SPECTRAL THEORY OF AUTOMORPHIC FORMS AND RELATED PROBLEMS

DIMITRIOS CHATZAKOS

Abstract. In this course we give an introduction to a modern area of analytic number theory. In particular, we will study the analytic theory of the most basic automorphic forms. These will be complex functions defined on the hyperbolic plane $\mathbb{H}$ being invariant under the action of Fuchsian groups $\Gamma \subset SL(2, \mathbb{R})$. After giving a short introduction to the main parts of the theory, we discuss trace formulas and their applications to counting problems on Riemann surfaces and arithmetic quantum chaos.

1. The hyperbolic plane and $GL(2, \mathbb{R})$-automorphic forms

Our goal in this course is to study the most classical example of nonholomorphic automorphic forms, the Maass cusp forms and the nonholomorphic Eisenstein series defined on Riemann surfaces of finite area. We will focus on some important topics of their spectral theory, mainly the spectral theorem and trace formulas. We will also discuss Weyl's and local Weyl's laws for the distribution of Laplace eigenvalues and eigenfunctions. We will apply this knowledge to study counting problems and QUE problems on arithmetic surfaces. Due to lack of time we omit totally some important parts of this beautiful theory, such as the Maass-Selberg relations, bounds for Fourier coefficients of Maass forms, Kloosterman sums, Kuznetsov formula and the sup norm problem. We also sketch only the basic ingredients about Hecke L-functions. Any reader interested to learn more should refer to some of the standard modern references for this area [2], [3], [4] or [5, Ch. 15].

1.1. Intuition coming from the Euclidean case.

The notion of automorphic forms unifies the different objects which look like 'waves' (periodic functions) defined on all different kind of manifolds. The main property of waves is that they correspond to eigenfunctions of a Laplace operator. The easiest such example are the periodic functions $\sin x$ and $\cos x$ on the real line. The main goal of harmonic analysis is to understand the decomposition of sufficiently good (in the L^2-sense) functions in terms of Laplace eigenfunctions.

The prototype of harmonic analysis is the classical Fourier analysis on Euclidean spaces. The n-dimensional Euclidean space

$$\mathbb{R}^n = \{\mathbf{x} = (x_1, x_2, ..., x_n) : x_i \in \mathbb{R}\} \tag{1.1}$$

is endowed with the standard Euclidean metric $ds^2 = dx_1^2 + ... + dx_n^2$ of constant curvature $K = 0$. The Laplace-Beltrami operator associated to ds^2 is

$$\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial x_i^2}. \tag{1.2}$$

Every sufficiently good function on $\mathbb{R}^n$ can be decomposed in terms of Laplace eigenfunctions. The periodic functions

$$\phi_{\mathbf{y}}(\mathbf{x}) = e(\langle \mathbf{x}, \mathbf{y} \rangle) = e^{2\pi i \langle \mathbf{x}, \mathbf{y} \rangle}, \quad \mathbf{y} \in \mathbb{R}^n \tag{1.3}$$

Date: February 22, 2020.

are eigenfunctions of $-\Delta$:

$$(\Delta + \lambda(\mathbf{y}))\phi_{\mathbf{y}}(\mathbf{x}) = 0 \tag{1.4}$$

with $\lambda = \lambda(\mathbf{y}) = 4\pi^2\|\mathbf{y}\|^2$. Then every $f \in \mathrm{L}^2(\mathbb{R}^n)$ has a Fourier integral expansion

$$f(\mathbf{x}) = \int_{\mathbb{R}^n} \hat{f}(\mathbf{y})e(-\langle \mathbf{y}, \mathbf{x}\rangle)d\mathbf{y}, \tag{1.5}$$

where the Fourier transform $\hat{f}$ is given by

$$\hat{f}(\mathbf{y}) = \int_{\mathbb{R}^n} f(\mathbf{x})e(\langle \mathbf{x}, \mathbf{y}\rangle)d\mathbf{x}. \tag{1.6}$$

If we are interested in periodic functions we naturally have to restrict our attention to the n-dimensional *flat torus*, i.e. the quotient space $\mathbb{R}^n/\mathbb{Z}^n$. Fourier analysis decomposes every $h \in \mathrm{L}^2(\mathbb{R}^n/\mathbb{Z}^n)$ in a Fourier expansion

$$h(\mathbf{x}) = \sum_{\mathbf{m}\in\mathbb{Z}^n} a_{\mathbf{m}} e(\langle \mathbf{x}, \mathbf{m}\rangle). \tag{1.7}$$

Now assume $f \in \mathcal{S}(\mathbb{R}^n)$ (the Schwarz space) and consider the integral opeator

$$L_f(g)(\mathbf{x}) = \int_{\mathbb{R}^n} f(\mathbf{x}-\mathbf{y})g(\mathbf{y})d\mathbf{y} = \int_{\mathbb{R}^n/\mathbb{Z}^n} k_f(\mathbf{x},\mathbf{y})g(\mathbf{y})d\mathbf{y} \tag{1.8}$$

where

$$k_f(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{m}\in\mathbb{Z}^n} f(\mathbf{x}-\mathbf{y}+\mathbf{m}). \tag{1.9}$$

The geometric calculation of the trace gives

$$Trace(L_f) = \int_{\mathbb{R}^n/\mathbb{Z}^n} k_f(\mathbf{x},\mathbf{x})d\mathbf{x} = \sum_{\mathbf{m}\in\mathbb{Z}^n} f(\mathbf{m}). \tag{1.10}$$

The Fourier expansion of k_f is given by

$$k_f(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{m}\in\mathbb{Z}^n} \hat{f}(\mathbf{m})e(\langle \mathbf{x},\mathbf{m}\rangle)\overline{e(\langle \mathbf{y},\mathbf{m}\rangle)}, \tag{1.11}$$

so if we calculate the trace of L_f spectrally we get

$$Trace(L_f) = \int_{\mathbb{R}^n/\mathbb{Z}^n} k_f(\mathbf{x},\mathbf{x})d\mathbf{x} = \sum_{\mathbf{m}\in\mathbb{Z}^n} \hat{f}(\mathbf{m}). \tag{1.12}$$

Thus we get the *Poisson formula*:

$$\sum_{\mathbf{m}\in\mathbb{Z}^n} f(\mathbf{m}) = \sum_{\mathbf{m}\in\mathbb{Z}^n} \hat{f}(\mathbf{m}), \tag{1.13}$$

which follows after computing the trace of an integral operator with two different ways, geometrically and spectrally.

The importance of this formula in classical and modern analytic number theory cannot be underestimated, as it has turned out to have numerous applications in various important arithmetic problems. Notice that for $n = 1$ if you apply (1.13) formally for $f(x) = |x|^{-s}$ (actually for a smooth approximation of f) you derive the functional equation of Riemann's zeta function $\zeta(s)$. Moreover it was the historically first example of a *trace formula*. The name is coming from the underlying principle that the proof of (1.13) was given by computing the trace of an operator (here that crucial operator was L_f) in two different ways. As we will see in the following sections, an analogous formula can be proved for hyperbolic surfaces. But this formula turns out to be much more complicated due to the fact that the group of automorphisms of the hyperbolic plane is *nonabelian*.

1.2. **The hyperbolic plane.** We proceed to the main topic of this course starting with a basic introduction to hyperbolic geometry. We will study automorphic forms defined on the hyperbolic plane $\mathbb{H}$, which can be viewed as the complex upper half-plane

$$\mathbb{H} = \{z = x + iy \in \mathbb{C} : y > 0\} = \{z \in \mathbb{C} : \Im(z) > 0\}.$$

Sometimes in the literature the hyperbolic n-space is denoted by $\mathbb{H}^n$ and the hyperbolic plane is denoted by $\mathbb{H}^2$ to indicate the dimension. Since in this course we will restrict our attention in dimension 2, we abbreviate the notation to $\mathbb{H}$. The hyperbolic plane is endowed with the hyperbolic metric

$$ds^2 = \frac{dx^2 + dy^2}{y^2} = \frac{|dz|^2}{\Im(z)^2},$$

i.e., for a path $\gamma = \{z(t) = x(t) + iy(t), t \in [0, 1]\}$ the hyperbolic length is given by

$$\ell(\gamma) = \int_0^1 \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2} \frac{dt}{y(t)}.$$

The distance $\rho(z, w)$ between two points $z, w \in \mathbb{H}$ is now defined as

$$\rho(z, w) = \min \ell(\gamma),$$

where the minimum is considered over all the paths γ from z to w. The distance function can be expressed in various elementary ways. The most useful for calculations usually is

$$\cosh \rho(z, w) = 1 + 2u(z, w),$$

where $u(z, w)$ is the point pair invariant function (we will explain this terminology later)

$$u(z, w) = \frac{|z - w|^2}{4\Im(z)\Im(w)}. \tag{1.14}$$

The hyperbolic metric induces a hyperbolic measure $d\mu$ on $\mathbb{H}$ given by

$$d\mu = \frac{dxdy}{y^2}.$$

We will denote the hyperbolic area of a set $A \subset \mathbb{H}$ by $\mu(A)$ or $\mathrm{vol}(A)$. The space $(\mathbb{H}, ds, d\mu)$ is a Riemannian manifold of constant negative curvature $K = -1$. The hyperbolic lines (geodesics) are represented by half-lines orthogonal to $\mathbb{R}$ and Euclidean semi-circles with center on $\mathbb{R}$. The hyperbolic circles are represented by Euclidean circles (but with different centers). Finally, hyperbolic triangles satisfy the Gauss defect (special case of the Gauss-Bonnet formula).

Theorem 1.1. *The hyperbolic area of a hyperbolic triangle A is given by*

$$\mathrm{vol}(A) = \pi - \alpha - \beta - \gamma,$$

where α, β, γ are the interior angles of A.

1.3. **Möbius transformations and the group of automorphisms.** A Möbius transformation $\gamma : \mathbb{C} \to \mathbb{C}$ is a conformal mapping with

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

where $a, b, c, d \in \mathbb{R}$. We identify the transformation γ with the matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

and we observe that we can obviously assume $\det(\gamma) = 1$. We notice also that γ and $-\gamma$ give the same transformation. An immediate calculation gives

$$\Im(\gamma z) = \frac{\Im(z)}{|cz+d|^2}, \tag{1.15}$$

hence every Möbius transformation acts on $\mathbb{H}$. Thus, if we denote by G the group

$$\mathrm{SL}_2(\mathbb{R}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : (a,b,c,d) \in \mathbb{R}^4, \quad ad - bc = 1 \right\}, \tag{1.16}$$

then the group M of Möbius transformations is isomorphic to the group $\mathrm{PSL}_2(\mathbb{R}) = G/\{\pm I\}$ (see exercise 5). A key fact is that the action of G is transitive and preserves the geometry of $\mathbb{H}$. To see this, notice

$$\gamma z - \gamma w = \frac{z-w}{(cz+d)(cw+d)},$$

for every $\gamma \in G$, hence

$$\frac{d\gamma z}{dz} = \frac{1}{(cz+d)^2}.$$

Using (1.15) we derive

$$\frac{|d\gamma z|}{\Im(\gamma z)} = \frac{|dz|}{\Im(z)}. \tag{1.17}$$

We conclude that Möbius transformations are isometries of the hyperbolic plane, i.e. $\rho(\gamma z, \gamma w) = \rho(z,w)$. In fact M is the group of isometries of $\mathbb{H}$ that preserve orientation. The whole group of isometries of $\mathbb{H}$ is generated by the Möbius transformations and the reflection map $z \to -\overline{z}$.

1.4. **Fuchsian groups and Riemann surfaces.** Since $\mathrm{PSL}_2(\mathbb{R})$ acts on $\mathbb{H}$, so does any subgroup $\Gamma \subset \mathrm{PSL}_2(\mathbb{R})$. However, the orbit space $\Gamma\backslash\mathbb{H}$ may not be sufficiently nice if we don't assume extra conditions on the group Γ. If we impose the condition $\Gamma\backslash\mathbb{H}$ to be a Hausdorff space, then it turns out that Γ must act discontinuously on $\mathbb{H}$, i.e. for any $z \in \mathbb{H}$ the orbit Γz has no limit points inside $\mathbb{H}$ (however the orbit Γz may have limit points in the boundary of $\mathbb{H}$). Poincaré proved the following fundamental theorem which classifies these groups.

