THE SELBERG TRACE FORMULA

EDGAR ASSING

Abstract. In this course, taught at the University of Bonn in the winter term 24/25, we discussed Selberg's trace formula. Note that these notes may contain typos and misunderstandings, for which I take full responsibility. For personal use only!

1. Introduction

Selberg developed *his* trace formula in [Se] as a *natural* non-commutative generalization of the Poisson summation formula. More than seventy years after its invention there exists a vast amount of literature on Selberg's trace formula (and probably even more on its generalizations). In these lectures we will mostly rely on the following:

- A extensive account that also includes some of Selberg's unpublished results is given in [He76, He83].
- A nice survey containing many interesting results is given in [Ve].
- For background on the spectral theory of automorphic forms we will often refer to [Iw].

But what is the trace formula all about? It is impossible to summarize this in a view words, but classically we can paint the following picture. Let X be compact Riemannian surface of (constant) negative curvature. The geometry of this space gives rise to the geodesic flow. This is a (physical) dynamical system describing the *classical world* and many interesting questions are connected with it. On the other hand we have the Laplace-Beltrami operator Δ_X acting on functions on X. The spectral data of this operator describes the *quantum world*. Many difficult problems surround the eigenvalues and eigenfunctions of Δ_X. The trace formula can be seen as a bridge between classical and quantum world. One side of the formula sees the spectral data of Δ_X, while the other one sees the geometry of X. Applications go both ways.

In these lectures we plan to cover the following topics:

- **The trace formula for compact quotients of the upper half plane.** This part of the course includes the necessary background concerning hyperbolic geometry and Fuchsian groups, spectral theory of the Laplace-Beltrami operator and a careful development of the trace formula.
- **First applications.** We will discuss Weyl laws, the prime geodesic theorem and the Selberg zeta function.

- **The trace formula for $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$.** This space is non-compact so that we have to revisit the spectral theory, which now features Eisenstein series. Also the development of the trace formula is more complicated and requires some regularization (or truncation).

We will not cover the trace formula beyond $\mathrm{SL}_2(\mathbb{R})$. In particular, we will not discuss any form of Arthur's trace formula and we can also not cover applications that rely on the comparison of different trace formulae. Furthermore, we will stick to the classical set-up and not talk about the adelic formulation of the trace formula.

We will end this introduction by discussing three examples.

1.1. Example A: The abstract trace formula.

Let H be a unimodular locally compact group.[1] Further, let $\Gamma \subseteq H$ be a discrete subgroup.[2] We consider the Hilbert space $L^2(\Gamma\backslash H)$ with respect to the Haar measure. We define the action

$$[R_h\phi](x) = \phi(xg) \text{ for } h \in H,\ x \in \Gamma\backslash H \text{ and } \phi \in L^2(\Gamma\backslash H).$$

Note that by invariance of the measure the inner product satisfies

$$\langle R_h\phi, R_h\psi\rangle = \int_{\Gamma\backslash H} \phi(xh)\overline{\psi(xh)}dx = \int_{\Gamma\backslash H} \phi(x)\overline{\psi(x)}dx = \langle \phi, \psi\rangle.$$

In other words, R is a unitary representation of H on $L^2(\Gamma\backslash H)$.[3] A fundamental problem in harmonic analysis (or representation theory) is to decompose R into irreducible components. This can be done by studying convolution operators (i.e. operators that arise when integrating R against suitable test functions). More precisely, given $f \in \mathcal{C}_c(H)$ we define

$$R(f) = \int_H f(y)R_y dy.$$

[1]This means that H is a locally compact topological space such that the maps $H \times H \ni (x,y) \mapsto xy \in H$ and $H \ni x \to x^{-1} \in H$ are continuous. It is well known that a locally compact group has (up to scaling) unique left and right Haar measure, See [vN]. We say that H is unimodular if the left and the right Haar measure agree. (The Haar measure is named after Alfréd Haar and has nothing to do with hair.)

[2]A subgroup of a topological group is discrete if the subspace topology is the discrete topology.

[3]We also require that the map $H \ni h \mapsto R_h\phi \in L^2(\Gamma\backslash H)$ is continuous, but this is true in our current setting.

This operator valued integral is to be understood as follows

$$\begin{aligned}[R(f)\phi](x) &= \int_H f(y)[R_y\phi](x)dy \\ &= \int_H f(y)\phi(xy)dy = \int_H f(x^{-1}y)\phi(y)dy \\ &= \int_{\Gamma\backslash H} \underbrace{\left(\sum_{\gamma\in\Gamma} f(x^{-1}\gamma y)\right)}_{K_f(x,y)} \phi(y)dy,\end{aligned}$$

for $\phi \in L^2(\Gamma\backslash H)$ and $x \in \Gamma\backslash H$. In particular, $R(f)$ is an integral operator (i.e. an operator that can be represented by integration against a suitable automorphic kernel function). Let us remark that

- Because we are assuming that f is compactly supported, the sum defining $K(x,y)$ is finite.
- We have $K_f(\gamma x, y) = K_f(x,y) = K_f(x,\gamma y)$.

From now on we will assume that $\Gamma\backslash H$ is compact. In this case we record the following two facts.

- The representation R decomposes discretely into irreducible representations with finite multiplicities:
$$R \cong \bigoplus \pi^{\oplus m(\pi,R)}. \tag{1}$$
Note that usually the representations π will be infinite dimensional.
- Under mild conditions on f the operator $R(f)$ is of trace class and the trace can be computed via
$$\operatorname{tr} R(f) = \int_{\Gamma\backslash H} K_f(x,x)dx. \tag{2}$$

We can now compute the trace of $R(f)$ in two ways. First, using the spectral expansion given in (1) we can decompose the trace

$$\operatorname{tr} R(f) = \sum_\pi m(\pi,R)\cdot \operatorname{tr}\pi(f),$$

where the operator $\pi(f)$ is defined by

$$\pi(f) = \int_H f(y)\pi(y)dy.$$

On the other hand we can use (2) and insert the definition of K_f. We obtain

$$\operatorname{tr} R(f) = \int_{\Gamma\backslash H} K(x,x)dx = \int_{\Gamma\backslash H}\sum_{\gamma\in\Gamma} f(x^{-1}\gamma x)dx.$$

We now write $\{\Gamma\}$ for a set of representatives of conjugacy classes in Γ. Furthermore, given any set $\Omega \subseteq H$ we write

$$\Omega_\gamma = \{\delta \in \Omega \colon \delta^{-1}\gamma\delta\}$$

for the centralizer of γ in Ω. This allows us to massage the expression as follows

$$\begin{aligned}
\operatorname{tr} R(f) &= \int_{\Gamma\backslash H} \sum_{\gamma\in\{\Gamma\}} \sum_{\delta\in\Gamma_\gamma\backslash\Gamma} f((\delta x)^{-1}\gamma\delta x)dx \\
&= \sum_{\gamma\in\{\Gamma\}} \int_{\Gamma_\gamma\backslash H} f(x^{-1}\gamma x)dx \\
&= \sum_{\gamma\in\{\gamma\}} \operatorname{Vol}(\Gamma_\gamma\backslash H_\gamma) \int_{H_\gamma\backslash H} f(x^{-1}\gamma x)dx.
\end{aligned}$$

Combining everything we obtain the abstract trace formula for compact quotients

$$\sum_\pi m(\pi, R)\cdot \operatorname{tr}\pi(f) = \sum_{\gamma\in\{\Gamma\}} \operatorname{Vol}(\Gamma_\gamma\backslash H_\gamma)\cdot \int_{H_\gamma\backslash H} f(x^{-1}\gamma x)dx.$$

The left hand side of this equality is the so called spectral side featuring *irreducible characters* weighted by multiplicities. On the right hand side, the so called *geometric side*, we have *orbital integrals* weighted by certain volumes. Note that, without further information on the terms in these sums, the trace formula is rather useless.

Exercise 1.1. *Take $H = \mathbb{R}^n$ and $\Gamma = \mathbb{Z}^n$ and interpret the Poisson summation formula*

$$\sum_{\mathbf{n}\in\mathbb{Z}^n} f(\mathbf{n}) = \sum_{\mathbf{n}\in\mathbb{Z}^n} \widehat{f}(\mathbf{n}) \ \textit{for}\ f\in\mathcal{C}_c^\infty(\mathbb{R}^n)$$

as trace formula.

Exercise 1.2. *Take H to be a finite group and let Γ be any proper subgroup. Use the trace formula to derive the Frobenius reciprocity formula.*

1.2. **Example B: The round sphere.** This example, taken from [Ma] is more concrete. Let $X = \mathbb{S}^2$ denote the round sphere. The Laplace-Beltrami operator is given by

$$\Delta_{\mathbb{S}^2} = -\frac{1}{\sin(\theta)}\frac{\partial}{\partial\theta}\left(\sin(\theta)\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin(\theta)^2}\frac{\partial^2}{\partial\phi^2}$$

in spherical coordinates $\theta\in[0,\pi)$ and $\phi\in[0,2\pi)$. In this case the spectral theory is well known. See for example [Tr, Satz 31.1]. We give a brief summary.

The eigenvalue problem

$$\Delta_{\mathbb{S}^2} f = \lambda f \tag{3}$$

has solutions precisely when $\lambda_l = l(l+1)$ for $l \in \mathbb{Z}_{\geq 0}$. It turns out to be convenient to write $\rho_l = \sqrt{\lambda_l + \frac{1}{4}} = l + \frac{1}{2}$. The corresponding eigenspace has dimension $m(\lambda_l) = m(\rho_l) = 2l + 1$ with eigenfunctions given by *spherical harmonics*

$$Y_l^m(\theta, \phi) = (-1)^m \left[\frac{2l+1}{4\pi} \cdot \frac{(l-m)!}{(l+m)!}\right]^{\frac{1}{2}} P_l^m(\cos(\theta))e^{im\phi},$$

for $m = -l, \dots, l$. Here P_l^m are so called Legendre polynomials of the first kind. See Figure 1 for an example.

We now choose a test function $h \in \mathcal{C}^2(\mathbb{R})$ satisfying

(1) h is even (i.e. $h(x) = h(-x)$);
(2) h extends to an analytic function on the strip $-\sigma < \operatorname{Im}(\rho) < \sigma$; and
(3) For $i = 0, 1, 2$ we have $|h^{(i)}(x)| \leq C_i(1 + |\operatorname{Re}(x)|)^{-2-\delta}$ for some $\delta > 0$.

This allows us to compute

$$\sum_{j=0}^{\infty} m(\rho_j)h(\rho_j) = \sum_{l=0}^{\infty}(2l+1)h(l+\frac{1}{2}) = \sum_{l\in\mathbb{Z}} |l + \frac{1}{2}|h(l + \frac{1}{2}).$$

Put $g(x) = |x + \frac{1}{2}|h(x + \frac{1}{2})$ and take the Fourier transform

$$\begin{aligned}\widehat{g}(y) = \int_{\mathbb{R}} g(x)e^{2\pi iyx}dx &= e^{-\pi iy}\int_{-\infty}^{\infty} |x|h(x)e^{2\pi xy}dx \\ &= e^{-\pi iy}\int_0^{\infty} xh(x)e^{2\pi xy}dx - e^{-\pi iy}\int_{-\infty}^0 xh(x)e^{2\pi ixy}dx.\end{aligned}$$

Note that g is not in $\mathcal{C}_c^\infty(\mathbb{R})$. Nonetheless, there is a sufficiently general version of the Poisson summation formula that we are allowed to apply. We obtain

$$\sum_{j=0}^{\infty} m(\rho_j)h(\rho_j) = \sum_{l\in\mathbb{Z}}(-1)^l \int_0^{\infty} xh(x)e^{-2\pi xi|l|}dx - \sum_{l\in\mathbb{Z}}(-1)^l \int_{-\infty}^0 xh(x)e^{-2\pi ix|l|}dx \tag{4}$$

After applying partial integration twice and using the assumption on h one verifies that the right hand side converges absolutely. We now consider the paths given in Figure 2.

We take the integral from 0 to infinity and deform it to the path $\mathcal{C}_1$. It can be seen that on this path of integration we can exchange sum and integration. We arrive at

$$\sum_{l\in\mathbb{Z}}(-1)^l \int_0^{\infty} xh(x)e^{2\pi ix|l|}dx = \int_{\mathcal{C}_1} zh(z)\sum_{l\in\mathbb{Z}}(-1)^l e^{-2\pi i|l|z}dz.$$

At this point we recall the well known expansion

$$\tan(z) = \frac{1}{i}\sum_{l\in\mathbb{Z}}(-1)^l e^{-2i|l|z}$$

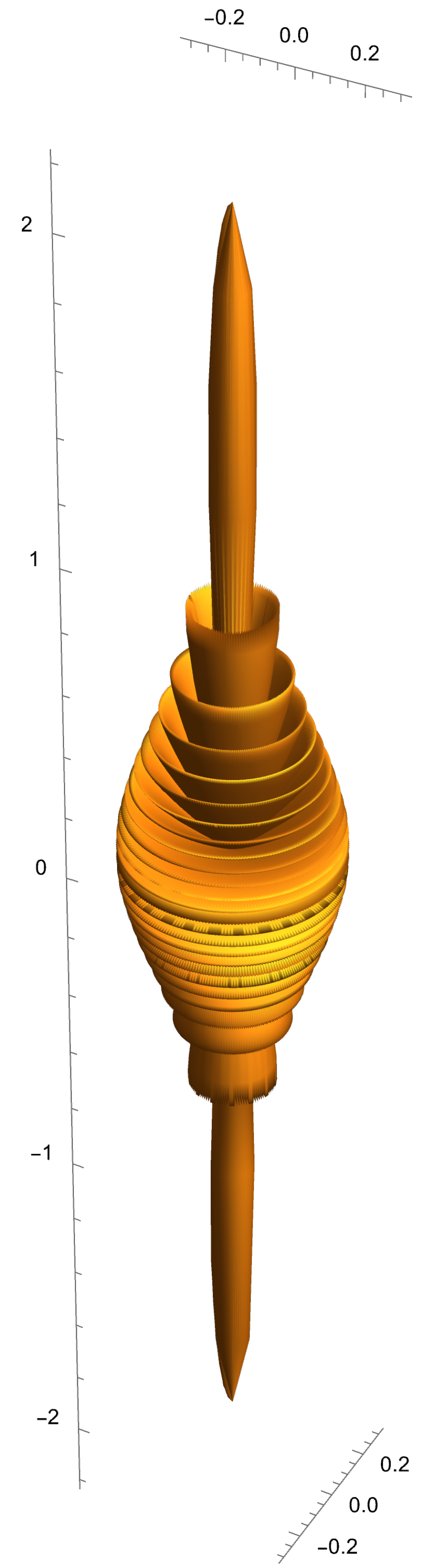

FIGURE 1. Picture of spherical harmonic. (Created by Dr. R. Toma.)

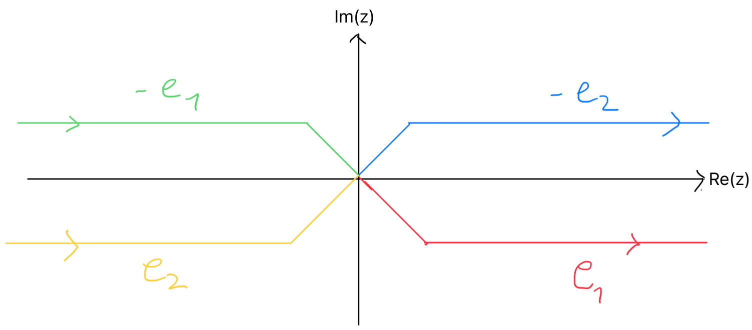

FIGURE 2. Integration paths $\mathcal{C}_1$ and $\mathcal{C}_2$.

and get

$$\sum_{l\in\mathbb{Z}}(-1)^l\int_0^\infty xh(x)e^{2\pi ix|l|}dx = \int_{\mathcal{C}_1} zh(z)\tan(\pi z)dz.$$

Similarly we obtain

$$-\sum_{l\in\mathbb{Z}}(-1)^l\int_{-\infty}^0 xh(x)e^{-2\pi ix|l|}dx = -i\int_{\mathcal{C}_2} zh(z)\tan(\pi z)dz = i\int_{\mathcal{C}_2^{-1}} zh(z)\tan(\pi z)dz,$$

where $\mathcal{C}_2^{-1}$ is the path $\mathcal{C}_2$ but with reversed direction. If we put

$$\mathcal{C} = \mathcal{C}_1 \cup (-\mathcal{C}_1) \cup \mathcal{C}_2^{-1} \cup (-\mathcal{C}_2)^{-1}, \tag{5}$$

then we obtain the formula

$$\sum_{j=0}^\infty m(\rho_j)h(\rho_j) = \frac{i}{2}\int_{\mathcal{C}} zh(z)\tan(\pi z)dz.$$

Let us summarize this in form of a theorem.

Theorem 1.1 (Trace formula for the sphere). *Let $h \in C^2(\mathbb{R})$ be a test function satisfying*

(1) *h is even (i.e. $h(x) = h(-x)$);*
(2) *h extends to an analytic function on the strip $-\sigma < \mathrm{Im}(\rho) < \sigma$; and*
(3) *For $i = 0,1,2$ we have $|h^{(i)}(x)| \le C_i(1+|\mathrm{Re}(x)|)^{-2-\delta}$ for some $\delta > 0$.*

Then we have

$$\sum_{j=0}^\infty m(\rho_j)h(\rho_j) = \frac{i}{2}\int_{\mathcal{C}} zh(z)\tan(\pi z)dz,$$

for the path $\mathcal{C}$ as defined in (5).

Exercise 1.3. *Give a direct proof of the Weyl law*

$$\sum_{\rho_j \leq X} m(\rho_j) = \frac{\mathrm{Vol}(\mathbb{S}^2)}{4\pi} X^2 + O(X)$$

for the sphere and show that the error term is essentially sharp. What can be said using the trace formula?

1.3. **Example C: The flat torus.** Let $\mathbb{T}_2 = \mathbb{R}^2/\mathbb{Z}^2$ be the flat two dimensional torus. It will be useful to identify $\mathbb{R}^2 = \mathbb{C}$ and view $\mathbb{Z}^2$ as the lattice generated by 1 and i. We consider the Laplace operator on $L^2(\mathbb{T}_2)$ given by

$$\Delta = -\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) \tag{6}$$

on $\mathcal{C}^\infty(\mathbb{T}_2)$, where we write $z = x + iy$. Write $e(x) = e^{2\pi i x}$ and note that $f_{m,n}(z) = e(mx + ny)$ satisfies

$$\Delta f_{m,n}(z) = 4\pi^2(m^2 + n^2) f_{m,n}(z). \tag{7}$$

The multiplicity of an eigenvalue $4\pi^2 k$ is given by

$$r(k) = \sharp\{m, n \in \mathbb{Z} \colon k = m^2 + n^2\}.$$

Further, recall that Δ is essentially self adjoint and consider its self adjoint extension A.[4] It should be clear from (7) that expansion in eigenfunctions of A is nothing but Fourier expansion.

