ARITHMETIC QUANTUM CHAOS

EDGAR ASSING

ABSTRACT. In this course, taught at the University of Bonn in the winter term 23/24, we discuss (arithmetic) quantum unique ergodicity. In particular, we want to outline the proof of Lindenstrauss' AQUE-Theorem. While preparing this material we were greatly influenced by the very nice 2010 lecture notes *Arithmetic Quantum Unique Ergodicity for* $\Gamma\backslash\mathbb{H}$ by M. Einsiedler and T. Ward. Note that these notes may contain typos and misunderstandings, for which I take full responsibility. For personal use only!

1. Introduction

In order to explain the goals of this lecture we will say a couple of words about *Chaos* and *Quantum Chaos*. This will lead us naturally to *Arithmetic Quantum Chaos*. Note that this introduction is not meant to be completely formal. Many of the notions touched upon will be revisited later in the course in the context of hyperbolic surfaces. We refer to the nice texts [Sar1,Sar2] for an alternative source of inspiration.

1.1. **Chaos.** While we have an intuitive feeling what chaos should entail, it is not immediately clear how to fully capture this notion mathematically. We will attempt to describe some aspects of chaos from the point of view fo dynamical systems. For more details we refer to [AA,EW] and also [Sog].

Definition 1.1. A classical dynamical system is a collection $(M, \mu, \{\phi_t\}_{t\in\mathbb{R}})$, where M is a smooth manifold, μ is a measure and $\phi_t\colon M \to M$ (with $t \in \mathbb{R}$) is a one parameter group[1] of measure preserving[2] diffeomorphism.

Remark 1.1. This notion is taken from [AA]. It is slightly different from the usual notion of an abstract dynamical system, but all dynamical systems we are interested in will actually be classical dynamical systems as above.

[1]By this we mean that

$$\phi_{t+s}(x) = \phi_t(\phi_s(x)) = \phi_s(\phi_t(x)),\ \phi_0(x) = x$$

holds and that $\mathbb{R} \times X \ni (t, x) \mapsto \phi_t(x) \in X$ is continuous.

[2]Let $\mathcal{B}$ be the Borel σ-algebra on M. We say that ϕ_t is measure preserving if for every $E \in \mathcal{B}$ we have

$$\mu(\phi_t^{-1}(E)) = \mu(E).$$

Note that $\phi_t^{-1}(E) \in \mathcal{B}$ is automatic, since ϕ_t is assumed to be a diffeomorphism.

Example 1.2. We give two examples:

(1) Let X be a compact Riemannian Manifold and let $M = T^1X$ be the unit tangent bundle. Then the geodesic flow $g_t \colon M \to M$ defines a classical dynamical system. We can also consider the geodesic flow on X, which we will also denote by g_t.
(2) Write $p_1, \dots, p_n, q_1, \dots, q_n$ for coordinates of $\mathbb{R}^{2n}$ and let $H \colon \mathbb{R}^{2n} \to \mathbb{R}$ be smooth. Then the equations
$$\frac{d\mathbf{q}}{dt} = \frac{\partial H}{\partial \mathbf{p}} \text{ and } \frac{d\mathbf{p}}{dt} = -\frac{\partial H}{\partial \mathbf{q}}$$
define the so called Hamiltonian Flow. (Here we shorten $\mathbf{p} = (p_1, \dots, p_n)$ and $\mathbf{q} = (q_1, \dots, q_n)$.) In particular, if $n = 3$ and $H(\mathbf{p}, \mathbf{q}) = \frac{1}{2m}(q_1^2+q_2^2+q_3^2)$, then the Hamiltonian Flow describes the motion of a single, freely moving particle of mass m in 3-dimensional space.

Let us return to the problem of formalizing chaos. A key property of chaotic systems should be some *unpredictability*. From this point of view it makes sense to partition M into *boxes*, say c_i and study the probabilities p_i of finding the trajectories $\phi_t(x)$ in them. We can think of these probabilities as proportional to *the (normalized) time $\phi_t(x)$ spends in c_i.*

This leads us to define the time averages
$$[\mathrm{Av}_T f](x) = \frac{1}{T} \int_0^T f(\phi_t(x))dt,$$
for some functions $f \colon M \to \mathbb{C}$, which will be specified later.

Turning back to our informal discussion above we would now expect that p_i is proportional to
$$\lim_{T\to\infty} [\mathrm{Av}_T \mathbb{1}_{c_i}](x),$$
given the limit exists. Here $\mathbb{1}_{c_i}$ is the indicator function on $c_i \subseteq M$. One aspect of chaotic flows is that one expects the probabilities p_i and thus the limits of the time averages to be independent of the starting point x of the flow. If this is true, then one should have
$$\lim_{T\to\infty} [\mathrm{Av}_T f](x) = \int_M f(x)d\mu(x). \tag{1}$$
Note that at the moment this is a formal statement, since we have not taken any convergence and regularity issues into account.

Definition 1.2. We call $(M, \mu, \{\phi_t\}_{t\in\mathbb{R}})$ ergodic if any Borel set E, which is invariant under the flow satisfies $\mu(E) \in \{0, 1\}$.

Theorem 1.3 (Mean Ergodic Theorem). *Let M be a compact smooth manifold with Borel probability measure μ and let $(M, \mu, \{\phi_t\}_{t\in\mathbb{R}})$ be a classical dynamical system. Then, if for $f \in L^2(X, \mu)$, $\pi(f)$ denotes the orthogonal projection of f*

onto the subspace $\mathcal{I} \subseteq L^2(X,\mu)$ of functions which are invariant under the flow, we have

$$\lim_{T\to\infty} \int_X |[Av_T f](x) - \pi(f)(x)|^2 \, d\mu(x) = 0.$$

Proof. Note that

$$\mathcal{I} = \{f \in L^2(X,\mu) \colon f(\phi_t(\cdot)) = f(\cdot) \text{ for all } t \in \mathbb{R}\}$$

is a closed subspace. Obviously, for $f \in \mathcal{I}$, the statement is trivially true.

Put

$$\mathcal{S} = \{f \in L^2(X,\mu) \colon f(\cdot) = h(\phi_s(\cdot)) - h(\cdot) \text{ for some } h \in L^2(X,\mu) \text{ and } s \in \mathbb{R}^\times\}.$$

Let $g(\cdot) = h(\phi_s(\cdot)) - h(\cdot) \in \mathcal{S}$ and $f \in \mathcal{I}$. Then we compute

$$\langle g, f\rangle = \langle h(\phi_s(\cdot)), f\rangle - \langle h, f\rangle = \langle h, f(\phi_{-s}(\cdot))\rangle - \langle h, f\rangle = 0.$$

Thus $\mathcal{S}$ is orthogonal to $\mathcal{I}$. It is easy to check that the conclusion of the theorem holds for $f \in \mathcal{S}$. Indeed, we have

$$\frac{1}{T}\int_0^T [h(\phi_{t+s}(\cdot)) - h(\phi_t(\cdot))] \, dt = \frac{1}{T}\int_T^{T+s} h(\phi_t(\cdot))dt - \frac{1}{T}\int_0^s h(\phi_t(\cdot))dt \to 0$$

as $T \to \infty$.

It remains to be seen that $\overline{\mathcal{S}+\mathcal{I}} = L^2(X,\mu)$. We argue by contradiction and assume that $0 \neq g \in L^2(X,\mu)$ is orthogonal to $\overline{\mathcal{I}+\mathcal{S}}$. In particular, we have $g \notin \mathcal{I}$ so that there is $t \in \mathbb{R}$ with $g(\phi_t(\cdot)) - g(\cdot) \neq 0$. Furthermore,

$$0 = \langle g, g(\Phi_t(\cdot)) - g\rangle = \langle g, g(\Phi_t(\cdot))\rangle - \|g\|_{L^2}^2.$$

This allows us to conclude

$$\|g(\phi_t(\cdot)) - g\|_{L^2}^2 = \|g(\phi_t(\cdot))\|_{L^2}^2 + \|g\|_{L^2}^2 - \langle g, g(\phi_t(\cdot))\rangle - \langle g(\phi_t(\cdot)), g\rangle = 0.$$

But this implies $g(\phi_t(\circ)) - g = 0$, which is the desired contradiction. □

It is a nice exercise to show that if ϕ_t is ergodic, then a function in $L^2(X,\mu)$ that is invariant under the flow must be constant. Thus we have

$$\pi(f) = f_{\mathrm{av}} = \int_M f(x)d\mu(x).$$

The upshot is that the theorem above makes (1), for ergodic systems, precise in an L^2-sense:

Corollary 1.4. *Let M be a compact smooth manifold with Borel probability measure μ and let $(M,\mu,\{\phi_t\}_{t\in\mathbb{R}})$ be an ergodic classical dynamical system. Then we have*

$$\lim_{T\to\infty} \int_X |[Av_T f](x) - f_{av}|^2 d\mu(x) = 0.$$

Ergodicity only partially captures the notion of chaos in the context of classical (and abstract) dynamical systems. Later we will encounter entropy which is a more quantitative invariant.

1.2. **Quantum Chaos.** The naive quantisation of the Hamiltonian Flow (with Hamiltonian $H\colon \mathbb{R}^{2n} \to \mathbb{R}$) is simply given by the operator

$$\mathcal{H} = H(p_1, \dots, p_n, \frac{\hbar}{i}\frac{\partial}{\partial p_1}, \dots, \frac{\hbar}{i}\frac{\partial}{\partial p_n})$$

where $\hbar$ is the Planck Constant. In our basic example when $n = 3$ and $H(\mathbf{p}, \mathbf{q}) = \frac{1}{2m}\|\mathbf{p}\|^2$, we simply get

$$\mathcal{H} = -\frac{\hbar^2}{2m} \cdot \Delta.$$

More generally the (naive) quantisation of the geodesic flow (on a Riemannian manifold X) is the Laplace Beltrami Operator Δ_X. Thus there should be a close connection between classical properties of the geodesic flow g_t and the spectrum of Δ_X. This is the context of Bohr's correspondence principle, which very roughly states that the classical dynamical system should be visible in high frequency quantum states.

Before we continue our discussion let us recall some properties of Δ_X in the special case of compact X:

- $-\Delta_X$ is an unbounded essentially self-adjoint positive operator on $L^2(X, \mu)$.
- The spectrum is a pure point spectrum
$$0 = \lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \dots$$
with only ∞ as accumulation point.
- There is an orthonormal basis of $L^2(X, \mu)$ consisting of eigenfunctions $v_j \in \mathcal{C}^\infty(X)$. In particular $-\Delta_X v_j = \lambda_j v_j$.

We define the measures μ_j (on X) by

$$\int_X f(x) d\mu_j(x) = \int_X f(x) \cdot |v_j(x)|^2 d\mu_X. \tag{2}$$

where we write μ_X for the probability volume measure. The physical interpretation of these measures is that $\mu_j(\Omega)$ is the probability of finding a particle of energy λ_j in $\Omega \subseteq X$. We call weak-$*$-limit points of the sequences$\{\mu_j\}_j$ quantum limits. It is a basic question to ask what measures are quantum limits.

Theorem 1.5 (Shnirelman's Theorem/Quantum Ergodicity)**.** *Let X be as in the preceding discussion and assume that the geodesic flow g_t on X is ergodic with respect to μ_X. Then there exists a density one subsequence*[3] $\{v_{j_k}\}$ *such that*

$$\int_X f(x) d\mu_{j_k}(x) \to \int_X f(x) d\mu_X(x) \text{ for all } f \in \mathcal{C}^\infty(X). \tag{3}$$

We say that $\{\mu_{j_k}\}$ *equidistributes in physical space.*

[3] By this we mean that $\frac{\sharp\{k\colon \lambda_{j_k} \leq R\}}{\sharp\{j\colon \lambda_j \leq R\}} \to 1$ as $R \to \infty$

Remark 1.6. There is a version of this theorem on the unique tangent bundle of X, but to formulate it appropriately we need to talk about so called microlocal lifts. This will be the content of a later chapter.

Note that Shnirelman's theorem, which was also independently proven by Zelditch and Colin de Verdière respectively, holds in great generality and is essentially best possible. Indeed, there are examples of manifolds with ergodic geodesic flow, where passing to a subsequence is necessary. Nonetheless there is the following conjecture:

Conjecture 1.1 (Rudnick-Sarnak 1994)**.** *Let X be a compact Riemann surface with constant curvature -1. Then μ_X is the only quantum limit. In other words,* (3) *holds without passing to a subsequence.*

This is the Quantum Unique Ergodicty (QUE) conjecture and it is in general still wide open.

1.3. **Arithmetic Quantum Chaos.** There are special hyperbolic surfaces, so called *arithmetic* surfaces, that come from number theory. The special feature of arithmetic surfaces is that they come with many additional (non-trivial) symmetries. These give rise to so called Hecke-Operators, which commute with the Laplace-Beltrami-Operator. It is therefore natural to consider simultaneous eigenfunctions of $-\Delta_X$ and (almost) all Hecke-Operators. We call such eigenfunctions simply *joint eigenfunctions*. Roughly speaking studying questions from Quantum Chaos in the setting of arithmetic manifolds and joint eigenfunctions can be called Arithmetic Quantum Chaos. A key result in this area is

Theorem 1.7 (Lindenstrauss 2006)**.** *Let X be an arithmetic hyperbolic surface and let $\{v_j\}$ be a basis of joint eigenfunctions for $L^2(X,\mu_X)$. Then*

$$\int_X f(x)d\mu_j(x) \to \int_X f(x)d\mu_X(x) \text{ for all } f \in \mathcal{C}^\infty(X).$$

Thus the full sequence of joint eigenfunctions equidistributes in physical space.

The goal of this lecture is to have a careful look at this theorem and to explain its proof. Following [BrL] we will actually prove a slightly stronger result.

Contents

2. The geometry of $\mathbb{H}$ and its quotients

This section closely follows [EW, Section 9]. Let $\mathbb{H} = \{x+iy \in \mathbb{C} \colon y > 0\}$ be the upper half plane. The tangent bundle is given by $T\mathbb{H} = \mathbb{H} \times \mathbb{C}$ and $T_z = \{z\} \times \mathbb{C}$ is the tangent plane at z. Given a differentiable function $\phi \colon [0,1] \to \mathbb{H}$ we define its derivative at t by

$$D\phi(t) = (\phi(t), \phi'(t)) \in T_{\phi(t)}\mathbb{H}.$$

The hyperbolic Riemannian metric is given by the collection of inner products

$$\langle (z,v), (z,w) \rangle_z = \frac{v \cdot \overline{w}}{\operatorname{Im}(z)^2}$$

for $z \in \mathbb{H}$ and $(z,v), (z,w) \in T_z\mathbb{H}$. This induces a hyperbolic metric as follows. First, given a piecewise differentiable curve $\phi \colon [0,1] \to \mathbb{H}$, we define the length by

$$L(\phi) = \int_0^1 \|D\phi(t)\|_{\phi(t)} dt, \text{ where } \|D\phi(t)\|_{\phi(t)}^2 = \langle D\phi(t), D\phi(t) \rangle_{\phi(t)}.$$

The desired metric is now given by

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

where the infimum is taken over all continuous, piecewise differentiable paths $\phi \colon [0,1] \to \mathbb{H}$ with $\phi(0) = z_0$ and $\phi(1) = z_1$. We compactify the upper half plane by adding the boundary $\partial\mathbb{H} = \mathbb{R} \cup \{\infty\}$. More precisely, $\overline{\mathbb{H}} = \mathbb{H} \cup \partial\mathbb{H}$. The metric can be extended naturally to $\overline{\mathbb{H}}$. The unit tangent bundle is

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

Let $\mathrm{SL}_2(\mathbb{R})$ be the special linear group and write $\mathrm{PSL}_2(\mathbb{R}) = \mathrm{SL}_2(\mathbb{R})/\{\pm 1\}$ for the projective special linear group. It acts on $\mathbb{H}$ via Möbius transformations:

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

The derivative action $Dg \colon T\mathbb{H} \to T\mathbb{H}$ of g is given by

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

One can verify the following facts:

- The action is isometric: $d(g(z_0), g(z_1)) = d(z_0, z_1)$ for all $z_0, z_1 \in \mathbb{H}$ and $g \in \mathrm{PSL}_2(\mathbb{R})$. Furthermore, the action of $g \in \mathrm{PSL}_2(\mathbb{R})$ on $T\mathbb{H}$ given by Dg preserves the Riemannian metric.
- The action of $\mathrm{PSL}_2(\mathbb{F})$ is transitive on $\mathbb{H}$.
- The stabilizer of $i \in \mathbb{H}$ is given by $\mathrm{PSO}_2 = \mathrm{SO}_2/\{\pm 1\}$. In particular we can identify $\mathbb{H} \cong \mathrm{PSL}_2(\mathbb{R})/\mathrm{PSO}_2$.
- Dg preserves the length of tangent vectors, so that $\mathrm{PSL}_2(\mathbb{R})$ acts on the unit tangent bundle $T^1\mathbb{H}$. This action is simply transitive and we have the identification $T^1\mathbb{H} \cong \mathrm{PSL}_2(\mathbb{R})$. (This identification sends $(i, i) \in T^1\mathbb{H}$ to the identity $I_2 \in \mathrm{PSL}_2(\mathbb{R})$.)

Proposition 2.1. *For any two points $z_1, z_2 \in \mathbb{H}$ there is a unique path*

$$\phi\colon [0, d(z_1, z_2)] \to \mathbb{H}$$

of unit speed with $\phi(0) = z_1$ and $\phi(d(z_1, z_2)) = z_2$. Moreover, there is a unique isometry $g \in PSL_2(\mathbb{R})$ such that $\phi(t) = g(e^t i)$.

Proof. We only give a sketch of the argument. The first step is to treat the case $z_1 = y_1 i$ and $z_2 = y_2 i$ with $0 < y_1 < y_2$. This can be done by direct computation. It turns out that $d(iy_1, iy_2) = \log(y_2) - \log(y_1)$ and the minimizing path is $\phi(t) = y_1 \cdot \left(\frac{y_2}{y_1}\right)^t i$.

In the general case it remains to find $g \in \mathrm{PSL}_2(\mathbb{R})$ such that $g^{-1}z_1 = i$ and $g^{-1}z_2 = yi$ for some $y > 1$. We leave this as an exercise to the reader. $\square$

From this is it easy to deduce that the geodesic curves in $\mathbb{H}$ are precisely vertical lines or half circles centered at the real line. We can now define the geodesic flow $g_t\colon T^1\mathbb{H} \to T^1\mathbb{H}$ as follows. We first set

$$g_t((i, i)) = (e^t i, e^t i) = Da_t^{-1}(i, i) \text{ for } a_t = \begin{pmatrix} e^{-t/2} & 0 \\ 0 & e^{t/2} \end{pmatrix}.$$

We extend this to an arbitrary point $(z, v) = g(i, i)$ by setting

$$g_t(z, v) = Dg(g_t(i, i)) = D(ga_t^{-1})(i, i).$$

In particular, under the identification $T^1\mathbb{H} = \mathrm{PSL}_2(\mathbb{R})$, the geodesic flow is described as right multiplication by the inverse matrix of a_t (i.e. $R_{a_t}(g) = ga_t^{-1}$). On the other hand the derivative action of $\mathrm{PSL}_2(\mathbb{R})$ on $T^1\mathbb{H} = \mathrm{PSL}_2(\mathbb{R})$ is given by left multiplication.

We introduce the following measures

$$\int_{T^1\mathbb{H}} f((z, v))d\mu_{T^1\mathbb{H}} = \int_{\mathbb{R}} \int_{\mathbb{R}_{>0}} \int_{S^1} f((x + iy, y\eta)d\eta \frac{dydx}{y^2}.$$

On the upper half plane we have

$$\int_{\mathbb{H}} f(z) d\mu_{\mathbb{H}} = \int_{\mathbb{R}} \int_{\mathbb{R}_{>0}} f(x+iy) \frac{dydx}{y^2}.$$

A direct computation reveals, that both these measures are invariant under the corresponding actions of $\mathrm{PSL}_2(\mathbb{R})$. In particular, using the identification $T^1\mathbb{H} = \mathrm{PSL}_2(\mathbb{R})$ the measure $\mu_{T^1\mathbb{H}}$ defines a (left) Haar measure of $\mathrm{PSL}_2(\mathbb{R})$. This Haar measure will be also be denoted by $\mu_{T^1\mathbb{H}}$ for now. We directly obtain

Lemma 2.2. *The geodesic flow $(T^1\mathbb{H}, \mu_{T^1\mathbb{H}}, g_t)$ is a classical dynamical system.*

Definition 2.1. A Fuchsian group is a discrete subgroup of $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$.

Definition 2.2. A fundamental set $\mathcal{E} \subseteq \mathrm{PSL}_2(\mathbb{R})$ for Γ is a measurable set of representatives of the set of orbits $\Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$. We define the co-colume of Γ as $\mu_{T^1\mathbb{H}}(\mathcal{E})$.[4]

Definition 2.3. A lattice (in $\mathrm{PSL}_2(\mathbb{R})$) is a Fuchsian group with finite co-volume. A lattice Γ is called uniform if $\Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ is compact.

Example 2.3. The group $\mathrm{PSL}_2(\mathbb{Z})$ is a lattice, which is not uniform.

From now on let $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$ be a lattice. We associate the spaces

$$M = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R}) = \Gamma\backslash T^1\mathbb{H} \text{ and } X = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})/\,\mathrm{PSO}_2 = \Gamma\backslash\mathbb{H}.$$

Most important for us is that M (resp. X) inherits very nice structural properties from $T^1\mathbb{H}$ (resp. $\mathbb{H}$). Essentially it will be sufficient for all our purposes to think of M as a quotient of the semisimple (real) Lie group $\mathrm{PSL}_2(\mathbb{R})$.