Theorem 1.2. *A group Γ' of $SL_2(\mathbb{R})$ is discrete in the norm topology if and only if the projection Γ of Γ' in $PSL_2(\mathbb{R}) \simeq M$ acts discontinuously on $\mathbb{H}$.*

Such a group is called a *Fuchsian group*. Not all of them are sufficiently nice for our purposes, as the quotient space $\Gamma\backslash\mathbb{H}$ can be arbitrarily large. For instance, consider the group $\Gamma = \langle \gamma \rangle$ with

$$\gamma = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

In that case the quotient space has infinite hyperbolic area. However, the most obvious example of a discrete subgroup of $\mathrm{PSL}_2(\mathbb{R})$ is $\mathrm{PSL}_2(\mathbb{Z})$ and in that case the quotient space has finite area.

Definition 1.3. A Fuchsian group Γ is called *cofinite* if the surface $\mathcal{M} = \Gamma\backslash\mathbb{H}$ satisfies

$$\mathrm{vol}(\Gamma\backslash\mathbb{H}) < \infty.$$

Further, if $\Gamma\backslash\mathbb{H}$ is compact then Γ is called *cocompact*. The area of the *modular surface* $\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}$ is equal to $\pi/3$ (see exercise 6) but the surface is not compact. There exists a more refined notion measuring the 'size' of a Fuchsian group in terms of geometry. We do not discuss the detailed definition here, but it turns out every Fuchsian group is either *of the first kind* (i.e. cofinite), *elementary* (i.e. trivial) or *of the second kind* (i.e. not sufficiently large).

The quotient space $\Gamma\backslash\mathbb{H}$ has a structure of a Riemann surface. That means $\Gamma\backslash\mathbb{H}$ is not only a Riemannian manifold (or orbifold, see subsection 1.6) of dimension 2 over $\mathbb{R}$, but it is also a complex manifold of dimension 1 over $\mathbb{C}$. From the point of view of algebraic topology the group Γ is isomorphic to the fundamental group $\pi_1(\mathcal{M})$ of $\mathcal{M} = \Gamma\backslash\mathbb{H}$. The distance function $\rho(z,w)$ in $\mathbb{H}$ induces a distance $\rho_\Gamma(z,w)$ in $\Gamma\backslash\mathbb{H}$ given by

$$\rho_\Gamma(z,w) = \inf_{\gamma\in\Gamma} \rho(\gamma z, w),$$

where each point in $\Gamma\backslash\mathbb{H}$ is identified with one of its representatives in $\mathbb{H}$.

1.5. **Classification of group elements.** An element $\gamma \in \mathrm{PSL}_2(\mathbb{R})$ acts as a rigid motion on the hyperbolic plane and this motion can be understood in terms of some invariants. First, the identity element $\gamma = I$ has trivial action, and we consider it as a special case. Except I, any other element fixes one or two distinct points on $\overline{\mathbb{H}} = \mathbb{H} \cup \mathbb{R} \cup \{i\infty\}$. If

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

and we consider separately the two cases $c = 0$ and $c \neq 0$. In the first case γ has a unique fixed point $z_0 = b/(d-a)$ in $\mathbb{R} \cup \{i\infty\}$. In the second case γ has two fixed points

$$z_{1,2} = \frac{a-d \pm \sqrt{(a+d)^2-4}}{2c} = \frac{a-d \pm \sqrt{\mathrm{tr}\,(\gamma)^2-4}}{2c}.$$

If $|\,\mathrm{tr}\,(\gamma)\,| < 2$ then $z_{1,2}$ are complex conjugates, if $|\,\mathrm{tr}\,(\gamma)\,| = 2$ then $z_1 = z_2 \in \mathbb{R}$ and if $|\,\mathrm{tr}\,(\gamma)\,| > 2$ then γ has two distinct fixed points in $\mathbb{R} \cup \{i\infty\}$. We have the following definition:

Definition 1.4. An element γ of Γ is classified as elliptic, parabolic or hyperbolic if $|\mathrm{tr}(\gamma)| < 2$, $|\mathrm{tr}(\gamma)| = 2$ or $|\mathrm{tr}(\gamma)| > 2$.

An important property of this definition is that it is invariant under conjugation. We can thus split the group Γ in elliptic, parabolic and hyperbolic conjugacy classes

$$\{\gamma\} = \{a\gamma a^{-1} : a \in G\}.$$

Parabolic and hyperbolic motions have infinite order, whereas elliptic motions have finite order. In fact, a parabolic motion acts by translation, as it is conjugate to a transformation of the form

$$z \to \begin{pmatrix} 1 & v \\ 0 & 1 \end{pmatrix} z = z + v \tag{1.18}$$

with $v \in \mathbb{R}$ (which fixes the point at infinity). A hyperbolic motion acts by dilation and it is conjugate to a transformation of the form

$$z \to \begin{pmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2} \end{pmatrix} z = pz \tag{1.19}$$

with $p \in \mathbb{R}$. This transformation fixes the two distinct points 0 and $i\infty$. Finally, an elliptic motion acts by rotation, it is conjugate to a transformation of the form

$$z \to \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} z \tag{1.20}$$

with $\theta \in \mathbb{R}$ (with unique fixed point in $\mathbb{H}$ the point $z = i$). A point $z \in \overline{\mathbb{H}}$ is called a *cusp* for Γ if it is the fixed point of a parabolic $\gamma \in \Gamma$. For example

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z}),$$

hence $i\infty$ is a cusp of the modular group $\mathrm{PSL}_2(\mathbb{Z})$. Further, if $z_{1.2} \in \mathbb{R} \cup \{i\infty\}$ are the two fixed points of a hyperbolic element γ then γ maps the whole geodesic connecting z_1 and z_2 to itseld (but not identically). For example, the transformation (1.19) maps the imaginary axis connecting 0 to $i\infty$ to itself. That defines a 1-1 correspondence between closed geodesics on $\Gamma\backslash\mathbb{H}$ and hyperbolic conjugacy classes of Γ.

For a fixed $z \in \overline{\mathbb{H}}$ and a Fuchsian group Γ the stabilizer $\Gamma_z \subset \Gamma$ is cyclic. An element γ_0 is called primitive if $\langle \gamma_0 \rangle = \Gamma_z$ for every fixed point z of γ_0. Any other $\gamma \neq I$ is a power of a unique primitive element $\gamma = \gamma_0^n$ for some $n \in \mathbb{Z}$.

1.6. **Fundamental domains.** Two points $z, w \in \overline{\mathbb{H}}$ are said to be Γ-equivalent if $w \in \Gamma z$. We have the following definition.

Definition 1.5. A set $\mathcal{F} \subset \mathbb{H}$ is called a fundamental domain for Γ if:
(a) $\mathcal{F}$ is a domain in $\mathbb{H}$,
(b) any two distinct points in $\mathcal{F}$ are not Γ-equivalent,
(c) any orbit of Γ contains a point in $\overline{\mathcal{F}}$.

A fundamental domain for Γ is a model for the quotient space $\Gamma\backslash\mathbb{H}$. The cusps of Γ are on the boundary of $\mathbb{H}$ and there exist only finitely many Γ-inequivalent cusps. They indicate whether a fundamental domain for Γ touches the topological boundary of $\mathbb{H}$. If Γ has elliptic points then $\Gamma\backslash\mathbb{H}$ has conical points and is not smooth (in that case $\Gamma\backslash\mathbb{H}$ is called *orbifold*). A fundamental domain for the modular group is given by the domain $\mathcal{F} = \{z \in \mathbb{H} : |z| > 1, \Re(z) < 1/2\}$, which touches $\overline{\mathbb{H}}$ only at $i\infty$.

Proposition 1.6. *A cofinite group Γ is cocompact if and only if $\Gamma\backslash\mathbb{H}$ does not have cusps, i.e. if and only if Γ does not contain parabolic elements.*

If $\Gamma\backslash\mathbb{H}$ has cusps then for every cusp $\mathfrak{a}$ there exists a matrix $\sigma_{\mathfrak{a}} \in \mathrm{PSL}_2(\mathbb{R})$ such that

$$\sigma_{\mathfrak{a}}\infty = \mathfrak{a}, \quad \sigma_{\mathfrak{a}}^{-1}\Gamma_{\mathfrak{a}}\sigma_{\mathfrak{a}} = \left\{ \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} / \{\pm I\} : n \in \mathbb{Z} \right\},$$

where $\Gamma_{\mathfrak{a}}$ is ths stabilizer of $\mathfrak{a}$ in Γ. Since $\Gamma_{\mathfrak{a}}$ is cyclic we write $\gamma_{\mathfrak{a}}$ to denote a fixed generator of $\Gamma_{\mathfrak{a}}$. For example, for $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ we have

$$\Gamma_{\mathrm{i}\infty} = \left\langle \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \right\rangle.$$

1.7. **Examples of arithmetic groups.** Among the whole zoo of Fuchsian groups, a special case is that of arithmetic groups. We already mentioned the most standard group of this kind, which is the modular group $\mathrm{PSL}_2(\mathbb{Z}) \subset \mathrm{PSL}_2(\mathbb{R})$. This is the projective quotient $\mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$ of the group

$$\mathrm{SL}_2(\mathbb{Z}) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : (a, b, c, d) \in \mathbb{Z}^4, \quad ad - bc = 1 \right\}. \tag{1.21}$$

The modular group is cofinite and the modular surface has a unique cusp at $i\infty$. We will not try here to define rigorously the notion of an arithmetic group. Instead, we will use some of the standard examples of arithmetic groups, such as $\mathrm{SL}_2(\mathbb{Z})$ and its *congruence* subgroups of level $N \geq 1$ defined by

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod N \right\} \subset \mathrm{SL}_2(\mathbb{Z}),$$

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : c \equiv 0 \mod N \right\} \subset \mathrm{SL}_2(\mathbb{Z}),$$

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : c \equiv 0 \mod N, \quad a \equiv d \equiv 1 \mod N \right\} \subset \mathrm{SL}_2(\mathbb{Z}).$$

Any of the groups Γ defined above have finite index $[\mathrm{SL}_2(\mathbb{Z}) : \Gamma]$ in $\mathrm{SL}_2(\mathbb{Z})$ and the quotient space $\Gamma\backslash\mathbb{H}$ has finite hyperbolic area equal to $[\mathrm{SL}_2(\mathbb{Z}) : \Gamma]\,\mathrm{vol}(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H})$. In particular we have

$$[\mathrm{SL}_2(\mathbb{Z}) : \Gamma_0(N)] = N \prod_{p|N}\left(1+\frac{1}{p}\right),$$
$$[\mathrm{SL}_2(\mathbb{Z}) : \Gamma(N)] = N^3 \prod_{p|N}\left(1-\frac{1}{p^2}\right).$$

Hence congruence subgroups are cofinite arithmetic Fuchsian groups.

There exist arithmetic groups which are cocompact. These groups arise from quaternion algebras. For a, b square-free integers with $a > 0$ let

$$H = \left(\frac{a,b}{\mathbb{Q}}\right) \tag{1.22}$$

be the indefinite quaternion algebra over $\mathbb{Q}$ linearly generated by $\{1, \omega, \Omega, \omega\Omega\}$, where $\omega^2 = a$, $\Omega^2 = b$ and $\omega\Omega + \Omega\omega = 0$. We say that H ramifies at p if $H_p = H \otimes_{\mathbb{Q}} \mathbb{Q}_p$ is a division algebra. Let $N(x)$ denote the norm of $x \in H$. For R a maximal order of H, denote by $R(n)$ the elements $x \in R$ of norm n. Denote also by ϕ the embedding that maps

$$x = x_0 + x_1\omega + x_2\Omega + x_3\omega\Omega = \xi + \eta\Omega \tag{1.23}$$

to

$$\phi(x) = \begin{pmatrix} \overline{\xi} & \eta \\ b\overline{\eta} & \xi \end{pmatrix}. \tag{1.24}$$

Then, the group

$$\Gamma_H := \phi(R(1)) \subset \mathrm{SL}_2(\mathbb{R}) \tag{1.25}$$

is a Fuchsian group of the first kind, and the quotient $\Gamma_H\backslash\mathbb{H}$ is a Riemann surface. The discriminant $D = D_H$ of the algebra is defined as the product of the ramified primes of H.