Next, we consider the resolvent

$$R(\lambda) = (A - \lambda)^{-1} \tag{8}$$

of A. This is an integral operator and we denote its kernel by $r(z, z'; \lambda)$. We can easily write down the spectral expansion:

$$r(z, z'; \lambda) = \sum_{m,n \in \mathbb{Z}} (4\pi^2(m^2 + n^2) - \lambda)^{-1} f_{m,n}(z) \overline{f_{m,n}(z')}. \tag{9}$$

On the other hand we can compute the Green's function $g_{\mathbb{C}}(z, z'; s)$ of the equation $\Delta u - s^2 u = 0$. It turns out that

$$g_{\mathbb{C}}(z, z'; s) = \frac{i}{4} H_0^2(|z - z'|s).$$

We arrive at

$$r(z, z'; \lambda) = \sum_{\gamma \in \mathbb{Z}^2} g_{\mathbb{C}}(z, z' + \gamma; s), \tag{10}$$

[4] Recall that we call an (unbounded) operator essentially self adjoint if its closure is self-adjoint. See [Tr, Definition 21.2]. To see that Δ is essentially self adjoint, one first shows that the unbounded operator Δ with domain $\mathcal{C}^\infty(\mathbb{T}_2)$ is symmetric and non-negative. This is due to Green's theorem. Now we can construct a self adjoint extension using Friedrich's theorem. See [Tr, Satz 17.11& 17.12].

for $\lambda = s^2$. Note that a standard asymptotic formula for $H_0^2(|x|s)$ implies that the right hand side of (10) converges absolutely if $\operatorname{Im}(s) > 0$. For other s it is to be understood by analytic continuation otherwise.

We are now ready to prove the following trace formula like identity.

Theorem 1.2 (Hardy-Voronoï formula)**.** *Let $h\colon \mathbb{R}_{\geq 0} \to \mathbb{C}$ be a function that analytic in a neighborhood of $\mathbb{R}_{\geq 0}$ and rapidly decreasing there (i.e. $h(\lambda) \ll_A |\operatorname{Re}(\lambda)|^{-A}$ for $A \geq 2$). Then we have*

$$\sum_{n=0}^{\infty} r(k)h(k) = \pi \sum_{l=0}^{\infty} r(l) \int_0^{\infty} h(x) J_0(2\pi\sqrt{lx})dx, \tag{11}$$

where $J_0(\cdot)$ is the classical J-Bessel function of index 0.

Proof. We consider the integral operator $h(A)$ obtained from h. Recall that

$$h(A) = -\frac{1}{2\pi i}\int_{\Omega} h(\lambda)R(\lambda)d\lambda,$$

where Ω is a contour enclosing the spectrum and lying in the region where h is analytic. Such a contour exists, by our assumption on h.

We compute the trace of $h(A)$ as

$$\operatorname{tr} h(A) = \sum_{m,n\in\mathbb{Z}} \langle h(A)f_{m,n}, f_{m,n}\rangle = \sum_{k\in\mathbb{Z}_{\geq 0}} r(k)\cdot h(4\pi^2 k).$$

In particular, the assumption that h is rapidly decaying ensures that the sum is defined and $h(A)$ is of trace class.

On the other hand we can compute the trace by integrating the kernel corresponding to $h(A)$ over the diagonal. One can justify the following computation

$$\operatorname{tr} h(A) = -\frac{1}{2\pi i}\int_{\mathbb{T}_2}\int_{\Omega} h(\lambda) r(z,z,\lambda)d\lambda = -\frac{1}{8\pi}\sum_{\gamma\in\mathbb{Z}^2}\int_{\Omega} h(\lambda)H_0^2(|\gamma|\sqrt{\lambda})d\lambda.$$

Note that we cross the branchcut of $\sqrt{\lambda}$. So that on the lower half of the contour we integrate over $H_0^2(-|\gamma|\sqrt{\lambda})$. Here we use the formula

$$H_0^2(-z) = H_0^2(e^{\pi i}z) = H_0^2(z) + 2J_0(z).$$

After deforming the contour to consist of the two pieces $(\infty, i0)$ and $(-i0, \infty)$ we find that the $H_0^2(z)$-parts cancel out and we are left with

$$\operatorname{tr} h(A) = -\frac{1}{4\pi}\sum_{\gamma\in\mathbb{Z}^2}\int_{(\infty,i0)} h(\lambda)J_0(|\gamma|\sqrt{\lambda})d\lambda.$$

Arranging the γ sum according to $|\gamma| = \sqrt{k}$ and making a change of variables yields

$$\operatorname{tr} h(A) = \pi \sum_{k\in\mathbb{Z}_{\geq 0}} r(k)\int_0^{\infty} h(4\pi^2\lambda)J_0(2\pi\sqrt{k\lambda})d\lambda.$$

Combining the two expressions for the trace completes the proof up to re-scaling h. □

Remark 1.3. This proof, taken from [Ve], is a bit convoluted. However, it gave us the opportunity to introduce some important ideas that will play a role later on. The biggest issue is the strong restrictions on h that we inherit from our strategy. Note that these are not necessary. Indeed, the equality (11) holds also for $h \in \mathcal{C}_c^\infty(\mathbb{R})$ (and even functions that can be nicely approximated by such functions).

The following corollary can be interpreted as the Weyl law for the torus. However, it is probably better known as the Gaußcircle problem.

Corollary 1.4. *We have*

$$\sharp\{(m,n) \in \mathbb{Z}^2 \colon m^2+n^2 \leq X\} = \pi X + R(X)$$

with $R(X) \ll X^{\frac{3}{8}}$.

Proof. We choose $h(x) = \max(X-x,0)$. Plugging this into the formula above we obtain

$$\begin{aligned} \sum_{0\leq k \leq X}(X-k)r(k) &= \pi \sum_{l=0}^{\infty} r(l) \int_0^X (X-x)J_0(2\pi\sqrt{lx})dx \\ &= \frac{\pi}{2}X^2\cdot r(0) + \sum_{l=1}^{\infty} r(l) \int_0^X (X-x)J_0(2\pi\sqrt{lx})dx \\ &= \frac{\pi}{2}X^2 + \frac{X}{\pi}\sum_{l=1}^{\infty}\frac{r(l)}{l}J_2(2\pi\sqrt{lX}). \end{aligned} \tag{12}$$

To compute the integrals for $l \neq 0$ one needs to know some formulae for J-Bessel functions or one can look up the integral.

We continue by observing that $J_2(z) \ll z^{-\frac{1}{2}}$ as $z \to \infty$. Thus we can estimate

$$\sum_{l=1}^{\infty}\frac{r(l)}{l}J_2(2\pi\sqrt{lX}) \ll X^{-\frac{1}{4}}\sum_{l=1}^{\infty}\frac{r(l)}{l^{\frac{5}{4}}} \ll X^{-\frac{1}{4}}. \tag{13}$$

We conclude that

$$\sum_{0\leq k\leq X}(X-k)r(k) = \frac{\pi}{2}X^2 + \widetilde{R}(X),$$

for $\widetilde{R}(X) \ll X^{\frac{3}{4}}$.

In order to deduce the desired statement we observe that

$$\widetilde{R}(X) = \int_0^X R(t)dt.$$

From the simple estimate

$$O(h) + \frac{1}{h}\int_{X-h}^{X} R(t)dt \leq R(X) \leq O(h) + \frac{1}{h}\int_{X}^{X+h} R(t)dt$$

we deduce that

$$R(X) \ll h + \frac{X^{\frac{3}{4}}}{h} \tag{14}$$

Choosing $h = X^{\frac{3}{8}}$ gives the desired result. □

Remark 1.5. Note that the main term πX is nothing but the volume of the disc with radius $\sqrt{X}$. An easy bound, due to Gauß, for the remainder is $R(X) \ll \sqrt{X}$. This estimate comes by observing that $R(X)$ is bounded by the circumference of the disc, which is $2\pi\sqrt{X}$. Using the Voronoï-formula as given above one can actually show that $R(X) \ll_\epsilon X^{\frac{1}{3}+\epsilon}$. We refer to [IK, Corolarry 4.9] for a proof. The current record is

$$R(X) \ll X^{\frac{131}{416}}.$$

This estimate is due to Huxley, see [Hu]. Note that it is conjectured that $R(X) \ll_\epsilon X^{\frac{1}{4}+\epsilon}$ for all $\epsilon > 0$. This is far out of reach of current technology and essentially best possible. The latter was demonstrated by Hardy and Landau (independently).

Exercise 1.4. *Give an alternative proof of the Voronoï summation formula using Poisson summation.*

2. The hyperbolic plane

The upper half plane is given by

$$\mathbb{H} = \{z = x + iy \colon x \in \mathbb{R},\, y \in \mathbb{R}_{>0}\}.$$

The boundary is given by

$$\partial\mathbb{H} = \mathbb{R} \cup \{\infty\},$$

and we set $\overline{\mathbb{H}} = \mathbb{H} \cup \partial\mathbb{H}$.

We will start by describing some basic structural properties of $\mathbb{H}$. Here we will follow the nice exposition in [EW]. Then we turn towards the relevant global spectral theory mostly relying on [Iw].

2.1. Basic hyperbolic geometry.

2.1. Basic hyperbolic geometry. Let us briefly talk about the structure of $\mathbb{H}$. To do so we have to consider the tangent bundle

$$T\mathbb{H} = \mathbb{H} \times \mathbb{C}.$$

We equip $T_z\mathbb{H} = \{z\} \times \mathbb{C}$ with the vector space structure inherited from $\mathbb{C} = \mathbb{R}^2$. The hyperbolic Riemannian metric is given by

$$\langle \cdot, \cdot \rangle_z \colon T_z\mathbb{H} \times T_z\mathbb{H} \to \mathbb{C},\ ((z,w),(z,u)) \mapsto \frac{1}{y^2} w \cdot \overline{u}.$$

In particular we have the norms $\|(z,v)\|_z^2 = \langle (z,v),(z,v)\rangle_z = \frac{\|v\|}{y^2}$. This allows us to define a metric on $\mathbb{H}$. Given $\phi\colon [0,1]\to\mathbb{H}$ we set

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

and

$$L(\phi) = \int_0^1 \|D\phi(t)\|_{\phi(t)}dt.$$

Finally, for $z_0, z_1 \in \mathbb{H}$ we set

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

where the infimum is taken over all continuous piecewise differentiable curves ϕ with $\phi(0) = z_0$ and $\phi(1) = z_1$.

Remark 2.1. Note that this metric introduces the same topology as the euclidean metric.

Remark 2.2. We can extend the metric to $\overline{\mathbb{H}}$ in the obvious way. Practically this means that the distance between any point $z\in\mathbb{H}$ and any point $\alpha\in\partial\mathbb{H}$ is infinite.

We now introduce a group that nicely acts on $\mathbb{H}$. Indeed, this group is nothing but

$$\mathrm{SL}_2(\mathbb{R}) = \left\{ g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : \det(g) = ad-bc = 1 \right\}$$

and also $\mathrm{PSL}_2(\mathbb{R}) = \{\pm I_2\}\backslash \mathrm{SL}_2(\mathbb{R})$. Here I_2 is the identity matrix. We will need the following important matrices

$$n(x) = \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix},\ a(y) = \begin{pmatrix} y^{\frac{1}{2}} & 0 \\ 0 & y^{-\frac{1}{2}} \end{pmatrix} \text{ and } k_\theta = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

We obtain the corresponding subgroups

$$N(\mathbb{R}) = \{n(x)\colon x\in\mathbb{R}\},\ A(\mathbb{R}) = \{a(y)\colon y\in\mathbb{R}_{>0}\} \text{ and } \mathrm{SO}_2(\mathbb{R}) = \{k_\theta\colon \theta[0,2\pi)\}.$$

The action on $z\in\mathbb{H}$ is given by

$$g.z = \frac{az+b}{cz+d} \text{ where } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}).$$

Let us first convince ourselves that this is well defined. Indeed, $cz+d=0$ would contradict $z\in\mathbb{H}$ or $g\in\mathrm{SL}_2(\mathbb{R})$. Furthermore, we compute

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

so that $g.z\in\mathbb{H}$.

Remark 2.3. The action can be easily extended to $\overline{\mathbb{H}}$. For example we have

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

Lemma 2.4. *We have the following important properties:*

(1) *The action satisfies*
$$d(g.z_0, g.z_1) = d(z_0, z_1) \text{ for all } g \in \mathrm{SL}_2(\mathbb{R}) \text{ and } z_0, z_1 \in \mathbb{H}.$$
(2) *The action is transitive (i.e. for any two points $z_0, z_1 \in \mathbb{H}$ there is $g \in \mathrm{SL}_2(\mathbb{R})$ with $g.z_0 = z_1$).*
(3) *The stabilizer of $i \in \mathbb{H}$ is $\mathrm{SO}_2(\mathbb{R})$.*

Proof. To see (1) we take $g \in \mathrm{SL}_2(\mathbb{R})$ and consider the resulting map
$$g\colon \mathbb{H} \to \mathbb{H},\ z \mapsto g.z.$$
Differentiating this yields $Dg\colon T\mathbb{H} \to T\mathbb{H}$ given by
$$Dg(z,v) = (g.z, g'(z)v) = \left(\frac{az+b}{cz+d}, \frac{v}{(cz+d)^2}\right).$$
Write $(Dg)_z$ for the map $T_z\mathbb{H} \to T_{g.z}\mathbb{H}$ obtained by projection on the second component. We claim that Dg preserves the Riemann metric. This is a simple computation:
$$\begin{aligned}\langle (Dg)_z v, (Dg)_z w\rangle_{g.z} &= \left(\frac{y}{|cz+d|^2}\right)^{-2}\left(\frac{v}{(cz+d)^2}\right)\overline{\left(\frac{w}{(cz+d)^2}\right)}\\ &= y^{-2}v\overline{w} = \langle v, w\rangle_z.\end{aligned}$$
With this at hand one uses the chain rule to compute $L(g \circ \phi) = L(\phi)$ and we are done with this part of the proof.

For (2) we define $g_z = n(x)n(y)$, for $z = x+iy \in \mathbb{H}$. Observe that $g_z.i = z$.

Finally, in order to compute the stabilizer of i we take $g.i = i$. By looking at the imaginary parts we find $|cz+d| = 1$. Thus there is θ with $c = \sin(\theta)$ and $d = \cos(\theta)$. Further solving $(ai+b)/(\sin(\theta)i + \cos(\theta)) = i$ we find that $a = \cos(\theta)$ and $b = -\sin(\theta)$. This completes the proof. □

Corollary 2.5. *We have*
$$\mathbb{H} \cong \mathrm{SL}_2(\mathbb{R})/\,\mathrm{SO}_2(\mathbb{R}) \cong \mathrm{PSL}_2(\mathbb{R})/\,\mathrm{PSO}_2(\mathbb{R}).$$

Proof. The first isomorphism is given by $z \mapsto g_z$. Everything else is easy to check. □

Remark 2.6. If we define
$$T^1\mathbb{H} = \{(z,v) \in T\mathbb{H}\colon \|v\|_z = 1\},$$
then this is preserved by the action of Dg. Even more, it turns out that
$$T^1\mathbb{H} \cong \mathrm{PSL}_2(\mathbb{R}).$$
The first isomorphism is given by
$$\mathrm{PSL}_2(\mathbb{R}) \ni g \mapsto Dg(i,i) \in T^1\mathbb{H}. \tag{15}$$

Exercise 2.1. *Show that the map given in* (15) *is really an isomorphism. Furthermore, check that the action of* $\mathrm{PSL}_2(\mathbb{R})$ *on* $T^1\mathbb{H}$ *is conjugate to the action of* $\mathrm{PSL}_2(\mathbb{R})$ *on* $\mathrm{PSL}_2(\mathbb{R})$ *by left multiplication.*

Our goal is now to compute the distance function in more detail. To do so we will use the following important fact.

Lemma 2.7. *For any two points* $z, w \in \mathbb{H}$ *there is* g *such that* $g.z = i$ *and* $g.w = iy$ *for some* $y \geq 1$*.*

Proof. Without loss of generality we can assume that $z = i$. Now we consider the orbit $k_\theta.w$ for $\theta \in [0, \pi)$. Since this is compact set there is an element with maximal imaginary part. Lets call it $w' = k_{\theta_0}.w$. One can easily check that w' must be of the desired shape. □

Lemma 2.8. *We have*

$$d(z_0, z_1) = \log\left[\frac{|z-\overline{w}| + |z-w|}{|z-\overline{w}| - |z-w|}\right]. \tag{16}$$

Proof. We know that the left hand side of (16) is invariant under the action of $\mathrm{SL}_2(\mathbb{R})$. A brute force computation shows that the same is true for the right hand side. Thus, by Lemma 2.7, it is sufficient to check equality for $z_0 = i$ and $z_1 = iy$ for $y \geq 1$. It is easy to see that in this case

$$\log\left[\frac{|i+iy| + |i-iy|}{|i+iy| - |i-iy|}\right] = \log(y).$$

On the other hand we can take a path $\phi(t) = \phi_1(t) + i\phi_2(t)dt$ with $\phi_1(0) = \phi_1(1) = 0$, $\phi_2(0) = 1$ and $\phi_2(1) = y$. We compute

$$L(\phi) = \int_0^1 \|\phi'(t)\| \frac{dt}{\phi_2(t)} \geq \int_0^1 |\phi_2'(t)| \frac{dt}{\phi_2(t)} \geq \int_0^1 \phi_2'(t) \frac{dt}{\phi_2(t)} = \log(y).$$

We see that the lower bound is obtained when taking $\phi_1(t) = 0$ and $\phi_2(t) = y^t$. This shows that

$$d(i, iy) = L(\phi) = \log(y)$$

and the proof is complete. □

It is often better to work with the $\mathrm{SL}_2(\mathbb{R})$-invariant function u defined by

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

One easily checks that

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

Indeed, we have

$$\begin{aligned}\cosh(d(i,iy)) = \cosh(\log(y)) &= \frac{y}{2}+\frac{1}{2y}\\ &= 1+2\frac{(y-1)^2}{4y} = 1+2\frac{|i-iy|^2}{4\operatorname{Im}(i)\operatorname{Im}(iy)} = 1+2u(i,iy).\end{aligned}$$

After verifying invariance once again we are done.

Next we define the measure μ on $\mathbb{H}$ by

$$\int_{\mathbb{H}} f(z)d\mu(z) = \int_0^\infty\int_{-\infty}^\infty f(x+iy)\frac{dxdy}{y^2}.$$

It is straight forward to see that this measure is $\mathrm{SL}_2(\mathbb{R})$-invariant. Indeed one only computes the Jacobian of $z \mapsto g.z$.

