma we will prove that the microlocal lift is close to ν when the eigenvalue r is large.

LEMMA 5.6. *For a K-finite function f we have*

$$\left| I_\varphi(f) - \int_{\Gamma \backslash G} f \, d\nu \right| = O_f(r^{-1/2}).$$

PROOF. We have

$$\int f d\nu = \langle f\psi, \psi\rangle = \frac{1}{2N+1}\sum_{n,m=-N}^{N}\langle f, \varphi_{2n}, \varphi_{2m}\rangle.$$

Suppose $f \in \bigoplus_{l=-L}^{L} A_{2l}$, then $\langle f\varphi_{2n}, \varphi_{2m}\rangle = 0$, when $|n-m| > L$ by orthogonality of the A_l.

Now if $|n-m| \leq L$, we have

$$\begin{aligned}\langle f\varphi_{2n}, \varphi_{2m}\rangle &= \frac{1}{ir+\frac{1}{2}+n-1}\langle fE^{+}\varphi_{2n-2}, \varphi_{2m}\rangle \\ &= \frac{1}{ir-\frac{1}{2}+n}\left(\langle E^{+}(f\varphi_{2n-2}), \varphi_{2m}\rangle - \langle E^{+}(f)\varphi_{2n-2}, \varphi_{2m}\rangle\right)\end{aligned}$$

The second term in the brackets is bounded by Cauchy-Schwarz and we get

$$\begin{aligned}\langle f\varphi_{2n}, \varphi_{2m}\rangle &= \frac{1}{ir-\frac{1}{2}+n}\langle f\varphi_{2n-2}, (-E^{-})\varphi_{2m}\rangle + O(r^{-1}) \\ &= -\frac{-ir+\frac{1}{2}-m}{ir-\frac{1}{2}+n}\langle f\varphi_{2n-2}, \varphi_{2m-2}\rangle + O(r^{-1}) \\ &= \langle f\varphi_{2n-2}, \varphi_{2m-2}\rangle + O(r^{-1})\end{aligned}$$

as we have $-\frac{-ir+\frac{1}{2}-m}{ir-\frac{1}{2}+n} = 1 + O\left(\frac{|n-m|}{r}\right)$. By iterating this process we obtain

$$\langle f\varphi_{2n}, \varphi_{2m}\rangle = \langle f\varphi_{2n-2m}, \varphi_0\rangle + O(Nr^{-1})$$

using that $m \leq N$. We then have $O(Nr^{-1}) = O(r^{-1/2})$.

Now

$$\begin{aligned}\langle f\psi, \psi\rangle &= \frac{1}{2N+1}\sum_{n,m=-N}^{N}\mathbf{1}_{|n-m|\leq L}\left(\langle f\varphi_{2n-2m}, \varphi_0\rangle + O(r^{-1/2})\right) \\ &= \sum_{l=-L}^{L}\frac{2N+1-|l|}{2N+1}\left(\langle f\varphi_{2l}, \varphi_0\rangle + O(r^{-1/2})\right) \\ &= \langle f\sum_{l=-L}^{L}\varphi_{2l}, \varphi_0\rangle + O(r^{-1/2}) \\ &= I_\varphi(f) + O(r^{-1/2})\end{aligned}$$

using that $\frac{2N+1-|l|}{2N+1} = 1 + O(N^{-1})$. □

This lemma tells us that any weak limit of (a subsequence of) I_φ when $\lambda \to \infty$ is a probability measure. The second lemma will give us the asymptotic invariance of this measure under the geodesic flow.

LEMMA 5.7. *There exists a differential operator L independent of r such that*

$$I_\varphi((rH+L)f) = 0.$$

In particular $I_\varphi(Hf) = O_f(r^{-1})$.

PROOF. Let us write

$$\lambda I_\varphi(f) = \lambda\langle f\varphi_\infty, \varphi_0\rangle = \langle f\varphi_\infty, \Omega\varphi_0\rangle = \langle f\varphi_\infty, E^-E^+\varphi_0\rangle = \langle E^-E^+(f\varphi_\infty), \varphi_0\rangle,$$

and then decompose

$$\begin{aligned}\langle E^-E^+(f\varphi_\infty), \varphi_0\rangle = \langle E^-E^+(f)\varphi_\infty, \varphi_0\rangle + \langle E^+(f)E^-(\varphi_\infty), \varphi_0\rangle \\ + \langle E^-(f)E^+(\varphi_\infty), \varphi_0\rangle + \langle fE^-E^+(\varphi_\infty), \varphi_0\rangle.\end{aligned}$$

To estimate these different terms we note that

$$E^-(\varphi_\infty) = \sum_{n\in\mathbb{Z}}\left(ir + \frac{1}{2} - n\right)\varphi_{2n-2} = \left(ir + \frac{i}{2}W - \frac{1}{2}\right)\varphi_\infty$$

$$E^+(\varphi_\infty) = \sum_{n\in\mathbb{Z}}\left(ir + \frac{1}{2} + n\right)\varphi_{2n+2} = \left(ir - \frac{i}{2}W - \frac{1}{2}\right)\varphi_\infty$$

$$E^-E^+(\varphi_\infty) = \left(\Omega - \frac{1}{4}W^2 - \frac{i}{2}W\right)\varphi_\infty = \lambda\varphi_\infty - \left(\frac{1}{4}W^2 + \frac{i}{2}W\right)\varphi_\infty.$$

and we obtain by cancelling the terms in λ and grouping the terms in r that

$$ir\left(\langle E^+(f)\varphi_\infty, \varphi_0\rangle + \langle E^-(f)\varphi_\infty, \varphi_0\rangle\right) = 2irI_\varphi(Hf)$$

is equal to a term independent of r containing a linear combination of terms of the form $I_\varphi(Lf)$ and $\langle L'fW^k(\varphi_\infty), \varphi_0\rangle$, where L and L' are some differential operators. Therefore to conclude we only need to note that

$$\langle f_1W(f_2), \varphi_0\rangle = -\langle W(f_1)f_2, \varphi_0\rangle.$$

Indeed since $W\varphi_0 = 0$ we have

$$0 = \langle f_1f_2, W\varphi_0\rangle = -\langle W(f_1f_2), \varphi_0\rangle = -\langle W(f_1)f_2, \varphi_0\rangle - \langle f_1W(f_2), \varphi_0\rangle.$$

So we can rewrite the terms

$$\langle L'fW^k(\varphi_\infty), \varphi_0\rangle = \langle L''f\varphi_\infty, \varphi_0\rangle = I_\varphi(L''f),$$

for some differential operator L''. □

Let $\tilde{\nu}$ be a probability measure obtained as a weak limit of I_φ when $\lambda \to \infty$. This lemma tells us that

$$\int_{\Gamma\backslash G} Hf\, d\tilde{\nu} = 0, \tag{5.1}$$

for any K-finite function and by density any smooth function. As H generates the geodesic flow, i.e. $a_t = \exp(tH)$, we obtain the invariance. To see this just notice that if $\tilde{f}(g) = f(ga_t)$, then $H\tilde{f} = \frac{d}{dt}f(ga_t)$, and apply (5.1) to $\tilde{f}$. We get

$$\frac{d}{dt}\int_{\Gamma\backslash G} f(ga_t)\, d\tilde{\nu} = 0.$$

The invariance then follows from integration in t.