Now let D, N be natural numbers such that: i) D is a product of an even number of different primes and, ii) $N > 1$ is a natural number with $(D, N) = 1$. If $R = \mathcal{O}(D, N)$ is an Eichler order of level N and ϕ a monomorphism : $H \to M_2(\mathbb{R})$, we denote the group Γ_H associated to $\mathcal{O}(D, N)$ and ϕ by $\Gamma(D, N)$. We have

$$\Gamma(D, N) \subset \mathrm{SL}_2(\mathbb{Q}(\sqrt{a})). \tag{1.26}$$

The theory of Shimura provides a canonical model $\mathcal{X}(D, N)$ for $\Gamma(D, N)\backslash\mathbb{H}$ and a modular interpretation. The canonical model $\mathcal{X}(D, N)$ is a projective curve defined over $\mathbb{Q}$. The quotient $\Gamma(D, N)\backslash\mathbb{H}$ is noncompact if $D = 1$ (the nonramified case), and in this case we have $\mathcal{X}(1, N) = \Gamma_0(N)\backslash\mathbb{H}$. If $D > 1$, then $\mathcal{X}(D, N)$ is a compact surface. The curves $\mathcal{X}(D, 1)$, for $D > 1$, can be viewed as the compact analogues of the modular surface $\mathcal{X}(1, 1) = \mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}$.

1.8. **The Laplace-Beltrami operator.** On every Riemannian manifold with arbitrary Riemannian metric

$$ds^2 = \sum_{i,j} g_{i,j}\,dx_i dx_j \tag{1.27}$$

we define a Laplace (or Laplace-Beltrami) operator Δ expressed in local coordinates x_i as

$$\Delta = -\frac{1}{\sqrt{\det(g_{ij})}} \sum_{i,j} \frac{\partial}{\partial x_i} \sqrt{\det(g_{ij})}\, g^{ij} \frac{\partial}{\partial x_j} \tag{1.28}$$

where g^{ij} are the entries of the inverse matrix of (g_{ij}). In the case of the hyperbolic plane the Laplace-Beltrami operator reduces to

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

This Laplacian implies a Laplacian on every surface $\Gamma\backslash\mathbb{H}$ and has a rich and beautiful spectral theory. In order to explain this, we need to consider functions living on the Riemann surfaces.

1.9. **Automorphic functions and automorphic forms.** The theory of automorphic functions had already been investigated since the time of Poincaré and many people had contributed to this, with Hecke being probably the main contributor.

Definition 1.7. A function $f : \mathbb{H} \to \mathbb{C}$ is said to be automorphic with respect to Γ if

$$f(\gamma z) = f(z)$$

for every $\gamma \in \Gamma$.

Clearly, such a function defines a function on the surface $f : \Gamma\backslash\mathbb{H} \to \mathbb{C}$. We denote the space of Γ-automorphic functions by $\mathcal{A}(\Gamma\backslash\mathbb{H})$. A crucial question is whether such functions exist and, if so, how can we construct them. The most obvious way to construct an automorphic function is by the method of images: if $p(z)$ is a function of rapid decay on the hyperbolic plane then

$$f(z) = \sum_{\gamma\in\Gamma} p(\gamma z) \tag{1.29}$$

is automorphic. Another way is averaging not over the whole group but only over cosets of Γ. Assume for a while that Γ is not cocompact and for a fixed cusp $\mathfrak{a}$ consider the element $\sigma_{\mathfrak{a}}$ from (1.6). Then the so called *Poincaré series* defined by

$$E_{\mathfrak{a}}(z|p) = \sum_{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma} p(\sigma_{\mathfrak{a}}^{-1}\gamma z) \tag{1.30}$$

is automorphic for any function $p(z)$ which is $\Gamma_{\mathrm{i}\infty}$-periodic and of moderate growth. For functions defined on the Riemann surface $\Gamma\backslash\mathbb{H}$ we have the *Petersson* inner product given by

$$\langle f, g\rangle = \int_{\Gamma\backslash\mathbb{H}} f(z)\overline{g(z)}\frac{dxdy}{y^2}.$$

The Hilbert space

$$\mathrm{L}^2(\Gamma\backslash\mathbb{H}) = \left\{ f \in \mathcal{A}(\Gamma\backslash\mathbb{H}) : \|f\|_2 := \langle f, f\rangle^{1/2} < \infty \right\} \tag{1.31}$$

will be the main object of interest in our analysis of automorphic forms. The group $\mathrm{PSL}_2(\mathbb{R})$ leaves invariant the Laplace-Beltrami operator in $\mathbb{H}$. That means

$$\Delta f_\gamma(z) = \Delta f(\gamma z)$$

for any fixed $\gamma \in \mathrm{PSL}_2(\mathbb{R})$, where $f_\gamma(z) := f(\gamma z)$. Moreover, any $\mathrm{PSL}_2(\mathbb{R})$-invariant differential operator on $\mathbb{H}$ is a polynomial on Δ with constant coefficients. That means the Laplacian generates the algebra of invariant differential operators.

An application of Stokes' theorem shows that Δ is symmetric with respect to the Petersson inner product in a dense subspace of $\mathrm{L}^2(\Gamma\backslash\mathbb{H})$:

$$\langle f, \Delta g\rangle = \langle \Delta f, g\rangle \tag{1.32}$$

for any f, g in the space

$$\mathcal{D}(\Gamma\backslash\mathbb{H}) := \{ f \in \mathcal{B}(\Gamma\backslash\mathbb{H}) : \Delta f \in \mathcal{B}(\Gamma\backslash\mathbb{H}) \} \tag{1.33}$$

where $\mathcal{B}(\Gamma\backslash\mathbb{H})$ denotes the space of bounded functions on $\Gamma\backslash\mathbb{H}$. One can also directly show that

$$\langle -\Delta f, f\rangle = \int_{\Gamma\backslash\mathbb{H}} |\nabla f|^2 dxdy \geq 0, \tag{1.34}$$

hence $-\Delta$ is non-negative in $\mathcal{D}(\Gamma\backslash\mathbb{H})$.

Analysis' goal in to understand function spaces defined on the manifolds of interest. The main object of harmonic analysis is to understand these function spaces through the spectral resolution of the Laplace-Beltrami operator, i.e. expanding sufficiently good functions in terms of Laplace eigenfunctions. This naturally leads to the definition of automorphic forms, given by Maass on 1949.

Definition 1.8 (Maass)**.** A smooth automorphic function $f \in \mathcal{A}(\Gamma\backslash\mathbb{H})$ that is an eigenfunction of $-\Delta$:

$$(\Delta + \lambda)f = 0$$

is called an automorphic form.

We denote by $\mathcal{A}_s(\Gamma\backslash\mathbb{H})$ the space of eigenfunctions of $-\Delta$ on $\Gamma\backslash\mathbb{H}$ with eigenvalue $\lambda = s(1-s)$. For notational reasons we also write $s = 1/2 + it$, hence $\lambda = 1/4 + t^2$. Notice that $t \in \mathbb{R}$ if and only if $\Re(s) = 1/2$, i.e. if and only if $\lambda \geq 1/4$. The eigenvalues $\lambda < 1/4$ will called *small* (or *exceptional*) eigenvalues of Γ.

To study the spectral resolution of Δ one has to understand the algebra of $\mathrm{PSL}_2(\mathbb{R})$-invariant integral operators on $\mathbb{H}$. These are operators of the form

$$L_k(f)(z) = \int_{\mathbb{H}} k(z,w)f(w)d\mu(w). \tag{1.35}$$

They are natural generalizations of the operator (1.8). When the kernel $k(z,w)$ is *point pair invariant* (meaning $k(\gamma z, \gamma w) = k(z,w)$ for all $\gamma \in \mathrm{PSL}_2(\mathbb{R})$) then $k(z,w)$ depends only on the function $u(z,w)$ defined in (1.14). We can then write $k = k(u)$ and moreover we can write

$$L_k(f)(z) = \int_{\Gamma\backslash\mathbb{H}} K(z,w)f(w)d\mu(w) \tag{1.36}$$

for

$$K(z,w) = \sum_{\gamma\in\Gamma} k(u(\gamma z, w)), \tag{1.37}$$

which is called the *automorphization* of kernel $k(u)$. We will not proceed in the proof of spectral resolution with details, although we will sketch the main arguments of the proof. The integral operators play a crucial role in the proof, but we will also refer to them later, when we will discuss the pre-trace formula and the hyperbolic lattice counting problem. For noncompact surfaces, a crucial part in the proof of the spectral theorem is the analytic continuation of Eisenstein series. We will retun to this topic in section 2.

1.10. **The spectral theorem for compact surfaces.** When the surface $\Gamma\backslash\mathbb{H}$ is compact, the Hilbert-Schmidt theory applies to sufficiently modified integral operators on $\Gamma\backslash\mathbb{H}$. This gives the spectral resolution of Δ in the space $\mathcal{D}(\Gamma\backslash\mathbb{H})$. Applying Friedrichs extension theorem and the spectral theorem for symmetric operators on Hilbert spaces we get a unique self-adjoint extension of the Laplacian on the whole L^2-space. The method of proof works in some generality; the following result holds for any compact Riemannian manifold.

Theorem 1.9. *Let Δ denote the Laplacian obtained as the Friedrichs extension of (1.8) on the $L^2(\Gamma\backslash\mathbb{H})$. Then the operator $-\Delta$ has only discrete spectrum $\{\lambda_j\}_{j=0}^{\infty}$ such that $\lambda_0 = 0$ corresponds to the constant eigenfunction and $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_j \leq \ldots$ with $\lambda_j \to \infty$ as $j \to \infty$.*

We denote by $\{u_j(z)\}_{j=0}^{\infty}$ a set of *Maass forms*, i.e. an orthogonal system of eigenfunctions of the Laplacian:

$$(\Delta + \lambda_j)\, u_j(z) = 0.$$

We will also assume that u_j are L^2-normalized, i.e. $\|u_j\|_2 = 1$. We thus conclude every $f \in \mathrm{L}^2(\Gamma\backslash\mathbb{H})$ has a spectral expansion of the form

$$f(z) = \sum_{\lambda_j \geq 0} \langle f, u_j\rangle u_j(z). \tag{1.38}$$

1.11. **The spectral resolution for noncompact surfaces: Eisenstein series.** Assume now that Γ is cofinite but not cocompact. We can use the Poincaré series from (1.30) to construct automorphic forms. Notice that if $p(z)$ is an eigenfunction of $-\Delta$ then so is $E_{\mathfrak{a}}(z|p)$ but due to convergence issues we have to assume growth conditions for $p(y)$. Choosing $p(y) = y^s$, which is an eigenfunction of $-\Delta$, then the *Eisenstein series* defined by

$$E_{\mathfrak{a}}(z,s) = \sum_{\gamma\in\Gamma_{\mathfrak{a}}\backslash\Gamma} \left(\Im(\sigma_{\mathfrak{a}}^{-1}\gamma z)\right)^s. \tag{1.39}$$

is an automorphic form by construction. As a complex function of $s \in \mathbb{C}$ the series is well-defined for $\Re(s) > 1$ thus it is not in $\mathrm{L}^2(\Gamma\backslash\mathbb{H})$. Notice that this construction is well-defined for every inequivalent cusp $\mathfrak{a}$ of Γ.

For the understanding of the L^2-space the meromorphic continuation of $E_{\mathfrak{a}}(z,s)$ is necessary to the whole complex plane. That was one of the main achievements of Selberg in 1956. We will sketch the proof of this result for the modular surface in the next section. As a summary, we conclude the operator $-\Delta$ has continuous spectrum covering the interval $[1/4,\infty)$ with multiplicity equal to the number of cusps. For a point $\lambda = s(1-s) = 1/4 + t^2 \geq 1/4$ in the continuous spectrum the corresponding eigenfunction is the Eisenstein series evaluated on the critical line:

$$\left(\Delta + \left(1/4 + t^2\right)\right) E_{\mathfrak{a}}\left(z, 1/2 + it\right) = 0.$$

For understanding the spectral theory in $\Gamma\backslash\mathbb{H}$ it is important that the only poles of the Eisenstein series $E_{\mathfrak{a}}(z,s)$ in $\Re(s) > 1/2$ are simple and real, see [3]. The residues are eigenfunctions of Δ in $L^2(\Gamma\backslash\mathbb{H})$. We will return to these residues in the next subsection.