Remark 2.4. In general X is not a Riemann surface. Indeed, if Γ contains elements of finite order (i.e. elliptic elements) then $\Gamma\backslash\mathbb{H}$ is not smooth. It turns out that X is an orbifold. In particular, M is strictly speaking not the unit tangent bundle of X in the classical sense. However, these technicalities will not play a role and we will mostly work with the quotients as such. Nonetheless, for motivational and visual purposes it is good to keep in mind that M is essentially the unit tangent bundle of X.

We equip M with a natural measure μ_M as follows. For $B \subseteq M$ we define

$$\mu_M(B) = \mu_{T^1\mathbb{H}}(\mathcal{E} \cap \pi^{-1}B)$$

where $\mathcal{E}$ is a fundamental set for Γ and $\pi\colon T^1\mathbb{H} \to M$ is the canonical projection coming from the quotient construction. Similarly one defines a measure μ_X on X.

The geodesic flow on M is given by

$$g_t x = R_{a_t^{-1}}(x) = x a_t^{-1},$$

for $x = \Gamma g \in M$ with $g \in \mathrm{PSL}_2(\mathbb{R})$. (Here we crucially use the identification $T^1\mathbb{H} = \mathrm{PSL}_2(\mathbb{R})$.)

[4]Note that this is independent of the choice of the fundamental set.

Lemma 2.5. *Let $\Gamma \subseteq PSL_2(\mathbb{R})$ be a lattice. Then μ_M is invariant under R_g for $g \in PSL_2(\mathbb{R})$. In particular, μ_M is an invariant measure for the geodesic flow on M.*

Proof. Exercise. □

Remark 2.6. More generally given a (compact) Riemannian manifold X with cotangent bundle T^*X one can consider the geodesic flow

$$g_t \colon T^*X \setminus \{0\} \to T^*X \setminus \{0\}.$$

By Liouville's theorem this flow preserves the Liouville measure μ on the cotangent bundle, so that $(T^*X, \mu, \{g_t\})$ is a classical dynamical system. Note that one can consider this flow also on the co-sphere bundle S^*X instead of T^*X.

In a little bit the groups

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

play an important role. We refer to U^- (resp. U^+) as the stable horocycle flow (resp. unstable horocycle flow.) The name is connected to the following little computation. Let $u^- = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix}$ and $u^+ = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix}$. Then we have

$$a_t u^- a_t^{-1} = \begin{pmatrix} 1 & se^{-t} \\ 0 & 1 \end{pmatrix} \to 1$$

as $t \to \infty$ and

$$a_t u^+ a_t^{-1} = \begin{pmatrix} 1 & 0 \\ se^t & 1 \end{pmatrix} \to 1$$

as $t \to -\infty$. We will also set $A = \{a_t \colon t \in \mathbb{R}\}$.

It is an important exercise to check that U^+ and U^- generate $\mathrm{PSL}_2(\mathbb{R})$. It is actually easier (and sufficient for our purposes) to see that U^+, U^- and A generate $\mathrm{PSL}_2(\mathbb{R})$. To see this let take $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{R})$. If $a \neq 0$, then it is easy to find $\alpha, \beta, t \in \mathbb{R}$ such that

$$\begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} \cdot a_t \cdot \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} e^{-\frac{t}{2}} & \beta e^{-\frac{t}{2}} \\ \alpha e^{-\frac{t}{2}} & \alpha\beta e^{-\frac{t}{2}} + e^{\frac{t}{2}} \end{pmatrix} = \pm g.$$

The case $a = 0$ is handled by choosing $\alpha, \beta \in \mathbb{R} \setminus \{0\}$ such that

$$\begin{pmatrix} 1 & 0 \\ \alpha & 1 \end{pmatrix} \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -\frac{1}{\beta} & 1 \end{pmatrix} = \begin{pmatrix} 0 & \beta \\ -\frac{1}{\beta} & 1 + \alpha\beta \end{pmatrix} = \pm g.$$

We are now ready to show that the geodesic flow on M is ergodic:

Proposition 2.7. *Let Γ be a lattice. Then the geodesic flow on $M = \Gamma \backslash PSL_2(\mathbb{R})$ is ergodic (w.r.t. μ_M).*

Proof. Let $f \in L^2(M, \mu_M)$ be invariant under the geodesic flow. We can imagine $f\colon \mathrm{PSL}_2(\mathbb{R}) \to \mathbb{C}$ with

$$\|f\|^2 = \int_{\mathcal{E}} |f(g)|^2 d\mu_{T^1\mathbb{H}} < \infty$$

and $f(\gamma g) = f(g)$ for all $\gamma \in \Gamma$ and all $g \in \mathrm{PSL}_2(\mathbb{R})$. The invariance under the geodesic flow translates into

$$[R_{a_t^{-1}} f](g) = f(R_{a_t^{-1}} g) = f(g a_t^{-1}) = f(g)$$

for all $t \in \mathbb{R}$ and all $g \in \mathrm{PSL}_2(\mathbb{R})$. In order to show that the geodesic flow is ergodic we have to show that f is constant. To do so we will show that f is U^+ and U^- invariant. This suffices because U^+ and U^- generate $\mathrm{PSL}_2(\mathbb{R})$, so that f can only be constant. Without loss of generality we assume that $\|f\| = 1$.

It will now be convenient to observe that we have a unitary representation of $\mathrm{PSL}_2(\mathbb{R})$ on the Hilbert space $L^2(M, \mu_M)$ given by $g \mapsto R_g$.[5] We see that since f is invariant under the geodesic flow it is fixed by A. We consider the auxiliary function

$$p(g) = \langle R_g f, f \rangle = \int_{\mathcal{E}} f(hg)\overline{f(h)} d\mu_{T^1\mathbb{H}}(h).$$

Of course we have

$$p(a_1 g a_2) = \langle R_{a_1 g a_2} f, f \rangle = \langle R_g R_{a_2} f, R_{a_1^{-1}} f \rangle = \langle R_g f, f \rangle = p(g)$$

for all $a_1, a_2 \in A$. Furthermore, by Cauchy-Schwarz we have

$$|p(g)| \leq \|h_g f\| \cdot \|f\| = \|f\|^2 = 1.$$

Note that equality holds if and only if $h_g f$ and f are linearly depend. In particular, we deduce that if $p(g) = 1$, then $R_g f = f$. Indeed, we obviously have $|p(g)| = 1$ in this case, so that the argument above yields $R_g f = \eta f$ for $\eta \in S^1$. On the other hand this implies $p(g) = \eta$, so that $\eta = 1$ is the only possibility.

We are now almost done. For any $u^+ \in U^+$ we see that

$$1 = \|f\|^2 = p(I_2) = \lim_{t \to -\infty} p(a_t u^+ a_t^{-1}) = p(u^+).$$

Thus by the argument above we must have $R_{u^+} f = f$, so that f is U^+ invariant. The argument for U^- is essentially the same. □

Remark 2.8. This argument relies on a very general observation of Margulis, which can be used to show such ergodicity results in great generality. Let us formulate this observation as a general fact. Let $\mathcal{H}$ be a Hilbert space equipped with an unitary action by a metric group G. Further, let $v_0 \in \mathcal{H}$ be a vector fixed by a subgroup $A \subseteq G$. Then v_0 is fixed by any other element $h \in G$ with

$$B_\delta^G(h) \cap A B_\delta^G A \neq \emptyset \tag{4}$$

[5] A unitary representation of a metrizable group G on a Hilbert space $\mathcal{H}$ is given by an action $G \times \mathcal{H} \to \mathcal{H}$ such that g acts unitarily on $\mathcal{H}$ and for every $v \in \mathcal{H}$ the map $g \mapsto g(v)$ is continuous.

for every $\delta > 0$. Here $B_\delta^G = \{g \in G\colon d_G(g,e) < \delta\}$ and $B_\delta^G(h) = hB_\delta^G$.

Remark 2.9. More generally it can be shown that the geodesic flow for compact Riemannian manifolds with negative curvature is ergodic.

3. Microlocal lifts

We start with a brief discussion of the classical theory in order to motivate the upcoming computations. A nice classical reference is [Sog].

Let X be a Riemannian manifold and let $M = S^*X$ be its co-sphere bundle. Given a classical observable $a(x,\xi) \in \mathcal{C}_c^\infty(T^*X)$ one associates the corresponding quantum observable

$$\mathrm{Op}_h(a) = a^w(x, \frac{h}{i}\partial_x)\colon L^2(X) \to L^2(X).$$

Here $0 < h < 1$ is the semmiclassical parameter (replacing the Planck constant $\hbar$ in the introduction) and w stands for the so called Weyl Quantization. It is not extremely important for us how this quantization is defined. For $X = \mathbb{R}^n$ one has

$$\mathrm{Op}_h(a)f(x) = (2\pi h)^{-n}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} e^{\frac{i}{h}(x-y)\xi} a(\frac{x+y}{2},\xi)f(y)dyd\xi.$$

For general one essentially uses charts and partitions of unity to transport this definition from $\mathbb{R}^n$ to a Riemannian Manifold X. Important for us are the following properties

$$\mathrm{Op}_h(a)\mathrm{Op}_h(b) = \mathrm{Op}_h(ab) + O(h) \text{ and } \mathrm{Op}_h(a)^* = \mathrm{Op}_h(\overline{a}) + O(h).$$

Suppose that X is compact and take an orthonormal basis of eigenfunctions $\{v_j\}$ of $-\Delta_X$ with corresponding eigenvalues $1 = \lambda_0 \leq \lambda_1 \leq \dots$. Put $h_j = \frac{1}{\sqrt{\lambda_j}}$. We can now define the micro local lift of μ_j^{ML} of the measure μ_j defined in (2). We put

$$\int_{T^*X} a(x,\xi)d\mu_j^{\mathrm{ML}}(x,\xi) = \langle \mathrm{Op}_{h_j}(a)v_j, v_j\rangle \text{ for } a \in \mathcal{C}_c^\infty(T^*X). \tag{5}$$

Definition 3.1. A probability measure σ on T^*X is called a quantum limit (also semiclassical measure) if there is a sequence $\{v_{j_k}\}_{k\in\mathbb{N}}$ of L^2-normalized eigenfunctions of $-\Delta_X$ such that

$$\int_{T^*X} a(x,\xi)d\mu_{j_k}^{\mathrm{ML}}(x,\xi) \to \int_{T^*X} a(x,\xi)d\sigma(x,\xi) \text{ for } a \in \mathcal{C}_c^\infty(T^*X)$$

as $k \to \infty$.

One can show that any quantum limit is invariant under the geodesic flow and supported on $M = S^*X$. Let us sketch the invariance under the geodesic flow. To do so we define the operators

$$U_h(t) = \exp(\frac{ith}{2}\Delta_X)\colon L^2(X) \to L^2(X)$$

and let g_t denote the geodesic flow on T^*X. Egorov's theorem now states that

$$U_h(-t)\mathrm{Op}_h(a)U_h(t) = \mathrm{Op}_h(a \circ g_t) + O_t(h).$$

Note that $U_{h_j}(t)v_j = e^{\frac{it\sqrt{\lambda_j}}{2}}v_j$. Given a quantum limit σ and $a \in \mathcal{C}_c^\infty(T^*X)$ we can now simply compute

$$\begin{aligned}\int_{T^*X}[a \circ g_t](x,\xi)d\sigma(x,\xi) &= \lim_{k\to\infty}\langle \mathrm{Op}_{h_{j_k}}(a \circ g_t)v_{j_k}, v_{j_k}\rangle \\ &= \lim_{k\to\infty}\langle \mathrm{Op}_{h_{j_k}}(a)U_{h_{j_k}}(t)v_{j_k}, U_{h_{j_k}}(t)v_{j_k}\rangle \\ &= \lim_{k\to\infty}\langle \mathrm{Op}_{h_{j_k}}(a)v_{j_k}, v_{j_k}\rangle \\ &= \lim_{k\to\infty}\int_{T^*X} a(x,\xi)d\mu_{j_k}^{\mathrm{ML}}(x,\xi) = \int_{T^*X} a(x,\xi)d\sigma(x,\xi).\end{aligned}$$

This establishes the desired invariance.

In this section we are going to establish microlocal lifts and their properties explicitly in the setting of the upper half plane and its quotients. Our construction will be slightly different, since for applications to arithmetic quantum chaos requires that the microlocal lifts are compatible with the Hecke Operators.

Through the rest of this section let Γ be a lattice. We will work with the corresponding quotients $M = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ and $X = \Gamma\backslash\mathbb{H}$.

3.1. **The Lie algebra and differential operators.** We will have to talk a bit about Lie Algebras and the corresponding differential operators. A good reference for this in general is [Kn].

The Lie algebra of $\mathrm{SL}_2(\mathbb{R})$ is given by

$$\mathfrak{sl}_2(\mathbb{R}) = \{m \in \mathrm{Mat}_{2\times 2}(\mathbb{R})\colon \mathrm{tr}(m) = 0\}$$

equipped with $[m_1, m_2] = m_1m_2 - m_2m_1$. When defining the exponential map using the usual series expansion one quickly verifies that

$$\det(\exp(m)) = \exp(\mathrm{tr}(m)) \text{ and } \exp(\mathrm{Ad}_g(m)) = g\exp(m)g^{-1}$$

where $\mathrm{Ad}_g(m) = gmg^{-1}$.

Each $m \in \mathfrak{sl}_2(\mathbb{R})$ gives rise to an differential operator on M. Indeed

$$[m.f](x) = \left[\frac{\partial}{\partial t}f(x\exp(tm))\right]_{t=0},$$

for $f \in \mathcal{C}^\infty(M)$, defines a new smooth function on M. Recall that $M = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ for a lattice Γ. It will be convenient to work with functions in the space

$$\begin{aligned}\mathcal{C}_b^\infty(M) = \{&f \in \mathcal{C}^\infty(M) \cap L^\infty(M)\colon \\ &m_1 \dots m_r.f \in \mathcal{C}^\infty(M) \cap L^\infty(M) \text{ for all } m_1, \dots, m_r \in \mathfrak{sl}_2(\mathbb{R})\}.\end{aligned}$$

Note that if Γ is uniform (i.e. M is compact) then $\mathcal{C}_c^\infty(M) = \mathcal{C}_b^\infty(M) = \mathcal{C}^\infty(M)$ and we are dealing with the usual space of test functions. We identify $m \in \mathfrak{sl}_2(\mathbb{R})$ with the first order differential operator defined above. These are unbounded operators on $L^2(M)$ with domain $\mathcal{C}_b^\infty(M)$. Of course one can compose these operators and one verifies that $m_1 \circ m_2 - m_2 \circ m_1 = [m_1, m_2]$. More precisely

$$m_1.(m_2.f) - m_2.(m_1.f) = ([m_1, m_2]).f \text{ for } f \in \mathcal{C}_b^\infty(M).$$

Furthermore we set

$$\mathrm{ad}_w(m) = [w, m] = \left[\frac{\partial}{\partial t}\mathrm{Ad}_{\exp(tw)}(m)\right]_{t=0}. \tag{6}$$

The universal enveloping algebra $\mathcal{U} = \mathcal{U}(\mathfrak{sl}_2(\mathbb{R}))$ is the (infinite dimensional) associative algebra generated by $m \in \mathfrak{sl}_2(\mathbb{R})$ and a unit element 1. We extend the action Ad_g of $\mathrm{SL}_2(\mathbb{R})$ on $\mathfrak{sl}_2(\mathbb{R})$ to $\mathcal{U}$ by defining

$$\mathrm{Ad}_g(m_1 \circ m_2) = \mathrm{Ad}_g m_1 \circ \mathrm{Ad}_g m_2.$$

This is well defined because

$$\mathrm{Ad}_g[u, v] = [\mathrm{Ad}_g u, \mathrm{Ad}_g v].$$

We want to extend the action ad_w of $\mathfrak{sl}_2(\mathbb{R})$ similarly by

$$\mathrm{ad}_w(m_1 \circ m_2) = [w, m_1 \circ m_2].$$

Recall that on the universal enveloping algebra the bracket operation is defined by the usual commutator bracket, so that

$$\begin{aligned}
[w, m_1 \circ m_2] &= w \circ m_1 \circ m_2 - m_1 \circ m_2 \circ w \\
&= w \circ m_1 \circ m_2 - m_1 \circ w \circ m_2 + m_1 \circ w \circ m_2 - m_1 \circ m_2 \circ w \\
&= [w, m_1] \circ m_2 + m_1 \circ [w, m_2].
\end{aligned}$$

In particular, in order for this extension to be be compatible with (6) we need the product rule

$$\left[\frac{\partial}{\partial t}\mathrm{Ad}_{\exp(tw)}(m_1 \circ m_2)\right]_{t=0} = \left[\frac{\partial}{\partial t}\mathrm{Ad}_{\exp(tw)}(m_1)\right]_{t=0} \circ m_2 + m_1 \circ \left[\frac{\partial}{\partial t}\mathrm{Ad}_{\exp(tw)}(m_2)\right]_{t=0}.$$

This can be verified directly from the definition of the action Lie algebra action on $\mathcal{C}_b^\infty$.

Important coordinates are

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

Note that $\mathcal{W} = \mathcal{U}^- - \mathcal{U}^+$. One easily verifies the identities

$$[\mathcal{H}, \mathcal{U}^\pm] = \pm \mathcal{U}^\pm \text{ and } [\mathcal{U}^+, \mathcal{U}^-] = 2\mathcal{H}.$$

Lemma 3.1 (and Definition)**.** *The element*

$$\Omega_c = \mathcal{H} \circ \mathcal{H} + \frac{1}{2}[\mathcal{U}^+ \circ \mathcal{U}^- + \mathcal{U}^- \circ \mathcal{U}^+]$$

is $SL_2(\mathbb{R})$ invariant. We call Ω_c the Casimir element/operator.

Proof. We start by observing that Ω_c is fixed under the action of $\mathrm{SL}_2(\mathbb{R})$ if and only if

$$\left[\frac{\partial}{\partial t}\mathrm{Ad}_{\exp(tw)}\Omega_c\right]_{t=0} = 0.$$

(This reduction uses the fact that $\mathrm{SL}_2(\mathbb{R})$ is connected and that $\mathrm{Ad}_{\exp((t_0+t)w)} = \mathrm{Ad}_{\exp(t_0 w)}\mathrm{Ad}_{\exp(tw)}$.) We consider several cases using (6) and the commutator relations:

- For $w = \mathcal{H}$ we get

$$\begin{aligned}\left[\frac{\partial}{\partial t}\mathrm{Ad}_{\exp(t\mathcal{H})}\Omega_c\right]_{t=0} &= [\mathcal{H},\mathcal{H}]\circ\mathcal{H} + \mathcal{H}\circ[\mathcal{H},\mathcal{H}] \\ &\quad + \frac{1}{2}\left([\mathcal{H},\mathcal{U}^+]\circ\mathcal{U}^- + \mathcal{U}^+\circ[\mathcal{H},\mathcal{U}^-] + [\mathcal{H},\mathcal{U}^-]\circ\mathcal{U}^+ + \mathcal{U}^-\circ[\mathcal{H},\mathcal{U}^+]\right) = 0.\end{aligned}$$

- For $w = \mathcal{U}^+$ we get

$$\begin{aligned}\left[\frac{\partial}{\partial t}\mathrm{Ad}_{\exp(t\mathcal{U}^+)}\Omega_c\right]_{t=0} &= [\mathcal{U}^+,\mathcal{H}]\circ\mathcal{H} + \mathcal{H}\circ[\mathcal{U}^+,\mathcal{H}] \\ &\quad + \frac{1}{2}\left(0 + \mathcal{U}^+\circ[\mathcal{U}^+,\mathcal{U}^-] + [\mathcal{U}^+,\mathcal{U}^-]\circ\mathcal{U}^+ + 0\right) = 0.\end{aligned}$$

- The computation for $w = \mathcal{U}^-$ is similar.

□

Recall that $X = M/\mathrm{SO}_2$. In particular, we can view a function $f \in \mathcal{C}_b^\infty(X)$ as a function $f\colon M \to \mathbb{C}$ with $f(xk) = f(x)$ when $k \in \mathrm{SO}_2$ and $x \in M$. In particular, since SO_2 is compact we can view $\mathcal{C}_b^\infty(X) \subseteq \mathcal{C}_b^\infty(M)$.

Proposition 3.2. *For $f \in \mathcal{C}_b^\infty(X)$ we have*

$$\Omega_c.f = y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)f = \Delta_X f.$$

The operator Δ_X is the Laplace-Beltrami Operator on X.