5.2. Quantum ergodicity

Let $G = \mathsf{PSL}(2,\mathbb{R})$ and $M = \Gamma\backslash\mathbb{H}$ be a compact hyperbolic surface. We fix $\{\varphi_j\}_{j\in\mathbb{N}}$ an orthonormal basis of eigenfunctions, with corresponding eigenvalues $\lambda_j = 1/4 + r_j^2$. We have defined in the previous section a distribution $I_{\varphi_j} = \langle \varphi_j, \mathrm{Op}(f)\varphi_j\rangle$, that corresponds to an operator $\mathrm{Op}(f)$ whose kernel can be made explicit. This kernel is not radial, but by studying its trace as we did for radial kernels to prove Selberg trace formula and Weyl law, it is possible to prove the following proposition [**Zel87**].

Proposition 5.8. *For any K-finite function $f \in C^\infty(\Gamma\backslash G)$ and any $\varepsilon > 0$,*

$$\frac{1}{N(L,\varepsilon)} \sum_{j:|\lambda_j - L|<\varepsilon} I_{\varphi_j}(f) \to \frac{1}{|\Gamma\backslash G|}\int_{\Gamma\backslash G} f(g)\,dg,$$

when $L \to +\infty$, where $N(L,\varepsilon) = \#\{j : |\lambda_j - L| < \varepsilon\}$.

This tells us that on average, the sequence of microlocal lifts converges to the normalised Haar measure. However, there could be cancellations in the average that allow this convergence. Using the ergodicity of the geodesic flow and the asymptotic invariance of the microlocal lifts, it can be show that these cancellations do not happen.

Theorem 5.9 (Quantum Ergodicity). *For any K-finite function $f \in C^\infty(\Gamma\backslash G)$ and any $\varepsilon > 0$,*

$$\frac{1}{N(L,\varepsilon)} \sum_{j:|\lambda_j - L|<\varepsilon} \left| I_{\varphi_j}(f) - \frac{1}{|\Gamma\backslash G|}\int_{\Gamma\backslash G} f(g)\,dg \right|^2 \to 0,$$

when $L \to +\infty$.

Proof. We assume without loss of generality that $\int_{\Gamma\backslash G} f(g)\,dg = 0$. Let us define the time average along the geodesic flow $M_T(f)$ by

$$M_T(f)(g) = \frac{1}{T}\int_0^T f(ga_t)\,dt,$$

for $T > 0$.

By Lemma 5.7 on the asymptotic invariance of I_{φ_j} under the geodesic flow, we have

$$I_{\varphi_j}(M_T(f)) = I_{\varphi_j}(f) + O_{f,T}\left(\frac{1}{L}\right),$$

if $|\lambda_j - L| \le \varepsilon$. Then

$$\frac{1}{N(L,\varepsilon)} \sum_{j:|\lambda_j - L|<\varepsilon} \left|I_{\varphi_j}(f)\right|^2 = \frac{1}{N(L,\varepsilon)} \sum_{j:|\lambda_j - L|<\varepsilon} \left|I_{\varphi_j}(M_T(f))\right|^2 + O_{f,T}\left(\frac{1}{L}\right).$$

Now by Lemma 5.6 we know that there exists a probability measure ν_j such that

$$|I_{\varphi_j}(f) - \nu_j(f)| = O(\lambda_j^{-1/2}),$$

and we get

$$\begin{aligned}
\frac{1}{N(L,\varepsilon)} \sum_{j:|\lambda_j - L|<\varepsilon} \left|I_{\varphi_j}(f)\right|^2 &= \frac{1}{N(L,\varepsilon)} \sum_{j:|\lambda_j - L|<\varepsilon} \left|\int M_T(f)d\nu_j\right|^2 + O_{f,T}\left(\frac{1}{\sqrt{L}}\right)\\
&\le \frac{1}{N(L,\varepsilon)} \sum_{j:|\lambda_j - L|<\varepsilon} \int |M_T(f)|^2\,d\nu_j + O_{f,T}\left(\frac{1}{\sqrt{L}}\right)\\
&= \frac{1}{N(L,\varepsilon)} \sum_{j:|\lambda_j - L|<\varepsilon} I_{\varphi_j}\left(|M_T(f)|^2\right)d\nu_j + O_{f,T}\left(\frac{1}{\sqrt{L}}\right).
\end{aligned}$$

where we used Cauchy-Schwarz theorem to get $\left|\int M_T(f)d\nu_j\right|^2 \leq \int |M_T(f)|^2\, d\nu_j$ and Lemma 5.6 a second time. By Proposition 5.8 we obtain

$$\limsup_{L\to\infty} \frac{1}{N(L,\varepsilon)} \sum_{j:|\lambda_j - L|<\varepsilon} \left|I_{\varphi_j}(f)\right|^2 = \frac{1}{|\Gamma\backslash G|} \int_{\Gamma\backslash G} |M_T f(g)|^2\, dg.$$

And the mixing of the geodesic flow, implying its ergodicity tells us that $\frac{1}{|\Gamma\backslash G|}\int_{\Gamma\backslash G} |M_T f(g)|^2\, dg \to 0$ when $T \to +\infty$, as $\int_{\Gamma\backslash G} f(g)\, dg = 0$. □

A corollary of this theorem is that most of the eigenfunctions tend to the uniform measure.

Corollary 5.10. *There exists a subsequence $\lambda_{j_k} \to \infty$ so that for any K-finite f*

$$I_{\varphi_{j_k}}(f) \to \frac{1}{|\Gamma\backslash G|} \int_{\Gamma\backslash G} f(g)\, dg$$

when $k \to +\infty$, and the subsequence is of density 1, meaning that

$$\lim_{L\to+\infty} \frac{\#\{\lambda_{j_k} \leq L\}}{L} = 1.$$

In particular we obtain that

$$|\varphi_{j_k}|^2\, d\mu \to d\mu$$

weakly, when $k \to +\infty$.

A conjecture of Rudnick and Sarnak formulated in 1994 states that the full sequence should converge. This phenomenon is called Quantum Unique Ergodicity (QUE). It is still an open question but a partial solution was given by Lindenstrauss in 2006 in the case of arithmetic lattices Γ and specific bases of eigenfunctions. Understanding this theorem is the object of the next chapter.

CHAPTER 6

Arithmetic Quantum Unique Ergodicity

Let $M = \Gamma\backslash\mathbb{H}$ be a finite area hyperbolic surface (i.e. Γ is a lattice in $\mathsf{PSL}(2,\mathbb{R})$). We will be interested in this chapter in the case where Γ is an arithmetic lattice. To avoid entering into the details of the general definition of arithmetic lattices, we will restrict ourselves to the special but important case of $\Gamma = \mathsf{PSL}(2,\mathbb{Z})$. In this case M is not compact and we have not proved that $L^2(M)$ has an orthonormal basis of eigenfunctions of the Laplacian. Indeed it does not. However, we will admit without proof that there exists an infinite orthonormal system of eigenfunctions

$$\Delta\varphi_j = \lambda_j\varphi_j, \quad \|\varphi_j\|_2 = 1,$$

with a non decreasing sequence of eigenvalues $\lambda_j \to \infty$, and we will be interested in these eigenfunctions in the large eigenvalue limit.

Arithmetic lattices come with additional structure in the form of Hecke operators, that we will define in this chapter. These operators commute with each other and with the Laplacian and it is therefore possible to consider bases of joint eigenfunctions of the Laplacian and Hecke operators.

Recall the definition of microlocal lift I_{φ_j} from the previous chapter Section 5.1.3. We have the following theorem.