1.12. **The discrete specrtum: Maass cusp forms.** If Γ has cusps, a straightforward calculation shows that an automorphic form in L^2 is orthogonal to the meromorphically continued Eisenstein series $E_{\mathfrak{a}}(z,s)$ if the zero Fourier coefficients in the Fourier expansion around cusp $\mathfrak{a}$ vanishes identically. Thus the eigenfunctions of Δ that are smooth functions in $\Gamma\backslash\mathbb{H}$ and have exponential decay at every cusp $\mathfrak{a}$ of Γ are called Maass *cusp forms*.

Maass cusp forms belong in $\mathrm{L}^2(\Gamma\backslash\mathbb{H})$. In that case, an orthonormal basis $\{u_j(z)\}$ of eigenfunctions for the (possibly infinite) discrete spectrum $\{\lambda_j\}$ consists of the Maass cusp forms and the residues of the Eisenstein series. Hence, for noncompact surfaces we have the spectral decomposition of the L^2-space in Maass cusp forms, residues of Eisenstein series at $\Re(s) > 1/2$ and Eisenstein series:

$$\mathrm{L}^2(\Gamma\backslash\mathbb{H}) = \mathrm{L}^2{}_{cusp}(\Gamma\backslash\mathbb{H}) \oplus \mathrm{L}^2{}_{res.}(\Gamma\backslash\mathbb{H}) \oplus \mathrm{L}^2{}_{Eisen.}(\Gamma\backslash\mathbb{H}). \tag{1.40}$$

Notice that we have included the constant eigenfunction corresponding to $\lambda_0 = 0$ in the cuspidal part.

1.13. **Exercises.** 1) Prove that if $f(\mathbf{x})$ is a distance function then so is $\hat{f}(\mathbf{y})$.

2) Prove the Fourier expansiom (1.11) (i.e. that the Fourier coefficient of $k_f(\mathbf{x},\mathbf{y})$ is indeed equal to $\hat{f}(\mathbf{m})$).

3) Prove that if $z = pi$ with $p > 0$ and $w = i$ then $\rho(z, w) = \log p$. Prove also that if

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

then $2\cosh \rho(\gamma i, i) = a^2 + b^2 + c^2 + d^2$.

4) Prove equation (1.15).

5) Prove that a hyperbolic circle of radius R and centre $z = i$ is also a Euclidean circle, of radius $r = \sinh R$ and centre $z = i \cosh R$.

6) Prove that the Möbius transformations form indeed a group M isomorphic to $\mathrm{PSL}_2(\mathbb{R})$.

7) Prove that the modular group is generated by the matrices

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

(Hint: You can assume $c \geq 0$. For $c = 0$ the matrix is a power of T. For $c \geq 1$ use the Euclidean algorithm on the pair a, c). Using that, prove that the domain $\mathcal{F} = \{z \in \mathbb{H} : |z| > 1, \Re(z) < 1/2\}$ is indeed a fundamental domain for the modular surface.

8) Assuming exercise 7 prove that

$$\mathrm{vol}(\mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}) = \frac{\pi}{3}.$$

9) Let $A \subset \mathbb{H}$ be any set and $\gamma \in \mathrm{PSL}_2(\mathbb{R})$ with $\gamma \neq \pm I$. Prove that for sufficiently small $\epsilon > 0$ we have

$$\mathrm{vol}(\{z \in A : \rho(\gamma z, z) < \epsilon\}) = 0,$$

if (i) γ is hyperbolic or if (ii) γ is parabolic and A is compact.

10) Show that

$$\sum_{(m,n)\in\mathbb{Z}^2\setminus\{0,0\}} \frac{1}{|mz + n|^{2s}}$$

converges absolutely for any $z \in \mathbb{H}$ and for any s with $\Re(s) > 1$.

2. Spectral theory of the hyperbolic Laplacian and Hecke operators

We already discussed a large part of the basic spectral theory on $\mathrm{L}^2(\Gamma\backslash\mathbb{H})$ and the spectral resolution of the Laplace-Beltrami operator. As we mentioned already, a key ingredient in the full spectral theorem is the analytic continuation of the Eisenstein series. In the case of the modular surface this is much simpler than the case of a general cofinite group.

2.1. **Analytic continuation of Eisenstein series.** We discuss the proof of the analytic continuation of the Eisenstein series $E_{\mathfrak{a}}(z, s)$ for $\Re(s) \leq 1$. We will first deal with the simple case $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$, where we have the Eisenstein series corresponding to the unique cusp $i\infty$ of the modular surface. After some analytic manipulation the Eisenstein series takes the form

$$
\begin{aligned}
E(z,s) &= \sum_{\gamma\in\Gamma_{i\infty}\backslash\Gamma} (\Im(\gamma z))^s \\
&= y^s + \sum_{c=1}^{\infty} \sum_{d\in\mathbb{Z},(d,c)=1} \frac{y^s}{|cz+d|^{2s}} \\
&= y^s + \phi(s)y^{1-s} + \sum_{n\neq 0} \phi(n,s) K_{s-1/2}(2\pi n y) e(nx)
\end{aligned} \tag{2.1}
$$

(this expression has to be slightly modified for $s = 1/2$). Here $K_\nu(x)$ stands for the modified Bessel function of the second kind and of order ν. This is the Fourier expansion of $E(z,s)$. The Fourier coefficient $\phi(s)$ is called the *scattering determinant* and is given by the quantity

$$
\phi(s) = \frac{\Lambda(2-2s)}{\Lambda(2s)}, \quad \Lambda(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s). \tag{2.2}
$$

The coefficients $\phi(n,s)$ are also given in terms of arithmetic functions:

$$
\phi(n,s) = \frac{\pi^s}{\Gamma(s)\zeta(2s)\sqrt{|n|}} \sum_{ab=|n|} \left(\frac{a}{b}\right)^{s-1/2}. \tag{2.3}
$$

The invariance of $\phi(s)$ under the change of variable $s \to 1-s$ is clear and the same holds for $\phi(n,s)$, although this is not so clear from (2.3). These properties imply the functional equation for $E(z,s)$; in particular they give

$$
E(z,s) = \phi(s)E(z,1-s). \tag{2.4}
$$

Examining poles and residues we verify that $\phi(s)$ is holomorphic for $\Re(s) \geq 1/2$ except one simple pole at $s = 1$. This pole has residue the constant function

$$
\frac{1}{2\Lambda(2)} = \frac{1}{\mathrm{vol}(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H})} = \frac{3}{\pi}.
$$

For a general cofinite group, the proof is much more sophisticated. A proof of Selberg using Fredholm theory of integral equations can be found in [3]. An important fact arising in the proof is that the functional equation of the Eisenstein series at the inequivalent cusps is related to *all* the Eisenstein series simultaneously; that means one has to treat the family of Eisenstein series $\mathcal{E}(z,s) = (E_{\mathfrak{a}}(z,s))_{\mathfrak{a}}$ as a single vector. The functional equation then reads

$$
\mathcal{E}(z,s) = \Phi(s)\mathcal{E}(z,1-s). \tag{2.5}
$$

The matrix $\Phi(s)$ is called the scattering matrix and determinant $\det\Phi(s) = \phi(s)$ is the scattering determinant.

2.2. **The Spectral theorem and the pre-trace formula.** We wrote the spectral decomposition of $\mathrm{L}^2(\Gamma\backslash\mathbb{H})$ in subsections (1.10) and (1.12). We now write down the spectral theorem explicitly in the following form:

Theorem 2.1 (Spectral theorem in $L^2(\Gamma\backslash\mathbb{H})$ (Selberg , Huber , Roelcke))**.** *Every function $f \in L^2(\Gamma\backslash\mathbb{H})$ has a spectral expansion*

$$
f(z) = \sum_{\lambda_j\geq 0} \langle f, u_j\rangle u_j(z) + \frac{1}{4\pi}\sum_{\mathfrak{a}} \int_{-\infty}^{\infty} \langle f, E_{\mathfrak{a}}(\cdot, 1/2+it)\rangle E_{\mathfrak{a}}(z,1/2+it)dt, \tag{2.6}
$$

which converges in the norm topology. Here, the second sum is over all the (finitely many) inequivalent cusps $\mathfrak{a}$ of Γ, if Γ contains parabolic elements. If $f(z)$ belongs to the domain $D(\Gamma\backslash\mathbb{H})$ of functions $f \in \mathcal{A}(\Gamma\backslash\mathbb{H})$ such that f and Δf are smooth and bounded, then the expansion (2.6) converges pointwise absolutely and uniformly on compact sets.

Around the same time this theorem was already known independently to Roelcke (and later Huber) but in the weaker version of the cocompact case (as they were missing the analytic continuation of Eisenstein series).

There is a special consequence of the spectral theorem we are particularly interested in. This is the spectral theorem for automorphic kernels, the so-called *pre-trace formula*. As in (1.35) let L_k be an integral operator in $\mathbb{H}$ defined by

$$L_k(f)(z) = \int_{\mathbb{H}} k(z,w) f(w) d\mu(w), \tag{2.7}$$

where $k : \mathbb{H} \times \mathbb{H} \to \mathbb{C}$ is the *kernel* of L and the functions k, f are such that the integral converges absolutely. The operator L is $\mathrm{SL}_2(\mathbb{R})$-invariant if and only if k is point-pair invariant kernel and thus if and only if k depends only on $u(z,w)$ given by formula (1.14).

Assume that $k(u)$ is smooth enough. The invariant integral operators commute with the Laplace operator. Hence, an eigenfunction of Δ in $\mathbb{H}$ is also an eigenfunction for all invariant integral operators. If

$$(\Delta + \lambda) f(z) = 0,$$

for $\lambda = 1/4 + t^2$ then we will have

$$\int_{\mathbb{H}} k(u(z,w)) f(w) d\mu(w) = h(t) f(z) \tag{2.8}$$

for some function $h(t)$ depending only on k. This function $h(t)$ is called the *Selberg/Harish-Chandra transform* of the kernel $k(u)$. For given $k(u)$ the transform $h(t)$ is the Fourier transform of the Abel transform of k; more precisely it can be computed in three steps by the formulas:

$$\begin{aligned} q(v) &= \int_v^{+\infty} \frac{k(u)}{\sqrt{u-v}} du, \\ g(r) &= 2q\left(\left(\sinh\frac{r}{2}\right)^2\right), \\ h(t) &= \int_{-\infty}^{+\infty} e^{irt} g(r) dr. \end{aligned} \tag{2.9}$$

The Selberg/Harish-Chandra transform is an even function of t. The required smoothness of $k(u)$ can be expressed in terms of $h(t)$: it must be holomorphic in the strip $|\Im t| \le 1/2 + \epsilon$ for an $\epsilon > 0$ and it must satisfy the bound

$$h(t) \ll (|t| + 1)^{-2-\epsilon} \tag{2.10}$$

inside this strip. Under these assumptions we can invert the process and compute $k(u)$ for a given $h(t)$:

$$\begin{aligned} g(r) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{irt} h(t) dt, \\ q(v) &= \frac{1}{2} g\left(2\log\left(\sqrt{v+1} + \sqrt{v}\right)\right), \\ k(u) &= -\frac{1}{\pi} \int_u^{+\infty} \frac{1}{\sqrt{v-u}} dq(v). \end{aligned} \tag{2.11}$$

The proof of relations (2.9) is rather elementary, since $h(t)$ in (2.8) does not depend on the test function f and the point z. We have thus freedom to choose the function and the point that bring (2.8) in a simpler form. Picking $fw) = (\Im(w))^s$ (with $s = 1/2 + it$) and $z = i$ we get

$$
\begin{aligned}
h(t) &= \int_{\mathbb{H}} k(u(i,w))(\Im(w))^s d\mu(w) \\
&= 2\int_0^\infty \int_0^\infty k\left(\frac{x^2+(y-1)^2}{4y}\right) y^{s-2} dx dy \qquad (2.12)
\end{aligned}
$$

and using the change of variables $x =$ and $y = e^r$ we derive (2.9).