Proof. The proof is a cumbersome computation using coordinates and we omit the details. □

Recall that

$$\mathrm{SO}_2 = \left\{k_\theta = \begin{pmatrix}\cos(\theta) & \sin(\theta)\\ -\sin(\theta) & \cos(\theta)\end{pmatrix} : \theta \in [0,2\pi)\right\}.$$

By classical Fourier Analysis we have the orthonormal basis

$$e_n(k_\theta) = e^{in\theta}$$

for $n \in \mathbb{Z}$. Every $f \in \mathcal{C}^\infty(\mathrm{SO}_2)$ has a Fourier expansion

$$f = \sum_{n\in\mathbb{Z}} c_f(n)e_n \text{ where } c_f(n) = \langle f, e_n\rangle_{L^2(\mathrm{SO}_2)}.$$

Lemma 3.3. *We have $|c_f(n)| \leq C_f \cdot n^{-2}$ for some constant $C_f \in \mathbb{R}_{>0}$ depending on f.*[6] *Furthermore,*

$$[f * e_n](k_\psi) = \int_{SO_2} f(k)e_n(k_\psi k^{-1})dk = c_f(n) \cdot e_n(k_\psi).$$

Proof. By integration by parts we have

$$|c_f(n)| = |\langle f, e_n\rangle_{L^2(\mathrm{SO}_2)}| = \frac{1}{n^2}|\langle f'', e_n\rangle_{L^2(\mathrm{SO}_2)}| \leq n^{-2}\|f''\|_{L^1(\mathrm{SO}_2)}.$$

For the second claim we simply compute

$$\int_{\mathrm{SO}_2} f(k)e_n(k_\psi k^{-1})dk = \int_{\mathrm{SO}_2} f(k)\overline{e_n(k)}dke_n(k_\psi) = e_n(k_\psi)\langle f, e_n\rangle_{L^2(\mathrm{SO}_2)}.$$

□

We now lift this construction to $\mathrm{SL}_2(\mathbb{R})$ (actually $\mathrm{PSL}_2(\mathbb{R})$). We define the spaces

$$\mathcal{A}_n(M) = \{f \in \mathcal{C}_b^\infty(M) \colon f(xk) = e_n(k)f(x)\}.$$

Remark 3.4. We note that $\mathcal{C}_b^\infty(X) = \mathcal{A}_0(M)$ by the identification above. Furthermore, $\mathcal{A}_n(M) \neq \{0\}$ if and only if n is even. Indeed since $M = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ we have $f(-x) = f(x)$ when lifted to $\mathrm{SL}_2(\mathbb{R})$. In particular, for $f \in \mathcal{A}_n(M)$ we must have

$$e_n(k_\theta)f(x) = f(xk_\theta) = f(-xk_\theta) = f(xk_{\theta+\pi}) = e_n(k_{\theta+\pi})f(x) = (-1)^n e_n(k_\theta)f(x).$$

Lemma 3.5. *The space $\mathcal{A}_n(M)$ is characterized by $\mathcal{W}.f = in \cdot f$ (i.e. $\mathcal{A}_n(M) = \{f \in \mathcal{C}_b^\infty(M) \colon \mathcal{W}.f = in \cdot f\}$).*

Proof. It is easy to verify that if $f \in \mathcal{A}_n(M)$, then $\mathcal{W}.f = in \cdot f$. Indeed, first compute that $\exp(t\mathcal{W}) = k_t$. Thus we get

$$\begin{aligned}\mathcal{W}.f(x) &= \left[\frac{\partial}{\partial t}f(x\exp(t\mathcal{W}))\right]_{t=0} = \left[\frac{\partial}{\partial t}f(xk_t)\right]_{t=0} \\ &= \left[\frac{\partial}{\partial t}f(x)e_n(k_t)\right]_{t=0} = \left[in \cdot f(x)e^{int}\right]_{t=0} = in \cdot f(x).\end{aligned}$$

On the other hand, if $\mathcal{W}.f = in \cdot f$, then we can compute

$$\frac{\partial}{\partial\theta}[e^{-in\theta}f(xk_\theta)]_{\theta=\psi} = -ine^{-in\psi}f(xk_\psi) + ine^{-in\psi}f(xk_\psi) = 0.$$

In particular $e^{in\theta}f(xk_\theta)$ is constant, so that $f \in \mathcal{A}_n(M)$ as desired. □

---

[6]In analytic number theory jargon this is written as $c_f(n) \ll_f n^{-2}$.

For $f \in \mathcal{C}_b^\infty(M)$ we define

$$f_n(x) = \int_{\mathrm{SO}_2} f(xk)\overline{e_n(k)}dk.$$

Note that $f \mapsto f_n$ is a projection on $\mathcal{A}_n(M)$. Indeed, a simple change of variables yields

$$f_n(xk_\psi) = \int_K f(xk_\psi k)\overline{e_n(k)}dk = e_n(k_\psi)\int_K f(xk)\overline{e_n(k)}dk = e_n(k_\psi)f_n(x).$$

Lemma 3.6. *Let $f \in \mathcal{C}_b^\infty(M)$ and $n \in \mathbb{Z}$. We have $\|f_n\|_\infty \leq C_f \cdot n^{-2}$ for some positive constant C_f depending on f. Furthermore, if $g \in \mathcal{A}_l(M)$ and $h \in \mathcal{A}_m(M)$ with $m \neq l$, then $\langle g, h\rangle_{L^2(M)} = 0$.*

Proof. We first observe that $\mathcal{W}$ is precisely the differential operator in direction SO_2. Replicating the computation above yields:

$$\begin{aligned}\|f_n\|_\infty &= \sup_{x\in M}\left|\int_{\mathrm{SO}_2} f(xk)\overline{e_n(k)}dk\right| \\ &= n^{-2}\cdot\sup_{x\in M}\left|\int_{\mathrm{SO}_2}[(\mathcal{W}\circ\mathcal{W}).f(xk)\overline{e_n(k)}dk\right| \\ &\leq n^{-2}\cdot\|(\mathcal{W}\circ\mathcal{W}).f\|_\infty.\end{aligned}$$

For the second part of the statement we choose $k \in \mathrm{SO}_2$ such that $e_l(k)\overline{e_m(k)} \neq 1$. This is possible since $l \neq m$. Then we have

$$\langle g, h\rangle_{L^2(M)} = \langle R_k g, R_k h\rangle_{L^2(M)} = e_l(k)\overline{e_m(k)}\langle g, h\rangle_{L^2(M)}.$$

□

Definition 3.2. A function $f \in \mathcal{C}_b^\infty(M)$ is called K-finite (or SO_2-finite) if there is $N \in \mathbb{N}$ such that

$$f \in \bigoplus_{n=-N}^{N}\mathcal{A}_n(M).$$

Put

$$\mathfrak{sl}_2(\mathbb{C}) = \mathrm{sl}_2(\mathbb{R}) + i\cdot\mathrm{sl}_2(\mathbb{R}).$$

We can extend the action of $\mathfrak{sl}_2(\mathbb{R})$ on $f \in \mathcal{C}_b^\infty(M)$ to $\mathfrak{sl}_2(\mathbb{C})$ linearly by setting

$$(m_1 + im_2).f = m_1.f + im_2.f.$$

The complex Lie-Bracket is as expected

$$[m, w] = mw - wm = ([m_1, w_1] - [m_2, w_2]) + i([m_1, w_1] + [m_2, w_1])$$
$$\text{where } m = m_1 + im_2 \text{ and } w = w_1 + iw_2.$$

By direct computation one verifies

$$(m\circ w - w\circ m).f = [m, w].f.$$

Definition 3.3. The lowering operator $\mathcal{E}^+ \in \mathfrak{sl}_2(\mathbb{C})$ is given by

$$\mathcal{E}^- = \frac{1}{2}\begin{pmatrix} -1 & i \\ i & 1 \end{pmatrix} = \mathcal{H} + \frac{i}{2}(\mathcal{U}^+ + \mathcal{U}^-).$$

Similarly one defines the raising operator by

$$\mathcal{E}^+ = \frac{1}{2}\begin{pmatrix} -1 & -i \\ -i & 1 \end{pmatrix} = \mathcal{H} - \frac{i}{2}(\mathcal{U}^+ + \mathcal{U}^-).$$

The name of these operators is justified by the following result:

Proposition 3.7. *The elements $\mathcal{E}^\pm$ define operators*

$$\mathcal{E}^+ \colon \mathcal{A}_n(M) \to \mathcal{A}_{n+2}(M) \text{ and } \mathcal{E}^- \colon \mathcal{A}_n(M) \to \mathcal{A}_{n-2}(M).$$

Proof. We start by observing that $\mathcal{W} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ can be diagonalized over $\mathbb{C}$. Indeed

$$m^{-1}\mathcal{W}m = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix} = 2i\mathcal{H} \text{ for } m = \begin{pmatrix} i & -i \\ 1 & 1 \end{pmatrix}.$$

We calculate

$$\mathcal{E}^+ m = \frac{1}{2}\begin{pmatrix} -2i & 0 \\ 2 & 0 \end{pmatrix} = m\mathcal{U}^+$$

and similarly $\mathcal{E}^- m = m\mathcal{U}^-$. This implies

$$[\mathcal{W}, \mathcal{E}^\pm] = 2im[\mathcal{H}, \mathcal{U}^\pm]m^{-1} = \pm 2im\mathcal{U}^\pm m^{-1} = \pm 2i\mathcal{E}^\pm.$$

Now we conclude the proof by a simple computation:

$$\begin{aligned} \mathcal{W}.[\mathcal{E}^\pm.f] &= \mathcal{E}^\pm.[\mathcal{W}.f] + [\mathcal{W}, \mathcal{E}^\pm].f \\ &= ni \cdot \mathcal{E}^\pm.f \pm 2i \cdot \mathcal{E}^\pm.f = (n \pm 2)i \cdot \mathcal{E}^\pm.f \end{aligned}$$

for $f \in \mathcal{A}_n(M)$. □

We record the following useful identities:

$$\begin{aligned} \overline{\mathcal{E}^+} &= \mathcal{E}^-, \\ \mathcal{E}^+ + \mathcal{E}^- &= 2\mathcal{H} \text{ and} \\ \Omega_c &= \mathcal{E}^- \circ \mathcal{E}^+ - \frac{1}{4}\mathcal{W} \circ \mathcal{W} - \frac{i}{2}\mathcal{W}. \end{aligned}$$

Verifying these is an easy exercise.

Lemma 3.8. *If $f \in C_b^\infty(M)$ is K-finite and $m \in \mathfrak{sl}_2(\mathbb{C})$, then $m.f$ is K-finite.*

Proof. This is easily verified for the basis $\mathcal{W}$, $\mathcal{E}^+$ and $\mathcal{E}^-$ of $\mathfrak{sl}_2(\mathbb{C})$. The statement follows by linearity. □

Proposition 3.9. *For $m \in \mathfrak{sl}_2(\mathbb{C})$ and $f_1, f_2 \in C_b^\infty(M)$ we have*

$$\langle m.f_1, f_2\rangle_{L^2(M)} = -\langle f_1, \overline{m}.f_2\rangle_{L^2(M)}.$$

Proof. We first consider functions $f_1, f_2 \in \mathcal{C}_c^\infty(\mathrm{SL}_2(\mathbb{R}))$. Our goal is to check that

$$\langle m.f_1, f_2\rangle_{L^2(\mathrm{SL}_2(\mathbb{R}))} = -\langle f_1, \overline{m}.f_2\rangle_{L^2(\mathrm{SL}_2(\mathbb{R}))},$$

where the L^2-space is defined with respect to the Haar measure $\mu_{\mathrm{SL}_2(\mathbb{R})}$. Note that by linearity it suffices to check this for the basis $\mathcal{W}$, $\mathcal{U}^+$ and $\mathcal{U}^-$. We will show the computation for $\mathcal{U}^-$, since the other cases are very similar. We first recall the well known identity

$$0 = \int_{\mathbb{R}} \frac{d}{dx}(g_1\overline{g_2})(x)dx = \langle g_1', g_2\rangle_{L^2(\mathbb{R})} + \langle g_1, g_2'\rangle_{L^2(\mathbb{R})},$$

for $g_1, g_2 \in \mathcal{C}_c^\infty(\mathbb{R})$. Now we note that

$$\exp(t\mathcal{U}^-) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix} = u_t.$$

Now, using a well known decomposition of the $\mathrm{SL}_2(\mathbb{R})$-Haar measure, we compute

$$\begin{aligned}
\langle \mathcal{U}^- f_1, f_2\rangle_{L^2(\mathrm{SL}_2(\mathbb{R}))} &= \int_{\mathrm{SL}_2(\mathbb{R})} \mathcal{U}^-.f_1(d)\overline{f_2(x)}d\mu_{\mathrm{SL}_2(\mathbb{R})}(x) \\
&= \int_{\mathrm{SO}_2 A}\int_{U^-} \mathcal{U}^-.f_1(kau_x)\overline{f_2(kau_x)}dxd\mu(ka) \\
&= \int_{\mathrm{SO}_2 A}\int_{U^-} \frac{d}{dx}f_1(kau_x)\overline{f_2(kau_x)}dxd\mu(ka) \\
&= -\int_{\mathrm{SO}_2 A}\int_{U^-} f_1(kau_x)\overline{\frac{d}{dx}f_2(kau_x)}dxd\mu(ka) \\
&= \int_{\mathrm{SL}_2(\mathbb{R})} f_1(x)\overline{\mathcal{U}^-.f_2(x)}d\mu_{\mathrm{SL}_2(\mathbb{R})}(x) = -\langle f_1, \mathcal{U}^-.f_2\rangle_{L^2(\mathrm{SL}_2(\mathbb{R}))}.
\end{aligned}$$

We now turn to the case of M, first considering $f_1, f_2 \in \mathcal{C}_c^\infty(M)$ with compact support. Let $C = C(f_1, f_2)$ be a compact subset of M containing the support of f_1 and f_2. We will use the following fact concerning discrete subgroups $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$. Given $x \in M = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ there is $r = r_x > 0$ such that

$$B_r^{\mathrm{PSL}_2(\mathbb{R})} \ni g \mapsto xg \in B_r^M(x). \tag{7}$$

is injective. Therefore we find $x_1, \dots, x_l \in M$ such that $C \subseteq O_1 \cup \dots \cup O_l$ for $O_i = B_r^M(x_i)$. Choose $\chi_j \in \mathcal{C}_c^\infty(M)$ with $\chi_j \geq 0$, $\mathrm{supp}(\chi_j) \subseteq O_j$ and $\sum_{j=1}^l \chi_j(x) = 1$ for $x \in C$. Further, choose $\psi_j \in \mathcal{C}_c^\infty(M)$ with $\mathrm{supp}(\psi_j) \in \mathcal{O}_j$ and $\psi_j(x) = 1$ for $x \in \sup(\chi_j)$. Then, for $m \in \mathrm{sl}_2(\mathbb{C})$ we have

$$\langle m.f_1, f_2\rangle_{L^2(M)} = \sum_{j=1}^l \langle m.f_1, \chi_j f_2\rangle_{L^2(M)} = \sum_{j=1}^l \langle m.\psi_j f_1, \chi_j f_2\rangle_{L^2(M)}.$$

Using the map from (7) we identify

$$\mathcal{C}_c^\infty(O_j) = \mathcal{C}_c^\infty(B_r^{\mathrm{PSL}_2(\mathbb{R})}) \subseteq \mathcal{C}_c^\infty(\mathrm{SL}_2(\mathbb{R})) \subseteq L^2(\mathrm{SL}_2(\mathbb{R})).$$

Since $\psi_j f_1, \chi_j f_2 \in \mathcal{C}_c^\infty(O_j)$ we can thus apply the previous case and obtain

$$\begin{aligned}\langle m.f_1, f_2\rangle_{L^2(M)} &= -\sum_{j=1}^{l}\langle \psi_j f_1, \overline{m}.\chi_j f_2\rangle_{L^2(M)}\\ . &= -\sum_{j=1}^{l}\langle f_1, \overline{m}.\chi_j f_2\rangle_{L^2(M)}\\ &= -\langle f_1, \overline{m}.f_2\rangle_{L^2(M)}.\end{aligned}$$

Finally we treat the case when M is non-compact and $f_1, f_2 \in \mathcal{C}_b^\infty(M)$. Given $\epsilon > 0$ there is $K \subseteq X$ compact such that

$$\|f_2 \mathbb{1}_{M\setminus K}\|_{L^2(M)} < \epsilon \text{ and } \|(m.f_2)\mathbb{1}_{M\setminus K}\|_{L^2(M)} < \epsilon.$$

We choose a smooth function $\phi \in \mathcal{C}_c(M)$ with $\phi(M) \subseteq [0,1]$ and $\phi|_K = 1$. It is possible to make sure that $\|\overline{m}.\phi\|_\infty \leq 1$.[7] Finally take $\psi \in \mathcal{C}_c^\infty(M)$ with $\psi(x) = 1$ for $x \in \operatorname{supp}(\phi)$. We now compute

$$\begin{aligned}\langle m.f_1, f_2\rangle_{L^2(M)} &= \langle m.f_1, \phi\cdot f_2\rangle_{L^2(M)} + O(\|m.f_1\|_2\|(1-\phi)\,\mathrm{f}_2\,\|_2)\\ &= \langle m.\psi f_1, \phi f_2\rangle + O_{m,f_1}(\epsilon)\\ &= -\langle \psi f_1, \overline{m}.\phi f_2\rangle + O_{m,f_1}(\epsilon)\\ &= -\langle f_1, (\overline{m}.\phi)f_2 + \phi(\overline{m}.f_2)\rangle + O_{m,f_1}(\epsilon)\\ &= -\langle f_1, \phi(\overline{m}.f_2)\rangle + O(\|f_1\|_2\|f_2\mathbb{1}_{M\setminus K}\|_2) + O_{m,f_1}(\epsilon)\\ &= -\langle f_1, \overline{m}.f_2\rangle + O(\|f_1\|_2\|\overline{m}.f_2\mathbb{1}_{M\setminus K}\|_2) + O_{m,f_1}(\epsilon)\\ &= -\langle f_1, \overline{m}.f_2\rangle + O_{m,f_1}(\epsilon)\end{aligned}$$

where we have used the statement for compactly supported test functions. Equality follows because we can choose $\epsilon > 0$ as small as desired. □

Remark 3.10. Of course the proposition above tells us that $-\overline{m}$ is the adjoint of the unbounded operator m with domain $\mathcal{C}_b^\infty(M)$ on $L^2(M)$. Note that this notion can sensitive to the domain.

Corollary 3.11. *We have*

$$\Omega_c^* = \Omega_c \textit{ and } (\mathcal{E}^\pm)^* = -\mathcal{E}^\mp.$$

Furthermore, if $f \in \mathcal{A}_n(M)$ satisfies $-\Omega_c.f = (\frac{1}{4}+r^2)f$, then

$$\|\mathcal{E}^\pm.f\|_2 = |ir + \frac{1}{2}(1\pm n)| \cdot \|f\|_2.$$

[7] We leave it as an exercise to construct such a function.

Proof. Since the other statements are easily verified we only compute the L^2-norms. This is done as follows:

$$\begin{aligned}\|\mathcal{E}^+.f\|_2^2 &= \langle \mathcal{E}^+.f, \mathcal{E}^+ f\rangle \\ &= -\langle \mathcal{E}^- \circ \mathcal{E}^+.f, f\rangle \\ &= -\langle (\Omega_c + \frac{1}{4}\mathcal{W}\circ\mathcal{W} + \frac{i}{2}\mathcal{W}).f, f\rangle \\ &= (\frac{1}{4} + r^2 + \frac{1}{4}n^2 + \frac{1}{2}n)\|f\|^2.\end{aligned}$$

An easy computation shows that

$$(\frac{1}{4} + r^2 + \frac{1}{4}n^2 + \frac{1}{2}n) = |ir + \frac{1}{2}(n+1)|^2.$$

The case $\mathcal{E}^-$ is done similarly. □

3.2. **The Microlocal-Lift a la Zelditch-Wolpert.** We now turn towards the construction of the (approximate) microlocal lift. This construction is nicely explained in the lecture notes by Einsiedler and Ward mentioned in the introduction. Relevant are also the works [Wo, Li01, Ze]. Recall that $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$ is a lattice, $M = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ and $X = M/\mathrm{PSO}_2$.

Definition 3.4. Let $\phi \in \mathcal{C}_b^\infty(X)$ be an eigenfunction of $-\Delta_X$ with eigenvalue $\lambda_\phi = \frac{1}{4} + r_\phi^2$, where $r_\phi \in \mathbb{R}_+$. Then we define

$$\phi_0(x) = \phi(xK) \in \mathcal{A}_0(M).$$

We continue inductively and set

$$\phi_{2n+2} = \frac{1}{ir + \frac{1}{2} + n}\mathcal{E}^+.\phi_{2n} \in \mathcal{A}_{2n+2}(M),$$

for $n \geq 0$ and

$$\phi_{2n-2} = \frac{1}{ir + \frac{1}{2} - n}\mathcal{E}^-.\phi_{2n} \in \mathcal{A}_{2n-2}(M),$$

for $n \leq 0$.

We start by recording some simple properties

Lemma 3.12. *For all $n \in \mathbb{Z}$ we have $\|\phi_{2n}\|_2 = 1$,*

$$\phi_{2n\pm 2} = \frac{1}{ir + \frac{1}{2} \pm n}\mathcal{E}^\pm.\phi_{2n},$$

and

$$(\mathcal{E}^- \circ \mathcal{E}^+).\phi_{2n} = (\lambda_\phi - n^2 - n)\phi_{2n}.$$

Proof. The proof is very easy and only uses properties of the operators Ω_c and $\mathcal{E}^\pm$ established in the previous subsection. We leave the details as an exercise. □

Definition 3.5. Let $\phi \in \mathcal{C}_b^\infty(X)$ be an eigenfunction of $-\Delta_X$ with eigenvalue λ_ϕ. For $N = N(\lambda_\phi)$ we set

$$\tilde{\phi}^{(N)} = \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{N} \phi_{2n}.$$

We call $\tilde{\phi}^{(N)}$ approximate (microlocal) lift.[8]

Obviously we have $\|\tilde{\phi}^{(N)}\|_2 = 1$. Indeed this is a consequence of the normalization of the ϕ_{2n}'s and the fact that the spaces $\mathcal{A}_{2n}(M)$ are mutually orthogonal. We will now study some properties of these approximate lifts.