Theorem 6.1 (Arithmetic QUE). *If we assume that the eigenfunctions φ_j are joint eigenfunctions of the Laplacian and a Hecke operator then the only limit of I_{φ_j} when $j \to +\infty$ is the normalised Haar measure. In particular $|\varphi_j|^2 d\mu \to d\mu$ when $j \to +\infty$.*

Let us assume first that M is compact. This theorem is then a consequence of a measure rigidity theorem of Lindenstrauss [**Lin06**].

Theorem 6.2 (Lindenstrauss). *Let Γ be an arithmetic lattice in $\mathsf{SL}(2,\mathbb{R})$ and μ a probability measure on $X = \Gamma\backslash\,\mathsf{SL}(2,\mathbb{R})$ such that*

1. *μ is invariant under the geodesic flow;*
2. *μ is p-Hecke recurrent for a prime number p;*
3. *μ has positive entropy on every ergodic component (for the geodesic flow);*

then μ is the normalised Haar measure m_X on X.

We have seen in the previous chapter that the distributions I_{φ_j} are asymptotically close to probability measures and are invariant under the geodesic flow. In the case where M is compact this implies that the possible limits of I_{φ_j} are invariant probability measures. However on a non-compact space, a weakly convergent sequence of probability measures does not necessarily converge to a probability measure: there can be some *escape of mass* at infinity such that the limit has a total mass less than 1. This possibility was excluded for eigenfunctions on $\mathsf{PSL}(2,\mathbb{Z})\backslash\mathbb{H}$ by Soundararajan [**Sou10**]. We thus have a sequence of distributions I_{φ_j} and we know that the limits are invariant probability measures. The goal of this chapter is to verify that these probability measures satisfy the two other conditions of the theorem, proving that there is only one possible limit.

6.1. Hecke recurrence

6.1.1. The p-adic extension of $\Gamma\backslash\,\mathsf{SL}(2,\mathbb{R})$. We first recall the definition and some properties of the p-adic numbers.

Let p be a prime number. For $r \in \mathbb{Q}$ we can write $r = p^k \frac{m}{n}$ with $p \nmid mn$ and $k \in \mathbb{Z}$. We define the *p-adic norm* as

$$|r|_p = p^{-k}.$$

The *field of p-adic numbers* $\mathbb{Q}_p$ is the completion of $\mathbb{Q}$ with respect to the p-adic norm. To get some intuition, let us note that a p-adic number can be represented by an infinite expansion

$$x = \sum_{k=-m}^{+\infty} x_k p^k \quad 0 \le x_k < p. \tag{6.1}$$

The p-adic norm encodes congruences modulo p^k for a certain power k. The higher this power, the closer the elements are:

$$x \equiv y \bmod p^k \Leftrightarrow |x-y|_p = p^{-k}.$$

Define the set of p-adic integers $\mathbb{Z}_p$ as

$$\mathbb{Z}_p = \{x \in \mathbb{Q}_p : |x|_p \le 1\}.$$

These are the p-adic numbers with an expansion in positive powers of p:

$$x \in \mathbb{Z}_p \Leftrightarrow x = \sum_{k=0}^{+\infty} x_k p^k.$$

The integers $\mathbb{Z}_p$ constitute a compact open subring of $\mathbb{Q}_p$. Finally we will consider the ring $\mathbb{Z}\left[\frac{1}{p}\right]$. It can be thought of in the following way

$$\mathbb{Z}\left[\frac{1}{p}\right] = \{x = \pm \sum_{k=-m}^{n} x_k p^k : m, n \in \mathbb{N} \quad 0 \le x_k < p\},$$

which has the advantage of making the density of $\mathbb{Z}\left[\frac{1}{p}\right]$ in $\mathbb{Q}_p$ (and $\mathbb{R}$) clear. Note however that the negative elements of $\mathbb{Z}\left[\frac{1}{p}\right]$ seen as a subset of $\mathbb{Q}_p$ have also a representation as an infinite expansion of the form (6.1).

Exercise 6.3. The diagonal embedding

$$\mathbb{Z}\left[\frac{1}{p}\right] \hookrightarrow \mathbb{R} \times \mathbb{Q}_p$$

given by $x \mapsto (x, x)$ is discrete and co-compact. We have the isomorphism

$$\mathbb{Z}\backslash\mathbb{R} \simeq \mathbb{Z}\left[1/p\right]\backslash\mathbb{R} \times \mathbb{Q}_p.$$

Definition 6.4. Let R be a ring, R^* be the group of units (invertible elements), then

$$\mathsf{PGL}_2(R) = \left\{\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in R \quad \det\gamma \in R^*\right\} / \sim,$$

where $\gamma \sim \gamma'$ if and only if there exists $r \in R^*$ such that $\gamma = r\gamma'$.

Exercise 6.5. Show that

$$\mathsf{SL}_2(\mathbb{Z})\backslash \mathsf{SL}_2(\mathbb{R}) \simeq \mathsf{PGL}_2(\mathbb{Z})\backslash \mathsf{PGL}_2(\mathbb{R}).$$

We will use the following notation:

$$G = \mathsf{PGL}_2(\mathbb{R}), \Gamma = \mathsf{PGL}_2(\mathbb{Z}),$$
$$G_p = \mathsf{PGL}_2(\mathbb{Q}_p), K_p = \mathsf{PGL}_2(\mathbb{Z}_p), \Gamma_p = \mathsf{PGL}_2(\mathbb{Z}[1/p]).$$

Proposition 6.6. *The diagonal embedding* $\mathsf{PGL}_2(\mathbb{Z}[1/p]) \hookrightarrow \mathsf{PGL}_2(\mathbb{R}) \times \mathsf{PGL}_2(\mathbb{Q}_p)$ *is a lattice and we have an isomorphism*

$$\mathsf{PGL}_2(\mathbb{Z})\backslash \mathsf{PGL}_2(\mathbb{R}) \simeq \mathsf{PGL}_2(\mathbb{Z}[1/p])\backslash \mathsf{PGL}_2(\mathbb{R}) \times \mathsf{PGL}_2(\mathbb{Q}_p)/ \mathsf{PGL}_2(\mathbb{Z}_p).$$

PROOF. With the previous notation, we want to show

$$\Gamma\backslash G \simeq \Gamma_p\backslash G \times G_p/K_p.$$

We denote by e the identity element in G and G_p. We will show that the action of G on the double quotient

$$g \cdot \Gamma_p(g', h)K_p = \Gamma_p(g'g, h)K_p$$

is transitive and that

$$\mathrm{Stab}_G(\Gamma_p(e, e)K_p) = \Gamma.$$

For the transitivity we want to show that $\{g \cdot \Gamma_p(e, e)K_p : g \in G\} = \Gamma_p\backslash G \times G_p/K_p$. In other words that for all $(g, h) \in G \times G_p$, there exist $\gamma \in \Gamma_p$ and $k \in K_p$ such that

$$\Gamma_p(g, h)K_p = \Gamma_p(\gamma g, \gamma hk)K_p = \Gamma_p(\gamma g, e)K_p.$$

So equivalently we want to show that for any $h \in G_p$, there exist $\gamma \in \Gamma_p$ and $k \in K_p$ such that $\gamma hk = e$. We start with a matrix $h = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G_p$ and multiply on the left and on the right by elements of Γ_p and K_p. First we can always multiply on the right by $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ and assume that $|a|_p \geq |b|_p$, so that $b/a \in \mathbb{Z}_p$. Then we multiply on the right by $\begin{pmatrix} 1 & -b/a \\ 0 & 1 \end{pmatrix}$ to obtain a matrix $\begin{pmatrix} a & 0 \\ c & d' \end{pmatrix}$. By multiplying on the left by $\begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix}$ we get $\begin{pmatrix} a & 0 \\ c' & d' \end{pmatrix}$, where $c' = a\alpha + c$. We can choose by density $\alpha \in \mathbb{Z}[1/p]$ close enough to $-c/a$ such that $|c'|_p \leq |d'|_p$, and we then multiply on the right by $\begin{pmatrix} 1 & 0 \\ -c'/d' & 1 \end{pmatrix}$ to get the diagonal matrix $\begin{pmatrix} a & 0 \\ 0 & d' \end{pmatrix}$. We conclude by using diagonal matrices in Γ_p and K_p to get the identity matrix. Note that we showed $G_p = \Gamma_p K_p$.