We thus get some flexibility in choosing appropriately $k(u)$ or $h(t)$. This flexibility becomes even more obvious in the pre-trace formula giving the spectral expansion of the automorphization kernel $K(z,w)$. If we restrict the domain of the operator L_k to Γ-automorphic functions f we get

$$(Lf)(z) = \int_{\Gamma\backslash\mathbb{H}} K(z,w) f(w) d\mu(w), \qquad (2.13)$$

where $K(z,w)$ is the automorphic kernel given by (1.37). We view $K(z,w)$ as a function of z. Then, we compute the Fourier coefficients of K, which are given by

$$
\begin{aligned}
\langle K(\cdot,w), u_j\rangle &= h(t_j)\overline{u_j(w)}, \qquad (2.14)\\
\langle K(\cdot,w), E_{\mathfrak{a}}(\cdot,1/2+it)\rangle &= h(t)\overline{E_{\mathfrak{a}}(w,1/2+it)}.
\end{aligned}
$$

The spectral theorem implies that the automorphic kernel K has the following spectral expansion.

Theorem 2.2 (Pre-trace formula)**.** *Assume the pair $k(u)$ and $h(t)$ is related by equations (2.9) and $h(t)$ satisfies (2.10). Then the automorphic kernel given by (1.37) has the spectral expansion*

$$
\begin{aligned}
K(z,w) &= \sum_j h(t_j) u_j(z)\overline{u_j(w)} \\
&+ \frac{1}{4\pi}\sum_{\mathfrak{a}} \int_{-\infty}^{\infty} h(t) E_{\mathfrak{a}}(z,1/2+it)\overline{E_{\mathfrak{a}}(w,1/2+it)} dt, \qquad (2.15)
\end{aligned}
$$

which converges absolutely and uniformly on compact sets.

An automorphic kernel $K(z,w)$ that is absolutely integrable on the diagonal $z = w$ is said to be *of trace class*. For those kernels one can go further and deduce the *Selberg trace formula*, which relates the spectrum of the Laplacian with the *length spectrum* of $\Gamma\backslash\mathbb{H}$. An immediate application of the Selberg trace formula is Weyl's law which, roughly speaking, counts the size of the spectrum up to height T. For instance, in the simple case that Γ is cocompact Weyl's law states that, as $T \to \infty$, we have the asymptotic formula

$$\#\{j : |t_j| \le T\} \sim \frac{\mathrm{vol}(\Gamma\backslash\mathbb{H})}{4\pi} T^2. \qquad (2.16)$$

There are specific arithmetic groups for which we know a stronger form of Weyl's law. Selberg has proved that for congruence groups we have

$$\#\{j : |t_j| \le T\} = \frac{\mathrm{vol}(\Gamma\backslash\mathbb{H})}{4\pi} T^2 + O(T\log T). \qquad (2.17)$$

We will give a proof of (2.16) in the next section as an application of Selberg's trace formula.

2.3. **Local Weyl's law.** For many applications the following local version of Weyl's law is sufficient since it implies an average bound for automorphic forms in the spectral limit.

Theorem 2.3 (Local Weyl's law)**.** *For every z, as $T \to \infty$, we have the asymptotic*

$$\sum_{|t_j|<T} |u_j(z)|^2 + \sum_{\mathfrak{a}} \frac{1}{4\pi} \int_{-T}^{T} |E_{\mathfrak{a}}(z,1/2+it)|^2 dt \sim cT^2, \qquad (2.18)$$

where $c = c(z)$ depends only on the number of elements of Γ fixing z.

Local Weyl's law can be proved using the automorphic heat kernel and a Tauberian argument. When z remains in a bounded region of $\mathbb{H}$ the constant $c(z)$ is uniformly bounded, depending only on Γ. For z not remaining in a compact set, instead of the asymptotic (2.18) one can refer to the following Bessel's inequality, where the second term depends on the height function of z:

$$\sum_{|t_j|<T} |u_j(z)|^2 + \sum_{\mathfrak{a}} \int_{-T}^{T} |E_{\mathfrak{a}}(z, 1/2+it)|^2 dt \ll T^2 + T y_\Gamma(z). \tag{2.19}$$

The height function $y_\Gamma(z)$ is defined as

$$y_\Gamma(z) = \max_{\mathfrak{a}} \max_{\gamma\in\Gamma} \{\Im(\sigma_{\mathfrak{a}}^{-1}\gamma z)\}. \tag{2.20}$$

This function is bounded for compact surfaces but increases as y escapes close to the cusps.

2.4. **Hecke operators and Hecke eigenvalues.** If λ_j is an eigenvalue of the cofinite group Γ then the Maass form u_j attached to λ_j has a Fourier expansion analogous to that of Eisenstein series (2.1). This Fourier expansion around the cusp $\mathfrak{a}$ is of the form

$$u_j(\sigma_{\mathfrak{a}} z) = \rho_{\mathfrak{a}j}(0) y^{1-s_j} + 2\sum_{n\neq 0} y^{1/2} \rho_{\mathfrak{a}j}(n) K_{it_j}(2\pi|n|y) e^{2\pi i n z}. \tag{2.21}$$

The zero coefficient vanishes if u_j is a Maass cusp form. Otherwise u_j is a linear combination of Eisenstein series at $s_j > 1/2$. The tail of the series behaves like an expansion in exponentials since the K-Bessel function finally smooths out. The coefficients $\rho_{\mathfrak{a}j}(n)$ determine the Maass form. What is perhaps the most important feature of the theory is that when the group is arithmetic then a natural commutative algebra of Hecke operators exists. These operators help us with the study of the spectral theory in that case. As in the case of classical modular forms, we can choose a basis of eigenfunctions in our function space which diagonalizes the Hecke operators. The coefficients $\rho_{\mathfrak{a}j}(n)$ in that case are related to the eigenvalues of the Hecke operators. To what follows we simplify the notation working with the subgroup $\Gamma_0(N)$ and the cusp $\mathfrak{a} = i\infty$.

The n-th Hecke operator T_n is a sum over cosets of the set

$$\Gamma_n = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : (a,b,c,d) \in \mathbb{Z}^4, \quad ad-bc = n \right\}. \tag{2.22}$$

Notice that $\Gamma_1 = \mathrm{SL}_2(\mathbb{Z})$. For $n \in \mathbb{N}$, the operator $T_n : \mathcal{A}(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}) \to \mathcal{A}(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H})$ is defined by

$$\begin{aligned} T_n(f)(z) \quad &:= \frac{1}{\sqrt{n}} \sum_{\gamma\in\Gamma_1\backslash\Gamma_n} f(\tau z) \\ &= \frac{1}{\sqrt{n}} \sum_{ad=n} \sum_{b(\mod d)} f\left(\frac{az+b}{d}\right). \end{aligned} \tag{2.23}$$

The n-th Hecke operator is bounded on $\mathrm{L}^2(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H})$, self-adjoint and commutes with the Laplacian Δ (we can say that Δ plays the role of an infinite analogue operator T_∞) and further every two Hecke operators commute. Hence we can choose a joint orthonormal basis u_j of Maass forms (we call them Hecke-Maass forms). We denote by $\lambda_j(n)$ the eigenvalue of T_n for $u_j(z)$, i.e.

$$T_n u_j(z) = \lambda_j(n) u_j(z), \tag{2.24}$$

and $\eta_t(n)$ for the Eisenstein series, i.e.

$$T_n E_\infty(z, 1/2+it) = \eta_t(n) E_\infty(z, 1/2+it), \tag{2.25}$$

where

$$\eta_t(n) = \sum_{ad=n} \left(\frac{a}{d}\right)^{it}.$$

For level $N \geq 1$ we have $\mathcal{A}(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}) \subset \mathcal{A}(\Gamma_0(N)\backslash\mathbb{H})$. Every Hecke operator T_n acts on $\mathcal{A}(\Gamma_0(N)\backslash\mathbb{H})$, but the important cases are those with $(n, N) = 1$; in that case we can similarly pick common eigenfunctions satisfying (2.24), (2.25).

The multiplicativity relations of Hecke operators (see exercise 5) pass to their eigenvalues $\lambda_j(n)$. They are *Hecke multiplicative*:

$$\lambda_j(m)\lambda_j(n) = \sum_{d|(m,n)} \lambda_j\left(\frac{mn}{d^2}\right).$$

This notion includes the usual definition of multiplicative functions and much more; it allows us to define L-functions attached to the coefficients $\lambda_j(n)$ having good Euler products. This will be discussed in subsection 4.1.

2.5. **Exercises.** 1) If $k(u)$ is the characteristic function of the interval $[0, X]$ prove that the Selberg/Harish-Chandra transform $h_X(t)$ behaves asymptotically as

$$h_X(t) \sim |t|^{-3/2} X^{1/2+it}. \tag{2.26}$$

as $X, |t| \to \infty$.

2) Prove relations (2.14).

3) Prove that the index of Γ_n in the modular group is equal to

$$[\mathrm{SL}_2(\mathbb{Z}) : \Gamma_n] = \sum_{d|n} d = \sigma(n).$$

Conclude that the norm of T_n in $\mathrm{L}^2(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H})$ satisfies

$$\|T_n\| \leq \frac{\sigma(n)}{\sqrt{n}}.$$

4) After choosing a set of specific representatives of $\Gamma_1\backslash\Gamma_n$ prove the second equation in (2.23)

5) Prove that

$$T_m T_n = \sum_{d|(m,n)} T_{\frac{mn}{d^2}},$$

hence $T_m T_n = T_n T_m$ for every m, n.

2.6. **Open problems.** Here we state some interesting and extremely hard open problems which still play a key role in the progress of modern research.

1) Selberg's eigenvalue conjecture: is it true that $\lambda_1(\Gamma(N)) > 1/4$? We don't know that, but Selberg proved that $\lambda_1(\Gamma(N)) > 3/16$ for every $N \geq 1$. Since then this result has been improved in various directions.

2) The Sarnak-Phillips theory: for a generic cofinite group Γ, there are few (possibly finite) Maass forms. This conjecture is interesting, since it predicts that the general case is in contrast with the case of arithmetic groups.

3) Ramanujan conjecture: for $(n, N) = 1$ we have $\lambda_j(n) \ll n^{\epsilon}$. This conjecture is proved in the case of holomorphic modular forms (by Deligne) but the best current bound for Maass forms is $\lambda_j(n) \ll n^{7/64+\epsilon}$. This conjecture is quite similar to the Selberg eigenvalue conjecture.

3. An introduction to Trace formulas and applications

3.1. **Intuition from the Trace formula in the Euclidean case.** Selberg trace formula was historically the first noncommutative trace formula. To understand this, we return to the Euclidean harmonic analysis. In the Euclidean space $\mathbb{R}^n$ the trace formula (1.13) has a very simple form (it is actually symmetric) and was proved computing the trace of an invariant integral operator in two different ways, geometrically and spectrally. A similar analysis in our case indicates that the negative curvature makes the geometric side much more difficult. To do the computations for $\Gamma\backslash\mathbb{H}$ one has to consider the contribution of different conjugacy classes of Γ in the geometric side; when the group of automorphisms is abelian these computations are trivial (as in the Euclidean case). However, we know that the Fuchsian groups are far away from being abelian, and the rich family of their conjugacy classes forces us to treat each of them separately.

When the cofinite group Γ is not cocompact, an extra difficulty enters the game: in this case the trace of the integral operator K diverges. One has to truncate the fundamental domain $\mathcal{F}$ of $\Gamma\backslash\mathbb{H}$ and compute the trace of the operator on the *central part* of $\mathcal{F}$. This truncation can be measured by a height Y (indicating how much of the domain we have cut). The trace of K in the geometric side diverges like $A_1 \log Y + B_1$ and in the spectral side like $A_2 \log Y + B_2$. Not very surprisingly the equation $A_1 = A_2$ is a tautology and in that case the trace formula is the equality $B_1 = B_2$.