Proposition 3.13. *Let $\phi \in \mathcal{C}_b^\infty(X)$ be an eigenfunction of $-\Delta_X$ with eigenvalue $\lambda_\phi = \frac{1}{4} + r_\phi^2$. For $f \in \mathcal{C}_c^\infty(X)$ we have*

$$\int_M f(x)|\tilde{\phi}^{(N)}(x)|^2 d\mu_M(x) = \langle f\phi, \phi\rangle_{L^2(X)} + O_f(\max(N^{-1}, Nr^{-1})).$$

More generally, if $f \in \mathcal{C}_c^\infty(M)$ is K-finite, then

$$\int_M f(x)|\tilde{\phi}^{(N)}(x)|^2 d\mu_M(x) = \langle f \sum_{n=-N}^{N} \phi_{2n}, \phi\rangle_{L^2(M)} + O_f(\max(N^{-1}, Nr^{-1})).$$

Proof. We first note that $\mathcal{A}_l(M) \cdot \mathcal{A}_m(M) = \mathcal{A}_{l+m}(M)$. Suppose that

$$f \in \sum_{l=-L}^{L} \mathcal{A}_{2l}(M).$$

By definition of $\tilde{\phi}^{(N)}$ we have

$$\langle f\tilde{\phi}^{(N)}, \tilde{\phi}^{(N)}\rangle_{L^2(M)} = \frac{1}{2N+1} \sum_{m,n=-N}^{N} \langle f\phi_{2m}, \phi_{2n}\rangle_{L^2(M)}.$$

The case $Nr_\phi^{-1} \geq 1$ is easily treated by showing that everything is $O(1)$. Indeed we obviously have

$$\int_M f(x)|\tilde{\phi}^{(N)}(x)|^2 d\mu_M(x) = O(\|f\|_\infty) = O_f(1).$$

On the other hand we see that for $n \notin [-L, L]$ the function

$$f\phi_{2n} \in \sum_{l=-L}^{L} \mathcal{A}_{2(l+n)}(M)$$

is orthogonal to $\phi_0 \in \mathcal{A}_0(M)$. We conclude that

$$\langle f \sum_{n=-N}^{N} \phi_{2n}, \phi\rangle = O(\|f\|_\infty L) = O_f(1).$$

[8]Attention: To the best of our knowledge this is not standard terminology.

We can turn towards the more interesting case $Nr_\phi^{-1} < 1$. We first compute

$$\begin{aligned}
\langle f\phi_{2m}, \phi_{2n}\rangle_{L^2(M)} &= \frac{1}{ir - \frac{1}{2} + m}\langle f\mathcal{E}^+.\phi_{2m-2}, \phi_{2n}\rangle \\
&= \frac{1}{ir - \frac{1}{2} + m}\left[\langle \mathcal{E}^+.(f\phi_{2m-2}), \phi_{2n}\rangle_{L^2(M)} - \langle (\mathcal{E}^+.f)\phi_{2m-2}, \phi_{2n}\rangle_{L^2(M)}\right] \\
&= \frac{-1}{ir - \frac{1}{2} + m}\langle f\phi_{2m-2}, \mathcal{E}^-.\phi_{2n}\rangle_{L^2(M)} + O_f(r^{-1}) \\
&= \frac{ir - \frac{1}{2} + n}{ir - \frac{1}{2} + m}\langle f\phi_{2m-2}, \phi_{2n-2}\rangle_{L^2(M)} + O_f(r^{-1}) \\
&= \langle f\phi_{2m-2}, \phi_{2n-2}\rangle_{L^2(M)} + O_f((|n-m|+1)r^{-1}).
\end{aligned}$$

Since $f \in \sum_{l=-L}^{L} \mathcal{A}_{2l}(M)$ we see that for $|n-m| > 2L$ we have

$$\langle f\phi_{2m}, \phi_{2n}\rangle_{L^2(M)} = 0 = \langle f\phi_{2m-2}, \phi_{2n-2}\rangle_{L^2(M)}.$$

Thus we find that

$$\langle f\phi_{2m}, \phi_{2n}\rangle_{L^2(M)} = \langle f\phi_{2m-2}, \phi_{2n-2}\rangle_{L^2(M)} + O_f(r^{-1}).$$

Suppose $n > 0$, then iterating this process yields

$$\langle f\phi_{2m}, \phi_{2n}\rangle_{L^2(M)} = \langle f\phi_{2(m-n)}, \phi_0\rangle_{L^2(M)} + O_f(Nr^{-1}).$$

The case $n < 0$ is treated similarly. Smming this over $m, n \in [-N, N]$ with $|n-m| \le 2L$ and dividing by $2N+1$ yields

$$\langle f\tilde{\phi}^{(N)}, \tilde{\phi}^{(N)}\rangle_{L^2(M)} = \frac{1}{2N+1}\sum_{m,n=-N}^{N}\langle f\phi_{2(m-n)}, \phi_0\rangle_{L^2(M)} + O_f(Nr^{-1}).$$

Note that $2N+1-|l|$ is the number of ways in which l can be written as $m-n$ with $m, n \in [-N, N]$. We get

$$\begin{aligned}
\langle f\tilde{\phi}^{(N)}, \tilde{\phi}^{(N)}\rangle_{L^2(M)} &= \sum_{l=-L}^{L}\frac{2N+1-|l|}{2N+1}\langle f\phi_{2l}, \phi_0\rangle_{L^2(M)} + O_f(Nr^{-1}) \\
&= \sum_{l=-L}^{L}\langle f\phi_{2l}, \phi_0\rangle_{L^2(M)} + O_f(N^{-1} + Nr^{-1}) \\
&= \langle f\sum_{l=-L}^{L}\phi_{2l}, \phi_0\rangle_{L^2(M)} + O_f(N^{-1} + Nr^{-1}).
\end{aligned}$$

Here we have used that $\frac{2N+1-|l|}{2N+1} = 1 + O_f(N^{-1})$ and $\langle f\phi_{2l}, \phi_0\rangle_{L^2(M)} = O_f(1)$. Finally, using orthogonality of the spaces $\mathcal{A}_*(M)$ one more time allows us to extend the sum from $l \in [-L, L]$ to $[-N, N]$ without introducing any new error. (The case $N < L$ can be easily handled separately.) $\square$

Proposition 3.14 (Zelditch). *Let $\phi \in \mathcal{C}_b^\infty(X)$ be a $-\Delta_X$ eigenfunction with corresponding eigenvalue $\lambda_\phi = \frac{1}{4} + r_\phi^2$. Further, let $f \in \mathcal{C}_c^\infty(M)$ be K-finite and suppose that N is sufficiently large (in terms of f). Then there is some fixed degree-two operator $\mathcal{V}$ such that*

$$\left\langle [(r_\phi \mathcal{H} + \mathcal{V}).f] \sum_{n=-N}^{N} \phi_{2n}, \phi_0 \right\rangle = 0.$$

In particular,

$$\left\langle (\mathcal{H}.f) \sum_{n=-N}^{N} \phi_{2n}, \phi_0 \right\rangle = O_f(\sqrt{N} r^{-1}).$$

Proof. The two identities

$$\mathcal{E}^- \circ \mathcal{E}^+ \phi_0 = \Omega_c.\phi_0 = -\lambda_\phi \phi_0 \text{ and } \Omega_c. \sum_{n=-N}^{N} \phi_{2n} = -\lambda_\phi \sum_{n=-N}^{N} \phi_{2n}$$

will be crucial for the proof. For convenience we will abbreviate

$$\psi = \sum_{n=-N}^{N} \phi_{2n}.$$

Let $f_1, f_2 \in \mathcal{C}_c^\infty(M)$. We start by observing that

$$0 = \langle f_1 f_2, \mathcal{W}.\phi_0 \rangle = \langle \mathcal{W}.(f_1 f_2), \phi_0 \rangle = \langle (\mathcal{W}.f_1) f_2, \phi_0 \rangle + \langle f_1 (\mathcal{W}.f_2), \phi_0 \rangle$$

can be rewritten as

$$\langle (\mathcal{W}.f_1) f_2, \phi_0 \rangle = -\langle f_1 (\mathcal{W}.f_2), \phi_0 \rangle. \tag{8}$$

Another observation is that

$$\mathcal{E}^+ \psi = (i r_\phi - \frac{1}{2}\mathcal{W} - \frac{1}{2}). \underbrace{\sum_{n=-N+1}^{N+1} \phi_{2n+2}}_{=\psi^+}$$

and similarly

$$\mathcal{E}^- \psi = (i r_\phi + \frac{1}{2}\mathcal{W} - \frac{1}{2}). \underbrace{\sum_{n=-N-1}^{N-1} \phi_{2n-2}}_{\psi^-}.$$

Taking a leap of faith we start computing

$$
\begin{aligned}
-\lambda_\phi\langle f\psi,\phi_0\rangle &= \langle f\psi, \mathcal{E}^-\circ\mathcal{E}^+.\phi_0\rangle \\
&= \langle(\mathcal{E}^-\circ\mathcal{E}^+.f)\psi + (\mathcal{E}^+.f)(\mathcal{E}^-.\psi) + (\mathcal{E}^-.f)(\mathcal{E}^+.\psi) + f(\mathcal{E}^-\circ\mathcal{E}^+.\psi),\phi_0\rangle \\
&= \langle(\mathcal{E}^-\circ\mathcal{E}^+.f)\psi + (\mathcal{E}^+.f)((ir_\phi+\frac{1}{2}\mathcal{W}-\frac{1}{2}).\psi^-) + (\mathcal{E}^-.f)((ir_\phi-\frac{1}{2}\mathcal{W}-\frac{1}{2}).\psi^+) \\
&\quad + f((\Omega_c+\frac{1}{2}\mathcal{W}\circ\mathcal{W}+\frac{i}{2}\mathcal{W}).\psi),\phi_0\rangle
\end{aligned}
$$

We observe that, since $\Omega_c\psi = -\lambda_\phi\psi$, each side of the equation features the term $\lambda_\phi\langle f\psi,\phi_0\rangle$. We cancel these terms. Next we note that once N is sufficiently large ($N \geq 2L$ should suffice) we can replace ψ^+ and ψ^- by ψ, since the additional terms are orthogonal to ϕ_0. Now we recall that $\mathcal{E}^+ + \mathcal{E}^- = 2\mathcal{H}$, so that the terms featuring r_ϕ can be combined to

$$
\langle(\mathcal{E}^+.f)ir_\phi\psi + (\mathcal{E}^-f)ir_\phi\psi,\phi_0\rangle = 2ir_\phi\langle(\mathcal{H}.f)\psi,\phi_0\rangle.
$$

Thus we get

$$
\begin{aligned}
0 = 2ir_\phi\langle(\mathcal{H}.f)\psi,\phi_0\rangle + \langle(\mathcal{E}^-\circ\mathcal{E}^+.f)\psi + (\mathcal{E}^+.f)((\frac{1}{2}\mathcal{W}-\frac{1}{2}).\psi) \\
+ (\mathcal{E}^-.f)((-\frac{1}{2}\mathcal{W}-\frac{1}{2}).\psi) + f((\frac{1}{2}\mathcal{W}\circ\mathcal{W}+\frac{i}{2}\mathcal{W}).\psi),\phi_0\rangle.
\end{aligned}
$$

Using (8) we can write the remainder as $2i\langle(\mathcal{V}.f)\psi,\phi_0\rangle$ as desired. (Note that $\mathcal{V}$ does not depend on ϕ, λ_ϕ or r_ϕ. Indeed it was somehow a miracle that the corresponding terms combined and canceled just as needed.)

The second formula simply follows by observing that $\|\psi\|_2 = \sqrt{2N+1}$ and $\|\mathcal{V}f\|_\infty = O_f(1)$. □

Finally we will need the following technical lemma:

Lemma 3.15. *Let $f \in \mathcal{C}_c^\infty(M)$. The K-finite approximations*

$$
f_{[-L,L]} = f * \sum_{l=-L}^{L} e_{2l} \in \sum_{l=-L}^{L}\mathcal{A}_{2l}(M)
$$

converge uniformly to f as $L\to\infty$. Furthermore, $\mathcal{H}.f_{[-L,L]}$ converges uniformly to $\mathcal{H}.f$ as $L\to\infty$.

Proof. The first claim easily follows from our previous observation $f_n(x) = [f * e_n](x)$ satisfies $\|f_n\| = O_f(n^{-2})$. We turn towards the second claim. As in the proof of Proposition 3.9 it is easy to reduce to the case when $f\in\mathcal{C}_c^\infty(\mathrm{SL}_2(\mathbb{R}))$. We will need to estimate

$$\begin{aligned}[\mathcal{H}.f_n](x) &= \left[\frac{\partial}{\partial t}\int_{\mathrm{SO}_2} f_n(x\exp(t\mathcal{H})k)e_n(k^{-1})dk\right]_{t=0} \\ &= \left[\frac{\partial}{\partial t}\int_{\mathrm{SO}_2} f_n(xk\exp(t\mathrm{Ad}_k^{-1}(\mathcal{H}))k)e_n(k^{-1})dk\right]_{t=0} \\ &= \int_{\mathrm{SO}_2}[\mathrm{Ad}_k^{-1}(\mathcal{H}).f](xk)e(k^{-1})dk.\end{aligned}$$

We now observe that $\mathrm{Ad}_k^{-1}(\mathcal{H}).f$ is a smooth function of x and ϕ.[9] The upshot is that we can argue as above to find

$$\|\mathcal{H}.f\|_\infty = O_f(n^{-2}).$$

We conclude that $\mathcal{H}.f_{[-L,L]}$ converges to some function $h \in \mathcal{C}_c^\infty(\mathrm{SL}_2(\mathbb{R}))$.

Given $g \notin \mathrm{supp}(f)\mathrm{SO}_2$ then

$$f_{[-L,L]}(ga_T) = \int_0^T [\mathcal{H}.f_{[-L,L]}](ga_t)dt \to \int_0^T h(ga_t)dt$$

as $L \to \infty$. Thus

$$f(ga_T) = \int_0^T h(ga_t)dt$$

and differentiating (w.r.t $\mathcal{H}$) once again shows that $\mathcal{H}.f = h$ as desired. □

3.3. **Summary.** Let $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$ be a lattice and let $M = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ and $X = M/\mathrm{PSO}_2$ as usual. Suppose that there is a sequence $\phi_i \in \mathcal{C}_b^\infty(X)$ of $-\Delta_X$ eigenfunctions with corresponding eigenvalues $\lambda_{\phi_i} = \frac{1}{4} + r_{\phi_i}^2 \to \infty$.[10]

We define the measures $\mu_i = |\phi_i(\cdot)|^2\mu_X$ as in (2). After passing to a sub-sequence if necessary we assume that σ is the weak-*-limit of the sequence μ_i.

We further choose a sequence $N_i = N_i(\lambda_{\phi_i})$ such that $N_i r_{\phi_i}^{-1} \to 0$ and $N_i^{-1} \to 0$. Then we define the measures $\tilde{\mu}_i = |\tilde{\phi_i}^{(N_i)}(\cdot)|^2\mu_M$ by

$$\int_M f(x)d\tilde{\mu}_i = \int_M f(x)|\tilde{\phi_i}^{(N_i)}(x)|d\mu_M(x),$$

for $f \in \mathcal{C}_c^\infty(M)$.

Definition 3.6. A weak-*-limit point of $\{\tilde{\mu}_i\}_i$ is called a micro local lift of σ or a quantum limit of (ϕ_i).

[9] It is a nice exercise to write $\mathrm{Ad}_k^{-1}(\mathcal{H}).f$ as a linear combination, with coefficients depending on θ of $[\mathcal{H}.f](x)$ and $[\mathcal{U}^\pm.f](x)$.

[10] If Γ is uniform, so that X is compact this is an easy consequence of the spectral theorem for $-\Delta_X$ acting on $L^2(X)$ with domain $\mathcal{C}_c^\infty(X)$. On the other hand, if X is non-compact, such a sequence does not always exist. However, in all the cases that are of interest to us (for example $\mathrm{PSL}_2(\mathbb{Z})$) the existence of such a sequence can be established as consequence of Selberg's trace formula.

In this context we have the following important theorem:

Theorem 3.16. *Let σ^{ML} be a quantum limit of (ϕ_i) (and a micro local lift of σ). Then σ^{ML} has the following properties:*

(1) *The measure σ^{ML} projects to σ. (i.e. $\int_X f(x)d\sigma(x) = \int_M f(\pi_{M\to X}(x))d\sigma^{ML}(x)$ for $f \in \mathcal{C}_c^\infty(X)$.)*
(2) *The measure σ^{ML} is invariant under the geodesic flow.*

Proof. The first property is a direct consequence of Proposition 3.13. The second property is established as follows. Without loss of generality we can pass to a sub sequence if necessary and assume that $\tilde{\mu}_i$ converges weak-* to σ^{ML}. Let $f \in \mathcal{C}_c^\infty(M)$ be K-finite. Thus $\mathcal{H}.f$ and $\mathcal{V}.f$ are also K-finite. We put

$$\psi_i = \sum_{n=-N_i}^{N_i} \phi_{i,2n}.$$

First, applying Proposition 3.13 yields

$$\int_M [\mathcal{H}.f](x)|\tilde{\phi}_i^{(N_i)}(x)|^2 d\mu_M(x) = \langle (\mathcal{H}.f)\psi_i, \phi_{i,0}\rangle + O_f(N_i^{-1} + N_i r_{\phi_i}^{-1}).$$

At this point we can apply and Proposition 3.14 to see that

$$\int_M [\mathcal{H}.f](x)|\tilde{\phi}_i^{(N_i)}(x) d\mu_M(x) = O_f(N_i^{-1} + N_i r_{\phi_i}^{-1}).$$

We directly obtain

$$\int_M [\mathcal{H}.f](x) d\sigma^{\mathrm{ML}}(x) = 0$$

after passing to the limit.

Now we drop the assumption that f is K-finite and consider general $f \in \mathcal{C}_c^\infty(M)$. It is an easy application of Lemma 3.15 to deduce that we also have

$$\int_M [\mathcal{H}.f](x) d\sigma^{\mathrm{ML}}(x) = 0.$$

For $T \in \mathbb{R}$ this leads to

$$f(xa_T) - f(x) = \int_0^T [\mathcal{H}.f](xa_t)dt.$$

Integrating over M yields

$$\int_M f(xa_T)d\sigma^{\mathrm{ML}}(x) = \int_M f(x)d\sigma^{\mathrm{ML}}(x) + \int_0^T \int_M [\mathcal{H}.f](xa_t)d\sigma^{\mathrm{ML}}(x)dt = \int_M f(x)d\sigma^{\mathrm{ML}}(x).$$

This is precisely the desired statement. $\square$

4. Quantum Ergodicity

Throughout this section let $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$ be a uniform lattice. In particular $M = \Gamma\backslash\mathrm{PSL}_2(\mathbb{R})$ and $X = M\backslash\mathrm{PSO}_2$ are compact. Thus we have the following:

Theorem 4.1 (Spectral Theorem)**.** *Let $X = \Gamma\backslash PSL_2(\mathbb{R})/PSO_2$ be compact. Then $-\Delta_X = -y(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$ has discrete spectrum*

$$\sigma(-\Delta_X) = \{0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots \to \infty\}. \tag{9}$$

The corresponding eigenfunctions $\{\phi_j\}_{j\in\mathbb{N}\cup\{0\}}$ form an orthonormal basis for $L^2(X)$.

More can be said about the density of the eigenvalues. It will be convenient to write $\lambda_j = \frac{1}{4} + r_j^2$ as usual. Then, in our current setting, the Weyl law reads

$$\sharp\{j \in \mathbb{N} \colon r_j \leq R\} = C_X \cdot R^2 + O_X(R).$$

However, similar statements hold in much greater generality.[11]

We are now ready to state the *phase space version* of Theorem 1.5 and sketch a proof.

Theorem 4.2 (Quantum Ergodicity)**.** *Let M and X be as above and let $\{\phi_j\}_{j\in\mathbb{N}}$ be a sequence of $-\Delta_X$ eigenfunctions with eigenvalues $\lambda_j = \frac{1}{4} + r_j^2$ tending to infinity. Then the uniform measure μ_M is a quantum limit of $\{\phi_j\}$ in the sense of Definition 3.6. Even more, there is a density one subsequence $\{\phi_{j_r}\}_{r\in\mathbb{N}}$ such that*

$$\int_M f(x)d\tilde{\mu}_{j_r}(x) \to \int_M f(x)d\mu_M(x) \text{ for } f \in \mathcal{C}_c^\infty$$

as $r \to \infty$.

The proof will require the following ingredients, which we take for granted:

- We first suppose that there is a suitable pseudodifferential calculus
$$\mathcal{C}_c^\infty(M) \ni a \mapsto \mathrm{Op}(a) \in \mathcal{B}(L^2(X)).$$
In the case of hyperbolic surfaces a very convenient calculus of this kind was developed by Zelditch. For obvious reasons we write $\mathrm{Op}^Z(a)$ for this specific calculus.
- This pseudo differential calculus allows us to define the distributions
$$\int_M a(x)d\mu_j^{\mathrm{ML}}(x) = \mu_j^{\mathrm{ML}}(a) = \langle \mathrm{Op}^Z(a)\phi_j, \phi_j\rangle_{L^2(X)}$$
in analogy to (5). We call μ_j^{ML} distributions, because they are not positive in general. (This is a shortcoming of the Op^Z-calculus.) In particular we

[11] In the case of (compact) quotients of the upper half plane the Weyl law can be established using the Selberg Trace Formula. We refer to the nice survey *Weyl's law in the theory of automorphic forms* by W. Müller for more information.

note that the the microlocal lifts μ_j^{ML} depend on the quantization procedure. The corresponding quantum limits will however be independent of this choice.

- Due to results by S. Zelditch and S. Wolpert we can replace the distributions μ_j^{ML} by the approximate microlocal lifts $\tilde{\mu}_j^{\mathrm{ML}}$ defined above. More precisely we have
$$d\mu_j^{\mathrm{ML}}(x) = \int_M a(x)d\tilde{\mu}_j^{\mathrm{ML}}(x) + O(r_j^{-\frac{1}{4}}).$$
The upshot is that $\tilde{\mu}_j^{\mathrm{ML}}$ is measure and we have seen its explicit construction in the previous section.
- Finally, we will use the following local Weyl law:
$$\lim_{R\to\infty} \frac{1}{\sharp\{r_j \in [R,2R)\}} \sum_{R\leq r_j\leq 2R} \langle \mathrm{Op}^Z(a)\phi_j, \phi_j\rangle = \int_M a(d)d\mu_M(x). \tag{10}$$
This can be proven using the trace formula. We refer for example to [Ze, Lemma 4.1].