Remark 2.9. With suitable normalizations the measure μ arises from the Haar measure of $\mathrm{SL}_2(\mathbb{R})$ under the identification $\mathbb{H} \cong \mathrm{SL}_2(\mathbb{R})/\mathrm{SO}_2(\mathbb{R})$.

So far we have only worked in standard rectangular coordinates. However, we can also define (geodesic) polar coordinates. Indeed, we can write

$$x+iy = k_\varphi e^{-r}i,$$

for $\varphi \in [0,\pi)$ and $r \geq 0$. One can compute that

$$y = (\cosh(r)+\sinh(r)\cos(2\varphi))^{-1} \text{ and } x = y\sinh(r)\sin(2\varphi). \tag{17}$$

$$d\mu(z) = 2\sinh(r)\cdot drd\varphi. \tag{18}$$

If we further write $\cosh(r) = 1+2u$, then

$$d\mu(z) = 4dud\varphi.$$

We turn towards studying the motions generated by elements of $\mathrm{SL}_2(\mathbb{R})$. An important feature is the fixed point structure. For $g \neq \pm I_2$ we have $g.z = z$ if and only if

$$cz^2+(d-a)z-b = 0.$$

Solving this shows that the fixed points are given by $b/(d-a) \in \partial\mathbb{H}$ if $c=0$ and by

$$\frac{1}{2c}\left[a-d\pm\sqrt{(a+d)^2-4}\right].$$

We read of the following classification:

- g has one fixed point in $\partial\mathbb{H}$. This happens exactly when $|\operatorname{Tr}(g)| = 2$. We say that g is parabolic.
- g has two distinct fixed points in $\partial\mathbb{H}$. This happens exactly when $|\operatorname{Tr}(g)| > 2$. We say that g is hyperbolic.
- g has one fixed point in $\mathbb{H}$ (and one in the negative half plane). This happens exactly when $\operatorname{Tr}(g) < 2$. We say that g is elliptic.

This is obviously invariant under conjugation, so that it makes sense to say that a conjugacy class is parabolic, hyperbolic or elliptic.

Note that each conjugacy class

$$\{g\} = \{hgh^{-1} \colon h \in \mathrm{SL}_2(\mathbb{R})\}$$

meets one of the groups $N(\mathbb{R})$, $A(\mathbb{R})$ or $\mathrm{SO}_2(\mathbb{R})$. It is easy to see that elements in $N(\mathbb{R})$ are parabolic. Their action is nothing but (horizontal) translation. The elements in $A(\mathbb{R})$ are hyperbolic and they act by dilation. Finally, elements in $\mathrm{SO}_2(\mathbb{R})$ are elliptic and they act by rotations.

Exercise 2.2. *Let $\pm I_2 \neq g, h \in \mathrm{SL}_2(\mathbb{R})$. Show that g and h commute if and only if they have the same set of fixed points. Conclude that, up to the central elements $\pm I_2$, the centralizer $C_g = \{h \in \mathrm{SL}_2(\mathbb{R}) \colon gh = hg\}$ of g in $SL_2(\mathbb{R})$ is given by all elements h with the same fixed points (in $\overline{\mathbb{H}}$).*

2.2. **Global spectral theory.** We turn towards the global spectral theory (i.e. the spectral theory on $\mathbb{H}$). The Laplace-Beltrami operator on $\mathbb{H}$ is given by

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

One can that Δ is invariant. More precisely, for all $g \in \mathrm{SL}_2(\mathbb{R})$ we have

$$[\Delta g f](z) = \Delta f(g.z), \text{ where } [gf](z) = f(g.z).$$

In (geodesic) polar coordinates we have

$$\Delta = -\left[\frac{\partial^2}{\partial r^2} + \frac{1}{\tanh(r)}\frac{\partial}{\partial r} + \frac{1}{4\sinh(r)^2}\frac{\partial^2}{\partial \varphi^2}\right].$$

If we further insert $\cosh(r) = 1 + 2u$ we obtain

$$\Delta = -\left[u(u+1)\frac{\partial^2}{\partial u^2} + (2u+1)\frac{\partial}{\partial u} + \frac{1}{16u(u+1)}\frac{\partial^2}{\partial \varphi^2}\right].$$

We first consider the equation $\Delta f = \lambda f$ with $\lambda \in \mathbb{C}$ and $f \colon \mathbb{H} \to \mathbb{C}$. We can guess the following pairs of solutions

$$f(z) = \mathrm{Im}(z)^s \text{ and } f(z) = \mathrm{Im}(z)^{1-s}$$

for $\lambda = s(1-s)$ and $s \neq \frac{1}{2}$. Note that for $s = \frac{1}{2}$ (i.e. $\lambda = \frac{1}{4}$) one has to replace the second solution by $\mathrm{Im}(z)^{\frac{1}{2}} \log(\mathrm{Im}(z))$.

Next we consider functions f that depend only on the variable u. As above we put $\lambda = s(1-s)$. We can write the equation $\Delta f = s(1-s)f$ as

$$u(u+1)f'' + (2u+1)f' + s(1-s)f = 0 \tag{19}$$

We define

$$G_s(u) = \frac{1}{4\pi}\int_0^1 t^{s-1}(1-t)^{s-1}(t+u)^{-s}dt \tag{20}$$

Note that in view of the integral representation

$$ {}_2F_1(\alpha,\beta;\gamma;z) = \frac{\Gamma(\gamma)}{\Gamma(\beta)\Gamma(\gamma-\beta)}\int_0^1 t^{\beta-1}(1-t)^{\gamma-\beta-1}(1-tz)^{-\alpha}dt $$

we can express $G_s(u)$ as the Gauß hypergeometric function[5]

$$ G_s(u) = \frac{\Gamma(s)^2}{4\pi\Gamma(2s)}u^{-s}{}_2F_1(s,s;2s;\frac{1}{u}). $$

Remark 2.10. That G_s solves the required eigenvalue equation can be seen in a variety of ways. For example, we can put $\sigma = \cosh(r/2)^2$. If we view G_s as a function of σ, then the eigenvalue equation gets

$$ \sigma(1-\sigma)G_s'' + (1-2\sigma)G_s' - s(1-s) = 0. $$

This looks like the defining equation for the Gauß hypergeometric function. However, we need a solution that is regular at ∞ (and not at 1). We choose the Kummer solution given by

$$ C_s\sigma^{-s}{}_2F_1(s,s;2s;\frac{1}{\sigma}). $$

At this point we recall the identity $2\cosh(x/2)^2 - 1 = \cosh(x)$. In particular, we find

$$ \sigma = u+1. $$

Taking $C_s = \frac{\Gamma(s)^2}{4\pi\Gamma(2s)}$ and using identities between hypergeometric functions produces the solution $G_s(u)$ given above. Below we will see that we have chosen the correct solution and normalized it correctly.

Lemma 2.11. *The integral defining $G_s(u)$ converges absolutely for $\mathrm{Re}(s) = \sigma > 0$ and it defines a function for $u \in \mathbb{R}_+$ which solves (19). Moreover we have*

$$ G_s(u) = \frac{1}{4\pi}\log(\frac{1}{u}) + O_s(1) \text{ as } u \to 0, \tag{21} $$

$$ G_s'(u) = -(4\pi u)^{-1} + O_s(1) \text{ as } u \to 0 \text{ and} \tag{22} $$

$$ G_s(u) \ll u^{-\sigma} \text{ as } u \to \infty. \tag{23} $$

Proof. We establish the first asymptotic as follows. Put

$$ \nu = (|s|+1)^{-1}u \text{ and } \eta = (|s|+1)^{-1}. $$

We can assume that u is sufficiently small so that $0 < \nu < \eta < 1$. We split the integral in ranges and estimate

$$ \int_0^\nu t^{s-1}(1-t)^{s-1}(t+u)^{-s}dt \ll u^\sigma \int_0^\nu t^{\sigma-1}dt \ll 1. $$

[5]The Gauß hypergeometric function ${}_2F_1(\alpha,\beta;\gamma,z)$ is one solution of the differential equation

$$ z(1-z)F'' - ((\alpha+\beta+1)z-\gamma)F' - \alpha\beta F = 0. $$

Similarly we have

$$\int_\eta^1 t^{s-1}(1-t)^{s-1}(t+u)^{-s}dt \ll \int_\eta^1 t^{-1}(1-t)^{\sigma-1}(1+\frac{u}{t})^{-\sigma}dt \ll \int_\eta^1 (1-t)^{\sigma-1}dt \ll 1.$$

The main term will come from the remaining piece of the integral. Here we will use the expansion

$$\left(\frac{t(1-t)}{t+u}\right)^{s-1} = \left(1-\frac{u+t^2}{u+t}\right)^{s-1} = 1+O_s\left(\frac{u+t^2}{u+t}\right).$$

With this at hand we compute

$$\begin{aligned}\int_\nu^\eta t^{s-1}(1-t)^{s-1}(t+u)^{-s}dt &= \int_\nu^\eta \frac{dt}{t+u}+O\left(\int_\nu^\eta \frac{u+t^2}{(u+t)^2}dt\right)\\ &= \log\left(\frac{u+\eta}{u+\nu}\right)+O(1) = \log(1/u)+O(1).\end{aligned}$$

The asymptotic for the derivative is obtained similarly and we omit the details. The last claim follows after trivially estimating

$$\begin{aligned}|G_s(u)| &= u^{-\sigma}\frac{1}{4\pi}\int_0^1 (t(1-t))^{\sigma-1}(t/u+1)^{-\sigma}dt\\ &\leq u^{-\sigma}\frac{1}{4\pi}\int_0^1 t^{\sigma-1}(1-t)^{\sigma-1}dt = u^{-\sigma}\frac{\Gamma(\sigma)^2}{4\pi\Gamma(2\sigma)},\end{aligned}$$

for $u>0$. Here we have recognized the Beta-integral. ☐

Working more carefully (or alternatively using the series expansion of ${}_2F_1(s,s;2s;u^{-1})$)) one can obtain the refined expansion

$$G_s(u) = \frac{1}{4\pi}\log(\frac{1}{u}) - \frac{1}{2\pi}(\Psi(s)+\gamma)+o(1), \tag{24}$$

where $\Psi(s)$ is the digamma function and γ is Euler's constant.[6]

Remark 2.12. One verifies by partial integration that the function

$$F_s(u) = \frac{1}{\pi}\int_0^\pi (2u+1+2\sqrt{u(u+1)}\cos(\theta))^{-s}d\theta \tag{25}$$

also solves (19). This solution can be written as

$$F_s(u) = F(s,1-s;1;u).$$

Note that analogously one can find F_s as solution to a (19) which is regular at 0. From the integral we directly see that $F_s(u)=1$. In particular, F_s and G_s form a complete system of linearly independent solutions to (19).

[6]One compares this to [Bo, (4.14)]. Doing so one should remember that $\sigma = u+1$.

For $\mathrm{Re}(s) > 1$ we define the integral operator

$$[R_s f](z) = \int_{\mathbb{H}} G_s(u(z,w)) f(w) d\mu(w).$$

We claim that R_s is the resolvent, so that $G_s(u(z,w))$ is the Green function on $\mathbb{H}$. This will follow from the following theorem.

Theorem 2.13. *For $f \colon \mathbb{H} \to \mathbb{C}$ smooth and bounded we have*

$$(\Delta - s(1-s))R_s f = f.$$

Proof. A crucial ingredient in the proof is the identity

$$(\Delta - s(1-s))R_s f(z) = \int_{\mathbb{H}} G_s(u(z,w))(\Delta - s(1-s)) f(w) d\mu(w). \tag{26}$$

We will not prove this here.[7]

Let U_ϵ be a small euclidean disc around z and let V_ϵ be its exterior in $\mathbb{H}$. Using (26) we write

$$\begin{aligned}(\Delta - s(1-s))R_s f(z) = \int_{U_\epsilon} & G_s(u(z,w))(\Delta - s(1-s)) f(w) d\mu(w) \\ &+ \int_{V_\epsilon} G_s(u(z,w))(\Delta - s(1-s)) f(w) d\mu(w).\end{aligned}$$

We need to show that this equals $f(z)$ and we do so by considering the limit of the right hand side as $\epsilon \to 0$. We will frequently use the asymptotic behavior of G_s at 0. First, note that

$$\lim_{\epsilon \to 0} \int_{U_\epsilon} G_s(u(z,w))(\Delta - s(1-s)) f(w) d\mu(w) = 0.$$

We still need to consider the remaining integral over V_ϵ. To treat this we recall Green's formula:

$$\int_{V_\epsilon} (g\Delta_{\mathrm{euc}} f - f\Delta_{\mathrm{euc}} g) d\mu_{\mathrm{euc}} = \int_{\partial V} (g\frac{\partial f}{\partial n} - f\frac{\partial g}{\partial n}) dl \tag{27}$$

where $\frac{\partial}{\partial n}$ is the outer normal derivative. Since $[\Delta_w - s(1-s)]G_s(u(z,w)) = 0$ we get

$$\int_{V_\epsilon} G_s(u(z,w))(\Delta - s(1-s)) f(w) d\mu(w) = \int_{\partial U_\epsilon} \left(G_s \frac{\partial f}{\partial n} - f \frac{\partial G_s}{\partial n} \right) dl.$$

We now take the limit $\epsilon \to 0$ and find

$$\lim_{\epsilon \to 0} \int_{V_\epsilon} G_s(u(z,w))(\Delta - s(1-s)) f(w) d\mu(w) = -\frac{1}{2\pi} \lim_{\epsilon \to 0} \int_{\partial U_\epsilon} f(w) \frac{\partial \log(|z-w|)}{\partial n} dl.$$

[7] For regular kernels this is a formal consequence of the symmetry of Δ. However G_s is singular on the diagonal, so that this argument does not work. Instead one has to use differential operators coming from the Lie-algebra of $\mathrm{SL}_2(\mathbb{R})$. This enables one to use invariance of Δ.

The remaining integral can be written in euclidean polar coordinates as

$$-\frac{1}{2\pi}\lim_{\epsilon\to 0}\int_{\partial U_\epsilon} f(w)\frac{\partial \log(|z-w|)}{\partial n}dl = \lim_{\epsilon\to 0}\frac{1}{2\pi\epsilon}\int_{\partial U_\epsilon} f(w)dl = f(z).$$

This completes the argument. □

On the other hand we can consider integral operators associated to a kernel $k\colon \mathbb{H}\times\mathbb{H}\to\mathbb{C}$ defined by

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

It turns out that T_k is invariant (i.e. $[T_k(gf)](z) = T_k f(g.z))$ if and only if $k(gz,gw) = k(z,w)$. A function with this property will be called a point-pair invariant. By abuse of notation we can write a point-pair invariant as[8]

$$k(z,w) = k(u(z,w)).$$

We start by proving some basic results concerning invariant integral operators.

Lemma 2.14. *Let $k(z,w)$ be a smooth point-pair invariant on $\mathbb{H}\times\mathbb{H}$. Then*

$$\Delta_z k(z,w) = \Delta_w k(z,w).$$

Proof. We work in geodesic polar coordinates with origin w and compute

$$\Delta_z k(z,w) = u(u+1)k''(u) + (2u+1)k'(u).$$

Using geodesic coordinates with origin z to compute $\Delta_w k(z,w)$ gives the same result and we are done. □

Theorem 2.15. *Let T_k be an integral operator associated to a smooth point-pair invariant $k(z,w)$. Then the invariant integral operator T_k commutes with Δ.*

Proof. Let $f\in\mathcal{C}_0^\infty(\mathbb{H})$.[9] We have

$$\begin{aligned}[\Delta T_k f](z) = \int_{\mathbb{H}} \Delta_z k(z,w)f(w)d\mu(w) &= \int_{\mathbb{H}} \Delta_w k(z,w)f(w)d\mu(w)\\ &= \int_{\mathbb{H}} k(z,w)\Delta_w f(w)d\mu(w) = [T_k\Delta f](z).\end{aligned}$$

Here we used partial integration to move Δ_w from k to f. □

Definition 2.1. We call a function $f(z,w)$ radial at w, if as a function in z it depends only on the distance of z to w.

[8]We use the letter k for the point-pair invariant $k\colon\mathbb{H}\times\mathbb{H}\to\mathbb{C}$ as well as for the corresponding function $k\geq\mathbb{R}_{\geq 0}\to\mathbb{C}$. It should always be clear from the context what is meant.

[9]The subscript 0 indicates that f vanishes on the boundary of $\mathbb{H}$.

Note that point-pair invariants are radial at all points. But this is not necessarily the case in general. We define the stabilizer of w in $\mathrm{SL}_2(\mathbb{R})$ as

$$G_w = \{g \in \mathrm{SL}_2(\mathbb{R}) \colon g.w = w\}. \tag{28}$$

This is a compact group, which we equip with the Haar measure. For $f \colon \mathbb{H} \to \mathbb{C}$ we define

$$f_w(z) = \frac{1}{\mathrm{Vol}(G_w)} \int_{G_w} f(g.z) dg.$$

If $\sigma.i = w$, then we have $G_w = \sigma\,\mathrm{SO}_2(\mathbb{R})\sigma^{-1}$. In particular, we can write

$$f_w(z) = \frac{1}{2\pi} \int_0^{2\pi} f(\sigma k_\theta \sigma^{-1}.z) d\theta.$$

We call the operator $f \mapsto f_w$ the mean-value operator.

Lemma 2.16. *The function $f_w(z)$ is radial at w and satisfies*

$$f_z(z) = f(z).$$

Furthermore, we have

$$[T_k f](z) = [T_k f_z](z) \tag{29}$$

for an invariant integral operator T_k.