3.2. **Preliminary estimates.** Assume that k is a sufficiently smooth kernel and assume for a moment that k is compactly supported. Consider the kernel from (1.35):

$$L_k(f)(z) = \int_{\mathbb{H}} k(z,w) f(w) d\mu(w). \tag{3.1}$$

Then for K as in (1.37):

$$K(z,w) = \sum_{\gamma\in\Gamma} k(u(\gamma z, w)), \tag{3.2}$$

we get

$$L_k(f)(z) = \int_{\Gamma\backslash\mathbb{H}} K(z,w) f(w) d\mu(w) \tag{3.3}$$

The trace of $L = L_k$ can be computed in two ways: geometrically it is equal to

$$Trace(L_k) = \int_{\Gamma\backslash\mathbb{H}} K(z,z) d\mu(z). \tag{3.4}$$

Spectrally, if we assume for a moment that Γ is cocompact, then using the pre-trace formula we get the expansion

$$K(z,z) = \sum_j h(t_j) |u_j(z)|^2 \tag{3.5}$$

thus we get directly

$$Trace(L_k) = \int_{\Gamma\backslash\mathbb{H}} K(z,z) d\mu(z) = \sum_j h(t_j). \tag{3.6}$$

The other option that we have is to integrate the geometric quantity (3.2). In that case we get

$$Trace(L_k) = \sum_{\gamma\in\Gamma} \int_{\Gamma\backslash\mathbb{H}} k(\gamma z, z) d\mu(z). \tag{3.7}$$

It was Selberg who first understood how one can proceed with (3.7) in order to get an explicit expression for the trace of L_k. His idea was to consider the partition of Γ into conjugacy classes and compute each of them separately. If we denote by $\mathcal{C} = \{g\}$ a conjugacy class of Γ, then we can write

$$\begin{aligned} Trace(L_k) &= \sum_{\mathcal{C}} \sum_{\gamma \in \mathcal{C}} \int_{\Gamma \backslash \mathbb{H}} k(\gamma z, z) d\mu(z) \\ &= \sum_{\mathcal{C}=\{g\}} \sum_{\tau g \tau^{-1} : \tau \in \Gamma} \int_{\Gamma \backslash \mathbb{H}} k(\tau g \tau^{-1} z, z) d\mu(z). \end{aligned} \tag{3.8}$$

Notice that two elements $\tau_1, \tau_2 \in \Gamma$ give the same element in the inner sum if and only if

$$\tau_1 g \tau_1^{-1} = \tau_2 g \tau_2^{-1} \iff \tau_2^{-1} \tau_1 g = g \tau_2^{-1} \tau_1 \iff \tau_2^{-1} \tau_1 \in Z(g), \tag{3.9}$$

where the set $Z(\gamma)$ is the *centralizer* of an element $\gamma \in \Gamma$ defined by

$$Z(\gamma) = \{\gamma' \in \Gamma : \gamma' \gamma = \gamma \gamma'\}.$$

Thus we can write

$$\begin{aligned} Trace(L_k) &= \sum_{\mathcal{C}=\{g\}} \sum_{\tau \in Z(g) \backslash \Gamma} \int_{\Gamma \backslash \mathbb{H}} k(\tau g \tau^{-1} z, z) d\mu(z) \\ &= \sum_{\mathcal{C}=\{g\}} \int_{Z(g) \backslash \mathbb{H}} k(gz, z) d\mu(z) \end{aligned} \tag{3.10}$$

where in the last line we unfolded the integral.

The importance of this computation is great because the quotient space $Z(g)\backslash\mathbb{H}$ is a much simpler domain in $\mathbb{H}$ and allows us to compute the integral directly. Of course there are some serious calculations that have to be done in order to arrive to the final version of Selberg's trace formula, but the main idea of the proof was the split in conjugacy classes. We also have to notice that the so-called *orbital integral* inside the final sum in (3.10) depends only the conjugacy class of g in the group $\mathrm{SL}_2(\mathbb{R})$. That allows us to pick the elements of Γ, perhaps after conjugating the group, in one of the special subgroups

$$\begin{aligned} N &= \left\{ \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix} : x \in \mathbb{R} \right\}, \\ A &= \left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}, a \in \mathbb{R}^+ \right\}, \\ K &= \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \theta \in \mathbb{R} \right\}. \end{aligned} \tag{3.11}$$

3.3. **Some calculations.** Let's look first at the spectral side. If Γ is cocompact then the spectral side is just (3.6). If we have also continuous spectrum, then we have to truncate the fundamental domain up to height Y, so that we stay away from the cusps. We denote this domain $\mathcal{F}_Y$ and we need to calculate

$$\begin{aligned} &\sum_j h(t_j) \int_{\mathcal{F}_Y} |u_j(z)|^2 d\mu(z) \\ &+ \frac{1}{4\pi} \sum_{\mathfrak{a}} \int_{-\infty}^{\infty} h(t) \int_{\mathcal{F}_Y} |E_{\mathfrak{a}}(z, 1/2 + it)|^2 dt. \end{aligned} \tag{3.12}$$

The discrete spectrum remains the same up to an error term $O(Y^{-\epsilon})$ and at the end we will leave $Y \to \infty$. The continuous contribution needs a much more technical computation. It finally takes the form

$$-\frac{1}{4\pi} \int_{-\infty}^{\infty} \frac{\phi'(1/2+it)}{\phi(1/2+it)} h(t) dt + \frac{h(0)}{4} \operatorname{tr}(\Phi)(1/2) + g(0) h \log Y + O(Y^{-1}), \tag{3.13}$$

where g is the Fourier transform of $h(t)$, Φ is the scattering matrix, h is the rank of this matrix and $\phi(s)$ is the scattering determinant. Keep in mind that the spectral side is the easy side of the trace formula!

For the geometric side, the easy part if the contribution of the identity class which is just equal to

$$\int_{\Gamma\backslash\mathbb{H}} k(z,z)d\mu(z) = k(0)\,\mathrm{vol}(\Gamma\backslash\mathbb{H}) = \frac{\mathrm{vol}(\Gamma\backslash\mathbb{H})}{4\pi}\int_{-\infty}^{\infty} t\tanh(\pi t)h(t)dt. \tag{3.14}$$

. The rest of the computations become much more technical and a full proof of Selberg's trace formula remains out of the goals of this course. However, we can state the trace formula for reasons of completion.

Theorem 3.1 (Selberg Trace Formula, 1956)**.** *Assume that $h(t)$ is an even function which is holomorphic in the strip $|\Im t| \le 1/2+\epsilon$ for some $\epsilon > 0$ and satisfies the bound*

$$h(t) \ll (|t|+1)^{-2-\epsilon} \tag{3.15}$$

inside the strip. Let g be the Fourier transform of h. Then

$$\begin{aligned}
&\sum_j h(t_j) - \frac{1}{4\pi}\int_{-\infty}^{\infty}\frac{\phi'(1/2+it)}{\phi(1/2+it)}h(t)dt \\
&= \frac{\mathrm{vol}(\Gamma\backslash\mathbb{H})}{4\pi}\int_{-\infty}^{\infty} t\tanh(\pi t)h(t)dt \\
&+\sum_{\{P\}}\sum_{\ell=1}^{\infty}\frac{g(\ell\log p)\log p}{p^{\ell/2}-p^{-\ell/2}} \\
&+\sum_{\{R\}}\sum_{\ell=1}^{m-1}(2m\sin(\pi\ell/m))^{-1}\int_{-\infty}^{\infty}h(t)\frac{\cosh\pi(1-2\ell/m)t}{\cosh\pi t}dt \\
&+\frac{h(0)}{4}\left(\mathrm{tr}\,(I) - tr\Phi(1/2)\right) - g(0)h\log 2 - \frac{h}{2\pi}\int_{-\infty}^{\infty}h(t)\psi(1+it)dt.
\end{aligned} \tag{3.16}$$

The left hand contains information about the spectrum of the Riemann surface $\Gamma\backslash\mathbb{H}$ whereas the right side contains information from the conjugacy classes of the Fuchsian group. The contribution of the parabolic classes is known as the length spectrum of Γ.

The Selberg trace formula is one of the most powerful tools in the spectral theory of automorphic forms. It gives an explicit relationship between the geometric information attached to a group and spectral data of the Laplace operator. The first corollary of Selberg's trace formula is Huber's theorem which explains this interplay between the Laplace spectrum and the length spectrum.

Theorem 3.2 (Huber)**.** *For $\Gamma\backslash\mathbb{H}$ smooth and compact the eigenvalue spectrum and the lentgh spectrum (considered with multiplicities) determine each other.*

Proof. In that case the group has only hyperbolic elements (and the identity element) hence the trace formula takes the form

$$\sum_j h(t_j) = \frac{\mathrm{vol}(\Gamma\backslash\mathbb{H})}{4\pi}\int_{-\infty}^{\infty} t\tanh(\pi t)h(t)dt + \sum_{\{P\}}\sum_{\ell=1}^{\infty}\frac{g(\ell\log p)\log p}{p^{\ell/2}-p^{-\ell/2}}. \tag{3.17}$$

The statement follows from the flexibility in choosing the pair (h,g). □

3.4. **Application to Weyl's law.** The first application Selberg gave for his formula was Weyl's law:

Theorem 3.3. *Let* Γ *be a cofinite group. Then*

$$\#\{j : |t_j| \leq T\} - \frac{1}{4\pi}\int_{-\infty}^{\infty} \frac{\phi'(1/2+it)}{\phi(1/2+it)} dt \sim \frac{vol(\Gamma\backslash\mathbb{H})}{4\pi} T^2. \tag{3.18}$$

Proof. Pick the test function $h(t) = e^{\delta t^2}$. The Fourier transform is $g(x) = (4\pi\delta)^{-1/2} e^{-x^2/4\delta}$. Here δ is a small positive parameter. Then, the hyperbolic and the elliptic elements contribute in the trace formula a bounded quantity. The identity element contributes

$$\frac{\mathrm{vol}(\Gamma\backslash\mathbb{H})}{4\pi}\delta^{-1} \tag{3.19}$$

and the parabolic elements contribute at most $O(\delta^{-1/2}\log\delta)$. The result follows using a Tauberian argument. ☐

3.5. **Application to prime geodesics counting.** The second major application of trace formula is the Prime geodesic theorem. This problem asks to count how many primitive closed geodesic on the Riemann surface $\Gamma\backslash\mathbb{H}$ (or general hyperbolic manifolds) have length $\leq \log X$. This problem has great similarities with the Prime number theorem counting the number of prime numbers up to a fixed height X. Huber and Selberg first discovered that the number of such closed geodesics behaves asymptotically like $X/\log X$. This is not just a coincidence since the general set-up of the two problems have striking similarities.

In the theory of prime numbers the Prime number theorem states

$$\psi(X) = \sum_{p^k \leq X} \log p \sim X. \tag{3.20}$$

The growth of the error term $E(X) = \psi(X) - X$ is related to the Riemann Hypothesis. The size of the error is governed by the exponential sum

$$E(X) \sim \sum_{\rho} \frac{X^\rho}{\rho}, \tag{3.21}$$

where the sum runs over all the nontrivial zeros ρ of the Riemann zeta function $\zeta(s)$; the conjectural bound $E(X) = O_\epsilon(X^{1/2+\epsilon})$ is equivalent to RH.

We will state the Prime geodesic theorem on $\Gamma\backslash\mathbb{H}$ in terms of the Chebyshev function

$$\psi_\Gamma(X) = \sum_{\{P\}: N(P) \leq X} \Lambda_\Gamma(N(P)), \tag{3.22}$$

where the sum in (3.22) is taken over the hyperbolic classes of Γ, $N(P)$ denotes the norm of the hyperbolic element P and the von Mangoldt function is given by $\Lambda_\Gamma(N(P)) = \log N(P_0)$ if P is a power of a primitive hyperbolic element P_0 and zero otherwise. The norm of a hyperbolic element is defined as follows: if $P = P_0^n$ with P_0 primitive and conjugate to a matrix of the form

$$\begin{pmatrix} p^{1/2} & 0 \\ 0 & p^{-1/2} \end{pmatrix}, \quad p > 1$$

then $N(P) = p$. The length of the P-invariant primitive closed geodesic on $\Gamma\backslash\mathbb{H}^2$ is equal to $\log N(P_0) = \log p$. We have the following result:

Theorem 3.4. *For any cofinite group* Γ *we have*

$$\psi_\Gamma(X) = \sum_{1/2 < s_j \leq 1} \frac{X^{s_j}}{s_j} + E_\Gamma(X), \tag{3.23}$$

where the error term satisfies the bound $E_\Gamma(X) = O(X^{3/4})$.