Proof. Our first job is to prove the technical estimate
$$\frac{1}{R^2}\sum_{R\leq r_j<2R}\left|\int_M f(x)d\tilde{\mu}_j(x)\right|^2 \leq C\|f\|^2_{L^2(M)} + o_f(1), \tag{11}$$
for some positive constant C as $R\to\infty$. This will be done by freely using the ingredients mentioned above as well as the Weyl law. We first note that
$$\begin{aligned}\left|\int_M f(x)d\tilde{\mu}_j(x)\right|^2 = |\mu_j^{\mathrm{ML}}(f)|^2 + O_f(r_j^{-\frac{1}{4}}) = |\langle \mathrm{Op}^Z(f)\phi_j,\phi_j\rangle|^2 + O_f(r_j^{-\frac{1}{4}})\\ \leq \|\mathrm{Op}^Z(f)\phi_j\|^2_{L^2(X)} + O_f(r_j^{-\frac{1}{4}}).\end{aligned}$$
According to the usual rules of symbol calculus we obtain
$$\begin{aligned}\|\mathrm{Op}^Z(f)\phi_j\|^2_{L^2(X)} &= \langle \mathrm{Op}^Z(f)^*\mathrm{Op}^Z(f)\phi_j,\phi_j\rangle_{L^2(X)}\\ &= \langle \mathrm{Op}^Z(\overline{f})\phi_j\mathrm{Op}^Z(f)\phi_j,\phi_j\rangle_{L^2(X)} + O_f(r_j^{-1})\\ &= \langle \mathrm{Op}^Z(|f|^2)\phi_j,\phi_j\rangle_{L^2(X)} + O_F(r_j^{-1}).\end{aligned}$$
Thus we have obtained
$$\frac{1}{R^2}\sum_{R\leq r_j<2R}\left|\int_M f(x)d\tilde{\mu}_j(x)\right|^2 \leq \frac{1}{R^2}\sum_{R\leq r_j<2R}\langle \mathrm{Op}^Z(|f|^2)\phi_j,\phi_j\rangle_{L^2(X)} + O_f(R^{-\frac{1}{4}}). \tag{12}$$
The result is a direct consequence of the local Weyl law (10).

We are now ready to prove the following averaged version of quantum ergodicity:

$$\frac{1}{R^2}\sum_{R\le r_j<2R}\left|\int_M f(x)d\tilde{\mu}_j(x)-\int_M f(x)d\mu_M(x)\right|^2\to 0 \text{ as } R\to\infty. \tag{13}$$

To shorten notation let us write $\tilde{\mu}_j(f)=\int_M f(x)d\tilde{\mu}_j(x)$. Similarly we put $\mu_M(f)=\int_M f(x)d\mu_M(x)$ and recall that previously we have used the short hand $\mu_M(f)=f_{\mathrm{av}}$. As in the proof of Theorem 3.16 one deduces from Proposition 3.14 that

$$\tilde{\mu}_j(f)=\tilde{\mu}_j(R_{a_t^{-1}}f)+O_t(r_j^{-\frac{1}{2}}).$$

In particular, after recalling that $[\mathrm{Av}_T f(x)=\frac{1}{T}\int_0^T f(xa_t^{-1})dt$, we obtain

$$\tilde{\mu}_j(f)=\tilde{\mu}_j(\mathrm{Av}_T f)+O_T(r_j^{-\frac{1}{2}}).$$

Since the geodesic flow on M is ergodic, see Proposition 2.7, we can apply Corollary 1.4 to see that

$$\|\mathrm{Av}_T f-f_{\mathrm{av}}\|^2_{L^2(M)}\to 0 \text{ as } T\to\infty. \tag{14}$$

With this and (11) at hand we compute

$$\begin{aligned}\frac{1}{R^2}\sum_{R\le r_j\le 2R}|\tilde{\mu}_j(f)-f_{\mathrm{av}}|^2&=\frac{1}{R^2}\sum_{R\le r_j\le 2R}|\tilde{\mu}_j(\mathrm{Av}_T f)-f_{\mathrm{av}}|^2+O_{f,T}(R^{-\frac{1}{2}})\\&\le C\|\mathrm{Av}_T f-f_{\mathrm{av}}\|^2_{L^2(M)}+o_{f,T}(1).\end{aligned}$$

Passing to the R-limit we get

$$\begin{aligned}0&\le\liminf_{R\to\infty}\frac{1}{R^2}\sum_{R\le r_j\le 2R}|\tilde{\mu}_j(f)-f_{\mathrm{av}}|^2\\&\le\limsup_{R\to\infty}\frac{1}{R^2}\sum_{R\le r_j\le 2R}|\tilde{\mu}_j(f)-f_{\mathrm{av}}|^2\le C\|\mathrm{Av}_T f-f_{\mathrm{av}}\|^2_{L^2(M)}.\end{aligned}$$

In view of (14) we can make the right hand side of the previous inequality as small as desired and (13) follows directly.

Deducing the statement of the theorem from (13) is rather standard. First, one applies Chebyshev's inequality to deduce that there is $\epsilon(R)\to 0$ as $R\to\infty$ such that $|\tilde{\mu}_j(f)-\mu_M(f)|\le\epsilon(R)$ for all $r_j\in[R,2R)$ outside some exceptional set $\mathfrak{E}_R$ of cardinality at most $\epsilon(R)\cdot R^2$. This allows us to select a density one subsequence with

$$\tilde{\mu}_{j_r}(f)\to\mu_M(f) \text{ as } r\to\infty.$$

To pass from one specific observable f to all possible observables one uses a standard diagonal argument. $\square$

Remark 4.3. This is only a rough sketch of one possible proof of Quantum Ergodicity. For more details we refer to the literature. Here we have focused on the case of compact hyperbolic surfaces, but the result holds much more generally and has many variations.

We will now briefly discuss some aspects of the Operators $\mathrm{Op}^Z(a)$ mentioned above. These were defined by Zelditch using Helgason's Fourier transform. Let us briefly describe this. First we recall that the upper half plane $\mathbb{H}$ is isomorphic to the Poincaré disc $\mathbb{D}$. The boundary $\partial\mathbb{D}$ of $\mathbb{D}$ is just the circle. Under the isomorphism $\mathbb{H} \cong \mathbb{D}$ the action of $\mathrm{PSL}_2(\mathbb{R})$ turns into an action of $\mathrm{PSU}(1,1)$.[12] Given a tuple $(z,b) \in \mathbb{D} \times \partial\mathbb{D}$ we can find a unique horocycle $h(b,z)$ passing through z with (forward) endpoint b. The (signed) hyperbolic distance between $0 \in \mathbb{D}$ and $h(z,b)$ will be denoted by $\langle z,b\rangle$. As in the case of the upper half plane we identify $\mathbb{D} \times S^1 \cong T^1\mathbb{D} \cong \mathrm{PSU}(1,1)$.

Theorem 4.4 (Helgason)**.** *Let $\phi \in \mathcal{C}^\infty(\mathbb{D})$ be an eigenfunction of $-\Delta_{\mathbb{D}}$ with eigenvalue $\frac{1}{4}+r^2$ of (at most) polynomial growth. Then there exists a distribution $T_\phi \in \mathcal{D}(\partial\mathbb{D})$ such that*

$$\phi(z) = \int_{\partial\mathbb{D}} e^{(\frac{1}{2}+ir)\cdot\langle z,b\rangle} dT_\phi(b).$$

With this at hand we can define the pseudodifferential operators

$$[\mathrm{Op}^Z(a)\phi](z) = \int_{\partial\mathbb{D}} a(z,b) e^{(\frac{1}{2}+ir)\cdot\langle z,b\rangle} dT_\phi(b).$$

Here we need to assume that $\phi \in \mathcal{C}^\infty(\Gamma\backslash\mathbb{D})$ is an eigenfunction of $-\Delta_{\mathbb{D}}$ and $\Gamma \subseteq \mathrm{PSU}(1,1)$ is a uniform lattice. Further, we assume the symbol $a \in \mathcal{C}^\infty(\Gamma\backslash\mathrm{PSU}(1,1))$ to be Γ-invariant. Note that we use the identification $\mathrm{PSU}(1,1) \cong \mathbb{D}\times S^1 = \mathbb{D}\times\partial\mathbb{D}$ to evaluate $a(z,b)$. It can be verified that the calculus $a \mapsto \mathrm{Op}^Z(a)$ satisfies the desired properties.

As mentioned above we obtain distributions

$$a \mapsto \langle \mathrm{Op}^Z(a)\phi, \phi\rangle,$$

which yield microlocal lifts of eigenfunctions ϕ. Since these are not measures Zelditch used Friedrichs symmetrisation to approximate these by measures. This construction was revisited by S. Wolpert and later also simplified by E. Lindenstrauss. These developments ultimately lead to the construction described in the previous section.

We now pass back to the familiar setting of $\mathbb{H}$ and $\mathrm{PSL}_2(\mathbb{R})$, where at the moment we are considering a uniform lattice $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$. Let $\phi \in \mathcal{C}_b^\infty(X)$ be an eigenfunction of $-\Delta_X$ with corresponding eigenvalue $\frac{1}{4}+r^2$, for $r \in \mathbb{R}_+$. As before lift ϕ to M via

$$\phi_0(x) = \phi(xK) \in \mathcal{A}_0(M) \subseteq L^2(M).$$

[12] Of course the two groups are isomorphic: $\mathrm{PSL}_2(\mathbb{R}) \cong \mathrm{PSU}(1,1)$.

Let

$$V_\phi = \overline{\langle R_g\phi_0 \colon g \in \mathrm{SL}_2(\mathbb{R})\rangle}.$$

Thus V_ϕ is a closed subspace of $L^2(M)$, which is invariant under the action of $\mathrm{SL}_2(R)$ via R. We now suppose that V_ϕ is irreducible. In this case we are able to construct an explicit model for V_ϕ as follows. Let V_{ir} be the closure of

$$\left\{ f \in \mathcal{C}^\infty(\mathrm{SL}_2(\mathbb{R})) \colon f\left(\begin{pmatrix} y & x \\ 0 & y^{-1}\end{pmatrix} g\right) = |y|^{1+2ir} f(g) \text{ for } x, y \in \mathbb{R} \text{ and } g \in \mathrm{SL}_2(\mathbb{R})\right\}.$$

The inner product is given by

$$\langle f_1, f_2\rangle = \int_{\mathrm{PSO}_2} f_1(k)\overline{f_2(k)}dk$$

and $\mathrm{SL}_2(\mathbb{R})$ acts by right multiplication $[R_g f](x) = f(xg)$ as usual. In view of the Iwasawa decomposition any $f \in V_{ir}$ is uniquely determined by its restriction to SO_2. In particular it contains the functions f_k defined by

$$f_n\left(\begin{pmatrix} y & x \\ 0 & y^{-1}\end{pmatrix} k\right) = |y|^{1+2ir} e_n(k)$$

for $n \in 2\mathbb{Z}$. It is easy to verify that

$$V_{ir} = \overline{\bigoplus_{n\in 2\mathbb{Z}} \mathbb{C}\cdot f_n}.$$

It turns out that (V_ϕ, R) and (V_{ir}, R) are isomorphic. Let us call the isomorphism ι. We can normalise everything so that $\iota(\phi_0) = f_0$. Let us denote $\tilde{f}^{(N)} = \iota(\tilde{\phi}^{(N)})$. Of course also $\tilde{f}^{(N)}$ is determined by its restriction to SO_2 and we compute

$$\begin{aligned}
\tilde{f}^{(N)}(k_\theta) &= \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{N} e_{2n}(k_\theta) \\
&= \frac{1}{\sqrt{2N+1}} \sum_{n=-N}^{N} e^{2Ni\theta} \\
&= \frac{1}{\sqrt{2N+1}} e^{-2Ni\theta} \sum_{n=0}^{2N} e^{2Ni\theta} \\
&= \frac{1}{\sqrt{2N+1}} e^{-2Ni\theta} \cdot \frac{1-e^{(4N+1)i\theta}}{1-e^{2i\theta}} \\
&= \frac{1}{\sqrt{2N+1}} \frac{\sin((2N+1)\theta)}{\sin(2\theta)}.
\end{aligned}$$

In particular we observe that if we take $N \to \infty$ then $\tilde{f}^{(N)}|_{\mathrm{SO}_2}$ quickly approximates the delta distribution δ_{I_2} on SO_2. This can be made formal and one sees that

$$e_{ir} = \lim_{N\to\infty} \tilde{f}^{(N)} = \sum_{n=-\infty}^{\infty} f_{2n}$$

and $e_{ir}|_{\mathrm{SO}_2} = \delta_{I_2}$.

Turning back to the operator calculus Op^Z. The upshot of the definition is that given an Γ-invariant symbol a the operator $\mathrm{Op}^Z(a)$ maps V_ϕ to V_ϕ. It turns out that[13]

$$\mathrm{Op}^Z(a)\phi(g) \approx \int_{\mathrm{PSO}_2} a(gk)[\iota^{-1}e_{ir}](gk)dk.$$

This concludes our discussion of microlocal lifts and (classical) Quantum Ergodicity. Let us stress that a key feature of the approximate construction presented in the previous section is that everything remains in the (irreducible) representation V_ϕ generated by our eigenfunction ϕ.

5. Arithmetic surfaces and Hecke operators

We start with a little intermezzo, which will turn out to be useful later on. Fix a prime p (for example $p = 2$) and consider the infinite $(p+1)$ regular tree $\mathcal{T}_{p+1}$. We choose some root e. Given two vertices $x, y \in \mathcal{T}_{p+1}$ we define $d_p(x,y) = p^k$ if the shortest path connecting x and y consists of k edges. With this notation at hand we can define the operators

$$[\widetilde{T}_{p^k} f](x) = \sum_{\substack{y\in\mathcal{T}_{p+1},\\ d_p(x,y)=p^k}} f(y)$$

acting on functions $f\colon \mathcal{T}_{p+1} \to \mathbb{C}$.

Example 5.1. If $k = 0$, then we have $\widetilde{T}_1 = \mathrm{Id}$. The most important operator for us is $\widetilde{T}_p$, which up to normalization is precisely the graph Laplacian.

Lemma 5.2. *The operators $\widetilde{T}_{p^k}$ satisfy the relations*

$$\widetilde{T}_{p^{k+1}} = \widetilde{T}_{p^k} \circ \widetilde{T}_p - (p + \delta_{k=1})\widetilde{T}_{p^{k-1}}.$$

In particular, $\widetilde{T}_p$ generates the algebra of all the operators $\widetilde{T}_{p^k}$ and this algebra is necessarily commutative.

Proof. The recursive relation is easily proven geometrically by looking at the graph. Inductively one then proceeds to show that for every k the operator $\widetilde{T}_{p^k}$ can be expressed as a polynomial in $\widetilde{T}_p$. $\square$

[13]The symbol $\approx$ indicates that this is up to suitable normalisations, which we have not worked out.

We define

$$S_y(p^k) = \{x \in \mathcal{T}_{p+1} \colon d_p(x,y) = p^k\}$$

to be the sphere of radius k around y in $\mathcal{T}_{p+1}$.

The graph Laplacian also allow us to define a p-adic version of the wave equation. Indeed, for a tuple (Φ_0, Ψ_0) of compactly supported functions on $\mathcal{T}_{p+1}$, which we refer to as initial data, we recursively define

$$\begin{aligned}\Phi_{n+1} &= \frac{1}{2\sqrt{p}}\widetilde{T}_p\Phi_n - \left(1 - \frac{1}{4p}\widetilde{T}_p^2\right)\Psi_n \text{ and}\\ \Psi_{n+1} &= \frac{1}{2\sqrt{p}}\widetilde{T}_p\Psi_n + \Phi_n.\end{aligned}$$

Let P_n and Q_n denote the Chebyshev polynomials of first and second kind respectively. These are defined by

$$P_n(\cos(\theta)) = \cos(n\theta) \text{ and } Q_{n-1}(\cos(\theta)) = \frac{\sin(n\theta)}{\sin(\theta)}.$$

In particular, $P_1(x) = x$ and $Q_0(x) = 1$. Important for us are the recursive relations

$$\begin{aligned}P_{n+1}(x) &= xP_n(x) - (1-x^2)Q_{n-1}(x)\\ Q_n(x) &= xQ_{n-1}(x) + P_n(x).\end{aligned}$$

This allows us to derive the following result:

Lemma 5.3. *Let (Φ_0, Ψ_0) be a tuple of compactly supported functions on $\mathcal{T}_{p+1}$. Then we have*

$$\begin{aligned}\Phi_n &= P_n[\frac{1}{2\sqrt{p}}\widetilde{T}_p]\Phi_0 - \left(1 - \frac{1}{4p}\widetilde{T}_p^2\right)Q_{n-1}[\frac{1}{2\sqrt{p}}\widetilde{T}_p]\Psi_0 \textit{ and}\\ \Psi_n &= P_n[\frac{1}{2\sqrt{p}}\widetilde{T}_p]\Psi_0 + Q_{n-1}[\frac{1}{2\sqrt{p}}\widetilde{T}_p]\Phi_0.\end{aligned}$$

Proof. We prove this by induction. Note that the formula holds for $n = 1$, since $P_1(x) = x$ and $Q_0(x) = 1$. Suppose the identity holds for $n = k$, then for $n = k+1$ we compute

$$\begin{aligned}\Psi_{k+1} &= \frac{1}{2\sqrt{p}}\widetilde{T}_p\Psi_k + \Phi_k\\ &= \left[\frac{1}{2\sqrt{p}}\widetilde{T}_pP_k(\frac{1}{2\sqrt{p}}\widetilde{T}_p) - (1 - \frac{1}{4p}\widetilde{T}_p^2)Q_{k-1}(\frac{1}{2\sqrt{p}}\widetilde{T}_p)\right]\Psi_0\\ &\quad + \left[P_k(\frac{1}{2\sqrt{p}}\widetilde{T}_p) + \frac{1}{2\sqrt{p}}\widetilde{T}_pQ_{k-1}(\frac{1}{2\sqrt{p}}\widetilde{T}_p)\right]\Phi_0\\ &= P_{k+1}(\frac{1}{2\sqrt{p}}\widetilde{T}_p)\Psi_0 + Q_k(\frac{1}{2\sqrt{p}}\widetilde{T}_p)\Phi_0,\end{aligned}$$

where we applied the recursion formula of the Chebyshev polynomials. Similarly we derive the other formula:

$$\begin{aligned}
\Phi_{k+1} &= \frac{1}{2\sqrt{p}}\widetilde{T}_p\Phi_k - (1-\frac{1}{4p}\widetilde{T}_p^2)\Psi_k \\
&= \left[\frac{1}{2\sqrt{p}}\widetilde{T}_pP_k(\frac{1}{2\sqrt{p}}\widetilde{T}_p) - (1-\frac{1}{4p}\widetilde{T}_p^2)Q_{k-1}(\frac{1}{2\sqrt{p}}\widetilde{T}_p)\right]\Phi_0 \\
&\quad - (1-\frac{1}{4p}\widetilde{T}_p^2)\left[P_k(\frac{1}{2\sqrt{p}}\widetilde{T}_p) + \frac{1}{2\sqrt{p}}\widetilde{T}_pQ_{k-1}(\frac{1}{2\sqrt{p}}\widetilde{T}_p)\right]\Psi_0 \\
&= P_{k+1}(\frac{1}{2\sqrt{p}}\widetilde{T}_p)\Phi_0 - (1-\frac{1}{4p}\widetilde{T}_p^2)Q_k(\frac{1}{2\sqrt{p}}\widetilde{T}_p)\Psi_0.
\end{aligned}$$

□

Corollary 5.4 (Propagation Lemma)**.** *If* $(\Phi_0, \Psi_0) = (\delta_e, 0)$ *and* n *even, then we have*

$$\Phi_n = P_n[\frac{1}{2\sqrt{p}}\widetilde{T}_p]\delta_e(x) = \begin{cases} 0 & \text{if } d_p(x,e) \text{ odd or } d_p(x,e) > n, \\ -\frac{p-1}{2p^{n/2}} & \text{if } d_p(x,e) < n \text{ is even,} \\ \frac{1}{2p^{n/2}} & \text{if } d_p(x,e) = n. \end{cases} \tag{15}$$

This can be restated as

$$2p^n \cdot P_{2n}[\frac{1}{2\sqrt{p}}\widetilde{T}_p] = \widetilde{T}_{p^{2n}} - (p-1)\sum_{k=0}^{n-1}\widetilde{T}_{p^{2k}}.$$

Proof. The identity $\Phi_n = P_n[\frac{1}{2\sqrt{p}}\widetilde{T}_p]$ follows directly from the previous lemma. On the other hand we cal also compute Φ_n inductively using the definition. We start by observing that

$$\frac{1}{2\sqrt{p}}\widetilde{T}_p(\delta_e) = \frac{1}{2\sqrt{p}}\mathbb{1}_{S_e(p)}.$$

Thus we obtain

$$(\Phi_1, \Psi_1) = (\frac{1}{2\sqrt{p}}\mathbb{1}_{S_e(p)}, \delta_e).$$

From this we compute

$$\Psi_2 = \frac{1}{2\sqrt{p}}\widetilde{T}_p\Psi_1 + \Phi_1 = \frac{1}{\sqrt{p}}\mathbb{1}_{S_e(p)}.$$

Similarly we find

$$\begin{aligned}\Phi_2 &= \frac{1}{2p}\widetilde{T}_p(\mathbb{1}_{S_e(p)}) - \delta_e \\ &= \frac{1}{2p}\mathbb{1}_{S_e(p^2)} + \frac{(p+1)}{2p}\delta_e - \delta_e \\ &= \frac{1}{2p}\mathbb{1}_{S_e(p^2)} - \frac{p-1}{2p}\delta_e.\end{aligned}$$

It is clear now how to proceed to establish the full claim.