Proof. Clearly we have $f_z(z) = f(z)$. To see that $f_w(z)$ is radial at w we take z_0 and z_1 with $u(z_1, w) = u(z_0, w)$. We need to show that $f_w(z_0) = f_w(z_1)$. To do so we take $g_1 \in G_w$ with $g_1.z_0 = z_1$. The existence of g_1 follows from (the proof of) Lemma 2.7. Now we simply compute

$$\begin{aligned} f_w(z_0) = \frac{1}{\mathrm{Vol}(G_w)} \int_{G_w} f(g.z_0) dg &= \frac{1}{\mathrm{Vol}(G_w)} \int_{G_w} f(gg_1.z) dg \\ &= \frac{1}{\mathrm{Vol}(G_w)} \int_{G_w} f(g.z_1) dg = f_w(z_1). \end{aligned}$$

To see the final property we compute

$$
\begin{aligned}
[T_k f_z](z) &= \int_{\mathbb{H}} k(z,w) f_z(w) \qquad (30)\\
&= \frac{1}{\mathrm{Vol}(G_w)} \int_{\mathbb{H}} \int_{G_z} k(z,w) f(g.w) dg d\mu(w)\\
&= \frac{1}{\mathrm{Vol}(G_w)} \int_{G_z} \int_{\mathbb{H}} k(z,w) f(g.w) d\mu(w) dg\\
&= \frac{1}{\mathrm{Vol}(G_w)} \int_{G_z} \int_{\mathbb{H}} k(z,g^{-1}.w) f(w) d\mu(w) dg\\
&= \frac{1}{\mathrm{Vol}(G_w)} \int_{G_z} \int_{\mathbb{H}} k(g.z,w) f(w) d\mu(w) dg\\
&= \int_{\mathbb{H}} k(z,w) f(w) d\mu(w) = T_k f(z).
\end{aligned}
$$

This is exactly what we wanted and the proof is complete. □

Lemma 2.17. *Let $\lambda = s(1-s) \in \mathbb{C}$ and $w \in \mathbb{H}$ be fixed. Then there is a unique function*

$$\mathbb{H} \ni z \mapsto \omega_s(z,w) \in \mathbb{C} \qquad (31)$$

such that

1. *$\omega_s(z,w)$ is radial at w;*
2. *We have $\Delta_z \omega_s(z,w) = \lambda \omega_s(z,w)$; and*
3. *$\omega_s(w,w) = 1$.*

This function is given by

$$\omega_s(z,w) = F_s(u(z,w))$$

where $F_s(u) = {}_2F_1(s, 1-s; 1; u)$.

Proof. By (1) we can write $\omega(z,w) = \phi(u(z,w))$ for some function ϕ. Computing (2) in polar coordinates leads to the second order ODE given in (19). As discussed in Remark 2.12 we can thus express ϕ as a linear combination of F_s and G_s. Finally, we note that (3) implies $\phi(0) = 1$, so that we must have $\phi = F_s$ as desired. □

Corollary 2.18. *Suppose $f\colon \mathbb{H} \to \mathbb{C}$ satisfies $\Delta f = s(1-s)f$, then we have*

$$f_w(z) = \omega_s(z,w) f(w).$$

Proof. We note that f_w satisfies properties (1) and (2) from Lemma 2.17. Uniqueness then implies that f_w must be a scalar multiple of $\omega_s(\cdot, w)$. □

Theorem 2.19. *Let $f\colon \mathbb{H} \to \mathbb{C}$ satisfy $\Delta f = s(1-s)f$ and let $k \in \mathcal{C}_c^\infty(\mathbb{R}_{\geq 0})$. Then there is $\Lambda = \Lambda(k,s)$ with*

$$T_k f = \Lambda f.$$

Proof. Using Lemma 2.16 and the definition of T_k we compute

$$[T_k f](z) = [T_k f_z](z) = \int_{\mathbb{H}} k(z,w) f_z(w) d\mu(w) = \underbrace{\int_{\mathbb{H}} k(z,w)\omega_s(w,z) d\mu(w)}_{=\Lambda} f(z).$$

This completes the proof. □

Remark 2.20. The theorem above is very important. It states that an eigenfunction of Δ is automatically an eigenfunction of all (sufficient regular) invariant integral operators.

Exercise 2.3. *Prove the contrary of Theorem 2.19. More precisely, show that if $T_k f = \Lambda_k f$ for all $k \in \mathcal{C}_c^\infty(\mathbb{R}_{\geq 0})$, then there is $\lambda \in \mathbb{C}$ with $\Delta f = \lambda f$.*

A key point in Theorem 2.19 is that Λ depends on the Δ-eigenvalue $s(1-s)$ and on k but **not** on f. This allows us to test our integral operators against suitable f's. This allows us to establish the following result.

Definition 2.2 (Selberg–Harish-Chandra Transform)**.** Given $k \in \mathcal{C}_c^\infty(\mathbb{R}_{\geq 0})$ we define the Selberg–Harish-Chandra transform h of k using three steps:

$$\begin{aligned} q(v) &= \int_v^\infty k(u)(u-v)^{-\frac{1}{2}} du, \\ g(r) &= 2q\left(\sinh(r/2)^2\right), \text{ and} \\ h(t) &= \int_{-\infty}^\infty g(r) e^{irt} dr. \end{aligned}$$

Theorem 2.21. *Let $k \in \mathcal{C}_c^\infty(\mathbb{R}_{\geq 0})$ and let $f\colon \mathbb{H} \to \mathbb{C}$ be such that $\Delta f = s(1-s) f$. Then we have*

$$T_k f = h(t) f \text{ for } t = -\frac{i}{2} - is$$

(i.e. $s = \frac{1}{2} + it \in \mathbb{C}$), where h is the Selberg–Harish-Chandra transform of k.

Proof. We take $f(z) = \mathrm{Im}(z)^s$. Recall that $\Delta f = s(1-s) f$. By Theorem 2.19 we have $T_k f = \Lambda f$. We need to show that $\Lambda = h(t)$. To do so we observe that

$$\begin{aligned} \Lambda = \Lambda f(i) = [T_k f](i) &= \int_{\mathbb{H}} f(w) k(i,w) d\mu(w) \\ &= 2\int_0^\infty \int_0^\infty y^{s-2} k\left(\frac{x^2 + (y-1)^2}{4y}\right) dx dy. \end{aligned}$$

After making the change of variables $x = 2\sqrt{uy}$ we have

$$\begin{aligned} \Lambda &= 2\int_0^\infty \int_0^\infty y^{s-\frac{3}{2}} k\left(u + \frac{(y-1)^2}{4y}\right) \frac{du}{\sqrt{u}} dy \\ &= 2\int_0^\infty q\left(\frac{(y-1)^2}{4y}\right) y^{s-\frac{1}{2}} \frac{dy}{y}. \end{aligned}$$

We now change $y = e^r$ and recall $s = \frac{1}{2} + it$ to obtain

$$\Lambda = 2 \int_{-\infty}^{\infty} q\left(\frac{(e^r-1)^2}{4e^r}\right) e^{irt} dr$$
$$= \int_{-\infty}^{\infty} g(r) e^{irt} dr = h(t).$$

This completes the proof. $\square$

Another handy incarnation of the Selberg–Harish-Chandra transform is given by

$$h(t) = 4\pi \int_0^{\infty} k(u) F_s(u) du. \tag{32}$$

To see this we test our integral operator as follows

$$h(t) = h(t)\omega_s(i,i) = [T_k \omega(\cdot, i)](i)$$
$$= \int_{\mathbb{H}} k(u(i,z)) F_s(u(z,i)) d\mu(z) = 4\pi \int_0^{\infty} k(u) F_s(u) du.$$

In the last step we have computed the integral in polar coordinates.

The integral transform in (32) is related to the the Mehler–Fock transform and it can be inverted. The inversion formula is given by

$$k(u) = \frac{1}{4\pi} \int_{-\infty}^{\infty} F_{\frac{1}{2}+it}(u) h(t) \tanh(\pi t) t dt. \tag{33}$$

We can also invert this transformation step

$$g(r) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-irt} h(t) dt,$$
$$q(v) = \frac{1}{2} g(2\log(\sqrt{v+1} + \sqrt{v})) \text{ and} \tag{34}$$
$$k(u) = -\frac{1}{\pi} \int_u^{\infty} (v-u)^{-\frac{1}{2}} dq(v). \tag{35}$$

The first two steps are relatively simple. We first apply Fourier inversion and then we reverse the substitution. To see that the final Riemann-Stieltjes integral inverts the transform $k \rightsquigarrow q$ requires a little argument.

Lemma 2.22. *For $k \in \mathcal{C}_c^{\infty}(\mathbb{R}_{\geq 0})$ the inversion formula given in (35) holds.*

Proof. Suppose that the support of k is contained in the interval $[0, A]$. In particular $q(v) = 0$ for $v > A$. It is sufficient to consider $0 \leq u \leq A$. Note that

$$q'(v) = \int_v^{\infty} \frac{k'(u)}{(u-v)^{\frac{1}{2}}} du = -2 \int_v^{A} (u-v)^{\frac{1}{2}} k''(u) du.$$

The first step follows from the Leibniz rule and the second equality is integration by parts.

We will now consider the integral

$$I_\epsilon = -\frac{1}{\pi}\int_{u+\epsilon}^{\infty}(v-u)^{-\frac{1}{2}}dq(v) = -\frac{1}{\pi}\int_{u+\epsilon}^{A}\frac{q'(v)}{(v-u)^{\frac{1}{2}}}dv.$$

Using the formula for the derivative above we compute

$$\begin{aligned} I_\epsilon &= \frac{2}{\pi}\int_{u+\epsilon}^{A}\int_{v}^{A}\frac{(x-v)^{\frac{1}{2}}}{(v-u)^{\frac{1}{2}}}k''(x)dxdv \\ &= \frac{2}{\pi}\int_{u+\epsilon}^{A}\int_{u+\epsilon}^{x}\frac{(x-v)^{\frac{1}{2}}}{(v-u)^{\frac{1}{2}}}k''(x)dvdx \\ &= \frac{2}{\pi}\int_{u+\epsilon}^{A}k''(x)\int_{u+\epsilon}^{x}\frac{(x-v)^{\frac{1}{2}}}{(v-u)^{\frac{1}{2}}}dvdx. \end{aligned}$$

Now we use the formula

$$\int_u^x (v-u)^{-\frac{1}{2}}(x-v)^{\frac{1}{2}}dv = (x-u)\int_0^1 t^{-\frac{1}{2}}(1-t)^{\frac{1}{2}}dt = (x-u)B(1/2,3/2) = \frac{\pi}{2}(x-u).$$

In particular, we can take the limit $\epsilon \to 0$ and obtain

$$-\frac{1}{\pi}\int_u^{\infty}(v-u)^{-\frac{1}{2}}dq(v) = \lim_{\epsilon\to 0} I_\epsilon = \int_u^A k''(x)(x-u)dx = -\int_u^A k'(x)dx = k(u).$$

This completes the argument. □

It will be useful to get a feeling for the regularity properties of the functions involved in this transforms. We start with an easy lemma.

Lemma 2.23. *Let $k \in \mathcal{C}_c^2(\mathbb{R}_{\geq 0})$. Then we have*

1. $q \in \mathcal{C}_c^2(\mathbb{R}_{\geq 0})$;
2. $g \in \mathcal{C}_c^2(\mathbb{R}_{\geq 0})$ *even; and*
3. $h \in \mathcal{C}^\infty(\mathbb{R})$ *even with* $h(t) \ll t^{-2}$.

One sees that h actually defines a holomorphic function on $\mathbb{C}$. This is because k is compactly supported.

Proof. That q is compactly supported is clear. Furthermore, by the Leibniz rule for parameter integrals we compute

$$q'(v) = \int_v^\infty \frac{k'(u)}{(u-v)^{\frac{1}{2}}}du \text{ and } q''(v) = \int_v^\infty \frac{k''(u)}{(u-v)^{\frac{1}{2}}}du.$$

This implies that $q \in \mathcal{C}_c^2(\mathbb{R}_{\geq 0})$. The second statement concerning g is clear. The properties of h are also easy to see. For example, the growth condition is obtained by applying integration by parts twice:

$$h(t) = -\frac{1}{t^2}\int_{\mathbb{R}} e^{irt}g''(r)dr \ll_g t^{-2}.$$

This completes the proof. □

Remark 2.24. The inversion formula holds for a wider class of functions h. Indeed it is sufficient to assume

- $h(t)$ is even;
- $h(t)$ is holomorphic in the strip $|\operatorname{Im}(t)| \le \frac{1}{2} + \epsilon$; and
- In the above strip we have $h(t) \ll (|t| + 1)^{-2-\epsilon}$.

This also gives rise to a bigger class of admissible kernel functions $k(u)$.

Exercise 2.4. *Show that if $k \in \mathcal{C}_c^2(\mathbb{R}_{\ge 0})$, then the Selberg–Harish-Chandra transform h of k even satisfies*

$$h(t) \ll t^{-\frac{5}{2}}.$$

To do so one should first prove that for $0 < \alpha < 1$ and $f \in \mathcal{C}_c(\mathbb{R})$ with $|f(x_1) - f(x_2)| \ll |x_1 - x_2|^\alpha$ one has $\int_{\mathbb{R}} f(x)e^{irx}dx \ll r^{-\alpha}$ as $r \to \infty$.

3. Fuchsian groups

We now turn towards the theory of Fuchsian groups. There are many good references for this. We use [Bo, Si].

We first note that $\mathrm{Mat}_{2\times 2}(\mathbb{R}) \cong \mathbb{R}^4$ carries an inner product given by

$$\langle g, h\rangle = \mathrm{Tr}(gh^\top).$$

The corresponding norm $\|g\|^2 = \langle g, g\rangle$ is the Frobenius norm and satisfies $\|gh\| \le \|g\| \cdot \|h\|$. The group $\mathrm{SL}_2(\mathbb{R})$ inherits a metric topology using the embedding $\mathrm{SL}_2(\mathbb{R}) \subseteq \mathrm{Mat}_{2\times 2}(\mathbb{R})$.

Definition 3.1. A group $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$ is called discrete if the induced topology on Γ is discrete.

Remark 3.1. Similarly we define discrete groups of $\mathrm{PSL}_2(\mathbb{R})$. These are called Fuchsian groups.

Given $\Gamma \in \mathrm{SL}_2(\mathbb{R})$ we define $\overline{\Gamma}$ to be its image in $\mathrm{PSL}_2(\mathbb{R})$. Note that Γ is discrete if and only if $\overline{\Gamma}$ is discrete.

Definition 3.2. Let X be a Hausdorff topological space (e.g. $X = \mathbb{H}$) and let Γ be a group of homeomorphisms acting on X (e.g. $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$). We say that Γ acts discontinuously on X if for any point $z \in X$ and any compact set $Y \subseteq X$ we have

$$\sharp\{\gamma \in \Gamma \colon \gamma.z \in Y\} < \infty.$$

Remark 3.2. If $\Gamma \subseteq \mathrm{PSL}_2(\mathbb{R})$ acts discontinuously on $\mathbb{H}$, then the quotient $\Gamma\backslash\mathbb{H}$ is a well defined metric space. We call $\Gamma\backslash\mathbb{H}$ an orbifold. Note that the quotient is only smooth if Γ acts without fixed points. Recall that only elliptic elements have fixed points in $\mathbb{H}$.

Proposition 3.3. *A subgroup $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$ is discrete if and only if $\overline{\Gamma}$ acts discontinuously on $\mathbb{H}$. (We can also say $\overline{\Gamma}$ is Fuchsian if and only if $\overline{\Gamma}$ acts discontinuously.)*

Proof. If Γ is discrete, then the orbit $\Gamma.z$ is discrete. Since the intersection of a discrete and compact set finite the group Γ must act discontinuously.

To see the opposite direction we first claim that there is $z_0 \in \mathbb{H}$ which is not fixed by any non-trivial element in $\overline{\Gamma}$. To see this we take $\gamma_0 \neq \pm 1$ and suppose that $\gamma_0.w = w$ for $w \in \mathbb{H}$. We compute

$$d(\gamma_0.z, z) \leq d(\gamma_0.z, \gamma_0.w) + d(\gamma_0.w, z) = 2d(z, w).$$

This shows is that, if $w \in B_r(z)$ is fixed by some non-trivial element $\gamma_0 \in \Gamma$, then $\gamma_0.z \in B_{2r}(z)$. Thus we have

$$\sharp\{w \in B_r(z) \colon w \text{ is fixed by some non-trivial element in } \overline{\Gamma}\} \leq \sharp(\Gamma.z \cap B_{2r}(z)) < \infty.$$

After excluding these finitely many fixed points we easily pick a point z_0 with the desired property. This shows the claim.

Returning to the actual proof we assume that Γ is not discrete. Then there exists a sequence $\gamma_n \to 1$ consisting of distinct elements. This ensures that $(\gamma_n.z_0)_{n\in\mathbb{N}}$ has only distinct elements. Since $\gamma_n.z_0 \to z_0$, we obtain a contradiction to the discontinuity of the action. □

While we will usually start with a discrete subgroup $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$ and then work with the quotient $\Gamma\backslash\mathbb{H}$, the following important theorem of Hopf tells us that we could similarly work with hyperbolic surfaces X. The place of Γ is then taken by the fundamental group of X.

Theorem 3.4. *For any hyperbolic surface X there is a Fuchsian group $\overline{\Gamma}$ with no elliptic elements and a $\overline{\Gamma}$-invariant Riemannian covering map $\pi\colon \mathbb{H} \to X$ realizing the isometry $X \cong \overline{\Gamma}\backslash\mathbb{H}$.*

In order to practically work (and visualize) the quotients $\Gamma\backslash\mathbb{H}$ we have to introduce the concept of a fundamental domain. We do so in two steps.

Definition 3.3. A (Borel) measurable set $\mathcal{G} \subseteq \mathbb{H}$ such that $\sharp(\mathcal{G} \cap \Gamma.z) = 1$ for all $z \in \mathbb{H}$ is called a fundamental set. We define

$$\mathrm{Vol}(\Gamma\backslash\mathbb{H}) = \mu(\mathcal{G}), \tag{36}$$

where μ is the hyperbolic measure on $\mathbb{H}$.

Note that the fundamental sets exist and that the volume $\mathrm{Vol}(\Gamma\backslash\mathbb{H})$ is well defined (i.e. independent of the choice of $\mathcal{G}$).

Definition 3.4. A fundamental domain $\mathcal{F}_\Gamma \subseteq \mathbb{H}$ for a discrete subgroup $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$ is a closed region such that

(1) $\Gamma\mathcal{F}_\Gamma = \bigcup_{\gamma\in\Gamma} \gamma.\mathcal{F}_\Gamma = \mathbb{H}$;
(2) For each $\gamma \in \Gamma \setminus \{\pm 1\}$ the interiors of $\mathcal{F}_\Gamma$ and $\gamma.\mathcal{F}_\Gamma$ do not intersect;
(3) $\mathcal{F}_\Gamma$ differs from a fundamental set by a set of measure 0 (i.e. a null set).

Remark 3.5. The final condition ensures that $\mathrm{Vol}(\Gamma\backslash\mathbb{H}) = \mu(\mathcal{F}_\Gamma)$.

To see that (nice) fundamental regions exist we introduce the Dirichlet domain

$$\mathcal{D}_w = \{z \in \mathbb{H} \colon d(z,w) \leq d(z,\gamma w) \text{ for all } \gamma \in \Gamma\}.$$

We have the following result.