Proof. Notice that instead of picking $h(t)$ we can pick its Fourier transform $g(x)$, which determines the Fourier pair h, g. Pick a test function of the form $g(x) = 2\cosh(x/2)q(x)$ where $q(x)$ is even, smooth and compactly supported in an interval $[-\log(X+Y), \log(X+Y)]$. Then for $s_j > 1/2$ we have

$$h(t_j) = \frac{X^{s_j}}{s_j} + O(Y + X^{1/2}). \tag{3.24}$$

The discrete eigenvalues with $\Re(s_j) = 1/2$ contribute $O\left(|s_j|X^{1/2}\min\{1, X^2Y^{-2}|s_j|^{-2}\}\right)$. The whole discrete spectrum gives the expected main term and an error $O(Y + X^{3/2}Y^{-1})$. The continuous spectrum contribution goes in the error term.

On the geometric side the identity class, the elliptic classes and the parabolic classes contribute at most $X^{3/2}Y^{-1}$. Comparing the two sides and choosing $q(x)$ close to the characteristic we get

$$\psi_\Gamma(X) = \sum_{N(P)\le X} \Lambda_\Gamma(N(P)) = \sum_{p^k \le X} \log p = \sum_{1/2 < s_j \le 1} \frac{X^{s_j}}{s_j} + O(Y + X^{3/2}Y^{-1}). \tag{3.25}$$

To get the desired error we pick the balance $Y = X^{3/4}$. □

For arithmetic groups Γ further improvements on the bound for the error term can be deduced as an application of the (Bruggeman-)Kuznetsov formula. Such improvements were first deduced for the modular group $\Gamma = \mathrm{PSL}(2,\mathbb{Z})$ by Iwaniec and Luo-Sarnak. The crucial step in these works is proving a non-trivial bound on a specific spectral exponential sums over the Laplacian eigenvalues λ_j.

In the case of the Prime number theorem the asymptotic (3.21) follows from Riemann-von Mangoldt explicit formula. For the modular group an analogous formula was derived by Iwaniec, stating that

$$E_\Gamma(X) = \sum_{0 < t_j \le T} \frac{X^{1/2+it_j}}{1/2 + it_j} + O\left(\frac{X}{T}\log X\right), \tag{3.26}$$

for any parameter $T \le X^{1/2}$. From (3.26) we reduce the study of the error term to the study of the spectral exponential sum

$$S(T, X) := \sum_{0 < t_j \le T} X^{it_j}. \tag{3.27}$$

By Weyl's law we can trivially bound $S(T, X) \ll T^2$. Together with Iwaniec's explicit formula this recovers Selberg's bound $X^{3/4}$. Iwaniec was the first one who proved a non-trivial bound for $S(T, X)$. Moreover, the conjectural bound

$$S(T, X) \ll_\epsilon T^{1+\epsilon} X^\epsilon \tag{3.28}$$

would give the optimal expected bound

$$E_\Gamma(X) = O_\epsilon(X^{1/2+\epsilon}). \tag{3.29}$$

The best results that we know towards this conjecture are due to Soundararajan-Young and Balkanova-Frolenkov for the pointwise estimate:

$$E_\Gamma(X) = O_\epsilon(X^{25/36+\epsilon}) \tag{3.30}$$

and due to Balog, Biró, Harcos and Maga on average:

$$\left(\frac{1}{X}\int_X^{2X} |E_\Gamma(x)|^2 dx\right)^{1/2} = O_\epsilon\left(X^{7/12+\epsilon}\right).$$

The Burggeman-Kuznetsov formula belongs in the family of *relative trace formulas*. The spectral side of this trace formula appears information for the Hecke eigenvalues (Fourier coefficients) of the Maass forms. It relates

$$\sum_j h(t_j)\lambda_j(n)\overline{\lambda_j(m)} \tag{3.31}$$

to arithmetic information of the group (encoded in sums of Kloosterman sums).

3.6. **Open problems.** 1) For the modular group $\mathrm{SL}_2(\mathbb{Z})$ even the pointwise bound

$$E_\Gamma(X) = O_\epsilon(X^{2/3+\epsilon}) \tag{3.32}$$

seems to be extremely difficult, as it follows from the Lindelöf Hypothesis for specific Dirichlet L-functions. On the other hand, for a general cofinite group we only know Selberg's general bound $O(X^{3/4})$ for the error term in the Prime geodesic theorem.

4. L-functions, counting problems and quantum unique ergodicity

In this last section we will discuss some problems in more details. These problems are in the centre of modern research in the field and we will only mention some basic results about each one of them. First we remind some basic material for L-functions.

4.1. ***L*-functions attached to Maass forms.** The theory of L-functions has its origin on the study of Riemann's zeta function $\zeta(s)$ and Dirichlet's L-functions $L(s,\chi)$, where χ is a Dirichlet character. Roughly speaking, an L-function is a function of the complex variable s given by a Dirichlet series

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s}. \tag{4.1}$$

The coefficients are such that the series (4.1) converges absolutely for $\Re(s) > A$ where A is some constant depending on the sequence $(a_n)_{n=1}^{\infty}$. In order to call the series (4.1) an L-function we ask for extra properties; in fact we want (4.1) to have an analytic continuation on the whole complex plane, to satisfy a functional equation relating the value of the function at s with the value of the function at $1-s$ and to have an Euler product over the primes. For the Riemann zeta function and Dirichlet's L-functions these properties follow using Poisson summation and the unique factorization in $\mathbb{Z}$. More precisely we have that

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p (1-p^{-s})^{-1} \tag{4.2}$$

satisfy the functional equation

$$\Lambda(s) = \Lambda(1-s) \tag{4.3}$$

where $\Lambda(s)$ is the *completed* Riemann zeta function from (2.2). Similarly $L(s,\chi)$ satisfies

$$L(s,\chi) := \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} = \prod_p (1-\chi(p)p^{-s})^{-1} \tag{4.4}$$

for $\chi : (\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$ is a Dirichlet character mod q. The functional equation is slightly more involved in that case and contains the *Gauss sum* $\tau(\chi)$.

For the rest of this section we focus only the case of $\mathrm{SL}_2(\mathbb{Z})$. We can thus pick a common system of Hecke-Maass forms, i.e. a family of eigenfunctions both of the Hecke operators and the Laplace operator. The Fourier expanion (2.21) of a cusp form u_j takes the form

$$u_j(z) = 2\sum_{n\neq 0} y^{1/2}\rho_j(n)K_{it_j}(2\pi|n|y)e^{2\pi i n z}. \tag{4.5}$$

Then the Fourier coefficients $\rho_j(n)$ are related to the Hecke eigenvalues $\lambda_j(n)$ in the following way:

$$\rho_j(n) = \rho_j(1)\lambda_j(n), \quad \frac{\rho_j(1)}{\sqrt{\cosh(\pi t_j)}} \asymp t_j^{\pm\epsilon}. \tag{4.6}$$

We define the Hecke L-function attached to the cusp form u_j to be

$$L(s, u_j) := \sum_{n=1}^{\infty} \frac{\lambda_j(n)}{n^s}, \quad \Re(s) > 1. \tag{4.7}$$

Then as a function of s this function has all the good properties that we expect from an L-function: it has an analytic continuation to an entire function on the complex plane, it satisfies a functional equation and it has an Euler product. If we define the completed L-function by

$$\Lambda(s, u_j) := \pi^{-s}\Gamma\left(\frac{s - it_j}{2}\right)\Gamma\left(\frac{s + it_j}{2}\right)L(s, u_j) \tag{4.8}$$

then the function equation takes the form

$$\Lambda(s, u_j) = \epsilon_j\Lambda(1 - s, u_j). \tag{4.9}$$

where $\epsilon_j = \pm 1$ depending on whether u_j is *even* or *odd* (i.e. $u_j(-\overline{z}) = \epsilon_j u_j(z)$). The Euler product is a bi-product of the Hecke-multiplicativity for the Fourier coefficients $\lambda_j(n)$:

$$L(s, u_j) = \prod_p (1 - \lambda_j(p)p^{-s} + p^{-2s})^{-1}. \tag{4.10}$$

The theory of these L-functions is of basic interest both itself but also for applications. They appear naturally to a variety of problems, such as the Quantum unique ergodicity that we will discuss at the end of this section. Before doing that we return to counting problems; in particular to the hyperbolic lattice counting problem which has many similarities but also important differences with the Prime geodesic theorem.

4.2. **Hyperbolic lattice point counting.** Lattice point problems appear in many different areas number theory. These kind of problems first appeared in the work of Gauss and Dirichlet. Gauss' circle problem asks to estimate the number of integer points inside a Euclidean circle D with center at the origin and radius $X^{1/2}$. If we denote by $N(X)$ the number of integer points inside this circle

$$N(X) = \#\left\{w = (a, b) \in \mathbb{Z}^2 : a^2 + b^2 \leq x\right\}, \tag{4.11}$$

then an elementary geometric packing method gives

$$N(X) = \sum_{m \leq X} r(m) = \pi X + O(X^{1/2})$$

as $X \to \infty$. Here $r(m) = \#\{(a, b) \in \mathbb{Z}^2 : a^2 + b^2 = m\}$. The key point in Gauss' proof is bounding the error term $E(X) = N(X) - \pi X$ by the area of a boundary strip

$$\left\{X^{1/2} - 1/\sqrt{2} < |z| < X^{1/2} + 1/\sqrt{2}\right\}.$$

Gauss' bound for the error term $E(X) = N(X) - \pi X$ is not optimal. Using Poisson summation Voronoi, Sierpinski, Landau and van der Corput independently derived the bound

$$E(X) = O(X^{1/3}).$$

This bound has slightly been improved several times. The optimal existing bound $E(X) = O\left(X^{517/1648+\epsilon}\right)$ is due to Bourgain and Watt (2018). To deal with the error term $E(X)$ we use its Fourier expansion (involving special functions) which can be simplified to the form

$$E(X) = \frac{X^{1/4}}{\pi} \sum_{n=1}^{\infty} \frac{r(n)}{n^{3/4}} \sin\left(2\pi\sqrt{nX} - \frac{\pi}{4}\right) + O(1). \tag{4.12}$$

Hardy conjectured that one should expect $E(X) = O_\epsilon(X^{1/4+\epsilon})$ for every $\epsilon > 0$. He supported this conjecture by proving lower bounds for the error term.

In a general topological space X one can consider a group Γ acting discontinuously on X. Let $D = \{D_i\}$ be a family of compact subsets $D_i \subset X$ and z a point in X. The lattice point problem is to estimate the number of points of the orbit $\Gamma z = \{\gamma z : \gamma \in \Gamma\}$ which meet D_i. In this setting, Gauss' problem is the special case $X = \mathbb{R}^2$, $\Gamma = \mathbb{Z}^2$ and D is the family of circles with center at the origin $(0,0)$ and radius $x \to \infty$. If X is a homogeneous space $X = G/K$ of a Lie group G, where K is a maximal compact subgroup of G, we take Γ to be a lattice in G. We may consider D be a family of well-shaped compact sets or more general well-rounded sets. The case of the hyperbolic plane is the special case $G = \mathrm{SL}_2(\mathbb{R})$, $K = \mathrm{SO}_2(\mathbb{R})$ since in that case we can identify:

$$\mathbb{H} = \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R}).$$

Thus the hyperbolic lattice point problem asks to estimate the number of points in the orbit Γz that belong in a disk of radius R and center w, i.e. to give an asymptotic formula for

$$\#\{\gamma \in \Gamma : \rho(\gamma z, w) \leq R\},$$

or, using the standard point-pair invariant function $u(z,w)$, for the quantity

$$N(X; z, w) = \#\{\gamma \in \Gamma : 4u(z, \gamma w) + 2 \leq X\}$$

where $X \sim e^R$. We have the following theorem.