The final claim is also clear, since we can expand

$$P_{2n}[\frac{1}{2\sqrt{p}}\widetilde{T}_p] = \sum_{i=1} c_i \widetilde{T}_{p^k}$$

and compute the coefficients from the identity

$$P_{2n}[\frac{1}{2\sqrt{p}}\widetilde{T}_p]\delta_e = \sum_{i=1} c_i \mathbb{1}_{S_e(p^k)}$$

and (15). Note that here we used $\widetilde{T}_{p^k}\delta_e = \mathbb{1}_{S_e(p^k)}$. ☐

Lemma 5.5. *Let $0 < \eta < \frac{1}{2}$. For any sufficiently large $N \in \mathbb{N}$ and any $\theta_0 \in [0, \pi]$, there exists an operator K_N on $\mathcal{T}_{p+1}$ such that:*

(1) *$K_N(\delta_x)$ is supported on the union of the spheres $S_x(p^j)$ up to distance $j \leq N$.*

(2) *K_N satisfies the inequality*

$$|K_N(f)(x)| \leq c \cdot p^{-N\delta} \sum_{j=0}^{N} \sum_{y \in S_x(p^j)} |f(y)|,$$

for positive constants $c = c(p)$ and $\delta = \delta(\eta)$.

(3) *Any $\frac{1}{2\sqrt{p}}\widetilde{T}_p$ eigenfunction is an eigenfunction of K_N with eigenvalue ≥ -1.*

(4) *Eigenfunctions with $\frac{1}{2\sqrt{p}}\widetilde{T}_p$-eigenvalue $\cos(\theta)$ with $|\theta - \theta_0| \leq \frac{1}{2N}$ as well as all non-tempered eigenfunctions (i.e. $\frac{1}{2\sqrt{p}}\widetilde{T}_p$ eigenvalue $\notin [-1, 1]$) have K_N eigenvalue $> \eta^{-1}$.*

Proof. We first note that the eigenvalues of $\frac{1}{2\sqrt{p}}\widetilde{T}_p$ must lie in the interval $[-\frac{p+1}{2\sqrt{p}}, \frac{p+1}{2\sqrt{p}}]$. Given an eigenvalue λ in this interval we write

$$\lambda = \cos(\theta) \text{ for } \theta \in \begin{cases} [0, \pi] & \text{if } \lambda \in [-1, 1], \\ -i \cdot (0, \log(\sqrt{p})) & \text{if } \lambda \in (1, \frac{p+1}{2\sqrt{p}}) \text{ and} \\ -i(0, \log(\sqrt{p})) - \pi & \text{if } \lambda \in (-\frac{p+1}{2\sqrt{p}}, -1). \end{cases}$$

If θ is real, then we call λ tempered.

The idea is to construct operators K as a polynomial in $\frac{1}{2\sqrt{p}}\widetilde{T}_p$. Suppose f is an $\frac{1}{2\sqrt{p}}\widetilde{T}_p$-eigenfunction with eigenvalue λ and associated angle θ, then f is a K-eigenfunction and the eigenvalue can be represented as a function of θ. We denote the K-eigenvalue of f by $h_K(\theta)$.

We start with the special case $\theta_0 = 0$. We define the Fejér kernel

$$F_L(\theta) = \frac{1}{L}\left(\frac{\sin(L\theta/2)}{\sin(\theta/2)}\right)^2.$$

We define

$$h_{K_N}(\theta) = F_L(q\theta) - 1$$

for appropriately chosen L and q. In particular, we will take q even. First observe that if L is sufficiently large, then F_L is non-negative and satisfies

$$F_L(\theta) \geq C \cdot L \text{ for } \theta \in (-\frac{1}{L}, \frac{1}{L}).$$

In particular, as long as $2N > L > C^{-1}\eta^{-1} + 1$, then properties (3) and (4) hold. (The non-tempered spectrum can be treated using standard properties of sinh.) We continue by recalling the expansion

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

By construction of the Chebyshev polynomial we already know that for $K = P_n(\frac{1}{2\sqrt{p}}\widetilde{T}_p)$ we must have $h_k(\theta) = \cos(n\theta)$. Thus we can estimate

$$|[K_N f](x)| \leq \sum_{j=1}^{L} \left| P_{jq}\left(\frac{1}{2\sqrt{p}}\widetilde{T}_p\right) f(x)\right|.$$

From the Propagation Lemma (i.e. Corollary 5.4) it is clear that (1) holds as soon as $qL \leq N$. (Note that this does not contradict our earlier assumption on L, q and N.) We also directly get the estimate

$$|[K_N f](x)| \leq c \cdot p^{-q/2} \sum_{j=0}^{N} \sum_{y \in S_x(p^j)} |f(y)|.$$

We can now pick our parameters. First take $L = \lceil C^{-1}\eta^{-1}\rceil + 1$. Then choose $q = 2\lfloor N/2L \rfloor$. Obviously $qL \leq N < 2N$ and property (2) holds with $\delta = q/2N \sim \eta$.

The same argument works for $\theta_\circ = \pi$. Thus it remains to consider $\theta_\circ \in (0, \pi)$. Let us for a change choose some parameters at the beginning:

$$L = \lfloor \eta^{-1} \rfloor \text{ and } Q = \lceil \frac{1}{8} N\eta \rceil.$$

By Dirichlet's theorem we find $q \leq Q$ such that

$$|q\theta_\circ \bmod 2\pi| < 2\pi Q^{-1}.$$

We can further find an integer l such that $q' = 2lq$ satisfies

$$\frac{1}{64}Q\eta \leq q' \leq 2Q.$$

For N large enough such that $Q\eta > 64$ we observe that

$$2l < \frac{1}{32}Q\eta \text{ and } |q'\theta_\circ \bmod 2\pi| < \frac{1}{16}\pi\eta.$$

We define the kernel by

$$h_{K_N}(\theta) = F_{2L}(q'\theta) - 1.$$

Note that the corresponding operator is given by

$$K_N = \sum_{j=1}^{2L} \frac{2L-j}{L} P_{jq'}\left(\frac{1}{2\sqrt{p}}\widetilde{T}_p\right).$$

As earlier we first check properties (3) and (4). Indeed (3) is a direct consequence of the positivity of the Fejér kernel on the spectrum. Furthermore, the part of (4) concerning non-tempered eigenfunctions follows as above. Thus it remains to consider $\theta \in [\theta_\circ - \frac{1}{2N}, \theta_\circ + \frac{1}{2N}]$. Note that

$$|q'(\theta_\circ - \theta)| \leq \frac{Q}{N} \leq \frac{\eta}{8} + \frac{1}{N} < \frac{\eta}{6}$$

and conclude that

$$|q'\theta \bmod 2\pi| < (\frac{\pi}{16} + \frac{1}{6})\eta < \frac{\pi}{8}\eta \leq \frac{\pi}{8L}.$$

This allows us to establish the estimate

$$F_{2L}(q'\theta) > L + 1, \tag{16}$$

which concludes the proof of (4). To see (16) we first estimate

$$F_{2L}(q'\theta) = \frac{1}{2L}\frac{\sin^2(Lq'\theta)}{\sin^2(q'\theta/2)} \geq \frac{2}{L}\frac{\sin^2(Lq'\theta)}{(q'\theta)^2} \geq 2L\left(\frac{\sin(Lq'\theta)}{Lq'\theta}\right)^2.$$

By construction we have $Lq'\theta \bmod 2\pi \in [-\frac{\pi}{8}, \frac{\pi}{8}]$. This implies

$$\left|\frac{\sin(Lq'\theta)}{Lq'\theta}\right| \geq \frac{2\sqrt{2}}{\pi}.$$

Inserting this above yields

$$F_{2L}(q'\theta) \geq 2L\frac{8}{\pi^2} > L + 1$$

as desired. Next we turn to (1). This is indeed pretty obvious, since $P_{jq'}(\frac{1}{2\sqrt{p}}\widetilde{T}_p)$ is supported within a union of spheres of radius $jq' \leq 4LQ < N$ as desired. A similar estimate as above lets us establish (2) with $\delta = q'/N \geq \eta^2/512$. $\square$

We now change gear and turn towards arithmetic lattices (or surfaces). For the Arithmetic Quantum Unique Ergodicity Theorem of Lindenstrauss that we are aiming at only certain special arithmetic surfaces will be treated. An important feature that we will use is that for such a surface, say X, we can find a prime p so that the tree $\mathcal{T}_{p+1}$ can be embedded conveniently. That this is possible is not immediately clear from the definition of arithmetic lattices. Let us still give a formal definition:

Definition 5.1 (Commensurable)**.** Two subgroups $G_1, G_2 \subseteq G$ of some ambient group G are called commensurable if $[G_i \colon G_1 \cap G_2] < \infty$ for $i = 1, 2$.

Definition 5.2. Let $G \subseteq \mathrm{GL}_n(\mathbb{Q})$ be a linear algebraic group. Then $\Gamma \subseteq G(\mathbb{R}) \subseteq \mathrm{GL}_n(\mathbb{R})$ is said to be arithmetic if it is commensurable with $G(\mathbb{Z})$.

Example 5.6. If $G = \mathrm{SL}_2(\mathbb{Q}) \subseteq \mathrm{GL}_2(\mathbb{Q})$, then $G(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z}) \subseteq \mathrm{SL}_2(\mathbb{R})$ is obviously arithmetic. However

$$\mathrm{SL}_2(\mathbb{Z})\backslash\mathrm{SL}_2(\mathbb{R}) \cong \mathrm{PSL}_2(\mathbb{Z})\backslash\mathrm{PSL}_2(\mathbb{Z})$$

is not compact. (In other words $\mathrm{PSL}_2(\mathbb{Z})$ is an arithmetic but non-uniform lattice.)

To construct examples of uniform arithmetic lattices we will use quaternion division algebras over $\mathbb{Q}$. Everything there is to know about quaternion algebras can be found in the book *Quaternion Algebras* by J. Voight. Here we will be very brief and only introduce the bare minimum necessary for our purposes.

For $a, b \in \mathbb{N}$ we define

$$H(\mathbb{Q}) = \{\alpha = x + iy + jz + ijw \colon x, y, z, w \in \mathbb{Q}\}$$

As a vector space over $\mathbb{Q}$ this is isomorphic to $\mathbb{Q}^4$, but we equip it with an algebra structure by setting

$$i^2 = a,\ j^2 = b \text{ and } ij = -ji.$$

Let $F = \mathbb{Q}(\sqrt{a}) \subseteq \mathbb{R}$ and define the embedding

$$\iota \colon H(\mathbb{Q}) \to \mathrm{Mat}_2(F) \subseteq \mathrm{Mat}_2(\mathbb{R}),\ \alpha \mapsto \begin{pmatrix} x + \sqrt{a}y & z + \sqrt{a}w \\ b(z - \sqrt{a}w) & x - \sqrt{a}y \end{pmatrix}.$$

We define

$$\overline{\alpha} = x - iy - jz - ijw$$

and compute

$$\begin{aligned} \mathrm{nr}(\alpha) &= \alpha \cdot \overline{\alpha} = x^2 - ay^2 - bz^2 + abw^2 = \det(\iota(\alpha)) \text{ and} \\ \mathrm{tr}(\alpha) &= \alpha + \overline{\alpha} = 2x = \mathrm{tr}(\iota(\alpha)). \end{aligned}$$

In particular we see that $H(\mathbb{Q})$ is an division algebra if and only if $\mathrm{nr}(\alpha) \neq 0$ for all $\alpha \neq 0$.

Definition 5.3 (Order)**.** An order $\mathcal{O} \subseteq H(\mathbb{Q})$ is a subring containing 1 with additive group of rank 4 and such that $\mathrm{tr}(\alpha) \in \mathbb{Z}$ for all $\alpha \in \mathcal{O}$.

Example 5.7. The basic example of an order is $\widetilde{\mathcal{O}} = H(\mathbb{Z})$. Note that orders are ordered by inclusion. In general $\widetilde{\mathcal{O}}$ is not maximal, but there exists an maximal order $\mathcal{O}$ such that $\widetilde{O} \subseteq \mathcal{O}$.

Given an order $\mathcal{O}$ we define

$$\mathcal{O}^1 = \{\alpha \in \mathcal{O} \colon \mathrm{nr}(\alpha) = 1\}.$$

This defines a group and we can use it do define a subgroup

$$\Gamma_{\mathcal{O}} = \iota(\mathcal{O}^1)$$

of $\mathrm{SL}_2(\mathbb{R})$.

From now on we will **always** make the following assumptions

(1) Let $b \equiv 1 \bmod 4$ be prime.
(2) Let $a \equiv 3 \bmod 4$ be prime and assume that a is not a quadratic residue modulo b.
(3) We will let $\mathcal{O}$ be a maximal order containing $H(\mathbb{Z})$.

Example 5.8. We can take for example $a = 3$ and $b = 5$.

We make the following observations:

- If $\alpha \in H(\mathbb{Q})$ and $\mathrm{nr}(\alpha) = 0$, then $\alpha = 0$. In particular $H(\mathbb{Q})$ is a division algebra. To see this we suppose that $\alpha = x+iy+jz+ijw$ satisfies $\mathrm{nr}(\alpha) = 0$. Without loss of generality we can assume that $x, y, z, w \in \mathbb{Z}$. Looking at the norm equation modulo b yields

  $$0 = \mathrm{nr}(\alpha) \equiv x^2 - ay^2 \bmod b.$$

  However, since a is not a quadratic residue modulo b this implies $x \equiv y \equiv 0$ mod b. We can now divide the original equation by b. Considering the so obtained equality modulo b lets us repeat this argument indefinitely. Thus $x = y = z = w = 0$ is the only possibility.
- If $\pm 1 \neq \alpha \in H(\mathbb{Z})^1$, then $\iota(\alpha) \in \mathrm{SL}_2(\mathbb{R})$ is hyperbolic. To see this we note that we have to show $|\mathrm{tr}(\alpha)| = |\mathrm{tr}(\iota(\alpha))| > 2$. However, since $x \in \mathbb{Z}$ and $\mathrm{tr}(\alpha) = 2x$ the only cases we need to exclude are $x = 0, \pm 1$. For $x = 0$ the equation $\mathrm{nr}(\alpha) = 1$ yields

  $$1 + ay^2 + bz^2 = abw^2.$$

  Looking at this modulo 4 we see that the left hand side is $\equiv 1 \bmod 4$ while the right hand side is $\equiv 3 \bmod 4$. This is a contradiction. Similarly one sees that the case $x = \pm 1$ would lead to

  $$ay^2 + bz^2 = abw^2.$$

  Again one gets a contradiction by looking modulo 4.

- $\iota(H(\mathbb{R})) = \mathrm{Mat}_2(\mathbb{R})$. This can be checked directly. In particular, the reduced discriminant of $H(\mathbb{Q})$ is $\mathrm{discrd}(H(\mathbb{Q})) = ab$. Since $\mathrm{discrd}(H(\mathbb{Z})) = 4ab$ we see that $H(\mathbb{Z})$ is not maximal. Since
$$4ab = \mathrm{discrd}(H(\mathbb{Z})) = [\mathcal{O}\colon H(\mathbb{Z})]^2 \cdot \mathrm{discrd}(\mathcal{O}) = [\mathcal{O}\colon H(\mathbb{Z})]^2 \cdot ab$$
we see that $H(\mathbb{Z})$ must have index 2 in the maximal order $\mathcal{O}$. It can be seen that also $[\mathcal{O}^1\colon H(\mathbb{Z})^1] < \infty$.[14]

We can now sketch the proof of the following result. For a more in depth treatment we refer to Chapter 38 in the book *Quaternion Algebras* by J. Voight.

Proposition 5.9. *Both $\Gamma_{\mathcal{O}}$ and $\Gamma_{H(\mathbb{Z})}$ are uniform arithmetic lattices in $SL_2(\mathbb{R})$.*

Proof. We sketch the proof for $H(\mathbb{Z})$, since it is particularly nice. (The case of $\mathcal{O}$ is similar. Indeed many properties can be directly transferred because the index of $H(\mathbb{Z})^1$ in $\mathcal{O}^1$ is finite.) First note that $H(\mathbb{Z})^1$ is non-abelian. Furthermore, by our observation above $\iota(H(\mathbb{Z})^1)$ contains only hyperbolic elements (except plus minus the identity). Therefore, Nielsen's theorem applies to $\Gamma_{H(\mathbb{Z})}$ and yields that $\Gamma_{H(\mathbb{Z})} \subseteq \mathrm{SL}_2(\mathbb{R})$ is discrete. The volume of $\Gamma_{H(\mathbb{Z})}\backslash\mathbb{H}$ can be computed explicitly and is in particular finite. It can also be shown that $\Gamma_{H(\mathbb{Z})}$ is finitely generated. Thus so far we have seen that $\Gamma_{H(\mathbb{Z})}$ is a lattice. But it does not contain any parabolic elements. Therefore $\Gamma_{H(\mathbb{Z})}$ must be uniform.

Finally we need to show that $H(\mathbb{Z})$ is arithmetic. To see this we consider the embedding $\eta\colon H(\mathbb{Q}) \to \mathrm{Mat}_4(\mathbb{Q})$ given by
$$\eta(i) = \begin{pmatrix} 0 & a & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & a \\ 0 & 0 & 1 & 0 \end{pmatrix} \text{ and } \eta(j) = \begin{pmatrix} 0 & 0 & b & 0 \\ 0 & 0 & 0 & -b \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}.$$
We see that $\eta(H(\mathbb{Q})^1) = G(\mathbb{Q}) \subseteq \mathrm{GL}_4(\mathbb{Q})$ is a linear algebraic group. In particular we observe that $G(\mathbb{R}) = \eta(H(\mathbb{R})^1) \cong \mathrm{SL}_2(\mathbb{R})$ and $G(\mathbb{Z}) = \eta(H(\mathbb{Z})^1)$. Thus $H(\mathbb{Z})^1 \subseteq H(\mathbb{R})^1$ is obviously arithmetic. □

Note that $\pm 1 \in \mathcal{O}^1$ and we set $\overline{\Gamma}_{\mathcal{O}} = \{\pm I_2\}\backslash\Gamma_{\mathcal{O}} \subseteq \mathrm{PSL}_2(\mathbb{R})$. We have
$$M = \overline{\Gamma}_{\mathcal{O}}\backslash\mathrm{PSL}_2(\mathbb{R}) = \Gamma_{\mathcal{O}}\backslash\mathrm{SL}_2(\mathbb{R}).$$
We define the sets
$$\mathcal{O}(m) = \{\alpha \in \mathcal{O}\colon \mathrm{nr}(\alpha) = m\}.$$
We will need this sets only for $m = p$ with $p \nmid ab$.[15] Note that $\mathcal{O}^1$ acts on $\mathcal{O}(p)$ from the left and from the right. Thus, given $x = \Gamma_{\mathcal{O}} g_x \in M$ we obtain the collection

[14]Actually the two indexes $[\mathcal{O}\colon H(\mathbb{Z})]$ and $[\mathcal{O}^1\colon H(\mathbb{Z})^1]$ are directly related, but this relationship is irrelevant for us.

[15]If one wants to work with $H(\mathbb{Z})$ instead of $\mathcal{O}$, then one needs to assume $p \nmid 2ab$. The rest of the discussion will then apply.

of points

$$S_x(p) = \left\{ \Gamma_{\mathcal{O}}[\frac{1}{\sqrt{p}}\iota(\alpha)]g_x \colon \alpha \in \mathcal{O}^1\backslash\mathcal{O}(p) \right\} \subseteq M.$$

We make some observations:

(1) The set $S_x(p)$ is well defined. Indeed, if $x = \Gamma_{\mathcal{O}} g'_x$, then $g'_x = \iota(\beta) g_x$ for $\beta \in \mathcal{O}^1$ so that

$$\Gamma_{\mathcal{O}}[\frac{1}{\sqrt{p}}\iota(\alpha)]g'_x = \Gamma_{\mathcal{O}}[\frac{1}{\sqrt{p}}\iota(\alpha\beta)]g_x.$$

However, the map $\alpha \mapsto \alpha\beta$ is a permutation of $\mathcal{O}^1\backslash\mathcal{O}(p)$.

(2) We have $\sharp S_x(p) = \sharp\mathcal{O}^1\backslash\mathcal{O}(p) < \infty$. That the quotient $\mathcal{O}^1\backslash\mathcal{O}(p)$ is finite will be (very) briefly discussed later on. To see the first equality we simply observe that

$$\Gamma_{\mathcal{O}}[\frac{1}{\sqrt{p}}\iota(\alpha_1)]g_x = \Gamma_{\mathcal{O}}[\frac{1}{\sqrt{p}}\iota(\alpha_2)]g_x$$

precisely when $\alpha_1 \in \mathcal{O}^1\alpha_2$.

We call $y \in S_x(p)$ the p-neighbours of x and set $d_p(x, y) = p$ if and only if $y \in S_x(p)$. Now lets take

$$y = \Gamma_{\mathcal{O}} g_y = \Gamma_{\mathcal{O}}[\frac{1}{\sqrt{p}}\iota(\alpha_1)]g_x \in S_x(p).$$

A point $z \in S_y(p)$ can be written as

$$z = \Gamma_{\mathcal{O}}[\frac{1}{p}\iota(\alpha_2\alpha_1)]g_x.$$

We see that basically two things can happen. First, if $\alpha_2 = \overline{\alpha_1}$, then $\alpha_2\alpha_1 = \mathrm{nr}(\alpha_1) = p$. In this case $z = x$. Second, if $\alpha_2 \notin \mathcal{O}^1\overline{\alpha_1}$, then it can be checked that

$$z \notin \bigcup_{y' \in S_x(p)\backslash\{y\}} S_{y'}(p).$$

This allows us to inductively construct a tree in M centred at x:

- The vertices of the tree are given by $V = \bigcup_k B_x(p^k)$ where the balls $B_x(p^k)$ of radius k are inductively defined by $B_x(0) = \{x\}$ and $B_x(p^k) = \bigcup_{y \in B_x(p^{k-1})} S_y(p)$.
- Two points $y', y \in V$ are neighbours if and only if $y' \in S_y(p)$.