Lemma 3.6. *Let w be not the fixed point of an elliptic element in Γ. Then we have the following:*

(1) *$\mathcal{D}_w$ is convex and bounded by a union of geodesics.*
(2) *$\mathcal{D}_w$ is a fundamental domain for Γ.*

Proof. We first ketch the proof of (1). Note that we can write

$$\mathcal{D}_w = \bigcap_{\gamma\in\Gamma} \underbrace{\{z \in \mathbb{H} \colon d(z,w) \leq d(z,\gamma w)\}}_{=H_\gamma(w)}.$$

We claim that $H_\gamma(w)$ are (hyperbolic) half planes and this will directly give the first statement. To see this we can use Lemma 2.7 to find $g \in \mathrm{SL}_2(\mathbb{R})$ such that $g.w = i$ and $g\gamma.i = yi$ with $y \geq 1$. We find that

$$g^{-1}H_\gamma(w) = H_{a(y)}(i).$$

One easily checks that

$$\partial H_{a(y)}(i) = \{z \in \mathbb{H} \colon |z| = \sqrt{y}\}. \tag{37}$$

This is a typical geodesic as desired.

We turn towards (2). Let $z_0 \in \mathbb{H}$ and choose $\gamma_{\min} \in \Gamma$ such that

$$d(\gamma_{\min} z_0, w) = \min_{\gamma\in\Gamma} d(\gamma.z_0, w).$$

We then obviously have $\gamma_{\min}.z_0 \in \mathcal{D}_w$. This shows that

$$\mathbb{H} = \bigcup_{\gamma\in\Gamma} \gamma.\mathcal{D}_w.$$

On the other hand suppose that $z \in \mathcal{D}_w$ and $\gamma.z \in \mathcal{D}_w$ for $\gamma \neq \pm 1$. Then we have

$$d(\gamma.z, w) \leq d(\gamma z, \gamma w) = d(z,w) \leq d(z, \gamma^{-1}.w) = d(\gamma z, w) \tag{38}$$

Thus we obtain $d(z,w) = d(z,\gamma^{-1}w)$. This implies that z must be on the boundary of $\mathcal{D}_w$.

Finally, that $\mathcal{D}_w$ differs from a fundamental set by a null set follows from (1). □

Definition 3.5. For a Fuchsian group $\overline{\Gamma} \subseteq \mathrm{PSL}_2(\mathbb{R})$ we define the limit set $\Lambda(\overline{\Gamma}) \subseteq \partial\mathbb{H}$ to be the set of limit points of all orbits $\overline{\Gamma}.z$ for $z \in \mathbb{H}$.

Remark 3.7. One can show that if $w \in \mathbb{H}$ is not an elliptic fixed point of $\overline{\Gamma}$, then $\Lambda(\overline{\Gamma})$ is equal to the set of limit points of the single orbit $\Gamma.w$. In particular, $\Lambda(\overline{\Gamma})$ is $\overline{\Gamma}$-invariant.

We combine the following important theorem with a definition.

Theorem 3.8 (Poincaré, Fricke–Klein)**.** *If $\overline{\Gamma}$ is a Fuchsian group, then there are the following three possibilities for its limit sets:*

(1) $\sharp\Lambda(\overline{\Gamma}) \in \{0,1,2\}$*. In this case we say that $\overline{\Gamma}$ is elementary.*
(2) $\Lambda(\overline{\Gamma}) = \partial\mathbb{H}$*. In this case we say that $\overline{\Gamma}$ is of the first kind.*
(3) $\Lambda(\overline{\Gamma})$ *is a perfect*[10] *nowhere dense subset of $\partial\mathbb{H}$. In this case we say that $\overline{\Gamma}$ is of the second kind.*

We are interested in a more restricted class of Fuchsian groups. The following definition turns out to be crucial.

Definition 3.6. A Fuchsian group is said to be geometrically finite if there exists a fundamental domain that is a finite sided (hyperbolic) polygon.

The following important theorem puts this definition into perspective.

Theorem 3.9. *Let $\overline{\Gamma}$ be a Fuchsian group. The following are equivalent:*

(1) $\overline{\Gamma}\backslash\mathbb{H}$ *is topologically finite (i.e. has finite Euler characteristic);*
(2) $\overline{\Gamma}$ *is finitely generated;*
(3) $\overline{\Gamma}$ *is geometrically finite.*

We will not prove this here. However let us mention that the implication (1) $\implies$ (2) is due to the fact that topological finite spaces have finitely generated fundamental groups. On the other hand one can obtain (3) $\implies$ (1) by observing that $\Gamma\backslash\mathbb{H}$ can be constructed from the fundamental domain, which is now a finite sided polygon, by identifying (i.e. gluing) equivalent edges. The hard part is to show that (2) $\implies$ (3). This follows from a careful analysis of the Dirichlet domain $\mathcal{D}_w$ and its relation to generators of the group $\overline{\Gamma}$.

Definition 3.7. A discrete subgroup $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$ (resp. a Fuchsian group $\overline{\Gamma} \subseteq \mathrm{PSL}_2$) is called co-compact if $\mathcal{F}_\Gamma$ is compact. We say that Γ (resp. $\overline{\Gamma}$) is co-finite if $\mathrm{Vol}(\mathcal{F}_\Gamma) < \infty$. (A discrete co-finite subgroup of $\mathrm{SL}_2(\mathbb{R})$ is often called a lattice.)

Of course, if Γ is co-compact, then it is automatically co-finite. On the other hand it is a theorem of Siegel that co-finite groups Γ are geometrically finite and of the first kind. It is also true that geometrically finite groups of the first kind are co-finite.[11]

We will almost exclusively focus on co-finite or even co-compact groups. The following results turn out to be very useful for us.

---

[10] A set is perfect if it is closed and has no isolated points.

[11] One needs to be a bit careful here, since there are examples of Fuchsian groups of the first kind that are non co-finite. These are then of course not geometrically finite. One can also easily construct geometricall finite $\overline{\Gamma}$ with infinite co-volume. These will be of the second kind or elementary.

Theorem 3.10. *A co-finite Fuchsian group is co-finite if and only if it contains no parabolic elements.*

Theorem 3.11 (Nielsen)**.** *A non-abelian group* $\overline{\Gamma} \subseteq \mathrm{PSL}_2(\mathbb{R})$ *with only hyperbolic elements acts discontinuously.*

Theorem 3.12. *Any co-finite Fuchsian group* $\overline{\Gamma} \subseteq \mathrm{PSL}_2(\mathbb{R})$ *can be generated by elements*

$$A_1, \dots, A_g, B_1, \dots, B_q, E_1, \dots, E_l, P_1, \dots, P_h$$

where

1. *The elements* $A_1, \dots, A_g$ *and* $B_1, \dots, B_g$ *are hyperbolic and* g *is the genus of* $\overline{\Gamma}\backslash\mathbb{H}$*;*
2. *The elements* $E_1, \dots, E_l$ *are elliptic and there are* $m_j \in \mathbb{N}$ *with* $E_j^{m_j} = 1$*;*
3. *The elements* $P_1, \dots, P_h$ *are parabolic and* h *is the number of cusps;*[12]
4. *We have the relation*

$$[A_1, B_1] \cdot \ldots \cdot [A_g, B_g] \cdot E_1 \cdot \ldots E_l \cdot P_1 \cdot \ldots \cdot P_h = 1.$$

By the Gauß-Bonnet formula we have

$$2g - 2 + \sum_{j=1}^{l} (1 - \frac{1}{m_j}) + h = \frac{1}{2\pi} \operatorname{Vol}(\overline{\Gamma}\backslash\mathbb{H}). \tag{39}$$

We turn towards an example. The arguably most famous co-finite Fuchsian group is $\overline{\Gamma} = \mathrm{PSL}_2(\mathbb{Z})$ (resp. $\Gamma = \mathrm{SL}_2(\mathbb{Z})$). Let us look at some properties.

Lemma 3.13. *We have*

1. $\mathrm{SL}_2(\mathbb{Z})$ *is generated by* $T = n(1)$ *and* $S = k_{\pi/2}$.
2. *A fundamental domain is given by*

$$\mathcal{F}_\Gamma = \mathcal{D}_{2i} = \{z \in \mathbb{H} \colon |\operatorname{Re}(z)| \leq \frac{1}{2} \text{ and } |z| \geq 1\}.$$

Proof. We will only proof (1). We first compute

$$T^n \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a + cn & b + dn \\ c & d \end{pmatrix} \text{ and}$$
$$S \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} -c & -d \\ a & b \end{pmatrix}.$$

Further observe that $S^2 = -1$.

Now take $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. First, if $c = 0$, then we have $a = d = \pm 1$. We have

$$g = \begin{cases} S^2 T^{-b} & \text{if } a = -1, \\ T^b & \text{if } a = 1. \end{cases}$$

[12]The notion of a cusp will be introduced a little bit later.

Thus we can assume $c \neq 0$ and after multiplying with S^2 if necessary we even have $c > 0$. Now we write

$$g' = T^m g = \begin{pmatrix} a' & b' \\ c & d \end{pmatrix},$$

where m is chosen such that $0 \leq a' < c$. Now we note that the matrix Sg' has left lower entry strictly smaller than c. Thus repeating this process leads to the case $c = 0$ after finitely many steps. □

Corollary 3.14. *We have* $\mathrm{Vol}(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}) = \frac{\pi}{3}$*. In particular,* $\mathrm{SL}_2(\mathbb{Z})$ *is co-finite but not co-compact.*

Proof. We compute

$$\begin{aligned} \mathrm{Vol}(\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}) = \mu(\mathcal{F}_\Gamma) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{\sqrt{1-x^2}}^{\infty} y^{-2} dy dx \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{\sqrt{1-x^2}} dx = \int_{-\pi/6}^{\pi/6} \frac{\cos(t)}{\sqrt{1-\sin(t)^2}} dt = \int_{-\pi/6}^{\pi/6} dt = \pi/3. \end{aligned}$$

To see that $\mathrm{SL}_2(\mathbb{Z})$ is not co-compact we can either just look at the fundamental domain or we can observe that T is parabolic and apply Theorem 3.10. □

Exercise 3.1. *Let* $\overline{\Gamma}$ *be a Fuchsian group and let* $z \in \overline{\mathbb{H}}$*. Show that the stabilizer*

$$\overline{\Gamma}_z = \{\gamma \in \overline{\Gamma} \colon \gamma.z = z\}$$

is cyclic.

This exercise can be nicely illustrated by the following example.

Example 3.15. The stabilizer of ∞ in $\mathrm{SL}_2(\mathbb{R})$ is given by

$$\left\{\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R})\right\} = \{\pm 1\} \cdot N(\mathbb{R}) \cdot A(\mathbb{R}).$$

Taking the intersection with $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ we see that

$$\Gamma_\infty = \langle -1, T\rangle = \left\{\begin{pmatrix} \pm 1 & m \\ 0 & \pm 1 \end{pmatrix} : m \in \mathbb{Z}\right\}.$$

This is not cyclic. However, once we are passing to $\overline{\Gamma} = \mathrm{PSL}_2(\mathbb{Z})$ we have $\overline{\Gamma}_\infty = \langle T\rangle$.

Given a general co-finite Fuchsian group $\overline{\Gamma}$ (resp. a discrete group Γ) we say that the cusps of $\overline{\Gamma}$ (resp. Γ) are equivalence classes of parabolic fixed points of Γ. These will be denoted by $\mathfrak{a}, \mathfrak{b}, \ldots$. One can choose a fundamental domain $\mathcal{F}_\Gamma$ whose cuspidal vertices are inequivalent parabolic fixed points. Once such a fundamental domain is fixed these cuspidal vertices become canonical representatives of the cusps. We then sometimes abuse notation and call them the cusps of Γ.

Example 3.16. The parabolic fixed points of $\mathrm{SL}_2(\mathbb{Z})$ are given by $\mathbb{Q} \cup \infty$. But these are all equivalent. Thus there is only one cusp. In the standard fundamental domain $\mathcal{F}_{\mathrm{SL}_2(\mathbb{Z})} = \mathcal{D}_{2i}$ the cuspidal vertex is ∞. Thus we often say ∞ is the cusp of $\mathrm{SL}_2(\mathbb{Z})$.

We return to some more geometric properties of the quotients $\Gamma\backslash\mathbb{H}$. Recall that the hyperbolic distance between i and iy was realized by the vertical path connecting these two points. We extend this path to infinity and obtain the vertical geodesic $\mathfrak{l}_\circ = i\mathbb{R}_{>0}$. We view this as an oriented geodesic, where we imagine the orientation as *traveling from* 0 *to* ∞.[13] All other geodesics are of the form $g.\mathfrak{l}_\circ$ with $g \in \mathrm{PSL}_2(\mathbb{R})$. These are all vertical lines or half circles orthogonal to the real axis.

Given two distinct points $a, b \in \partial\mathbb{H}$, there is a unique oriented geodesic connecting a and b. Indeed, if $a = 0$ and $b = \infty$, then this geodesic is $\mathfrak{l}_\circ$. In general, we can find a matrix $g \in \mathrm{PSL}_2(\mathbb{R})$ such that $g.a = 0$ and $g.b = \infty$. The desired geodesic is then given by $g^{-1}.\mathfrak{l}_\circ$.

Given a hyperbolic matrix $h \in \mathrm{PSL}_2(\mathbb{R})$, then it is conjugate to an element in $A(\mathbb{R})$. More precisely,

$$ghg^{-1} = a(e^l) \text{ for } l \geq 0. \tag{40}$$

The two fixed points a, b of h are mapped to $0, \infty$. We call the fixed point mapped to 0 repelling and the one mapped to ∞ attracting. Applying our observation above there is a unique geodesic connecting a and b oriented so that it travels from repelling fixed point to attracting fixed point:

$$\alpha(h) = g^{-1}.\mathfrak{l}_\circ.$$

The oriented geodesic $\alpha(h) \subseteq \mathbb{H}$ is called the axis of h. We further define the displacement length $l(\gamma) = l$ for l as in (40). Note that we have

$$\mathrm{tr}(\gamma) = 2\cosh(\frac{l(\gamma)}{2}).$$

One also checks that

$$l(\gamma) = \min_{z \in \mathbb{H}} d(z, \gamma z).$$

This minimum is achieved by any element $z \in \alpha(h)$.

We now take a Fuchsian group $\overline{\Gamma}$ and consider the projection

$$\pi \colon \mathbb{H} \to \overline{\Gamma}\backslash\mathbb{H}.$$

Given an oriented geodesic $\mathfrak{l} \subseteq \mathbb{H}$ we obtain an oriented geodesic $\pi(\mathfrak{l}) \subseteq \overline{\Gamma}\backslash\mathbb{H}$. These are paths that locally minimize the hyperbolic distance on the quotient space. These can behave very differently. For now we single out closed oriented geodesics. The following classifying result is very important.

[13] Note that as a set $S.\mathfrak{l}_\circ = i\mathbb{R}_{<0}$, but has opposite orientation.

Lemma 3.17. *There is a one to one correspondence between closed oriented geodesics on $\overline{\Gamma}\backslash\mathbb{H}$ and hyperbolic conjugacy classes in $\overline{\Gamma}$. Under this correspondence the length of the close oriented geodesic corresponds to the displacement length of elements in the conjugacy class.*

Proof. Suppose $\gamma \in \Gamma$ is hyperbolic. We claim that $\mathfrak{l} = \pi(\alpha(\gamma))$ is a closed oriented geodesic of length $l(\gamma)$. By construction $\mathfrak{l}$ is oriented. If we write $t \mapsto \mathfrak{l}(t)$ for the parametrization of $\mathfrak{l}$ with unit speed, then we observe that $\mathfrak{l}(t + l(\gamma)) = \gamma.\mathfrak{l}(t)$. This shows that $\mathfrak{l}$ is closed. Now take $T \in \overline{\Gamma}$. We observe that

$$\alpha(T\gamma T^{-1}) = T\alpha(\gamma).$$

Thus we obtain

$$\pi(\alpha(T\gamma T^{-1})) = \pi(T.\alpha(\gamma)) = \pi(\alpha(\gamma)) = \mathfrak{l}.$$

This shows that our construction depends only on the conjugacy class of γ.

Let $t \mapsto \mathfrak{l}(t)$ be a oriented closed geodesic of length l. We parametrize with unit speed so that $\mathfrak{l}(t + l) = \mathfrak{l}(t)$. We lift $\mathfrak{l}$ to a geodesic $\widetilde{\mathfrak{l}}$ on $\mathbb{H}$. It is easy to see that there is a unique element $\gamma \in \mathrm{PSL}_2(\mathbb{R})$ with $\alpha(\gamma) = \widetilde{\mathfrak{l}}$ and $l(\gamma) = l$. We claim that $\gamma \in \overline{\Gamma}$. To see this we observe that since $\mathfrak{l}(t_0) = \mathfrak{l}(t_0 + l)$ there is $\kappa \in \overline{\Gamma}$ so that

$$\widetilde{\mathfrak{l}}(t_0) = \kappa.\widetilde{\mathfrak{l}}(t_0 + l) = \kappa\gamma.\widetilde{\mathfrak{l}}(t_0).$$

Thus $\kappa\gamma$ fixed the point $\widetilde{\mathfrak{l}}(t_0)$. After suitably choosing $\mathfrak{t}_o$ this implies that $\gamma = \kappa^{-1} \in \overline{\Gamma}$. Note that we made a choice when lifting $\mathfrak{l}$ to $\mathbb{H}$. It can be seen that a different choice $\widetilde{\mathfrak{l}}'$ produces a hyperbolic element γ', which is conjugate to γ (in $\overline{\Gamma}$). ☐

Definition 3.8. A primitive closed oriented geodesic in $\overline{\Gamma}\backslash\mathbb{H}$ is a closed oriented geodesic that is not an iterate of a shorter closed geodesic. Similarly we call an element $\gamma \in \overline{\Gamma}$ (resp. the conjugacy class generated by γ) primitive if it is not a positive power of another element.

Lemma 3.18. *Let $\overline{\Gamma}$ be Fuchsian. Each $\gamma \in \overline{\Gamma}$ can be written uniquely as*

$$\gamma = \gamma_0^k$$

for γ_0 primitive and $k \geq 1$. Furthermore, the centralizer $\overline{\Gamma}_\gamma$ of γ in $\overline{\Gamma}$ is given by $\langle \gamma_0 \rangle$.

Proof. We prove this for hyperbolic γ. The proof for parabolic and elliptic elements is similar.

After conjugating with an element in $\mathrm{PSL}_2(\mathbb{R})$ we can assume that $\gamma = a(e^l)$ with $l > 0$. We can solve the equation $g\gamma = \gamma g$ directly one obtains that $g \in A(\mathbb{R})$, so that $\overline{\Gamma}_\gamma \subseteq A(\mathbb{R})$. We restrict logarithm

$$\log\colon A(\mathbb{R}) \mapsto \mathbb{R},\, a(e^h) \mapsto h,$$

to $\overline{\Gamma}_\gamma$. The image is discrete. Thus, we have

$$\log(\overline{\Gamma}_\gamma) = l_0\mathbb{Z} \subseteq \mathbb{R}.$$

Now we must have $l = l_0 k$ and we get $\gamma = a(e^{l_0})^k$. □

Definition 3.9. The length spectrum of a hyperbolic orbifold $\overline{\Gamma}\backslash\mathbb{H}$ is defined to be the set

$$\mathcal{L}_{\overline{\Gamma}\backslash\mathbb{H}} = \{l(\gamma) \colon \{\gamma\} \text{ primitve hyperbolic conjugacy class in } \overline{\Gamma}\}.$$

In other words, $\mathcal{L}_{\overline{\Gamma}\backslash\mathbb{H}}$ is the set of all lengths of primitive closed oriented geodesics.