To show now that $\mathrm{Stab}_G(\Gamma_p(e, e)K_p) = \Gamma$, we write

$$\Gamma_p(g, e)K_p = \Gamma_p(e, e)K_p \Leftrightarrow \exists \gamma \in \Gamma_p, k \in K_p, \gamma g = \gamma k = e \Leftrightarrow g \in \Gamma_p \cap K_p = \Gamma.$$

□

Let us now fix a point Γg on $\Gamma\backslash G$. It is identified to

$$\Gamma_p(g, e)K_p \in \Gamma_p\backslash G \times G_p/K_p.$$

The orbit of this element under G_p

$$\{\Gamma_p(g, h)K_p : h \in G_p\}$$

can be identified to $G_p/K_p = \mathsf{PGL}_2(\mathbb{Q}_p)/\mathsf{PGL}_2(\mathbb{Z}_p)$. Indeed it can be checked easily that

$$\mathrm{Stab}_{G_p}(\Gamma_p(g, e)K_p) = K_p.$$

So we have a foliation of $\Gamma\backslash G$ where the leaves are these orbits: for $\Gamma g \in \Gamma\backslash G$, all the elements $\Gamma g'$ where $\Gamma_p(g', e)K_p = \Gamma_p(g, h)K_p$ for some $h \in G_p$. We will show that the leaves G_p/K_p have a natural tree structure.

6.1.2. The tree $\mathsf{PGL}_2(\mathbb{Q}_p)/\mathsf{PGL}_2(\mathbb{Z}_p)$. A lattice in $\mathbb{Q}_p^2$ is a discrete subgroup $L \subset \mathbb{Q}_p^2$ of the form $L = \mathbb{Z}_p v_1 + \mathbb{Z}_p v_2$ where $\{v_1, v_2\}$ is a basis of $\mathbb{Q}_p^2$. We define an equivalence relation between lattices by

$$L_1 \sim L_2 \Leftrightarrow L_1 = \alpha L_2, \alpha \in \mathbb{Q}_p^*.$$

We define X_p to be the set of equivalence classes $[L]$ of lattices in $\mathbb{Q}_p^2$. We see this set as the set of *vertices* of a graph. Two distinct vertices $[L_1]$ and $[L_2]$ are *neighbours* or *adjacent* if for some representatives L_1, L_2 we have

$$pL_1 \subset L_2 \subset L_1.$$

Note that this definition is symmetric because if $pL_1 \subset L_2 \subset L_1$ then $pL_2 \subset pL_1 \subset L_2$ and $pL_1 \in [L_1]$.

Remark 6.7. An equivalent definition of adjacency is to ask that for some representatives L_1 and L_2 we have $L_2 \subset L_1$ and

$$[L_1 : L_2] = p,$$

where $[L_1 : L_2]$ is the index of L_2 in L_1, that is the cardinality of L_1/L_2. Indeed if $pL_1 \subset L_2 \subset L_1$, then if we take the quotient by pL_1 we obtain

$$\{0\} \subset L_2/pL_1 \subset L_1/pL_1 \simeq (\mathbb{Z}_p/p\mathbb{Z}_p)^2 \simeq (\mathbb{Z}/p\mathbb{Z})^2.$$

By the correspondence theorem in group theory we get that the subgroups L_2 such that $pL_1 \subsetneq L_2 \subsetneq L_1$ are in bijection with the subgroups H such that $\{0\} \subsetneq H \subsetneq (\mathbb{Z}/p\mathbb{Z})^2$ and $[L_1 : L_2] = [(\mathbb{Z}/p\mathbb{Z})^2 : H] = p$.

Note that because there are $p+1$ subgroups H of index p in $(\mathbb{Z}/p\mathbb{Z})^2$ (or equivalently, $p+1$ one dimensional subspaces of $(\mathbb{Z}/p\mathbb{Z})^2$), each vertex of X_p has $p+1$ neighbours and X_p is therefore a $p+1$ regular graph.

The group $G_p = \mathsf{PGL}_2(\mathbb{Q}_p)$ acts transitively on X_p. Indeed we can write any lattice as

$$L = g\mathbb{Z}_p^2 \quad g \in G_p.$$

It can be checked that the stabilizer of the lattice $\mathbb{Z}_p^2$ is $\mathsf{PGL}_2(\mathbb{Z}_p) = K_p$. Hence the identification

$$X_p \simeq G_p/K_p = \mathsf{PGL}_2(\mathbb{Q}_p)/\,\mathsf{PGL}_2(\mathbb{Z}_p).$$

To understand X_p, it is useful to write the elements of $\mathsf{GL}_2(\mathbb{Q}_p)$ according to the Cartan decomposition.

Proposition 6.8 (Cartan decomposition of $\mathsf{GL}_2(\mathbb{Q}_p)$).

$$\mathsf{GL}_2(\mathbb{Q}_p) = \mathsf{GL}_2(\mathbb{Z}_p)\left\{\begin{pmatrix} p^m & 0 \\ 0 & p^n \end{pmatrix} : m, n \in \mathbb{Z}, m \leq n\right\}\mathsf{GL}_2(\mathbb{Z}_p).$$

Proof. We first apply row and column operations, as in the proof of Proposition 6.6, to reduce any element of $\mathsf{GL}_2(\mathbb{Q}_p)$ to a diagonal form. We then use that any $a \in \mathbb{Q}_p$ can be written $a = p^n\alpha$ with $\alpha \in \mathbb{Z}_p^*$ for some $n \in \mathbb{Z}$. □

Lemma 6.9. *Let L_1, L_2 be two lattices in $\mathbb{Q}_p^2$, then there exists a basis $\{v_1, v_2\}$ of L_1 such that $\{p^m v_1, p^n v_2\}$ is a basis of L_2 for some $m, n \in \mathbb{Z}, m \leq n$. In particular there exists $L_2' \in [L_2]$ such that $\{v_1, p^n v_2\}$ is a basis of L_2' for some $n \in \mathbb{N}$.*

Proof. If L_1, L_2 are two lattices in $\mathbb{Q}_p^2$, fixing a basis for each corresponds to choosing two elements $g_1, g_2 \in \mathsf{GL}_2(\mathbb{Q}_p)$ where the basis vectors are the columns of the corresponding matrices. We have $L_i = g_i\mathbb{Z}_p^2$ for $i = 1, 2$. Multiplying g_i on the right by an element of $\mathsf{GL}_2(\mathbb{Z}_p)$ does not change the lattice and corresponds to a change of basis.