We extend the p-distance in the obvious way and define the k-spheres by

$$S_x(p^k) = \{y \in M \colon d_p(x, y) = p^k\}.$$

We claim that for $p \nmid ab$ one has

$$\sharp\mathcal{O}^1\backslash\mathcal{O}(p) = p + 1. \tag{17}$$

Furthermore,

$$S_x(p^k) = \{\Gamma_{\mathcal{O}}[p^{-k/2}\iota(\alpha)]g_x \colon \alpha \in \mathcal{O}^1\backslash\mathcal{O}(p^k) \text{ primitive } \} \subseteq M.$$

This is not obvious and we will take it for granted. The upshot is that we have indeed constructed an embedding

$$\epsilon_x \colon \mathcal{T}_{p+1} \to M,\ e \mapsto x.$$

In particular, this justifies our suggestive notation because

$$\epsilon_x(S_e(p^k)) = S_x(p^k) \subseteq M.$$

Definition 5.4 (Hecke Operators)**.** We now define the operators

$$[\widetilde{T}_{p^k} f](x) = \sum_{y\in S_x(p^k)} f(y) \in \mathcal{C}_b^\infty(M),$$

for $f \in \mathcal{C}_b^\infty(M)$.

Remark 5.10. Note that we can write

$$[\widetilde{T}_p f](x) = \sum_{\alpha\in\mathcal{O}^1\backslash\mathcal{O}(p)} f\left(\frac{1}{\sqrt{p}}\iota(\alpha)x\right).$$

This is well defined and appears to be the more classical definition of the p-th Hecke Operator.

Proposition 5.11. *Let $a, b \in \mathbb{N}$ and $\mathcal{O}$ be as above and let $p \nmid ab$ be a prime. Then we have*

(1) *The operators $\widetilde{T}_{p^k}$ acting on $\mathcal{C}_b^\infty(M)$ inherit all the properties of the operators acting on functions on $\mathcal{T}_{p+1}$.*
(2) *Let $m \in \mathfrak{sl}_2(\mathbb{C})$, then*
$$m.[\widetilde{T}_{p^k} f] = \widetilde{T}_{p^k}[m.f]$$
for $f \in \mathcal{C}_b^\infty(M)$.
(3) *The operators $\widetilde{T}_{p^k}$ are self-adjoint.*

Proof. The first and the second property follow directly from the definitions. The third property is follows from an appropriate change of variables in the definition of the inner product. We omit the details. □

In particular we see that the operators $\widetilde{T}_{p^k}$ commute with Ω_c, $\mathcal{E}^+$ and $\mathcal{E}^-$. Furthermore, they leave the space $\mathcal{A}_0(M)$ invariant. Since we identify the latter space with $\mathcal{C}_b^\infty(X)$ we can let the operators $\widetilde{T}_{p^k}$ act on this space as well. Here they will commute with the Laplace-Beltrami Operator Δ_X.

Definition 5.5. A function $\phi \in \mathcal{C}_b^\infty(X)$ is called a joint eigenfunction (more precisely a $(\widetilde{T}_p, \Delta_X)$-joint eigenfunction) if ϕ is an eigenfunction of $\widetilde{T}_p$ and Δ_X.

We make the following observations:

- Such joint eigenfunctions exist, because $\widetilde{T}_p$ and Δ_X commute and are both self-adjoint.
- A joint eigenfunction is automatically an eigenfunction of all the operators $\widetilde{T}_{p^k}$ for $k \in \mathbb{N}$.
- If ϕ is a joint eigenfunction, then the approximate microlocal lifts $\tilde{\phi}^{(N)}$ are eigenfunctions of all the operators $\widetilde{T}_{p^k}$.

Definition 5.6 (Arithmetic Quantum Limit)**.** A measure σ on M is called an arithmetic quantum limit if the underlying sequence (ϕ_i) consists of joint eigenfunctions.

Remark 5.12. Note that we fix a single prime p and only consider the two operators $\widetilde{T}_p$ and Δ_X. One can also take more Hecke Opertors into account, but it turns out that for our purposes one is sufficient.

We conclude by a lengthy remark that somehow is trying to explain (17):

Remark 5.13. Suppose that $p \nmid 2ab$, so that we can work with $H(\mathbb{Z})$ instead of $\mathcal{O}$. Let $\mathbb{Z}_p$ denote the p-adic integers and write $\mathbb{Q}_p$ for their quotient field. Then the assumption $p \nmid 2ab$ ensures that we have an isomorphism

$$H(\mathbb{Q}_p) \cong \mathrm{Mat}_2(\mathbb{Q}_p) \text{ and } H(\mathbb{Z}_p) \cong \mathrm{Mat}_2(\mathbb{Z}_p).$$

It is well known that the quotient $\mathrm{PGL}_2(\mathbb{Q}_p)/\mathrm{PGL}_2(\mathbb{Z}_p)$ carries the structure of a $p+1$-regular tree. Indeed the neighbours of $\mathrm{PGL}_2(\mathbb{Z}_p)$ are given by

$$\left\{ \begin{pmatrix} 1 & 0 \\ 0 & p \end{pmatrix}, \begin{pmatrix} p & j \\ 0 & 1 \end{pmatrix} : 0 \leq j \leq p-1 \right\} \cdot \mathrm{PGL}_2(\mathbb{Z}_p).$$

Thus we can identify the infinite $p+1$ regular tree with

$$\mathcal{T}_{p+1} = \mathrm{PGL}_2(\mathbb{Q}_p)/\mathrm{PGL}_2(\mathbb{Z}_p) = PH(\mathbb{Q}_p)^\times/PH(\mathbb{Z}_p)^\times.$$

The key is now that we also have an natural isomorphism

$$\Gamma_{H(\mathbb{Z})}\backslash \mathrm{SL}_2(\mathbb{R}) \cong PH(\mathbb{Z}[\frac{1}{p}])^\times \backslash (PH(\mathbb{R})^\times \times PH(\mathbb{Q}_p)^\times)/PH(\mathbb{Z}_p)^\times.$$

The embedding of the infinite $p+1$ regular tree $\mathcal{T}_{p+1}$ on the left and side of this isomorphism is obtained by looking at an $PH(\mathbb{Q}_p)^\times$-orbit of a point in $PH(\mathbb{Z}[\frac{1}{p}])^\times \backslash (PH(\mathbb{R})^\times \times PH(\mathbb{Q}_p)^\times)$ on the right hand side.

6. Hecke Recurrence

Let $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$ be a uniform arithmetic lattice constructed as in the previous subsection. As usual we put $M = \Gamma\backslash \mathrm{PSL}_2(\mathbb{R})$ and $X = M/\mathrm{PSO}_2$. Further fix a prime $p \notin \{a, b\}$.

Definition 6.1. A measure μ on M is said to be T_p-recurrent if for any subset $A \subseteq M$ with $\mu(A) > 0$, for μ-almost every $x \in A$ there is a sequence $k_i \to \infty$ for which $S_{p^{k_i}}(x) \cap A \neq \emptyset$.

The goal of this section is to show that every arithmetic quantum limit σ is T_p recurrent. To establish this we first need some technical lemmas.

Lemma 6.1. *Let ϕ_j be a $\widetilde{T}_p$ eigenfunction. For $x \in M$ we have*

$$|\widetilde{\phi}_j^{(N_j)}(x)|^2 \leq \frac{c_p}{N} \cdot \sum_{d_p(y,x) \leq p^N} |\tilde{\phi}_j^{(N_j)}(y)|^2,$$

for some absolute constant $c_p > 0$ and N sufficiently large.

Proof. Let L be large and let $\lambda_j(p)$ be the $\frac{1}{2\sqrt{p}}\widetilde{T}_p$ eigenvalue of ϕ_j. We define

$$\theta_j = \begin{cases} \arccos(\lambda_j(p)) & \text{if } \lambda_j \in [-1,1], \\ 0 & \text{if } \lambda_j(p) > 1, \\ \pi & \text{if } \lambda_j(p) < -1. \end{cases}$$

By the pigeon hole principle we can find $q_j \in \{1, \ldots, 100L\}$ such that

$$|q_j \theta_j \bmod 2\pi| \leq \frac{\pi}{50L}.$$

We set

$$K_j = \sum_{l=1}^{L} P_{2q_j l}\left(\frac{1}{2\sqrt{p}}\widetilde{T}_p\right).$$

Note that

$$K_j \tilde{\phi}_j^{(N_j)} = a_j \cdot \tilde{\phi}_j^{(N_j)}$$

for

$$a_j = \sum_{l=1}^{L} P_{2q_j l}(\lambda_j(p)).$$

From now on suppose that $\lambda_j(p) \in [-1,1]$. The remaining cases can be treated similarly and we omit the details. In this case we see that

$$a_j = \sum_{l=1}^{L} P_{2q_j l}(\cos(\theta_j)) = \sum_{l=1}^{L} \cos(2q_j l \theta_j) \geq \frac{1}{2} \cdot L$$

for L sufficiently large.

On the other hand we have the estimate

$$
\begin{aligned}
|K_j\tilde{\phi}_j^{(N_j)}(x)| &\leq (p-1)\sum_{l=1}^{L}\sum_{p^{2q_j(l-1)}<d_p(y,x)\leq p^{2q_jl}} p^{-q_jl}|\tilde{\phi}_j^{(N_j)}(y)| \\
&\leq (p-1)\sum_{l=1}^{L}\Bigg(\underbrace{\sharp\{y\colon p^{2q_j(l-1)}<d_p(y,x)\leq p^{2q_jl}\}\cdot p^{-2q_jl}}_{\leq C_p'}\cdot\sum_{p^{2q_j(l-1)}<d_p(y,x)\leq p^{2q_jl}}|\tilde{\phi}_j^{(N_j)}(y)|^2\Bigg)^{\frac{1}{2}} \\
&\leq C_p\cdot L^{\frac{1}{2}}\Bigg(\sum_{d_p(y,x)\leq p^{100L^2}}|\tilde{\phi}_j^{(N_j)}(y)|^2\Bigg)^{\frac{1}{2}} \quad \text{for some constants } C_p, C_p'>0.
\end{aligned}
$$

Here we used Lemma 5.4 and Cauchy-Schwarz twice. We conclude by observing that

$$\frac{L^2}{4}|\widetilde{\phi}_j^{(N_j)}(x)|^2 \leq a_j^2|\widetilde{\phi}_j^{(N_j)}(x)|^2 = |K_j\widetilde{\phi}_j^{(N_j)}(x)|^2 \leq C_p\cdot L\cdot\sum_{d_p(y,x)\leq p^{100L^2}}|\tilde{\phi}_j^{(N_j)}(y)|^2.$$

□

Proposition 6.2. *Every arithmetic quantum limit σ^{ML} is T_p-recurrent.*

Proof. We give a simplified proof ignoring some minor measure theoretic subtleties. Suppose σ^{ML} is not T_p-recurrent. This allows us to find a Borel measurable subset $A \subseteq M$ with $\sigma^{\mathrm{ML}}(A) > 0$ such that $\epsilon_x(\mathcal{T}_{p+1}) \cap A$ is finite for (σ-almost) every $x \in A$. We partition A into disjoint subsets

$$A_K = \{x \in A\colon S_x(p^K) \cap A \neq \emptyset \text{ and } S_x(p^l) \cap A = \emptyset \text{ for } l > K\}.$$

We claim that the sets A_K are measurable and $A = \bigsqcup_{K\in\mathbb{N}} A_K$ (up to a set of measure 0). Thus there is K_0 such that $\sigma(A_{K_0}) > 0$. In particular, there is a constant C depending only on K_0 and p such that

$$\sharp(\epsilon_x(\mathcal{T}_{p+1}) \cap A_{K_0}) \leq C,$$

for all $x \in A_{K_0}$. We can now find another subset A' of A_{K_0} such that $\sigma(A') > 0$ and $\epsilon_x(\mathcal{T}_{p+1}) \cap A' = \{x\}$ for all $x \in A'$. This is the set we will work with.

Let $f = \mathbb{1}_{A'}$ be the characteristic function on A'. By construction we have

$$\widetilde{T}_{p^k} f = 0 \text{ for } k \in \mathbb{N}.$$

For sufficiently large N we compute

$$\begin{aligned}\sigma^{\mathrm{ML}}(A') &= \int_M f(z)d\sigma^{\mathrm{ML}}(z)\\ &= \lim_{j\to\infty}\sum_{l=0}^{N}\int_M [\widetilde{T}_{p^l}f](z)\cdot|\tilde{\phi}_j^{(N_j)}(z)|d\mu_M(z)\\ &= \lim_{j\to\infty}\sum_{l=0}^{N}\int_M f(z)[\widetilde{T}_{p^l}|\tilde{\phi}_j^{(N_j)}|^2](z)d\mu_M(z)\\ &= \lim_{j\to\infty}\int_{A'}\sum_{d_p(z,y)\le p^N}|\tilde{\phi}_j^{(N_j)}(y)|^2 d\mu_M(z).\end{aligned}$$

At this point we apply Lemma 6.1 to get

$$\sigma^{\mathrm{ML}}(A') \ge \frac{N}{c_p}\cdot\lim_{j\to\infty}\int_{A'}|\tilde{\phi}_j^{(N_j)}(z)|^2 d\mu_M(z) = \frac{N}{c_p}\sigma^{\mathrm{ML}}(A').$$

By taking N large enough so that $\frac{N}{c_p}>1$ we conclude that $\sigma^{\mathrm{ML}}(A')=0$. This is a contradiction. □

7. Entropy

Let us start by recalling the definition and some properties of entropy. We can do so in the setting of $(X,\mathcal{B},\mu)$ be a probability space equipped with a measure preserving map $T\colon X\to X$. Given a countable measurable partition $\mathcal{P}=\{A_1,A_2\dots\}$ of X we write $\bigvee_{i=0}^{n-1}T^{-i}\mathcal{P}$ for the common refinement of the partitions $\mathcal{P},T^{-1}\mathcal{P},\dots,T^{-(n-1)}\mathcal{P}$. We can now make the following definition:

Definition 7.1. The static entropy of $\mathcal{P}$ is defined by

$$H_\mu(\mathcal{P}) = -\sum_{A\in\mathcal{P}}\mu(A)\log(\mu(A)).$$

We define the dynamical entropy of $\mathcal{P}$ with respect to T by[16]

$$h_\mu(T,\mathcal{P}) = \lim_{n\to\infty}\frac{1}{n}H_\mu\left(\bigvee_{i=0}^{n-1}T^{-i}\mathcal{P}\right).$$

Finally, the entropy of T is defined by

$$h_\mu(T) = \sup_{\mathcal{P}} h_\mu(T,\mathcal{P}).$$

Example 7.1. Let $X=S^1\cong[0,1)$ be the circle equipped with the Borel σ-algebra and the Lebesgue measure μ. Further let T_α be the rotation with angle α. Then $h_\mu(T)=0$. We leave it as an exercise to show this. (It is particularly easy to see when T_α is periodic.)

[16]It is easy to verify the existence of the limit.

Recall the definition of ergodicity from Definition 1.2. In our current setting, where we have a discrete dynamical system $T\colon X \to X$, this definition can be interpreted as follows. The T-invariant probability measure μ on $(X, \mathcal{B})$ is ergodic if and only if any T-invariant set $E \in \mathcal{B}$ satisfies $\mu(E) \in \{0,1\}$. Often there are many ergodic T-invariant measures. Note that given two ergodic T-invariant measures, say μ_1 and μ_2, we can define

$$\mu_s = s\mu_1 + (1-s)\mu_2 \text{ for } s \in [0,1]. \tag{18}$$

Obviously this is a T-invariant probability measure. It turns out that for $s \in (0,1)$ it will not be ergodic.

Suppose from now on that X is a metric space and $\mathcal{B}$ is the Borel σ-algebra. Furthermore, we assume that $T\colon X \to X$ is continuous. Denote by $\mathcal{M}(X)$ the space of all Borel probability measures. Then one can show that

$$\mathcal{M}^T(X) = \{\mu \in \mathcal{M}(X)\colon T_*\mu = \mu\}$$

is a closed convex.[17]

Example 7.2. We can consider the metric space $M = \Gamma\backslash \mathrm{PSL}_2(\mathbb{R})$ for a lattice $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$ and the map T given by the following the geodesic flow for one time step. (Thus T corresponds to $x \mapsto xa_1^{-1}$.) We have seen the uniform measure μ_M is invariant under the geodesic flow. Thus $\mu_M \in \mathcal{M}^T(M)$.

In general we have the following result:

Theorem 7.3. *Let X be a compact metric space and let $T\colon X \to X$ be a continuous map. Then*

- *The set $\mathcal{M}^T(X)$ is non-empty.*
- *The extreme points of $\mathcal{M}^T(X)$ are precisely the measures that are ergodic with respect to T. We denote this set by $\mathcal{E}^T(X)$.*
- *If $\mu_1, \mu_2 \in \mathcal{E}^T(X)$ and $\mu_1 \neq \mu_2$, then μ_1 and μ_2 are mutually singular. (In other words there are disjoint measurable sets A, B such that $X = A \cup B$ and $\mu_1(A) = \mu_2(B) = 0$.)*

Proof. See Section 4.1 in the book *Ergodic Theory with a view towards Number Theory* by M. Einsiedler and T. Ward for proofs. □

This leads us to the following structure theorem:

Theorem 7.4 (Ergodic decomposition)**.** *Let X be a compact metric space and let $T\colon X \to X$ be a continuous map. Then for any $\mu \in \mathcal{M}^T(X)$ there is a unique probability measure λ defined on the Borel subsetes of the compact metric space $\mathcal{M}^T(X)$ such that*

- *$\lambda(\mathcal{E}^T(X)) = 1$, and*
- *$\int_X f(x)d\mu(x) = \int_{\mathcal{E}^T(X)} \left(\int_X f(x)d\nu(x)\right) d\lambda(\nu)$ for any $f \in \mathcal{C}(X)$.*

[17] We define $T_*\mu$ by $\int_X f(x)dT_*\mu(x) = \int_X f(T(x))d\mu(x)$.

Proof. The result is a direct consequence of Choquet's Theorem. (It can be thought of as generalising (18).) □

Let us return to entropy. It can be seen that entropy behaves well with respect to the ergodic decomposition described above:

Lemma 7.5. *Let X be a compact metric space and let $T\colon X \to X$ be a continuous transformation. Then we have*

$$h_\mu(T) = \int_{\mathcal{E}^T(X)} h_\nu(T) d\lambda(\nu).$$

Proof. We show the result for μ of the form (18). In this case we can use simple properties of the logarithm to see that

$$sH_{\mu_1}(\mathcal{P}) + (1-s)H_{\mu_2}(\mathcal{P}) \leq H_{\mu_s}(\mathcal{P}) \leq sH_{\mu_1}(\mathcal{P}) + (1-s)H_{\mu_2}(\mathcal{P}) + O(1).$$

This inequality directly implies that

$$h_{\mu_s}(T, \mathcal{P}) = sh_{\mu_1}(T, \mathcal{P}) + (1-s)h_{\mu_2}(T, \mathcal{P})$$

and we are done. The general case is more intricate and we omit the proof. □

Definition 7.2. A measure $\mu \in \mathcal{M}^T(X)$ is said to have positive entropy for almost all ergodic components if $h_\nu(T) > 0$ for λ-almost all $\nu \in \mathcal{E}^T(X)$.

This is the condition we have to check for our quantum limits. In order to do so we will need some technical condition, which is checkable in practice. The following technical result is what we need:

Lemma 7.6. *Let X be a compact metric space and let $T\colon X \to X$ be a continuous map and let $\mathcal{P}$ be a finite partition. Suppose that for any $\eta > 0$ there is $\delta = \delta(\eta) > 0$ such that for all sufficiently large N we have that any collection of distinct partition elements from $\bigvee_{i=0}^{\lfloor 2N \log(p) \rfloor - 1} T^{-i}\mathcal{P}$ with total μ-mass $> \eta$ must contain at least $p^{\delta N}$ partition elements. Then almost every ergodic component of μ has positive entropy.*

Proof. The idea is that the statement implies that there is at most μ-measure η on ergodic components of entropy less than δ. Taking $\eta \to 0$ gives the result. □

Remark 7.7. A similar, but stronger, condition one can check is the following: *Let X be a compact metric space and let $T\colon X \to X$ be continuous. Furthermore let $\mathcal{P}$ be a countable partition of X with $H_\mu(X) < \infty$. If there are constants $c, h > 0$ such that $\mu(A) \leq e^{-nh+c}$ for all $A \in \bigvee_{i=0}^{n-1} T^{-i}\mathcal{P}$ and all $n \geq 0$, then almost every ergodic component of μ has positive entropy. More precisely, one has $h_\nu(T, \mathcal{P}) > h$ for λ-almost every $\nu \in \mathcal{E}^T(X)$.* Indeed, going from this condition to the one stated in the lemma is a matter of comparing volumes.

Let us now turn towards our usual set-up. Thus we take a quaternion algebra $H(\mathbb{Q})$ over $\mathbb{Q}$ given by $i^2 = a$ and $j^2 = b$ for two primes a and b such that $a \equiv 4 \bmod 4$ and $b \equiv 1 \bmod 4$. Further we assume that a is not a quadratic residue modulo b. Finally take a maximal order $\mathcal{O}$ containing $H(\mathbb{Z})$ and associate the uniform arithmetic lattice $\Gamma_{\mathcal{O}} = \iota(\mathcal{O}^1)$. As usual we will work with the quotient $M = \Gamma_{\mathcal{O}} \backslash \mathrm{SL}_2(\mathbb{R})$, which plays the role of the unit tangent bundle of $X = M/\mathrm{PSO}_2$.