Remark 3.19. Note that because we chose to work with oriented geodesics we have $l \in \mathcal{L}_{\overline{\Gamma}\backslash\mathbb{H}}$ if and only if $-l \in \mathcal{L}_{\overline{\Gamma}\backslash\mathbb{H}}$.

We define the counting function

$$\pi_{\overline{\Gamma}\backslash\mathbb{H}}(t) = \sharp\{l \in \mathcal{L}_{\overline{\Gamma}\backslash\mathbb{H}} \colon |l| \leq t\}.$$

One of our first applications of the trace formula will be to show an asymptotic formula for this counting function. This will be the so called *Prime Geodesic Theorem*. For now we can only show the following easy upper bound.

Proposition 3.20. *Let $\overline{\Gamma}$ be a geometrically finite Fuchsian group. Then we have*

$$\pi_{\overline{\Gamma}\backslash\mathbb{H}}(t) = O_{\overline{\Gamma}}(e^t).$$

Proof. We prove this for co-compact $\overline{\Gamma}$.[14] Let $\mathcal{F}_{\overline{\Gamma}} = \mathcal{D}_w$ be a fundamental domain for $\overline{\Gamma}$. Recall in the case at hand this is a compact convex set with finitely many sides. Let $d > 0$ be the diameter of $\mathcal{F}_{\overline{\Gamma}}$.

Let $\mathfrak{l}$ be a primitive closed oriented geodesic and write it as $\mathfrak{l} = \pi(\alpha(\kappa))$ for a hyperbolic element $\kappa \in \overline{\Gamma}$. Let $z \in \alpha(\kappa) \cap \mathcal{F}_{\overline{\Gamma}}$. We compute

$$d(w, \kappa w) \leq 2d(w, z) + d(z, \kappa z) \leq 2d + l(\kappa).$$

This allows us to count

$$\begin{aligned}
\pi_{\overline{\Gamma}\backslash\mathbb{H}}(t) &\leq \sharp\{\kappa \in \overline{\Gamma} \colon d(w, \kappa w) \leq t + 2d\} \\
&\leq \sharp\{\kappa \in \overline{\Gamma} \colon \kappa\mathcal{F}_{\overline{\Gamma}} \cap B_{t+2d}(w) \neq \emptyset\} \\
&\leq \{\kappa \in \overline{\Gamma} \colon \kappa\mathcal{F}_{\overline{\gamma}} \subseteq B_{t+3d}(w) \neq \emptyset\} \\
&\leq \frac{\mathrm{Vol}(B_{t+3d}(w))}{\mathrm{Vol}(\mathcal{F}_{\overline{\Gamma}})}.
\end{aligned}$$

The volume of the hyperbolic disc $B_{t+3d}(w)$ of radius $t + 3d$ and center w is easily computed using polar coordinates and the desired upper bound follows directly. □

The same argument actually also gives the following counting result.

[14]The general proof uses the classification of hyperbolic ends of geometrically finite Fuchsian groups and the resulting definition of the compact core. We refer to [Bo, Proposition 2.19] for details.

Proposition 3.21. *For a geometrically finite Fuchsian group* $\overline{\Gamma}$ *and* $z, w \in \mathbb{H}$ *we have*

$$\sharp\{\gamma \in \overline{\Gamma} \colon d(z, \gamma w) \leq t\} = O_{\overline{\Gamma}}(e^t).$$

Definition 3.10. The exponent of convergence of a Fuchsian group $\overline{\Gamma}$ is

$$\delta_{\overline{\Gamma}} = \inf\{s \geq 0 \colon \sum_{\gamma \in \overline{\Gamma}} e^{-sd(z,\gamma w)} < \infty\},$$

for some $z, w \in \mathbb{H}$.

Remark 3.22. The definition of $\delta_{\overline{\Gamma}}$ is independent of the choice of z and w. Furthermore, Proposition 3.21 shows that for geometrically finite $\overline{\Gamma}$ we have $\delta_{\overline{\Gamma}} \leq 1$. It turns out that for co-finite $\overline{\Gamma}$ we have $\delta_{\overline{\Gamma}} = 1$.[15]

Finally we define the Selberg-Zeta function

$$Z_{\overline{\Gamma}\backslash\mathbb{H}}(s) = \prod_{l \in \mathcal{L}_{\overline{\Gamma}\backslash\mathbb{H}}} \prod_{k=0}^{\infty} (1 - e^{-(s+k)l}),$$

for $\operatorname{Re}(s) > \delta$. A major application of Selberg's trace formula will be the meromorphic continuation of this function.

4. Spectral theory for compact quotients

Throughout this section let $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$ be a co-compact discrete subgroup and let $\mathcal{F}_\Gamma$ be a fundamental domain. Note that $\mathcal{F}_\Gamma$ is compact. We canonically identify

$$L^2(\Gamma\backslash\mathbb{H}) = L^2(\mathcal{F}_\Gamma).$$

In particular, we write the inner product on $L^2(\Gamma\backslash\mathbb{H})$ as

$$\langle f, g\rangle = \int_{\mathcal{F}_\Gamma} f(z)\overline{g(z)} d\mu(z).$$

We can also easily make sense of the spaces

$$\mathcal{C}^k(\Gamma\backslash\mathbb{H}) = \mathcal{C}^k(\mathcal{F}_\Gamma).$$

We write $\Delta \colon L^2(\Gamma\backslash\mathbb{H}) \to L^2(\Gamma\backslash\mathbb{H})$ for the unbounded operator given by $\Delta = -y^2(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2})$ with domain $\mathcal{C}^\infty(\Gamma\backslash\mathbb{H})$.

Lemma 4.1. *The operator* Δ *is symmetric and non-negative.*

[15] One can also show that $\delta_{\overline{\Gamma}} = 0$ for elementary Fuchsian groups and $0 < \delta_{\overline{\Gamma}} < 1$ for geometrically finite Fuchsian groups of the second kind.

Proof. We use Stokes' Theorem to write
$$\int_{\mathcal{F}_\Gamma} \Delta f \overline{g} d\mu = \int_{\mathcal{F}_\Gamma} \nabla f \overline{\nabla g} d\mu_{\mathrm{euc}} - \int_{\partial \mathcal{F}_\Gamma} \frac{\partial f}{\partial n} \overline{g} dl.$$
Here ∂n is the outer normal derivative and dl is the euclidean length. By identifying sides of $\mathcal{F}_\Gamma$ that are equivalent under the Γ-action one verifies that the boundary integral vanishes. This shows that
$$\langle \Delta f, g \rangle = \int_{\mathcal{F}_\Gamma} \nabla f \overline{\nabla g} d\mu_{\mathrm{euc}}$$
and completes the proof. □

As a result we find that Δ has a unique self-adjoint extension. This is the so called Friedrichs extensions. We will slightly abuse notation and denote this extension also by Δ. Our goal in this section is to study the spectrum (and in particular the eigenvalues) of this unbounded self-adjoint operator. We start with some easy observations.

Corollary 4.2. *The following statements hold.*
(1) *The eigenvalues of Δ are non-negative.*
(2) *If $\Delta f = 0$, then f is constant.*
(3) *Two eigenfunctions with distinct eigenvalues are orthogonal.*

We define the *automorphic Green function* for Γ as
$$G_{\Gamma,s}(z,w) = \sum_{\gamma \in \overline{\Gamma}} G_s(u(\gamma z, w)),$$
for $z \notin \Gamma.w$ and $\mathrm{Re}(s) > 1$. Note that absolute convergence follows from Proposition 3.21. We record the following properties:
(1) We have $\overline{G_{\Gamma,s}(z,w)} = G_{\Gamma,\overline{s}}(z,w)$ and $G_{\Gamma,s}(z,w) = G_{\Gamma,s}(w,z)$.
(2) For $\gamma_1, \gamma_2 \in \Gamma$ we have $G_{\Gamma,s}(\gamma_1.z, \gamma_2.w) = G_{\Gamma,s}(z,w)$. In particular, we can view $G_{\Gamma,s}$ as a function $G\colon \Gamma\backslash\mathbb{H} \times \Gamma\backslash\mathbb{H} \to \mathbb{C}$ with singularities on the diagonal. (In view of our usual identification this is also interpreted as a function on $\mathcal{F}_\Gamma \times \mathcal{F}_\Gamma$.)
(3) Away from the diagonal (i.e. from (z,w) with $\Gamma.z = \Gamma.w$) the function $G_{\Gamma,s}$ is smooth.
(4) As $w \to z$ we have
$$G_{\Gamma,s}(z,w) = -\frac{\sharp\Gamma_z}{4\pi}\log(u(z,w)) + O_{\Gamma,s}(1). \tag{41}$$
This follows directly from Lemma 2.11 and Proposition 3.21.
(5) For $z \in \mathcal{F}_\Gamma$ we have
$$\int_{\mathcal{F}_\Gamma} |G_{\Gamma,s}(z,w)|^2 d\mu(w) \ll_{\Gamma,s} 1. \tag{42}$$

To see this it is sufficient to show that the integral over a small (hyperbolic) ball $B_r(z)$ is uniformly bounded. The resulting integral is estimated using (41).

Lemma 4.3. *For all $f \in \mathcal{C}^\infty(\Gamma\backslash\mathbb{H})$ we have*

$$(\Delta - s(1-s))R_{\Gamma,s}f = f \text{ for } [R_{\Gamma,s}f](z) = \int_{\mathcal{F}_\Gamma} G_{\Gamma,s}(z,w)f(w)d\mu(w).$$

Proof. This will directly follow from Theorem 2.13 and the unfolding trick. Indeed we compute

$$\begin{aligned}[R_{\Gamma,s}f](z) &= \sum_{\gamma\in\overline{\Gamma}} \int_{\mathcal{F}_\Gamma} G_s((u\gamma.z,w))f(w)d\mu(w) \\ &= \sum_{\gamma\in\overline{\Gamma}} \int_{\gamma^{-1}\mathcal{F}_\Gamma} G_s(u(z,w))f(\gamma.w)d\mu(w) \\ &= \sum_{\gamma\in\overline{\Gamma}} \int_{\gamma^{-1}\mathcal{F}_\Gamma} G_s(u(z,w))f(w)d\mu(w) \\ &= \int_{\mathbb{H}} G_s(u(z,w))f(w)d\mu(w) = [F_sf](z).\end{aligned}$$

This shows that

$$R_s|_{\mathcal{C}^\infty(\Gamma\backslash\mathbb{H})} = R_{\Gamma,s}$$

and the result is immediate. □

We now turn to the analysis of invariant integral operators. We start with some general remarks concerning integral operators on $L^2(\Gamma\backslash\mathbb{H})$. Note that we can view $\mathcal{F}_\Gamma$ as a domain in $\mathbb{R}^2$, so that in view of our identification $L^2(\Gamma\backslash\mathbb{H}) = L^2(\mathcal{F}_\Gamma)$ the theory is very well represented in the classical literature. We will only give the (relevant) highlights.

Recall that an integral operator on $L^2(\Gamma\backslash\mathbb{H})$ is given by

$$[Kf](z) = \int_{\Gamma\backslash\mathbb{H}} K(z,w)f(w)d\mu(w) = \int_{\mathcal{F}_\Gamma} K(z,w)f(w)d\mu(w),,$$

where $K\colon \Gamma\backslash\mathbb{H} \times \Gamma\backslash\mathbb{H} \to \mathbb{C}$ is a suitable kernel function. As usual we may think of K as a function on $\mathcal{F}_\Gamma \times \mathcal{F}_\Gamma$.

Lemma 4.4. *Let $K \in L^2(\Gamma\backslash\mathbb{H}\times\Gamma\backslash\mathbb{H})$, then the associated operator K is a bounded compact linear operator on $L^2(\Gamma\backslash\mathbb{H})$. The adjoint operator is given by the kernel*

$$K^*(z,w) = \overline{K(w,z)}.$$

Proof. We compute

$$\begin{aligned}\|Kf\|^2 &= \int_{\mathcal{F}_\Gamma} \left| \int_{\mathcal{F}_\Gamma} K(z,w) f(w) d\mu(w) \right|^2 d\mu(z) \\ &\leq \int_{\mathcal{F}_\Gamma} \int_{\mathcal{F}_\Gamma} |K(z,w)|^2 d\mu(z) d\mu(w) \cdot \|f\|^2.\end{aligned}$$

By assumption on K this shows that the operator is bounded. More precisely,

$$\|K\|_{\mathrm{Op}} \leq \|K\|_{L^2(\mathcal{F}_\Gamma \times \mathcal{F}_\Gamma)}.$$

To see compactness we argue as follows. We first approximate K by a sequence of smooth kernel $K \in \mathcal{C}^\infty(\mathcal{F}_\Gamma \times \mathcal{F}_\Gamma)$ (i.e. $K_n \to K$ in L^2-norm topology). Using Hölder one easily shows that

$$|[K_n f](z_1) - [K_n f](z_2)| \leq \mathrm{Vol}(\mathcal{F}_\Gamma)^{\frac{1}{2}} \cdot \sup_{w \in \mathcal{F}_\Gamma} |K_n(z_1,w) - K_n(z_2,w)| \cdot \|f\|_{L^2(\mathcal{F}_\Gamma}.$$

With this at hand one can apply Arzelá-Ascoli to show that the operator

$$K_n \colon L^2(\mathcal{F}_\Gamma) \to \mathcal{C}(\mathcal{F}_\Gamma)$$

is compact. Since the embedding $\mathcal{C}(\mathcal{F}_\Gamma) \to L^2(\mathcal{F}_\Gamma)$ is continuous we see that $K_n \colon L^2(\mathcal{F}_\Gamma) \to L^2(\mathcal{F}_\Gamma)$ is compact. Compactness of K follows, since $\|K - K_n\|_{\mathrm{Op}} \to 0$ as $n \to \infty$.

Finally, we have to compute the adjoint operator. We do so directly

$$\langle Kf, g \rangle = \int_{\Gamma\backslash\mathbb{H}} \int_{\Gamma\backslash\mathbb{H}} K(z,w) f(w) \overline{g(z)} d\mu(w) d\mu(z) = \langle f, K^* g \rangle.$$

This completes the proof. □

In particular, if the Kernel is square integrable and satisfies $K(z,w) = \overline{K(w,z)}$, then it defines a self-adjoint operator compact operator.

Lemma 4.5 (Bessel inequality)**.** *Let $K \in L^2(\Gamma\backslash\mathbb{H} \times \Gamma\backslash\mathbb{H})$ and let $\phi_1, \dots, \phi_n$ be orthonormal eigenfunctions with eigenvalues λ_i. Then we have*

$$\int_{\Gamma\backslash\mathbb{H}} \left| K(z,w) - \sum_{i=1}^n \lambda_i \phi(z) \overline{\phi_i(w)} \right|^2 d\mu(w) = \int_{\Gamma\backslash\mathbb{H}} |K(z,w)|^2 - \sum_{i=1}^n |\lambda_i|^2 |\phi_i(z)|^2.$$

for almost all z. Furthermore,

$$\begin{aligned}&\int_{\Gamma\backslash\mathbb{H}} \int_{\Gamma\backslash\mathbb{H}} \left| K(z,w) - \sum_{i=1}^n \lambda_i \phi(z) \overline{\phi_i(w)} \right|^2 d\mu(w) d\mu(z) \\ &\qquad = \int_{\Gamma\backslash\mathbb{H}} \int_{\Gamma\backslash\mathbb{H}} |K(z,w)|^2 d\mu(w) d\mu(z) - \sum_{i=1}^n |\lambda_i|^2.\end{aligned}$$

Proof. We open $|\cdot|^2$ and simply compute

$$\int_{\Gamma\backslash\mathbb{H}}\left|K(z,w)-\sum_{i=1}^{n}\lambda_i\phi(z)\overline{\phi_i(w)}\right|^2 d\mu(w)$$
$$=\int_{\Gamma\backslash\mathbb{H}}|K(z,w)|^2 d\mu(w)-\sum_{i=1}^{n}\lambda_i\phi_i(z)\cdot\overline{[K\phi_i](z)}-\sum_{i=1}^{n}\overline{\lambda_i\phi_i(z)}\cdot[K\phi_i](z)$$
$$+\sum_{i=1}^{n}\sum_{j=1}^{n}\lambda_i\overline{\lambda_j}\phi_i(z)\overline{\phi_j(z)}\langle\phi_i,\phi_j\rangle.$$

We are done since $K\phi_i=\lambda_i\phi_i$ and $\langle\phi_i,\phi_j\rangle=\delta_{i,j}$ by assumption. The second identity now follows by integrating the first one. $\square$

Corollary 4.6. *Let $K\in L^2(\Gamma\backslash\mathbb{H}\times\Gamma\backslash\mathbb{H})$ and let $\lambda_1,\ldots,\lambda_n$ be eigenvalues (that may occur with multiplicity). Then we have*

$$\sum_{j=1}^{n}|\lambda_j|^2\le\int_{\Gamma\backslash\mathbb{H}}\int_{\Gamma\backslash\mathbb{H}}|K(z,w)|^2 d\mu(z)d\mu(w).$$

To conclude this interlude we recall the spectral theorem for compact self-adjoint operators.[16]

Theorem 4.7. *Let T be a compact self-adjoint operator on a Hilbert space $\mathcal{H}$. Then the following statement hold.*

- *The spectrum of T consists of countably many eigenvalues.*
- *All non-zero eigenvalues have finite multiplicities and can accumulate only at 0.*
- *If $|\lambda_1|\ge|\lambda_2|\ge\ldots\to 0$ is a complete list of non-trivial eigenvalues (listed with multiplicity) and $x_1,x_2,\ldots$ is a corresponding system of (orthonormal) eigenvectors, then we have*

$$Tx=\sum_{i=1}^{\infty}\lambda_i\langle x,x_i\rangle\cdot x_i.$$

Let us return to the automorphic Green function $G_{\Gamma,s}$ for Γ and the corresponding integral operator $R\Gamma,s$. We take $s=2$ and observe that $R_{\Gamma,2}$ is a bounded, compact and self-adjoint operator. Compactness follows from (42) and Lemma 4.4. Using Lemma 4.3 it is easy to see that $R_{\Gamma,s}$ has dense range and shares eigenfunctions with Δ. This leads to the following important result.