Now there exists $g \in \mathsf{GL}_2(\mathbb{Q}_p)$ such that $g_1 g = g_2$. We use the Cartan decomposition to write

$$g = k_1 \begin{pmatrix} p^m & 0 \\ 0 & p^n \end{pmatrix} k_2.$$

We have

$$g_1 k_1 \begin{pmatrix} p^m & 0 \\ 0 & p^n \end{pmatrix} = g_2 k_2^{-1}.$$

We take for the basis v_1, v_2 of L_1 the columns of the matrix $g_1 k_1$. It is then clear that the columns of $g_2 k_2^{-1}$, giving a basis of L_2 are $p^m v_1, p^n v_2$ as we wanted. □

The previous lemma is the main ingredient of the following proposition.

PROPOSITION 6.10. *$X_p = \mathsf{PGL}_2(\mathbb{Q}_p)/\mathsf{PGL}_2(\mathbb{Z}_p)$ is a $p+1$-regular tree.*

PROOF. We already saw in Remark 6.7 that X_p is a $p+1$ regular graph. In order to prove it is a tree we will show that for any two vertices there exists a unique path connecting them. Take two classes of lattices $[L]$ and $[L_0]$ corresponding to two vertices of the tree. We deduce from Lemma 6.9 that for any representative L of the first class, we can find a unique representative L_0 of the second class, and respective bases for L and L_0 written as the columns of the matrices $g, g_0 \in \mathsf{GL}_2(\mathbb{Q}_p)$ (i.e. $L = g\mathbb{Z}_p^2$ and $L_0 = g_0\mathbb{Z}_p^2$) such that:

$$g_0 = g\begin{pmatrix}1 & 0\\ 0 & p^n\end{pmatrix} \quad n \geq 0.$$

From this we get that $L_0 \subset L$ and $L/L_0 \simeq \mathbb{Z}/p^n\mathbb{Z}$. As will become clear in what follows, n is the distance in the tree between L and L_0.

To find a path between L and L_0 we use the following simple lemma from group theory whose proof we leave as an exercise.

LEMMA 6.11. *Let $G \simeq \mathbb{Z}/p^n\mathbb{Z}$ for some $n \geq 1$ (i.e. G is a cyclic p-group). Then G has a unique finite composition series, that is there exists a unique chain of subgroups*

$$G_0 = \{0\} \subsetneq G_1 \subsetneq \ldots \subsetneq G_n = G,$$

with $|G_k| = p^k$ and $[G_{k+1} : G_k] = p$ for any $0 \leq k \leq n$.

By the lemma and the correspondence theorem we have a unique chain

$$L_0 \subsetneq L_1 \subsetneq \ldots \subsetneq L_n = L,$$

with $[L_{k+1} : L_k] = p$ for any $0 \leq k \leq n$, which corresponds to a unique path on the tree. □

REMARK 6.12. An alternative way to see what we did is to write the Cartan decomposition for $\mathsf{PGL}_2(\mathbb{Q}_p)$

$$\mathsf{PGL}_2(\mathbb{Q}_p) = \mathsf{PGL}_2(\mathbb{Z}_p)\left\{\left[\begin{pmatrix}1 & 0\\ 0 & p^n\end{pmatrix}\right] : n \in \mathbb{N}\right\}\mathsf{PGL}_2(\mathbb{Z}_p).$$

Then we can write

$$X_p \simeq \mathsf{PGL}_2(\mathbb{Q}_p)/\mathsf{PGL}_2(\mathbb{Z}_p) \simeq \mathsf{PGL}_2(\mathbb{Z}_p)\left\{\left[\begin{pmatrix}1 & 0\\ 0 & p^n\end{pmatrix}\right] : n \in \mathbb{N}\right\}.$$

The vertices at distance n in the tree from the origin $[\mathbb{Z}_p^2]$ are the classes of lattices

$$\left[k\begin{pmatrix}1 & 0\\ 0 & p^n\end{pmatrix}\mathbb{Z}_p^2\right],$$

for $k \in \mathsf{PGL}_2(\mathbb{Z}_p)$.

6.1.3. Hecke operators. For each point $x \in X = \mathsf{PGL}_2(\mathbb{Z})\backslash\mathsf{PGL}_2(\mathbb{R})$, we have now defined a set of points $\mathfrak{X}_p(x) \subset X$ with a tree structure called the *Hecke tree*. We denote by $d_p(x_1, x_2)$ the distance in the tree between $x_1, x_2 \in \mathfrak{X}_p(x)$ such that $d_p(x_1, x_2) = 1$ if and only if x_1 and x_2 are neighbours in $\mathfrak{X}_p(x)$. We write

$$\mathfrak{X}_p^n(x) = \{y \in \mathfrak{X}_p(x) : d_p(x, y) = n\}.$$

and define the *Hecke operators* T_{p^n} for any $n \geq 1$ as

$$T_{p^n} f(x) = \sum_{y \in \mathfrak{X}_p^n(x)} f(y),$$

for any function $f : X \to \mathbb{C}$.

We have the following relations that we leave as an exercise.

LEMMA 6.13. (1) $T_p^2 = T_{p^2} + (p+1)Id$
(2) $T_p T_{p^n} = T_{p^{n+1}} + pT_{p^{n-1}}$, *for* $n \geq 2$

Using the isomorphism $\Gamma\backslash G \simeq \Gamma_p\backslash G \times G_p/K_p$ and Remark 6.12 we can rewrite the definition of the Hecke operators in the following way. Let m_p be the Haar measure on G_p, normalised so that $m_p(K_p) = 1$. For $f : X \to \mathbb{C}$,

$$T_{p^n} f(\Gamma g) = \int_{B_n} f(\Gamma_p(g,h)K_p)\, dm_p(h),$$

where

$$B_n = K_p \begin{pmatrix} 1 & 0 \\ 0 & p^n \end{pmatrix} K_p,$$

and by abuse of notation we identify $f(\Gamma g)$ and $f(\Gamma_p(g,e)K_p)$. Let us state without proof some important properties of the Hecke operators.

PROPOSITION 6.14. *For any* $n \geq 1$
(1) T_{p^n} *is self adjoint.*
(2) T_{p^n} *commutes with the action of* G *and hence with all differential operators.*

By taking the quotient on the right by $K = \mathsf{SO}(2)$, the isomorphism $\Gamma\backslash G \simeq \Gamma_p\backslash G \times G_p/K_p$ gives an isomorphism between the modular surface $\Gamma\backslash G/K$ and $\Gamma_p\backslash G \times G_p/K \times K_p$ and the Hecke operators are well defined as operators on this quotient.

Let us give an explicit formula for T_p on the modular surface. We have seen that the neighbours of $[\mathbb{Z}_p^2]$ are of the form

$$\left[k \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \mathbb{Z}_p^2\right],$$

for $k \in \mathsf{PGL}_2(\mathbb{Z}_p)$. And we know that there are $p+1$ possible classes, corresponding to index p subgroups of $\mathbb{Z}_p^2$. It can be checked that the $p+1$ classes correspond to $[g\mathbb{Z}_p^2]$ with g among the matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix} \text{ and } \begin{pmatrix} p & -b \\ 0 & 1 \end{pmatrix} \text{ with } 0 \leq b < p.$$

For functions $f : \Gamma\backslash\mathbb{H} \to \mathbb{C}$ we therefore obtain the formula

$$T_p f(z) = f(pz) + \sum_{b=0}^{p-1} f\left(\frac{z+b}{p}\right).$$

6.1.4. Recurrence of eigenfunctions. The Hecke operator T_p can be seen as the analog of the Laplacian on the $p+1$-regular tree. We begin with a proposition about eigenfunctions on abstract regular trees.