Theorem 4.1 (Selberg, Günther, Good)**.** *Let z, w be two fixed points in $\mathbb{H}$ and Γ be a cocompact or cofinite Fuchsian group. Then, as $X \to \infty$, we have*

$$N(X; z, w) = \sum_{1/2 < s_j \leq 1} \pi^{1/2} \frac{\Gamma(s_j - 1/2)}{\Gamma(s_j + 1)} u_j(z)\overline{u_j(w)} X^{s_j} + E(X; z, w),$$

where the error term satisfies the bound

$$E(X; z, w) = O(X^{2/3}).$$

In the Euclidean circle problem, Gauss' argument works because the area of a large Euclidean disc dominates the length of the Euclidean circle which bounds the error term $E(X)$. The isoperimetric inequality for a Riemannian surface of constant curvature takes the form

$$4\pi A - KA^2 \leq L^2,$$

where A is the area of a domain D, L is the length of the bounday of D and K is the curvature of the surface. In the Euclidean plane we have $K = 0$ and Gauss' argument applies. However, in $\mathbb{H}$ the isoperimetric inequality gives $A \leq L$. Indeed, the area of a hyperbolic disc of radius R is $4\pi \sinh^2(\frac{R}{2}) \sim \pi e^R$ as $R \to \infty$ and the length of the circumference is $2\pi \sinh R \sim \pi e^R$ as $R \to \infty$. Hence, the area of the disc and the length of the boundary have the same order of growth. This explains the reason one cannot estimate the error term $E(X; z, w)$ using on an elementary geometric argument.

We must also emphasize that the $O(X^{2/3})$-bound should be regarded as the analogue of the $O(X^{1/3})$-bound in the Euclidean case; however, it has not been improved for any group Γ or any pair of points z, w. Selberg was the first who proved the bound $O(X^{2/3})$, but he didn't publish it. For Γ cofinite, Patterson obtained the bound $O(X^{3/4})$. Earlier, Fricker had already deduced the analogue of Theorem 4.1 for the 3-dimensional hyperbolic space $\mathbb{H}^3$. This result finally was proved by Günther for all rank one symmetric spaces. Good used a different approach to give a new proof of the $O(X^{2/3})$-bound. He proved a general sum formula that covers many cases of decompositions of the group $G = \mathrm{SL}_2(\mathbb{R})$. One of these cases corresponds to the classical lattice problem.

We briefly describe the idea of the proof of Theorem 4.1, which is quite close to the ideas used by Voronoi in the Euclidean problem. Assume Γ is cocompact. Let $k(u)$ be the characteristic function

$$k(u) = k_X(u) = \chi_{[0,(X-2)/4]}(u). \tag{4.13}$$

One can easily see that

$$N(X; z, w) = \sum_{\gamma\in\Gamma} k(u(\gamma z, w)). \tag{4.14}$$

The naive approach in the proof of Theorem 4.1 is to apply the pre-trace formula for this kernel $k(u)$. However, $k(u)$ is not smooth enough. If we let $h(t) = h_X(t)$ be the Selberg/Harish-Chandra transform of the kernel $k(u)$, then for t and X big enough the Selberg/Harish-Chandra transform behaves like

$$h_X(t) \sim |t|^{-3/2} X^{1/2+it}. \tag{4.15}$$

In this case the pre-trace formula implies a formula of the form

$$E(X; z, w) = \sum_{t_j\in\mathbb{R}} h_X(t_j) u_j(z)\overline{u_j(w)} + o(X^{1/2}). \tag{4.16}$$

In view of (4.15) and local Weyl's law (Theorem 2.3), the series in the expansion (4.16) diverges, and hence this choice of kernel fails to give an upper bound for $E(X; z, w)$. In order to have good estimates about the Selberg/Harish-Chandra transform one has to work with smooth approximations of the kernel $k(u)$. We define the kernels $k_\pm(u)$ by

$$k_+(u) = \begin{cases} 1, & \text{for} \quad u \le \frac{X-2}{4}, \\ \dfrac{-4u}{Y} + \dfrac{X+Y-2}{Y}, & \text{for} \quad \frac{X-2}{4} \le u \le \frac{X+Y-2}{4}, \\ 0, & \text{for} \quad \frac{X+Y-2}{4} \le u, \end{cases} \tag{4.17}$$

$$k_-(u) = \begin{cases} 1, & \text{for} \quad u \le \frac{X-Y-2}{4}, \\ \dfrac{-4u}{Y} + \dfrac{X-2}{Y}, & \text{for} \quad \frac{X-Y-2}{4} \le u \le \frac{X-2}{4}, \\ 0, & \text{for} \quad \frac{X-2}{4} \le u. \end{cases} \tag{4.18}$$

We obtain the upper bound

$$E(X; z, w) \ll \sum_{t_j\neq 0} h_\pm(t_j) u_j(z)\overline{u_j(w)} + O(X^{1/2}\log X + Y), \tag{4.19}$$

where the Selberg/Harish-Chandra transform $h_\pm(t)$ of $k_\pm(u)$ for $t \neq 0$ satisfies the bound

$$h_\pm(t) \ll |t|^{-5/2} \left\{\min\{|t|, X/Y\}\right\} X^{1/2}. \tag{4.20}$$

The upper bound (4.19) implies $E(X; z, w) = O(Y + XY^{-1/2} + X^{1/2}\log X)$ and the choice $Y = X^{2/3}$ implies the bound of Theorem 4.1. The general case of a cofinite group Γ does not affect the general argument of the proof, as the contribution of the Eisenstein series in this problem can be treated in the same way and goes to the error term.

Numerical investigations indicate that the bound of Theorem 4.1 is far from being optimal; in fact one should expect square root cancellation for the error term

$$E(X; z, w) = O_\epsilon(X^{1/2+\epsilon}) \tag{4.21}$$

for every $\epsilon > 0$. In the Euclidean circle problem, we know the eigenvalues $r(n)$ and the eigenfunctions appearing in the spectral expansion of the error term explicitly. This allows to improve the bound $O(X^{1/3})$. However, in the hyperbolic case only few things are known for the eigenfunctions $u_j(z)$ and the eigenvalues λ_j in general. For this reason any improvement of the bound $O(X^{2/3})$ towards the conjecture (4.21) is a much more difficult problem, since it depends on capturing cancellation on the spectral sum

$$\sum_{t_j>0} |t_j|^{-3/2} X^{1/2+it_j} u_j(z)\overline{u_j(w)}. \tag{4.22}$$

However we know that (4.21) is optimal: Rudnick and Philips showed that for Γ cocompact or the modular surface we have

$$E(X;z,z) = \Omega_\epsilon\left(X^{1/2}(\log\log X)^{1/4-\epsilon}\right) \tag{4.23}$$

for every $\epsilon > 0$.

4.3. **Quantum ergodicity.** Let $\mathcal{M}$ be a compact Riemannian manifold with Laplace-Beltrami operator $\Delta_{\mathcal{M}}$, dv the Riemannian volume on $\mathcal{M}$ and $\{\phi_j\}_{j=0}^{\infty}$ an L^2-normalized sequence of Laplace eigenfunctions with eigenvalues $\{\lambda_j\}_{j=0}^{\infty}$ tending to ∞. The Quantum ergodicity (QE) theorem of Schnirelman, Colin de Verdiere and Zelditch asserts that if the geodesic flow on the unit cotangent bundle $S^*\mathcal{M}$ is ergodic then there exists a density one subsequence of the measures $dv_j := |\phi_j|^2 dv$ that converges weakly to dv in the spectral limit $\lambda_j \to \infty$. This is equivalent with

$$\frac{1}{\mathrm{vol}(B)}\int_B |\phi_{j_k}|^2 dv \to \frac{1}{\mathrm{vol}(\mathcal{M})} \tag{4.24}$$

for a density one subsequence $\{\lambda_{j_k}\}$ and for every continuity set $B \subset \mathcal{M}$. In particular, the Quantum ergodicity theorem holds for manifolds of negative curvature. Zelditch extended this result to the case of noncompact surfaces. The main igredient in Zelditch's proof is a nontrivial upper bound for the quantum variance of the measures

$$\sum_{\lambda_j\le\Lambda}\left|\frac{1}{\mathrm{vol}(B)}\int_B |\phi_j|^2 dv - \frac{1}{\mathrm{vol}(\mathcal{M})}\right|^2 = o_B(\mathcal{N}(\Lambda)) \tag{4.25}$$

where $\mathcal{N}(\Lambda) = \#\{\lambda_j \le \Lambda\}$ and B is fixed.

For manifolds of negative curvature the *Quantum unique ergodicity* (QUE) conjecture of Rudnick and Sarnak predicts that there is no exceptional subsequence:

$$\frac{1}{\mathrm{vol}(B)}\int_B |\phi_j(z)|^2 dv(z) \to \frac{1}{\mathrm{vol}(\mathcal{M})}, \tag{4.26}$$

as $\lambda_j \to \infty$ for any fixed continuity set B on $\mathcal{M}$. This problem was resolved in the case ϕ_j are Hecke–Maaß forms on compact arithmetic Riemann surfaces $\Gamma\backslash\mathbb{H}$ by Linderstrauss with ergodic methods and by Soundararajan in the cofinite arithmetic case.

Much before the proof of the QUE conjecture, for the special case of $\Gamma = \mathrm{PSL}_2(\mathbb{Z})$ a striking relation between the conjecture and estimates for L-functions was discovered by Watson. His *triple product formula*, together with the spectral theorem, implies that (4.26) follows from an estimate for an L-function of degree 6.

Theorem 4.2 (Watson)**.** *For two fixed Hecke-Maass forms u, u_j of the modular surface the triple product*

$$\int_{\Gamma\backslash\mathbb{H}} |u(z)|^2 u_j(z)\, d\mu(z) \tag{4.27}$$

decomposes as a product of L-functions arising from u and u_j. A good subconvexity bound for the higher degree L-function of this product implies the QUE conjecture. In particular the Generalized Riemann Hypothesis implies the QUE conjecture (with the optimal rate).

Watson's formula is one of the most beautiful examples in modern analytic number theory where the connection between equidistribution problems and estimates for L-functions are very closely related (and quite often almost equivalent).

4.4. **Exercises.** 1) When we restrict a Maass form u_j on a fixed closed geodesic ℓ on $\Gamma\backslash\mathbb{H}$, the average size of the *period integral* of u_j defined by

$$\int_\ell u_j(z)ds(z)$$

is smaller that the average size of u_j (here ds denotes the arc length measure on the geodesic ℓ). That can been seen from the upper bound of the second moment

$$\sum_{t_j\le T}\left|\int_\ell u_j(z)ds(z)\right|^2 \ll T. \tag{4.28}$$

Assume for simplicity that Γ is cocompact. Using (4.28) prove that the error term of the lattice counting problem satisfies the average upper bound

$$\int_\ell E(X,z,w)ds(z) \ll_w X^{1/2+\epsilon}.$$

2) Using the Hecke multiplicativity properties of $\lambda_j(n)$ prove the Euler product (4.10).

4.5. **Open problems.** 1) What is the correct order of growth for $E_\Gamma(X,z,w)$? The bound $O(X^{2/3})$ has never been improved for any cofinite Fuchsian group Γ and for any pair of points z, w.

2) For the QUE on the modular surface the conjectural rate of convergence is

$$\frac{1}{\mathrm{vol}(B)}\int_B |u_j(z)|^2 d\mu(z) - \frac{1}{\mathrm{vol}(\Gamma\backslash\mathbb{H})} = O_{B,\epsilon}\left(t_j^{-1/2+\epsilon}\right). \tag{4.29}$$

Up to this moment, there is no effective bound for (4.29).

References

[1] D. Borthwick, *Introduction to Spectral Theory on Hyperbolic Surfaces.* Notes available online

[2] D. Goldfeld, *Automorphic forms and L-functions for the group GL(n,R).* With an appendix by Kevin A. Broughan. Cambridge Studies in Advanced Mathematics, 99. Cambridge University Press, Cambridge, 2006. xiv+493 pp. ISBN: 978-0-521-83771-2; 0-521-83771-5

[3] H. Iwaniec, *Spectral methods of automorphic forms.* Second edition. Graduate Studies in Mathematics, 53. American Mathematical Society, Providence, RI; Revista Matemática Iberoamericana, Madrid, 2002. xii+220 pp.

[4] H. Iwaniec, *Topics in classical automorphic forms.* Graduate Studies in Mathematics, 17. American Mathematical Society, Providence, RI, 1997. xii+259 pp.

[5] H. Iwaniec and E. Kowalski, *Analytic number theory.* American Mathematical Society Colloquium Publications, 53. American Mathematical Society, Providence, RI, 2004. xii+615 pp. ISBN: 0-8218-3633-1

[6] P. Sarnak, *Spectra of hyperbolic surfaces.* Bull. Amer. Math. Soc. (N.S.) 40 (2003), no. 4, 441–478.

[7] H. Then *Spectral theory of automorphic forms.* Notes, based on notes by A. Booker