Theorem 4.8 (Spectral theorem for compact quotients)**.** *Let $\Gamma\subseteq\mathrm{SL}_2(\mathbb{R})$ be a co-compact discrete subgroup. Then the operator Δ has discrete spectrum*

$$0=\lambda_0<\lambda_1\le\lambda_2\le\ldots\to\infty \tag{43}$$

[16] See for example [Tr, Theorem 18.4].

and the corresponding orthonormal system of eigenfunctions $\phi_0, \phi_1, \ldots$ forms a basis of $L^2(\Gamma\backslash\mathbb{H})$. If f is in the domain of Δ, then

$$f(z) = \sum_{i=0}^{\infty} \langle f, \phi_i \rangle \cdot \phi_i(z)$$

converges pointwise absolutely and uniformly on compacta.[17]

Note that whenever we list the eigenvalues $\lambda_0, \lambda_1, \ldots$ of Δ this list is ordered (i.e. $\lambda_0 = 0$) and each eigenvalues appears with the appropriate multiplicity. The corresponding eigenfunctions $\phi_0, \phi_1, \ldots$ (i.e. $\Delta\phi_i = \lambda_i\phi_i$) will always be assumed to be orthonormal. Note that since Δ is an elliptic operator the eigenfunctions are smooth. (This is the celebrated elliptic regularity theorem.)

Corollary 4.9. *Let $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$ be co-compact and let $\lambda_0, \lambda_1, \ldots$ be a list of eigenvalues of Δ on $L^2(\Gamma\backslash\mathbb{H})$. We have*

$$N_\Gamma(T) = \{i \in \mathbb{Z}_{\geq 0} \colon \lambda_i \leq T\} = O_\Gamma(T^2)$$

as $T \to \infty$.

Proof. Note that

$$R_{\Gamma,2}\phi_i = \frac{1}{\lambda_i + 2} \cdot \phi_i.$$

In particular, by Corollary 4.6 we have

$$S \colon = \sum_{i=0}^{\infty} \frac{1}{(\lambda_i + 2)^2} < \infty. \tag{44}$$

Because

$$N_\Gamma(T) \leq (T+2)^2 \cdot S, \tag{45}$$

we are done. □

Later we will use Selberg's trace formula to prove an asymptotic formula, a so called Weyl law, for $N_\Gamma(T)$.

Let $k \colon \mathbb{R}_{\geq 0} \to \mathbb{C}$ be a function such that its Selberg - Harish–Chandra transform h satisfies the properties in Remark 2.24. (This is for example guaranteed if k is smooth and compactly supported.) Recall that T_k is the corresponding invariant integral operator acting on functions from $\mathbb{H}$ to $\mathbb{C}$. When applied to the eigenfunctions ϕ_i of Δ we find that

$$T_k\phi_i = h(t_j)\phi_i,$$

[17] This convergence statement does not follow directly from the spectral theorem for $R_{\Gamma,2}$ and we will use it as a black box.

for $\lambda_i = \frac{1}{4} + t_i^2$ (or similarly $\lambda_i = s_i(1-s_i)$ and $s_i = \frac{1}{2} + it_i$). This follows from Theorem 2.21. We define the *automorphic kernel* by

$$k_\Gamma(z,w) = \sum_{\gamma\in\overline{\Gamma}} k(u(\gamma.z,w)). \tag{46}$$

It can be seen that the γ-sum converges absolutely for the k's in question.[18] By the unfolding trick we find that

$$[T_k f](z) = \int_{\mathcal{F}_\Gamma} k_\Gamma(z,w) f(w) d\mu(w),$$

for $f \in L^2(\Gamma\backslash\mathbb{H})$. We see that $T_k\colon L^2(\Gamma\backslash\mathbb{H}) \to L^2(\Gamma\backslash\mathbb{H})$ is an integral operator. As observed earlier the operators T_k commute with Δ and are all simultaneously *diagonalized* by the eigenfunctions ϕ_i of Δ. We have the following important expansion, which is sometimes called *pretrace equality.*

Theorem 4.10. *Let $k\colon \mathbb{R}_{\geq 0} \to \mathbb{C}$ be such that its Selberg-Harish–Chandra transform h satisfies the conditions from Remark 2.24. Then we have*

$$k_\Gamma(z,w) = \sum_{i=0}^{\infty} h(t_i)\cdot \phi_i(z)\overline{\phi_j(w)}.$$

Here $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$ is a list of eigenvalues of the Laplacian written as $\lambda_i = \frac{1}{4}+t_i^2$ and $\phi_0, \phi_1, \dots$ is a corresponding orthonormal system of eigenfunctions. The expansion converges absolutely and uniformly on compacta.

Remark 4.11. Note that sometimes the automorphic kernel (46) is defined by summing over Γ instead of $\overline{\Gamma}$. We have

$$\sum_{\gamma\in\Gamma} k(u(\gamma.z,w)) = c_\Gamma \cdot k_\Gamma(z,w),$$

where $c_\Gamma = 1$ if $-1 \notin \Gamma$ and $c_\Gamma = 2$ if $-1 \in \Gamma$. Note that when applying the unfolding trick it is more natural if the sum is taken over $\overline{\Gamma}$. Otherwise one needs to account for the multiplicity c_Γ with which the translates of the fundamental domain occur.

5. The trace formula for compact quotients

As the section title suggests we will now develop the trace formula for compact quotients. Doing so we will mostly follow [He76, Chapter One].

Throughout this section $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$ will denote a co-compact discrete group and we let $\overline{\Gamma}$ be its image in $\mathrm{PSL}_2(\mathbb{R})$. Furthermore, we fix a convenient fundamental domain $\mathcal{F}_\Gamma \subseteq \mathbb{H}$. We let

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \cdots \to \infty$$

[18] In particular, if k is compactly supported, then the sum is finite for fixed z and w. This is a direct consequence of the discontinuity of the action of Γ on $\mathbb{H}$.

denote a full list (with multiplicities) of eigenvalues of Δ. These eigenvalues will also be written as

$$\lambda_i = s_i(1-s_i) = \frac{1}{4}+t_i^2 \text{ and } s_i = \frac{1}{2}+it_i.$$

A corresponding orthonormal system of eigenfunctions is denoted by $\phi_0, \phi_1, \ldots$.

We start by developing the trace formula for $k \in \mathcal{C}_0^\infty(\mathbb{R}_{\geq 0})$, but the same reasoning works as soon as $k \in \mathcal{C}_0^2(\mathbb{R}_{\geq 0})$. Our goal is to compute the trace of the associated invariant integral operator $T_k\colon L^2(\Gamma\backslash\mathbb{H}) \to L^2(\Gamma\backslash\mathbb{H})$ in two ways. We start with the spectral expansion.

Lemma 5.1 (Spectral trace)**.** *Let $k \in \mathcal{C}_c^\infty(\mathbb{R}_{\geq 1})$, then we have*

$$\mathrm{Tr}(T_k) = \sum_{i=0}^{\infty} h(t_i).$$

and the right hand side is absolutely convergent.

Proof. We first sow absolute convergence. To do so we recall that by standard argument (see Lemma 2.23) one shows that

$$|h(t)| \ll_{k,A} (|t|+1)^{-A} \tag{47}$$

for $A \in \mathbb{N}$ arbitrarily large. After choosing A sufficiently large absolutely convergence follows from Corollary 4.6 and the observation that $|t_j| \asymp \sqrt{\lambda_j}$.

Once convergence is established we can compute the trace according to its definition:

$$\mathrm{Tr}(T_k) = \sum_{j=0}^{\infty} \langle T_k\phi_j, \phi_j\rangle = \sum_{j=0}^{\infty} h(t_j).$$

This completes the proof. $\square$

On the other hand, we can use Theorem 4.10 with $w = z$. Integrating over $\mathcal{F}_\Gamma$ yields

$$\mathrm{Tr}(T_k) = \int_{\mathcal{F}_\Gamma} k_\Gamma(z,z)d\mu(z).$$

Starting from this we compute the following unrefined formula for the geometric trace.

Lemma 5.2 (Unrefined geometric expansion)**.** *For $k \in \mathcal{C}_c^\infty(\mathbb{R}_{\geq 0})$ we have*

$$\mathrm{Tr}(T_k) = \sum_{\{\gamma\}} \int_{\mathcal{F}_\gamma} k(u(\gamma.z,z))d\mu(z),$$

where the $\{\gamma\}$-sum ranges over distinct conjugacy classes in $\overline{\Gamma}$ and $\mathcal{F}_\gamma$ is a fundamental domain for the centralizer $\overline{\Gamma}_\gamma$ of γ in $\overline{\Gamma}$. In particular, the expression on the right is independent of the choice of $\mathcal{F}_\gamma$.

Proof. Inserting the definition of the automorphic kernel in our expression for the trace yields

$$\begin{aligned}\mathrm{Tr}(T_k) &= \sum_{\gamma\in\Gamma}\int_{\mathcal{F}_\Gamma} k(u(\gamma.z,z))d\mu(z)\\ &= \sum_{\{\gamma\}}\sum_{\sigma\in\{\gamma\}}\int_{\mathcal{F}_\Gamma} k(u(\sigma.z,z))d\mu(z).\end{aligned}$$

An element $\sigma\in\{\gamma\}$ can be written as $\sigma = T^{-1}\gamma T$ for $T\in\overline{\Gamma}$. Note that $T^{-1}\gamma T = R^{-1}\gamma R$ if and only if $R\in\overline{\Gamma}_\gamma T$. Thus, we get

$$\mathrm{Tr}(T_k) = \sum_{\{\gamma\}}\sum_{T\in\overline{\Gamma}_\gamma\backslash\overline{\Gamma}}\int_{\mathcal{F}_\Gamma} k(u(T^{-1}\gamma T.z,z))d\mu(z) = \sum_{\{\gamma\}}\sum_{T\in\overline{\Gamma}_\gamma\backslash\overline{\Gamma}}\int_{T.\mathcal{F}_\Gamma} k(u(\gamma z,z))d\mu(z).$$

We put

$$\mathcal{D}_\gamma = \bigcup_{T\in\overline{\Gamma}_\gamma\backslash\overline{\Gamma}} T.\mathcal{F}_\Gamma,$$

so that

$$\mathrm{Tr}(T_k) = \sum_{\{\gamma\}}\int_{\mathcal{D}_\gamma} k(u(\gamma z,z))d\mu(z).$$

We claim that $\mathcal{D}_\gamma$ is a fundamental domain for $\overline{\Gamma}_\gamma$. To see we check:

- Let $z\in\mathbb{H}$. Then we write $z=\sigma.w$ for $w\in\mathcal{F}_\Gamma$ and $\sigma\in\overline{\Gamma}$. Further, put $\sigma = T\delta$ for $\delta\in\overline{\Gamma}_\gamma\backslash\overline{\Gamma}$ and $T\in\overline{\Gamma}_\gamma$. Then we have $z = T\delta.w\in T.\mathcal{D}_\gamma$.
- Now suppose $z_1,z_2\in\mathcal{D}_\Gamma$ and $Tz_1 = z_2$ for $T\in\overline{\Gamma}_\gamma$. We rewrite this as $\sigma_2^{-1}T\sigma_1 w_1 = w_2$ for $w_1,w_2\in\mathcal{F}_\Gamma$ and $\sigma_1,\sigma_2\in\overline{\Gamma}_\gamma\backslash\overline{\Gamma}$. Thus $w_1,w_2\in\partial\mathcal{F}_\Gamma$ and one concludes that z_1,z_2 are not in the interior of $\mathcal{D}_\gamma$.
- Finally, we note that the difference to a (measurable) fundamental set is contained in the union of the boundaries of the *tiles* $T.\mathcal{F}_\Gamma$. Thus it must be a set of measure zero.

Finally, since $z\mapsto k(u(\gamma.z,z))$ is $\overline{\Gamma}_\gamma$-invariant, we find that the choice of the fundamental domain is irrelevant. ☐

The integrals

$$I(\{\gamma\}) = \int_{\mathcal{F}_\gamma} k(u(\gamma z,z))d\mu(z),$$

where $\mathcal{F}_\gamma$ is a fundamental domain for the centralizer $\overline{\Gamma}_\gamma$, are called orbital integrals. We are no going to compute convenient expressions for them. This is best done by distinguishing different types of conjugacy classes.

Remark 5.3. Since Γ is assumed to be co-compact we do not have to consider parabolic conjugacy classes. This is due to Theorem 3.10.

The easiest, but very important, contribution comes from the identity element.

Lemma 5.4 (The identity contribution)**.** *We have*

$$I(\{1\}) = k(0)\operatorname{Vol}(\Gamma\backslash\mathbb{H}) = \operatorname{Vol}(\Gamma\backslash\mathbb{H}) \cdot \frac{1}{4\pi}\int_{\mathbb{R}} th(t)\tanh(\pi t)dt.$$

Proof. For $\gamma = 1$ we have $\overline{\Gamma}_\gamma = \overline{\Gamma}$ and $\mathcal{F}_\gamma = \mathcal{F}_\Gamma$. We compute

$$I(\{1\}) = \int_{\mathcal{F}_\Gamma} k(u(z,z))d\mu(z) = k(0)\operatorname{Vol}(\mathcal{F}_\Gamma).$$

The second expression follows from (33) and $F_s(0) = 1$. □

Exercise 5.1. *Prove the identity*

$$k(0) = \frac{1}{4\pi}\int_{\mathbb{R}} th(t)\tanh(\pi t)dt$$

using the 3*-step (inverse) Selberg-Harish–Chandra transform (i.e.* $h \rightsquigarrow g \rightsquigarrow q \rightsquigarrow k$*).*

We turn towards hyperbolic conjugacy classes. Recall that by Lemma 3.18 every hyperbolic conjugacy class can be written as

$$\{\gamma\} = \{\gamma_0^k\}$$

for γ_0 primitive and $k \geq 0$. With this in mind we prove the following lemma.

Lemma 5.5 (Hyperbolic contributions)**.** *Let* $\gamma_0 \in \overline{\Gamma}$ *be a primitive hyperbolic element with (signed) displacement length* $l(\gamma_0)$*. For* $k \geq 1$ *we have*

$$I(\{\gamma_0^k\}) = \frac{l(\gamma_0)}{2\sinh(kl(\gamma_0)/2)} g(kl(\gamma_0)).$$

Proof. We start by writing $p = e^{|l(\gamma_0)|} > 1$ and note that there is $g \in \mathrm{PSL}_2(\mathbb{R})$ such that

$$g^{-1}\gamma_0 g = a(p) = \operatorname{diag}(\sqrt{p}, 1/\sqrt{p}).$$

Also recall that the centralizer of γ_0^k is the cyclic group $\overline{\Gamma}_{\gamma_0} = \langle\gamma_0\rangle$ generated by γ_0. Put

$$\Lambda = \langle a(p)\rangle \subseteq g\overline{\Gamma}g^{-1}.$$

In particular, Λ is the centralizer of $a(p)$ in $g\overline{\Gamma}g^{-1}$ and $g^{-1}\Lambda g = \overline{\Gamma}_{\gamma_0}$. Since $a(p).z = p \cdot z$ and $p > 1$ it is easy to see that

$$\mathcal{F}_\Lambda = \{z \in \mathbb{H} \colon 1 \leq \operatorname{Im}(z) \leq p\}$$

is a fundamental domain for Λ. We see that $g.\mathcal{F}_\Lambda$ is a fundamental domain for $\overline{\Gamma}_{\gamma_0}$.

Choosing this particular domain in the definition of the orbital integral allows us to compute

$$\begin{aligned} I(\{\gamma_0^k\}) &= \int_{g.\mathcal{F}_\Lambda} k(u(\gamma_0^k.z, z))d\mu(z) \\ &= \int_{\mathcal{F}_\Lambda} k(u(\gamma_0^k gz, gz))d\mu(z) \\ &= \int_{\mathcal{F}_\Lambda} k(u(p^k \cdot z, z))d\mu(z). \end{aligned}$$

Now we recall that

$$u(p^k \cdot z, z) = \frac{|p^k z - z|^2}{4p^k \cdot \mathrm{Im}(z)^2} = \underbrace{\frac{(p^k - 1)^2}{4p^k}}_{=N^2} \frac{|z|^2}{\mathrm{Im}(z)^2}.$$

We can proceed by calculating the integral directly. Changing $x \to yx$ yields

$$\begin{aligned} I(\{\gamma_0^k\}) &= \int_1^p \int_{\mathbb{R}} k\left(N^2 \frac{x^2 + y^2}{y^2}\right) \frac{dxdy}{y^2} \\ &= \int_{\mathbb{R}} k\left(N^2(x^2+1)\right) \int_1^p \frac{dy}{y} dx \\ &= 2\log(p) \int_0^\infty k\left(N^2(x^2+1)\right) dx. \end{aligned}$$

At this point we change variables $u = N^2(x^2+1)$ (e.g. $x = (\frac{u}{N^2} - 1)^{\frac{1}{2}}$). We arrive at

$$I(\{\gamma_0^k\}) = \frac{\log(p)}{N} \int_{N^2}^\infty \frac{k(u)}{\sqrt{u - N^2}} du = \frac{\log(p)}{N} q(N^2).$$

Here we have recalled the transform from k to q given in Definition 2.2. We conclude the argument by replacing q with g using (34). To do so we observe that

$$2\log(\sqrt{N^2} + \sqrt{N^2+1}) = k\log(p). \tag{48}$$

We get

$$I(\{\gamma_0^k\}) = \frac{\log(p)}{2N} g(k\log(p)).$$

After recalling that $N = \frac{1}{2}(p^{k/2} - p^{-k/2})$ and $p = e^{|l(\gamma_0)|}$ we find that

$$I(\{\gamma_0^k\}) = \frac{|l(\gamma_0)|}{2\sinh(k|l(\gamma_0)|/2)} g(k|l(\gamma_0)|).$$

After recalling that g is even we see that we can drop the absolute values around the signed displacement length $l(\gamma_0)$. This concludes the proof. □

Note that Lemma 5.4 and Lemma 5.5 are sufficient for the trace formula of compact Riemann surfaces. This is because the corresponding Fuchsian groups do not have elliptic elements. However, since we do not want to exclude elliptic elements we have to compute their contribution.

Note that there are only finitely many elliptic conjugacy classes and these can be written as powers of primitive classes. Indeed, a primitive elliptic element $\gamma_0 \in \overline{\Gamma}$ has finite order $m(\gamma_0)$ and all elements in $\overline{\Gamma}$ with the same fixed point in $\mathbb{H}$ is given by γ_0^l for $1 \leq l < m(\gamma_0)$. Thus the following result is sufficient to cover the elliptic contribution.