PROPOSITION 6.15. *Let* f *be a function on a* $p+1$*-regular tree, and* T_p *be the operator defined by*

$$T_p f(x) = \sum_{y : d_p(x,y)=1} f(y),$$

where d_p *is the distance on the tree. We assume that there exists* λ *such that*

$$T_p f = \lambda f.$$

Then there exists a constant $c > 0$ *such that*

$$\sum_{y : d_p(x,y) \leq n} |f(y)|^2 \geq cn|f(x)|^2,$$

for any $n \geq 1$.

Proof. We only give an idea of the proof (for a complete one see for example [**EW10**]). By Cauchy-Schwarz inequality we have

$$\left|\sum_{i=0}^{n} T_{p^i} f(x)\right| = \left|\sum_{y:d_p(x,y)\leq n} f(y)\right| \leq \left(\sum_{y:d_p(x,y)\leq n} |f(y)|^2\right)^{1/2} \#\{y : d_p(x,y) \leq n\}^{1/2}$$
$$\leq \operatorname{const} p^{n/2} \left(\sum_{y:d_p(x,y)\leq n} |f(y)|^2\right)^{1/2}.$$

By Lemma 6.13, f is an eigenfunction of all T_{p^i} with eigenvalue λ_i satisfying the recurrence relation

$$\lambda_{i+1} = \lambda\lambda_i - p\lambda_{i-1}$$
$$\lambda_1 = \lambda \quad \lambda_2 = \lambda^2 - p - 1.$$

This can be solved to estimate

$$\left|\sum_{i=0}^{n} \lambda_i\right|,$$

and then we use this estimate in the inequality

$$\sum_{y:d_p(x,y)\leq n} |f(y)|^2 \geq \operatorname{const} p^{-n} \left|\sum_{i=0}^{n} \lambda_i\right|^2 |f(x)|^2.$$

□

Definition 6.16. A measure ν on X is called *Hecke recurrent* if for every ν-measurable set $B \subset X$ and ν-almost every $x \in B$,

$$\mathfrak{X}_p^n(x) \cap B \neq \varnothing,$$

for infinitely many $n \in \mathbb{N}$.

The following theorem can be applied to the microlocal lift of joint eigenfunctions of the Laplacian and a Hecke operator T_p for some prime p to verify the Hecke recurrence assumption of Lindenstrauss's theorem.

Theorem 6.17. *Let φ_k be a sequence of eigenfunctions of T_p, $\|\varphi_k\|_2 = 1$, and write*

$$d\nu_k = |\varphi(g)|^2\, dg.$$

Then any limit ν of ν_k is Hecke recurrent.

Proof. From Proposition 6.15 and the fact that the T_{p^i} are self-adjoint, we have

$$\langle \sum_{i=0}^{n} T_{p^i} f, |\varphi_k|^2\rangle = \langle f, \sum_{i=0}^{n} T_{p^i} |\varphi_k|^2\rangle \geq \operatorname{const}\, n\langle f, |\varphi_k|^2\rangle.$$

Passing to the limit we obtain

$$\int_X \left(\sum_{i=0}^{n} T_{p^i} f\right) d\nu \geq \operatorname{const}\, n \int_X f\, d\nu. \tag{6.2}$$

By approximation, this stays true for any function $f \geq 0$ measurable.

Now let $B \subset X$ be a measurable set and define

$$B_k = \{x \in B : B \cap \mathfrak{X}_p^k(x) = \varnothing\},$$
$$C_l = \bigcap_{k\geq l} B_k.$$

Then $\cup_{l\geq 1} C_l$ is the set of points in B that do not come back to B infinitely many times. We want to show that this set is of ν-measure 0. For that fix $l \geq 1$ and apply (6.2) to $f = \chi_{C_l}$ the characteristic function of the set C_l. For any $z \in X$, $\mathfrak{X}_p(z) \cap C_l$ contains at most $(p+1)p^{l-1}$ vertices (the number of vertices in a ball of radius l). Otherwise there is $y \in B$ so that $\mathfrak{X}_p(y) = \mathfrak{X}_p(z)$ and $\mathfrak{X}_p^k(y) \cap C_l \neq \varnothing$, contradiction. So

$$\sum_{i=0}^{\infty} T_{p^i} \chi_{C_l} \leq (p+1)p^{l-1},$$

and sending n to $+\infty$ in (6.2) gives $\nu(C_l) = 0$. We thus have by sub-additivity

$$\nu\left(\cup_{l\geq 1} C_l\right) \leq \sum_{l\geq 1} \nu(C_l) = 0.$$

□

6.2. Positive entropy

As before let $X = \mathsf{SL}(2,\mathbb{Z})\backslash \mathsf{SL}(2,\mathbb{R})$. We also fix a prime p and denote by T_p the associated Hecke operator on X. For convenience, we will use a slightly different normalisation of the Hecke operator and define

$$T_p f(x) = \frac{1}{\sqrt{p}} \sum_{y : d_p(x,y)=1} f(y).$$

Note that changing the normalisation does not change the eigenfunctions.

As T_p is self-adjoint, we know its spectrum is real. For an element $\lambda \in \mathbb{R}$ of the spectrum we will also use the parameter $\theta \in \mathbb{C}$ defined by $\lambda = 2\cos\theta$.

We now want to verify the condition of positive entropy on almost every ergodic component for quantum limits. We will follow the paper of Brooks and Lindenstrauss [**BL14**] where the positive entropy is proved for joint eigenfunctions of the Laplacian and only one Hecke operator (the original proof by Bourgain and Lindenstrauss [**BL03**] was using joint eigenfunctions of the Laplacian and the whole algebra of Hecke operators, meaning T_p for every prime p).

Instead of defining the notion of entropy in general, we give an equivalent formulation of what it means for an invariant measure to have "positive entropy on almost every ergodic component". Let $T = a_1$ be the time 1 map associated with the geodesic flow, and $\mathcal{P} = \{E_1, E_2, \ldots, E_K\}$ be a partition of X. The $2N$*-th refinement* of $\mathcal{P}$ with respect to T is the set

$$\bigvee_{i=-N}^{N} T^{-i}\mathcal{P} = \{T^{N-1}E_{j_{-N}} \cap \ldots \cap E_{j_1} \cap T^{-1}E_{j_2} \cap \ldots \cap T^{-(N-1)}E_{j_N} : j_\alpha \in \{1, \ldots K\}\}$$

Definition 6.18. Let ν be a Borelian probability measure on X invariant by a measurable map T. Let $\mathcal{P}$ be a partition of X. We say that ν has *positive entropy on almost every ergodic component* if for any $\eta > 0$, there exists $\delta(\eta) > 0$ such that for all N sufficiently large and for any subset $\mathcal{P}'_N \subset \bigvee_{i=-N}^{N} T^{-i}\mathcal{P}$ we have

$$\nu\left(\bigcup_{E\in\mathcal{P}'_N} E\right) > \eta \quad \Rightarrow \quad |\mathcal{P}'_N| > e^{\delta N}.$$

This means that any element of a partition refinement by T has to be of small ν-measure with a quantitative estimate of this fact. It is a way of saying that the action of the map T is chaotic on the probability space (X, ν).

To prove this property we will use that the measures we are interested in arise from eigenfunctions of a Hecke operator. We will build a general operator on the $p+1$-regular tree in the next section that will be our main tool.