Our goal is to prove the following important result:

Proposition 7.8. *Let σ be an arithmetic quantum limit on M, then almost all ergodic components of σ have positive entropy for the geodesic flow.*

Proving this requires some preparation. First we recall the matrices

$$u^+(s) = \begin{pmatrix} 1 & 0 \\ s & 1 \end{pmatrix},\ u^-(s) = \begin{pmatrix} 1 & s \\ 0 & 1 \end{pmatrix} \text{ and } a_t = \begin{pmatrix} e^{-t/2} & 0 \\ 0 & e^{t/2} \end{pmatrix}.$$

We define the neighborhoods

$$B(\epsilon, \tau) = \{a_t u^-(s_-) u^+(s_+) \colon t \in (-\tau, \tau) \text{ and } s_-, s_+ \in (-\epsilon, \epsilon)\}.$$

One checks that

$$B(\epsilon,\tau)^{-1} B(\epsilon,\tau) \subseteq B(4\epsilon, 3\tau).$$

Lemma 7.9. *For τ small but fixed, there is a constant $c = c(\tau) > 0$ such that for any $x, z \in M$ and any $0 < \epsilon < cp^{-2N}$ the tube $zB(\epsilon,\tau) \subseteq M$ contains at most $O(N)$ of the Hecke points $\bigcup_{j \le N} S_x(p^j)$.*

Furthermore, for any $x \in M$ there are at most $O(N)$ points $y \in \bigcup_{j\le N} S_z(p^j)$ such that $xB(cp^{-2N},\tau) \cap yB(cp^{-2N},\tau) \ne \emptyset$.

Proof. Let $\mathcal{F}$ be a fundamental domain for $\Gamma_{\mathcal{O}}$. We write $x', z' \in \mathcal{F}$ for appropriate lifts of $x, z \in M$ to the fundamental domain.

Suppose $y_i \in zB(\epsilon,\tau)$, for $i = 1,\ldots,k$ are distinct Hecke points in $S_x(p^{j_i})$ with $j_i \le N$. By construction we can write $y_i = \Gamma_{\mathcal{O}} \alpha_i' x'$ for $\alpha_i' = p^{-j_i/2}\iota(\alpha_i)$ and $\alpha_i \in \mathcal{O}(p^{j_i})$. We can choose a representative for α_i such that $y' = \alpha_i' x' \in z' B(\epsilon,\tau)$.

First we note that

$$(\alpha_1')^{-1}\alpha_{i+1}' \in x' B(4\epsilon, 3\tau)(x')^{-1}.$$

On the other hand we have $(\alpha_1')^{-1}\alpha_{i+1}' = p^{-(j_1+j_{i+1})/2}\iota(\beta_i) = \beta_i'$, for $\beta_i = \overline{\alpha_1}\alpha_{i+1} \in \mathcal{O}(p^{j_1+j_{i+1}})$. Of course $j_1 + j_{i+1} \le 2N$. We observe the estimates

$$\mathrm{Tr}(\beta_i) \ge \mathrm{Nr}(\beta_i)^{\frac{1}{2}}(2 - \frac{\epsilon}{c_1}) \text{ and}$$
$$\mathrm{Tr}(\beta_i^2) \ge \mathrm{Nr}(\beta_i)(2 - \frac{\epsilon}{c_1}).$$

Since the traces and the norms are integers we see that for $\epsilon < c_1 p^{-2N}$ we must have $\mathrm{Tr}(\beta_i^2) \ge 2\mathrm{Nr}(\beta_i)$. Even more, since

$$\mathrm{Tr}(\beta_i^2) = \mathrm{Tr}(\beta_i)^2 - \mathrm{Nr}(\beta_i)$$

we also get $\mathrm{Tr}(\beta_i) \ge 2\mathrm{Nr}(\beta_i)^{\frac{1}{2}}$. This implies that $L_i = \mathbb{Q}(\beta_i) \subseteq H(\mathbb{Q})$ is a real quadratic field. Note that every element that commutes with β_i is contained in L_i.

Similar estimates yield

$$|\mathrm{Nr}(\beta_i\beta_j - \beta_j\beta_i)| < \frac{1}{c_1}\mathrm{Nr}(\beta_i\beta_j)^{\frac{1}{2}}\epsilon.$$

Now recall that $\beta_i\beta_j - \beta_j\beta_i \in \mathcal{O}$ so that $\mathrm{Nr}(\beta_i\beta_j - \beta_j\beta_i) \in \mathbb{Z}$. By our assumption on ϵ we must have $\mathrm{Nr}(\beta_i\beta_j - \beta_j\beta_i) = 0$. Since $H(\mathbb{Q})$ is a division algebra the only element with trivial norm is 0. We conclude that $\beta_i\beta_j = \beta_j\beta_i$. Thus have $L = L_1 = \ldots = L_{k-1}$.

For any $\beta \in L \cap \mathcal{O}$ we have $\beta \in \mathcal{O}_L$. (Here $\mathcal{O}_L$ is the ring of integers in the field L.) Thus we get principal ideals

$$J_i = \beta_i \cdot \mathcal{O}_L.$$

These ideals have norm bounded by p^{2N} and we can re-scale them by writing $J_i' = p^{-l_i} J_i$ so that J_i' is not divisible by p.

Recall that all the β_i are distinct. Using $\beta_i' \in x' B(4\epsilon, 3\tau)(x')^{-1}$ once again we can ensure that all the ideals J_i' are distinct. (At least for sufficiently small τ.)

In summary we can bound the number k of distinct Hecke points y_i (up to level N) from above by the number of ideals in $\mathcal{O}_L$ of norm dividing p^{2N}. This number can be estimated by $4N$ using methods from classical algebraic number theory.

The second seemingly stronger statement follows directly by shrinking the constant. This is a consequence of the geometry of the tubes $B(\epsilon, \tau)$. This completes the proof. □

We are finally ready to establish positive entropy of almost every ergodic component:

Proof of Proposition 7.8. Take a partition $\mathcal{P}$ of M such that $\sigma(\partial P) = 0$ for every $P \in \mathcal{P}$. Further we assume that $\max_{P\in\mathcal{P}} \mathrm{diam}(P)$ is sufficiently small. Any partition element of the $\lfloor 2N\log(p)\rfloor$-th refinement is contained in a union of at most $O_c(1)$ tubes of the form $x_l B(cp^{-2N}, \tau)$, where $x_l \in M$ are some points. We can make c small enough, so that the statement from Lemma 7.9 holds.

As we have seen from our discussion above, in particular Lemma 7.6, we have to show the following. For any $\eta > 0$ there exists $\delta(\eta) > 0$ such that, for all N sufficiently large, any collection of distinct partition elements of the $\lfloor 2N\log(p)\rfloor$-th refinement of $\mathcal{P}$ whose union has mass $> \eta$, must contain at least $p^{\delta N}$ partition elements.

With this goal in mind we take any collection $\{E_1, \ldots, E_k\}$ of distinct partition elements of the $\lfloor 2N\log(p)\rfloor$-th refinement of $\mathcal{P}$. Put

$$\mathcal{E} = \bigcup_{l=1}^{k} E_l$$

and assume $\sigma(\mathcal{E}) > \eta$. We need to show that $k \geq p^{\delta\cdot N}$ (for sufficiently large N).

At this point we have to define some functions. Let $f_l = \mathbb{1}_{E_l}$ be the characteristic function on E_l for $1 \leq l \leq k$. Set

$$f_{\mathcal{E}} = \sum_{l=1}^{k} f_l = \mathbb{1}_{\mathcal{E}}.$$

Given E_k we associate tubes $B_{l,i} = x_{l,i}B(\epsilon, \tau)$ with $1 \leq i \leq C_l$ so that $E_l \subseteq \bigcup_{i=1}^{C_l} B_{l,i}$. Here $\epsilon = cp^{-2N}$. Finally write $E_{l,i} = E_l \cap x_{l,i}B(\epsilon,\tau)$ and $f_{l,i} = \mathbb{1}_{E_{l,i}}$.

Next we exploit that σ is an arithmetic quantum limit. Indeed we find a joint eigenfunction ϕ_j such that

$$\|\tilde{\phi}_j^{(N_j)} f\|^2 = \langle f, |\tilde{\phi}_j^{(N_j)}|^2 \rangle = \tilde{\mu}_j(\mathcal{E}) > \eta. \tag{19}$$

Let K_N denote the kernel constructed in Lemma 5.5. We consider the quantity

$$Q = \langle K_N(\tilde{\phi}_j^{(N_j)} f), \tilde{\phi}_j^{(N_j)} f \rangle_{L^2(M)}.$$

We will study this correlation from two perspectives.

First we expand f:

$$Q = \sum_{l,l'} \sum_{i,i'} \langle K_N(\tilde{\phi}_j^{(N_j)} f_{l,i}), \tilde{\phi}_j^{(N_j)} f_{l',i'} \rangle_{L^2(M)}.$$

Recall that by the geometric properties of the kernel K_N we have

$$|[K_N(\tilde{\phi}_j^{(N_j)} f_{l,i})](y)| \leq c \cdot p^{-N\delta} \sum_{r=0}^{N} \sum_{z \in S_y(p^r) \cap E_{l,i}} |\tilde{\phi}_j^{(N_j)}(z)|.$$

We obtain

$$\begin{aligned}
&\langle K_N(\tilde{\phi}_j^{(N_j)} f_{l,i}), \tilde{\phi}_j^{(N_j)} f_{l',i'} \rangle_{L^2(M)} \\
&\qquad\qquad \leq c \cdot p^{-N\delta} \int_{E_{l',i'}} \sum_{r \leq N} \sum_{y \in E_{l,i} \cap S_z(p^r)} |\tilde{\phi}_j^{(N_j)}(z)| \cdot |\tilde{\phi}_j^{(N_j)}(y)| d\mu_M(z).
\end{aligned}$$

The r and the y-sum can now be estimated by $O(N)$ using Lemma 7.9. We obtain

$$\langle K_N(\tilde{\phi}_j^{(N_j)} f_{l,i}), \tilde{\phi}_j^{(N_j)} f_{l,i} \rangle_{L^2(M)} \leq c' \cdot Np^{-N\delta} \cdot \|\tilde{\phi}_j^{(N_j)}\|_{L^2(E_{l,i})} \cdot \|\tilde{\phi}_j^{(N_j)}\|_{L^2(E_{l',i'})},$$

where we have used Cauchy-Schwarz. Summing over l, l', i, i' yields

$$Q \leq c' \cdot Np^{-N\delta} \cdot \left(\sum_{l,i} \|\tilde{\phi}_j^{(N_j)}\|_{L^2(E_{l,i})} \right)^2 \leq c' \cdot NKp^{-N\delta}.$$

Second we can apply the spectral decomposition to $\tilde{\phi}_j^{N(N_j)} f$. We get

$$\tilde{\phi}_j^{(N_j)} f = \sum_{\varpi \in \mathrm{ONB}} \langle \tilde{\phi}^{(N_j)} f, \varpi \rangle_{L^2(M)} \varpi.$$

Here ONB stands for some orthonormal basis of joint eigenfunctions of $L^2(M)$. Without loss of generality we can assume that $\tilde{\phi}^{(N_j)} \in$ ONB. We get

$$\begin{aligned} Q &= \sum_{\varpi \in \mathrm{ONB}} |\langle \tilde{\phi}_j^{(N_j)} f, \varpi \rangle_{L^2(M)}|^2 \langle K_N \varpi, \varpi \rangle_{L^2(M)} \\ &> \eta^{-1} \cdot |\langle \tilde{\phi}_j^{(N_j)} f, \tilde{\phi}_j^{(N_j)} \rangle_{L^2(M)}|^2 - \sum_{\substack{\varpi \in \mathrm{ONB}, \\ \varpi \neq \tilde{\phi}_j^{(N_j)}}} |\langle \tilde{\phi}_j^{(N_j)} f, \varpi \rangle_{L^2(M)}|^2. \end{aligned}$$

Here we have used the spectral properties of the kernel K_N. Note that

$$|\langle \tilde{\phi}^{(N_j)} f, \tilde{\phi}_j^{(N_j)} \rangle_{L^2(M)}|^2 = \|\tilde{\phi}_j^{(N_j)} f\|^4.$$

Thus we can use our assumption (19) to estimate

$$\sum_{\substack{\varpi \in \mathrm{ONB}, \\ \varpi \neq \tilde{\phi}_j^{(N_j)}}} |\langle \tilde{\phi}_j^{(N_j)} f, \varpi \rangle_{L^2(M)}|^2 = \|\tilde{\phi}_j^{(N_j)} f\|^2 - \|\tilde{\phi}_j^{(N_j)} f\|^4 < \|\tilde{\phi}_j^{(N_j)} f\|^2 (1-\eta).$$

Inserting this above yields

$$Q > \|\tilde{\phi}_j^{(N_j)} f\|^4 \eta^{-1} - \|\tilde{\phi}_j^{(N_j)} f\|^2 (1-\eta) > \eta^2 \eta^{-1} - \eta(1-\eta) = \eta^2.$$

Combining both estimates yields

$$N p^{-\delta N} k \geq C \cdot \eta^2,$$

for some constant $C > 0$. This can be rewritten as

$$k \geq C \cdot \eta^2 N^{-1} p^{\delta N} \geq p^{\delta' N}$$

for some new $\delta' = \delta'(\eta) > 0$. This completes the proof. $\square$

8. Measure classification and (A)QUE

We will now state the famous measure classification theorem of Lindenstrauss, which is the main input necessary to establish arithmetic quantum unique ergodicity.

Theorem 8.1 (Lindenstrauss). *Let $\Gamma = \Gamma_{\mathcal{O}}$ be a uniform arithmetic lattice over $\mathbb{Q}$ coming from a (maximal) order $\mathcal{O} \subseteq H(\mathbb{Q})$. Let μ be a measure on $M = \Gamma \backslash SL_2(\mathbb{R})$ with the following properties:*

1. *μ is invariant under the geodesic flow;*
2. *There is a prime p such that μ is T_p-recurrent;*
3. *The entropy of every ergodic component of μ is positive for the geodesic flow.*

Then $\mu = \mu_M$ is the uniform measure.

Taking this for granted we can directly obtain the main result of this lecture:

Theorem 8.2 (Brooks-Lindenstrauss 2011)**.** *Let $\Gamma = \Gamma_{\mathcal{O}}$ be a uniform arithmetic lattice over $\mathbb{Q}$ coming from a (maximal) order $\mathcal{O} \subseteq H(\mathbb{Q})$. The uniform measure μ_M is the only arithmetic quantum limit on $M = \Gamma \backslash SL_2(\mathbb{R})$.*

Proof. We simply have to apply the measure classification result of Lindenstrauss to an arithmetic quantum limit σ on M. To do so we note that:

- By Theorem 3.16 the measure σ is invariant under the geodesic flow.
- By Proposition 6.2 the measure σ is T_p-recurrent (for the prime p implicit in the arithmeticity assumption).
- The entropy condition follows from Proposition 7.8.

□

Remark 8.3. Originally Lindenstrauss proved arithmetic quantum unique ergodicty in 2006 for quantum limits coming from sequences of joint eigenfunctions taking Hecke Operators at infinitely many primes into account. See [Li06]. Here we followed an argument of Brooks and Lindenstrauss given in [BrL], which needs only the Hecke operator $\widetilde{T}_p$ at a single prime p.

Remark 8.4. While we have only worked with uniform lattices the measure classification result also applies to non-uniform congruence lattices such as $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. However, in this case one obtains the slightly weaker statement, that the measure is a multiple of the uniform measure μ_M. To show that one gets the uniform measure on the nose one needs to show that no mass escapes into the cusps. This requires an additional argument supplied by Soundararajan in 2009. See [Sou]. Thus arithmetic quantum unique ergodicty is also known in this context.

Remark 8.5. The strategy of Brooks and Lindenstrauss, which we followed here, works for so called quasimodes. More precisely a $w(t)$-quasimode is a function ϕ on $X = M/\mathrm{PSO}_2$ such that

$$\|(\Delta_X + (\frac{1}{4} + t^2))\phi\|_2 \leq tw(t)\|\phi\|_2.$$

The growth of the function $w(\cdot)$ crucially controls how well a quasimode approximates a true eigenfunction. Surprisingly Brooks and Lindenstrauss could show a version of QUE for joint $o(1)$-quasimoodes. (At this approximation level one does not expect equidistribution without the arithmeticity assumption.)

Going into the proof of the measure classification result would go beyond the scope of this lecture. Instead we will try to give some context.

Definition 8.1. Let G denote the real points of a linear algebraic group over $\mathbb{R}$ and suppose that $\Gamma \subseteq G$ is a lattice. We call a measure μ on $\Gamma \backslash G$ homogeneous if there is a closed subgroup $L \subseteq G$ such that μ is the natural L-invariant measure supported on a single closed orbit on L.

Conjecture 8.1 (Furstenberg, Katok-Spatzier, Margulis)**.** *Let A be the subgroup of diagonal matrices in $SL_n(\mathbb{R})$. For $n \geq 3$ any A-invariant ergodic probability measure $SL_n(\mathbb{Z})\backslash SL_n(\mathbb{R})$ is homogeneous.*

Theorem 8.6 (Ratner)**.** *Let $H \subseteq G$ be an algebraic subgroup generated by one parameter unipotent subgroups. Any H-invariant ergodic probability measure is homogeneous.*

Let us return to the setting where $G = \mathrm{SL}_2$. Note that in this case the above conjecture does not apply and this is for good reason. Here as always the (arguably) most important measure is the uniform measure, which comes from the Haar measure on $\mathrm{SL}_2(\mathbb{R})$. One has the following important theorem:

Theorem 8.7 (Bowen)**.** *Let Γ be a uniform lattice and put $M = \Gamma\backslash SL_2(\mathbb{R})$. Denote the time one geodesic flow by T. Then, for $\mu \in \mathcal{M}^T(M)$ we have*

$$h_\mu(T) \leq 1.$$

Furthermore, equality holds if and only if $\mu = \mu_M$.

Remark 8.8. Note that, if one can show that a quantum limit has entropy ≥ 1, then Quantum Unique Ergodicity would follow. However, doing so seems very hard. For comparison, it was shown by Bourgain and Lindenstrauss in [BoL], that (almost every) ergodic component of an arithmetic quantum limit has entropy larger than $\frac{1}{9}$. Note that they use an infinite family of Hecke operators. On the other hand, here we followed Brooks and Lindenstrauss. This allowed us to use only one Hecke Operator at the cost of having no quantitative lower bound on the entropy.

Another natural class of (ergodic) measures are precisely those that are supported on a closed geodesic. (These are homogeneous according to our definition above.) More precisely suppose that $l_x = \Gamma\backslash(\Gamma x A)$ is a periodic geodesic. Let μ_{l_x} denote the corresponding natural A-invariant measure. Then

$$h_{\mu_{l_x}}(T) = 0.$$

In particular, such measures can not occur as ergodic components of arithmetic quantum limits due to Proposition 7.8.

Remark 8.9. Historically scarring on closed geodesics was ruled out by Rudnick and Sarnak in [RS] using ideas related to recurrence. They considered this as evidence for their Quantum Unique Ergodicity conjecture. Even more, their argument may be considered the starting point for the entropy bounds derived by Bourgain und Lindenstrauss.

This is however not the end of the story. There is a big zoo of ergodic measures with positive entropy that are supported on fractal like sets. These can be constructed abstractly using the equivalence of the geodesic flow on a (compact) hyperbolic surface with certain Bernoulli shifts. Showing that, given the condition from Theorem 8.1, these can not appear is the main achievement in the proof of the measure classification theorem. We refer to [EL] for a nice survey.

References

[AA] V. I. Arnold, A. Avez, *Ergodic problems of classical mechanics*, Translated from the French by A. Avez. W. A. Benjamin, Inc., New York-Amsterdam, 1968.

[BoL] J. Bourgain, E. Lindenstrauss, *Entropy of quantum limits*, Comm. Math. Phys. 233 (2003), no. 1, 153–171.

[BrL] S. Brooks, E. Lindenstrauss, *Joint quasimodes, positive entropy, and quantum unique ergodicity*, Invent. Math. 198 (2014), no. 1, 219–259.

[EL] M. Einsiedler, E. Lindenstrauss, *Diagonal actions on locally homogeneous spaces*, Homogeneous flows, moduli spaces and arithmetic, 155–241, Clay Math. Proc., 10, Amer. Math. Soc., Providence, RI, 2010.

[EW] M. Einsiedler, T. Ward, *Ergodic theory with a view towards number theory*, Graduate Texts in Mathematics, 259. Springer-Verlag London, Ltd., London, 2011.

[Kn] A. W. Knapp, *Representation theory of semisimple groups. An overview based on examples*, Reprint of the 1986 original. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 2001.

[Li01] E. Lindenstrauss, *On quantum unique ergodicity for* $\Gamma\backslash\mathbb{H}\times\mathbb{H}$, Internat. Math. Res. Notices 2001, no. 17, 913–933.

[Li06] E. Lindenstrauss, *Invariant measures and arithmetic quantum unique ergodicity*, Ann. of Math. (2) 163 (2006), no. 1, 165–219.

[RS] Z. Rudnick, P. Sarnak, *The behaviour of eigenstates of arithmetic hyperbolic manifolds*, Comm. Math. Phys. 161 (1994), no. 1, 195–213.

[Sar1] P. Sarnak, *Arithmetic quantum chaos*, The Schur lectures (1992) (Tel Aviv), 183–236, Israel Math. Conf. Proc., 8, Bar-Ilan Univ., Ramat Gan, 1995.

[Sar2] P. Sarnak, *Recent progress on the quantum unique ergodicity conjecture*, Bull. Amer. Math. Soc. (N.S.) 48 (2011), no. 2, 211–228.

[Sog] C. D. Sogge, *Hangzhou lectures on eigenfunctions of the Laplacian*, Annals of Mathematics Studies, 188. Princeton University Press, Princeton, NJ, 2014.

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

[Wo] S. A. Wolpert, *Semiclassical limits for the hyperbolic plane*, Duke Math. J. 108 (2001), no. 3, 449–509.

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