6.2.1. Wave propagation on the tree. We define the *Chebyshev polynomials of the first kind* by the recurrence relation

$$\begin{aligned} P_0(x) &= 1 \\ P_1(x) &= x \\ P_{n+1}(x) &= 2xP_n(x) - P_{n-1}(x) \end{aligned}$$

These polynomial have an equivalent definition as the unique polynomials satisfying

$$P_n(\cos\theta) = \cos(n\theta) \quad \theta \in \mathbb{C}, n \in \mathbb{N}.$$

As P_n is a polynomial, we can define an operator $P_n(T_p/2)$ without ambiguity by replacing the indeterminate by T_p in the expression of P_n. We see $P_n(T_p/2)$ as some kind of wave propagation operator, with a discrete time n. The action of this operator on an eigenfunction Φ of T_p of eigenvalues $\lambda = 2\cos\theta$ is simply

$$P_n(T_p/2)\Phi = \cos(n\theta)\Phi.$$

This tells us how the operator P_n acts *spectrally*, we now want to know how it acts *geometrically*. In other words we want to describe its kernel. In the same way as we defined invariant operators on the hyperbolic plane, we can define operators on the vertices of a $p+1$ regular tree $\mathfrak{X}_p$ in the following way. Let $K : \mathbb{N} \to \mathbb{C}$ be a compactly supported function. We see this function as a *radial kernel* and define the operator associated by

$$K(f)(x) = \sum_{y \in \mathfrak{X}_p} K(d_p(x,y))f(y).$$

Where by abuse of notation we denoted by K both the kernel and the operator. We also call this type of operator a *convolution operator*. Such an operator defines an operator on functions of $X = \mathsf{SL}(2,\mathbb{Z})\backslash \mathsf{SL}(2,\mathbb{R})$ in an obvious way by summing over the points in the Hecke tree.

The following lemma can be proved by induction from the recurrence relation for the Chebyshev polynomials.

LEMMA 6.19 ([**BL14**]). *The operator $P_n(T_p/2)$ is associated with an invariant kernel*

$$K_n(m) = \begin{cases} 0 & \text{if } m \text{ is odd or } m > n \\ \frac{1-p}{2p^{n/2}} & \text{if } m < n \text{ and } m \text{ is even} \\ \frac{1}{2p^{n/2}} & \text{if } |x| = n \end{cases}$$

This operator is used to build another one which spectrally emphasises a particular eigenfunction, and geometrically spreads the mass across the tree.

LEMMA 6.20. *Let $\eta > 0$. There exists $N_0 = N_0(\eta) \in \mathbb{N}$ such that for any $N \geq N_0$ and any T_p eigenfunction Φ_j there exists a convolution operator K_N satisfying:*

(1) *The kernel $K_N : \mathbb{N} \to \mathbb{C}$ is supported on $[0,N]$,*
(2) *There exists $\delta > 0$ and a constant $C > 0$ such that $K_N(m) \leq Cp^{-N\delta}$ for any $m \in [0,N]$,*
(3) *Any T_p eigenfunction on X is also an eigenfunction of K_N, of eigenvalue ≥ -1 and Φ_j has K_N-eigenvalue $> \eta^{-1}$.*

Proof. We only give an outline of the proof (for details see [**BL14**]). The idea is to superpose several operators $P_n(T_p/2)$ for different times n in order to get the right spectral action. More precisely we take

$$K_N = \sum_{j=1}^{L} \frac{2(L-j)}{L} P_{jq}(T_q/2),$$

for a good choice of parameter q. We have

$$K_N\Phi = \sum_{j=1}^{L} \frac{2(L-j)}{L} \cos(jq\theta)\Phi = F_L(q\theta) - 1,$$

if Φ is an eigenfunction of T_p with eigenvalue $2\cos\theta$, and where $F_L(t)$ is the Fejér kernel. The Fejér kernel is a positive function and an approximation of the identity when $L \to +\infty$. The idea is to choose q such that $q\theta$ is close to $0 \bmod 2\pi$ which can be done by Dirichlet's approximation theorem.

On the other hand we have, denoting by $[P_n(T_p/2)]$ the radial kernel of the operator $P_n(T_p/2)$

$$\begin{aligned} K_N(m) &\le \sum_{j=1}^{L} 2\left|\left[P_{jq}(T_p/2)\right](m)\right| \\ &\lesssim \sum_{j=1}^{L} p^{-jq/2} \\ &\lesssim p^{-q/2} \end{aligned}$$

for any $m \le qL$, $K(m) = 0$ otherwise. And the result can be obtained by adjusting the parameters L, q. □

6.2.2. Proof of the positive entropy. Define the neighbourhood of the identity in $\mathsf{SL}(2,\mathbb{R})$

$$B(\varepsilon,\tau) = \{a_t u^-_{s_-} u^+_{s_+} : t \in (-\tau,\tau), s_-, s_+ \in (-\varepsilon,\varepsilon)\},$$

where a_t, u^-_s and u^+_s are respectively the geodesic, unstable horocycle and stable horocycle flows.

We need two lemmas that we are not proving.

Lemma 6.21. *Let $\mathcal{P}$ be a partition of X and consider the $[2N\log p]$-th refinement by the geodesic flow a_1. Any element of the refinement is contained in the union of $O_c(1)$ tubes of the form $xB(cp^{-2N},\tau)$ for $x \in X$.*

To simplify notation we will assume that the elements of the refinement are contained in only one of those tubes. We also need an estimate on the number of points of the Hecke tree contained in one of these small tubes.

Lemma 6.22. *For a fixed τ small enough, there exists a constant $c = c(\tau)$ such that for any $x, z \in X$ and any $\varepsilon < cp^{-2N}$, the tube $zB(\varepsilon,\tau)$ contains at most $O(N)$ of the Hecke points in a ball of radius N around x, i.e. $\bigcup_{j\le N} \mathfrak{X}^j_p(x)$.*

Proof of the positive entropy. Let $\mathcal{P}$ be a partition of X and let $\{E_1, E_2, \dots, E_K\}$ be a collection of distinct elements of the $[2N\log p]$-th refinement of $\mathcal{P}$ by the geodesic flow a_1, such that the union $\mathcal{E} = \cup_{k=1}^K E_k$ we have

$$\nu(\mathcal{E}) > \eta$$

for a quantum limit ν. Recall that ν is a limit of the microlocal lifts $|\Phi_j|^2(g)dg$ on X of joint eigenfunctions of the Laplacian and T_p. The Φ_j are joint eigenfunctions of T_p and the Casimir operator Ω on X. We deduce that there exists a T_p eigenfunction Φ_j such that

$$\nu_j(\mathcal{E}) = \|\Phi_j \mathbf{1}_{\mathcal{E}}\|_2^2 > \eta.$$

From Lemma 6.20 we take the kernel K_N such that the K_N eigenvalue of Φ_j is $> \eta^{-1}$ and consider the correlation

$$\langle K_N(\Phi_j \mathbf{1}_{\mathcal{E}}), \Phi_j \mathbf{1}_{\mathcal{E}}\rangle$$

We decompose

$$\langle K_N(\Phi_j \mathbf{1}_{\mathcal{E}}), \Phi_j \mathbf{1}_{\mathcal{E}}\rangle = \sum_{k=1}^{K}\sum_{k'=1}^{K} \langle K_N(\Phi_j \mathbf{1}_{E_k}), \Phi_j \mathbf{1}_{E_{k'}}\rangle.$$

We have

$$\begin{aligned} |K_N(\Phi_j \mathbf{1}_{E_k})(x)| &\lesssim p^{-N\delta} \sum_{j\leq N} \sum_{y\in \mathcal{X}_p^j(x)} |\Phi_j(y)|\mathbf{1}_{E_k}(y) \\ &= p^{-N\delta} \sum_{j\leq N} \sum_{\alpha\in R(p^j)} |\Phi_j(\alpha x)|\mathbf{1}_{E_k}(\alpha x), \end{aligned}$$

where we denote by $R(p^j)$ the set of transformations that sends a point x on the points $\mathfrak{X}_p^j(x)$. Now

$$\langle K_N(\Phi_j \mathbf{1}_{E_k}), \Phi_j \mathbf{1}_{E_{k'}}\rangle \lesssim p^{-N\delta} \sum_{j\leq N} \sum_{\alpha\in R(p^j)} \int_{\alpha^{-1}E_k\cap E_{k'}} |\Phi_j(\alpha x)||\Phi_j(x)|dx.$$

By Lemma 6.21 and Lemma 6.22, there are $O(N)$ elements $\alpha \in \cup_{j\leq N} R(p^j)$ so that $\alpha^{-1}E_k \cap E_{k'} \neq \varnothing$, so

$$\begin{aligned} \langle K_N(\Phi_j \mathbf{1}_{E_k}), \Phi_j \mathbf{1}_{E_{k'}}\rangle &\lesssim p^{-N\delta} N \left(\int_{\alpha^{-1}E_k} |\Phi_j(\alpha x)|^2 dx\right)^2 \left(\int_{E_{k'}} |\Phi_j(x)|^2 dx\right)^2 \\ &= p^{-N\delta} N \|\Phi_j\|_{L^2(E_k)} \|\Phi_j\|_{L^2(E_{k'})} \end{aligned}$$

We obtain that

$$\begin{aligned} \langle K_N(\Phi_j \mathbf{1}_{\mathcal{E}}), \Phi_j \mathbf{1}_{\mathcal{E}}\rangle &\lesssim p^{-N\delta} N \left(\sum_k \|\Phi_j\|_{L^2(E_k)}\right)^2 \\ &\lesssim p^{-N\delta} NK \sum_k \|\Phi_j\|_{L^2(E_k)}^2 \\ &\lesssim p^{-N\delta} NK, \end{aligned}$$

since $\|\Phi_j\|_2 = 1$.

On the other hand, we decompose $\Phi_j \mathbf{1}_{\mathcal{E}}$ on an orthonormal basis of $L^2(X)$ of T_p eigenfunctions $\{\psi_i\}$ containing Φ_j.

$$\Phi_j \mathbf{1}_{\mathcal{E}} = \langle \Phi_j \mathbf{1}_{\mathcal{E}}, \Phi_j\rangle \Phi_j + \sum_{\psi_i\neq\Phi_j} \langle \Phi_j \mathbf{1}_{\mathcal{E}}, \psi_i\rangle \psi_i.$$

We have

$$\langle \Phi_j \mathbf{1}_{\mathcal{E}}, \psi_i\rangle = \|\Phi_j \mathbf{1}_{\mathcal{E}}\|_2^2 \geq \eta,$$

and we get by the orthonormality

$$\sum_{\psi_i\neq\Phi_j} |\langle \Phi_j \mathbf{1}_{\mathcal{E}}, \Phi_j\rangle|^2 = \|\Phi_j \mathbf{1}_{\mathcal{E}}\|_2^2 - \|\Phi_j \mathbf{1}_{\mathcal{E}}\|_2^4 < \|\Phi_j \mathbf{1}_{\mathcal{E}}\|_2^2 (1-\eta).$$

By Lemma 6.20 the K_N eigenvalues of ψ_i are > -1, so

$$\begin{aligned}\langle K_N(\Phi_j \mathbf{1}_{\mathcal{E}}), \Phi_j \mathbf{1}_{\mathcal{E}}\rangle &= \sum_{\psi_i} |\langle \Phi_j \mathbf{1}_{\mathcal{E}}, \psi_i\rangle|^2 \langle K_N \psi_i, \psi_i\rangle \\ &\geq |\langle \Phi_j \mathbf{1}_{\mathcal{E}}, \Phi_j\rangle|^2 \langle K_N \Phi_j, \Phi_j\rangle - \sum_{\psi_i \neq \Phi_j} |\langle \Phi_j \mathbf{1}_{\mathcal{E}}, \psi_i\rangle|^2 \\ &\geq \|\Phi_j \mathbf{1}_{\mathcal{E}}\|_2^4 \langle K_N \Phi_j, \Phi_j\rangle - \|\Phi_j \mathbf{1}_{\mathcal{E}}\|_2^2 (1-\eta) \\ &\geq \|\Phi_j \mathbf{1}_{\mathcal{E}}\|_2^2 (\|\Phi_j \mathbf{1}_{\mathcal{E}}\|_2^2 \eta^{-1} - (1-\eta)) \\ &\geq \eta(\eta\eta^{-1} - 1 + \eta) = \eta^2 > 0.\end{aligned}$$

So we have

$$Np^{-N\delta} K \gtrsim \langle K_N(\Phi_j \mathbf{1}_{\mathcal{E}}), \Phi_j \mathbf{1}_{\mathcal{E}}\rangle \gtrsim \eta^2$$

and we get

$$K \gtrsim \eta^2 N^{-1} p^{\delta N},$$

which is the type of lower bound we wanted for the number of elements in the collection. □

Bibliography

[Ber16] Nicolas Bergeron, *The spectrum of hyperbolic surfaces*, Universitext, Springer, Cham; EDP Sciences, Les Ulis, 2016, Appendix C by Valentin Blomer and Farrell Brumley, Translated from the 2011 French original by Brumley [2857626]. MR 3468508

[BL03] Jean Bourgain and Elon Lindenstrauss, *Entropy of quantum limits*, Comm. Math. Phys. **233** (2003), no. 1, 153–171. MR 1957735

[BL14] Shimon Brooks and Elon Lindenstrauss, *Joint quasimodes, positive entropy, and quantum unique ergodicity*, Invent. Math. **198** (2014), no. 1, 219–259. MR 3260861

[EW10] Manfred Einsiedler and Thomas Ward, *Arithmetic quantum unique ergodicity*, Lecture notes Arizona Winter School, 2010.

[EW11] ______, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, vol. 259, Springer-Verlag London, Ltd., London, 2011. MR 2723325

[Hej76] Dennis A. Hejhal, *The Selberg trace formula for* $\mathrm{PSL}(2, R)$*. Vol. I*, Lecture Notes in Mathematics, Vol. 548, Springer-Verlag, Berlin-New York, 1976. MR 0439755

[Kat92] Svetlana Katok, *Fuchsian groups*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 1992. MR 1177168

[Lin06] Elon Lindenstrauss, *Invariant measures and arithmetic quantum unique ergodicity*, Ann. of Math. (2) **163** (2006), no. 1, 165–219. MR 2195133

[Mar12] Jens Marklof, *Selberg's trace formula: an introduction*, Hyperbolic geometry and applications in quantum chaos and cosmology, London Math. Soc. Lecture Note Ser., vol. 397, Cambridge Univ. Press, Cambridge, 2012, pp. 83–119. MR 2885182

[Sou10] Kannan Soundararajan, *Quantum unique ergodicity for* $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, Ann. of Math. (2) **172** (2010), no. 2, 1529–1538. MR 2680500

[Zel87] Steven Zelditch, *Uniform distribution of eigenfunctions on compact hyperbolic surfaces*, Duke Math. J. **55** (1987), no. 4, 919–941. MR 916129