Lemma 5.6 (Elliptic contributions)**.** *Let γ_0 be a primitive elliptic element of order $m(\gamma_0)$ and let $1 \leq k \leq m(\gamma_0)$. Then we have*

$$\begin{aligned} I(\{\gamma_0^k\}) &= \frac{1}{m(\gamma_0)} \int_0^\infty \frac{g(r)\cosh(r/2)}{\cosh(r) - \cos(2\pi k/m(\gamma_0))} dr \\ &= (2m(\gamma_0)\sin(\pi k/m(\gamma_0)))^{-1} \int_{\mathbb{R}} h(r) \frac{\cosh(\pi(1-2k/m(\gamma_0))r)}{\cosh(\pi r)} dr. \end{aligned}$$

Proof. First we take $g \in \mathrm{PGL}_2(\mathbb{R})$ such that $g^{-1}\gamma_0 g = k_\theta \in \mathrm{SO}_2$ with $\theta = \pi/m(\gamma_0)$. Let $\Lambda = \langle k_\theta \rangle$ and note that k_θ acts on $\mathbb{H}$ by rotation with angle 2θ around i. We choose $\mathcal{F}_\Lambda$ to be a corresponding sector. Arguing as in the hyperbolic case we need to compute

$$I(\{\gamma_0^k\}) = \int_{\mathcal{F}_\Lambda} k(u(k_\theta^k.z, z)) d\mu(z) = \frac{1}{m(\gamma_0)} \int_{\mathbb{H}} k(u(k_\theta^k.z, z)) d\mu(z).$$

In the second step we have used that it takes exactly $m(\gamma_0)$ copies of $\mathcal{F}_\Lambda$ to cover the full upper half plane.

From here we can compute the desired expression using polar coordinates as in (18). We obtain

$$\begin{aligned} I(\{\gamma_0^k\}) &= \frac{2}{m(\gamma_0)} \int_0^\pi \int_0^\infty k(u(k_\theta^k k_\varphi a(e^{-r}).i, k_\varphi a(e^{-r}).i)) \sinh(r) dr d\varphi \\ &= \frac{2\pi}{m(\gamma_0)} \int_0^\infty k(u(k_\theta^k a(e^{-r}).i, a(e^{-r}).i)) \sinh(r) dr. \end{aligned}$$

We compute

$$u(k_\theta^k z, z) = \frac{\sin(k\theta)^2}{4y^2} |z^2 + 1|^2.$$

If $z = e^{-r} i$, then this becomes

$$u(k_\theta^k e^{-r} i, e^{-r} i) = \sinh(r)^2 \sin(k\theta)^2.$$

Inserting this above and putting $u = \sinh(r)^2$ yields

$$\begin{aligned} I(\{\gamma_0^k\}) &= \frac{\pi}{m(\gamma_0)} \int_0^\infty k(u \cdot \sin(k\theta)^2) \frac{du}{\sqrt{u+1}} \\ &= \frac{\pi}{m(\gamma_0)\sin(k\theta)} \int_0^\infty \frac{k(u)}{\sqrt{u+\sin(k\theta)^2}} du. \end{aligned}$$

For notational simplicity we temporarily write $a = \sin(k\theta)$. In view of (35) we can compute

$$\begin{aligned} I(\{\gamma_0^k\}) &= -\frac{1}{m(\gamma_0)a} \int_0^\infty \int_u^\infty q'(v)(v-u)^{-\frac{1}{2}}(u+a^2)^{-\frac{1}{2}} dv du \\ &= -\frac{1}{m(\gamma_0)a} \int_0^\infty q'(v) \int_0^v ((v-u)(u+a^2))^{-\frac{1}{2}} du dv \\ &= -\frac{1}{m(\gamma_0)a} \int_0^\infty q'(v) \int_0^{v/(v+a^2)} (u(1-u))^{-\frac{1}{2}} du dv \\ &= \frac{1}{m(\gamma_0)} \int_0^\infty q(v)(v+a^2)^{-1} v^{-\frac{1}{2}} dv \\ &= \frac{1}{2m(\gamma_0)} \int_0^\infty \frac{g(r)\cosh(r/2)}{\sinh(r/2)^2 + a^2} dr. \end{aligned}$$

One directly verifies that, for $a = \sin(k\theta)$, we have

$$\sinh(r/2)^2 + a^2 = \frac{1}{2}(\cosh(r) - \cos(2k\theta)).$$

Inserting this above yields

$$I(\{\gamma_0^k\}) = \frac{1}{m(\gamma_0)} \int_0^\infty \frac{g(r)\cosh(r/2)}{\cosh(r) - \cos(2k\theta)} dr$$

as desired. We omit the proof of the second identity for the integral. □

We can now summarize everything and arrive at the first form of the trace formula.

Theorem 5.7. *Let $\Gamma \in \mathrm{SL}_2(\mathbb{R})$ be a co-compact discrete subgroup and let $k \in \mathcal{C}_c^\infty(\mathbb{R}_{\geq 0})$ and let h denote its Selberg-Harish–Chandra transform. We write the eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots$ of Δ as $\lambda_j = \frac{1}{4} + t_j^2$. Recall that g is the* *(inverse) Fourier transform of h. Then we have*

$$\begin{aligned}\sum_{j=0}^{\infty} h(t_j) = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi} \int_{\mathbb{R}} th(t)\tanh(\pi t)dt \\ + \frac{1}{2} \sum_{\substack{\{\gamma_0\} \\ \textit{primitive} \\ \textit{hyperbolic}}} l(\gamma_0) \sum_{k=1}^{\infty} \frac{g(kl(\gamma_0))}{\sinh(kl(\gamma_0)/2)} \\ + \sum_{\substack{\{\gamma_0\} \\ \textit{primitive} \\ \textit{elliptic}}} \frac{1}{m(\gamma_0)} \sum_{k=1}^{m(\gamma_0)-1} \int_0^{\infty} \frac{g(r)\cosh(r/2)}{\cosh(r) - \cos(2\pi k/m(\gamma_0))} dr,\end{aligned}$$

where $l(\gamma_0)$ is the displacement length of hyperbolic γ_0 and $m(\gamma_0)$ is the order of elliptic γ_0. The infinite sums in this expressions are absolutely convergent.

Proof. We compute the trace of T_k on $L^2(\Gamma\backslash\mathbb{H})$ in two ways. First, spectrally using Lemma 5.1. This gives the left hand side of our formula. Next we use the unrefined geometric expansion from Lemma 5.2. The conjugacy classes are split up according to their type. The identity contribution is computed in Lemma 5.4. The remaining conjugacy classes are sorted according to the underlying primitive class. The expression for hyperbolic classes is then given in Lemma 5.5, while for the elliptic case we refer to Lemma 5.6.

Absolute convergence on the spectral side (i.e. the left hand side) is part of Lemma 5.1. On the geometric side (i.e. the right hand side) the only potentially infinite sum is the sum involving hyperbolic terms. However, since g is compactly supported this sum is actually finite. $\square$

We now want to extend the trace formula to a wider class of functions. We define

$$\tilde{\delta} = \inf\{\sigma \geq 1 \colon \sum_{j=1}^{\infty} \lambda_j^{-\sigma} < \infty\}.$$

By Corollary 4.6 we have $\tilde{\delta} \leq 2$. We define the following class of functions.[19]

Definition 5.1 (Sufficiently regular h)**.** We say a function h is sufficiently regular if, for some fixed $\epsilon_0 > 0$, we have

(1) h is even;
(2) $h(t)$ is holomorphic in the strip $|\operatorname{Im}(t)| \leq \frac{1}{2} + \epsilon_0$; and
(3) $h(t) \ll (|\operatorname{Re}(t)| + 1)^{-2\tilde{\delta}-\epsilon_0}$.

Later we will see that $\tilde{\delta} = 1$, so that sufficiently regular functions are precisely those described in Remark 2.24.

[19]Disclaimer: The following definition is not standard terminology.

Lemma 5.8. *If h is sufficiently regular, then its Fourier transform $g = \widehat{h}$ satisfies*

$$g(u) \ll e^{-(\frac{1}{2}+\epsilon_0)|u|}.$$

Proof. Without loss of generality we can assume that $u < 0$. Shifting contour in the definition of the Fourier transform leads to

$$g(u) = \frac{1}{2\pi}\int_{-\infty}^{\infty} h(t)e^{-itu}dt = \int_{(\frac{1}{2}+\epsilon_0)i-\infty}^{(\frac{1}{2}+\epsilon_0)i+\infty} h(t)e^{-itu}dt.$$

We bound the right hand side trivially by

$$\int_{-\infty}^{\infty} |h(x+i(\frac{1}{2}+\epsilon_0))|e^{(\frac{1}{2}+\epsilon_0)u}dx \ll e^{-(\frac{1}{2}+\epsilon_0)|u|}\int_{-\infty}^{\infty}(1+|x|)^{-2\tilde{\delta}-\epsilon_0}dx \ll e^{-(\frac{1}{2}+\epsilon_0)|u|}.$$

This proves the statement. ☐

Corollary 5.9. *Suppose h is sufficiently regular, then the sum*

$$\sum_{\substack{\{\gamma_0\}\\ \textit{primitive}\\ \textit{hyperbolic}}} l(\gamma_0)\sum_{k=1}^{\infty}\frac{g(kl(\gamma_0))}{\sinh(kl(\gamma_0)/2)}$$

is absolutely convergent.

Proof. Note that there is a minimal length $l_\Gamma > 0$ so that $|l(\gamma_0)| \geq l_\Gamma$ for all primitive hyperbolic elements γ_0. First, let us fix a primitive hyperbolic element γ_0. Without loss of generality we can take $l(\gamma_0) > 0$. Then we can estimate the k-sum as

$$\sum_{k=1}^{\infty}\frac{g(kl(\gamma_0))}{\sinh(kl(\gamma_0)/2)} \ll \sum_{k=1}^{\infty} e^{-(1+\epsilon_0)kl(\gamma_0)} \ll \frac{e^{-(1+\epsilon_0)l(\gamma_0)}}{1-e^{-(1+\epsilon_0)l(\gamma_0)}} \ll_\Gamma e^{-(1+\epsilon_0)l(\gamma_0)}.$$

Thus we are left with estimating

$$\sum_{\substack{\{\gamma_0\}\\ \text{primitive}\\ \text{hyperbolic}}} |l(\gamma_0)|e^{-(1+\epsilon_0)|l(\gamma_0)|}.$$

In view of Proposition 3.21 this is easily seen to be absolutely convergent. ☐

We can now state the main theorem of this chapter.

Theorem 5.10 (Selberg trace formula I)**.** *Let $\Gamma \in \mathrm{SL}_2(\mathbb{R})$ be a co-compact discrete subgroup and let h be a sufficiently regular function. We write the eigenvalues*

$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$ *of* Δ *as* $\lambda_j = \frac{1}{4} + t_j^2$. *Then we have*

$$\sum_{j=0}^{\infty} h(t_j) = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi} \int_{\mathbb{R}} th(t)\tanh(\pi t)dt$$

$$+ \frac{1}{2} \sum_{\substack{\{\gamma_0\} \\ \textit{primitive} \\ \textit{hyperbolic}}} l(\gamma_0) \sum_{k=1}^{\infty} \frac{g(kl(\gamma_0))}{\sinh(kl(\gamma_0)/2)}$$

$$+ \frac{1}{2} \sum_{\substack{\{\gamma_0\} \\ \textit{primitive} \\ \textit{elliptic}}} \frac{1}{m(\gamma_0)} \sum_{k=1}^{m(\gamma_0)-1} \sin(\pi k/m(\gamma_0))^{-1} \int_{\mathbb{R}} h(r) \frac{\cosh(\pi(1-2k/m(\gamma_0))r)}{\cosh(\pi r)} dr,$$

where $d(\gamma_0)$ *is the displacement length of hyperbolic* γ_0 *and* $m(\gamma_0)$ *is the order of elliptic* γ_0. *The integrals and series involved in this formula are all absolutely convergent.*

Proof. We first assume that

$$h(t) \ll e^{-c|\operatorname{Re}(t)|} \text{ for some } c > 0. \tag{49}$$

For such h, the proof of Lemma 5.8 is easily modified to yield

$$g^{(k)}(u) \ll_k e^{-(\frac{1}{2}+\epsilon_0)|u|} \text{ for } k \in \mathbb{Z}_{\geq 0}.$$

Fix a smooth, monotonically decreasing function $\varphi\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ with support in $[0,2]$ and such that $\varphi|_{[0,1]} = 1$. We define

$$\varphi_m(x) = \begin{cases} 1 & \text{if } |x| \leq m, \\ \varphi(|x|-m) & \text{for } |x| > m. \end{cases}$$

We define

$$g_m(u) = g(u)\varphi_m(u)$$

and we let h_m, q_m, k_m denote the transforms obtained from g_m via (35).[20] We check that

$$g_m^k(u) \ll_k e^{-(\frac{1}{2}+\epsilon_0)|u|}$$

uniformly in m. On the other hand we compute

$$h_m = \widehat{\varphi_m} * h.$$

This can be used to show that $h_m(t) \ll (1+|t|)^{-5}$ for $t \in \mathbb{R}$ and $\lim_{m\to\infty} h_m(t) = h(t)$ uniformly for $\operatorname{Im}(t) \leq \frac{1}{2}$. We can thus apply Theorem 5.7 and take the limit $m \to \infty$ on both sides of the formula. This establishes the desired formula for functions h satisfying (49).

[20]Note that one has to be carefully when defining derivatives of q_m at 0, but this is only a little technicality.

Finally lets take h sufficiently regular. We define

$$h_\epsilon(t) = h(t)e^{-\epsilon t^2}.$$

This is now a function for which (49) holds. After applying the trace formula for h_ϵ it is easy to justify, that one can take the limit $\epsilon \to 0$. This completes the proof. □

6. Applications I

We now turn towards basic applications of the trace formula. We are doing so following the beginning of [He76, Chapter Two].

Throughout this section we will fix a co-compact discrete subgroup $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$. We enumerate the eigenvalues of Δ on $L^2(\Gamma\backslash\mathbb{H})$ by

$$0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots \to \infty$$

and write $\lambda_j = \frac{1}{4} + t_j^2$. Important invariants of Γ are the co-volume $\mathrm{Vol}(\Gamma\backslash\mathbb{H})$ and the length of the shortest closed geodesic

$$l_\Gamma = \min \mathcal{L}_{\overline{\Gamma}\backslash\mathbb{H}}.$$

Our first goal is to establish a Weyl-law. This is an asymptotic formula for

$$N_\Gamma(T) = \sharp\{j \in \mathbb{Z}_{\geq 0} \colon \lambda_j \leq T\}.$$

Next we will prove the Prime geodesic theorem, which captures the asymptotic growth of

$$\pi_{\overline{\Gamma}\backslash\mathbb{H}}(t) = \{l \in \mathcal{L}_{\overline{\Gamma}\backslash\mathbb{H}} \colon |l| \leq t\}.$$

Finally, we will study analytic properties of the Selberg-Zeta function.

6.1. The Weyl law for co-compact quotients.

Two very nice surveys on Weyl's law are [Iv, Mü]. We have seen the Weyl law for the sphere in Exercise 1.3 and the Weyl law for the torus in Corollary 1.4. We now turn towards the Weyl law for compact Riemann surfaces of constant negative curvature.

We start with an easy estimate.

Proposition 6.1. *For $T > 0$ we have*

$$\sum_{j=0}^{\infty} e^{-t_j^2 T} = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi T} + O_\Gamma(1) \tag{50}$$

as T approaches 0.

Proof. We take

$$h(t) = e^{-r^2 T}.$$

Note that the Fourier transform is given by

$$g(r) = \widehat{h}(r) = \frac{1}{\sqrt{4\pi T}} e^{-\frac{r^2}{4T}}.$$

Applying the trace formula to h we obtain

$$\sum_{j=0}^{\infty} e^{-Tt_j^2} = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi} \int_{-\infty}^{\infty} te^{-t^2T} \tanh(\pi t)dt + \text{ Hyperbolic Terms} + \text{Elliptic Terms}.$$

We first handle the identity contribution. Here we will use that $\tanh(\pi t) = 1 + O(e^{-2\pi t})$ for $t \geq 0$. This allows us to write

$$\begin{aligned} \int_{-\infty}^{\infty} te^{-t^2T} \tanh(\pi t)dt &= 2\int_0^{\infty} te^{-t^2T}dt + O\left(\int_0^{\infty} te^{-t^2T-2\pi t}\right) \\ &= \frac{1}{T} + O(1). \end{aligned}$$

The elliptic contribution can be estimated trivially and is seen to be $O_\Gamma(1)$. Thus, it remains to treat the hyperbolic contribution. We first estimate

$$\begin{aligned} \text{Hyperbolic Terms} &= \frac{1}{2} \sum_{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}} l \sum_{k=1}^{\infty} \frac{1}{\sqrt{4\pi T}} e^{-\frac{k^2l^2}{4T}} \sinh(kl/2)^{-1} \\ &\ll_\Gamma \frac{1}{\sqrt{T}} \sum_{\substack{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}, \\ l>0}} le^{\frac{-l^2}{4T}} \sum_{k=1}^{\infty} e^{-kl_\Gamma/2} \ll_\Gamma \frac{1}{\sqrt{T}} \sum_{\substack{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}, \\ l>0}} le^{\frac{-l^2}{4T}}. \end{aligned}$$

The function $x \mapsto x \cdot e^{-\frac{x^2}{4T}}$ has a global maximum $\sqrt{2Te}$ at $x = \sqrt{2T}$. Thus we can estimate

$$\text{Hyperbolic Terms} \ll_\Gamma \pi_{\Gamma\backslash\mathbb{H}}(100\sqrt{T}) + \frac{1}{\sqrt{T}} \sum_{\substack{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}, \\ l>100\sqrt{T}}} le^{\frac{-l^2}{4T}}.$$

Estimating the remaining sum trivially yields

$$\text{Hyperbolic Terms} \ll_\Gamma \pi_{\Gamma\backslash\mathbb{H}}(100\sqrt{T}) + 1.$$

Recall that, for T sufficiently small, we have $\pi_{\Gamma\backslash\mathbb{H}}(100\sqrt{T}) = 0$. This completes the proof. $\square$

This proposition allows us to deduce Weyl's law in asymptotic form, but without a meaningful error term.

Theorem 6.2 (Weyl law I). *Let $\Gamma \subseteq \mathrm{SL}_2(\mathbb{R})$ be a co-compact discrete subgroup. Then we have*

$$N_\Gamma(T) = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi} T(1 + o_\Gamma(1)).$$

Proof. This follows at once from a classical Tauberian theorem.[21] We temporarily write

$$\widetilde{N}_\Gamma(T) = \sharp\{j \in \mathbb{Z}_{\geq 0} \colon t_j^2 \leq T\}.$$

Clearly we have $N_\Gamma(T) = \widetilde{N}_\Gamma(T) + O(1)$. We write the result from Proposition 6.1 as

$$\int_0^\infty e^{-Tt} d\widetilde{N}_\Gamma(T) \sim \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi} T^{-1},$$

for $T \to 0$. Using the Tauberian theorem we conclude that $\widetilde{N}_\Gamma(T) \sim \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi} T$ as $T \to \infty$ and are done. □

Remark 6.3. In particular, we have $\tilde{\delta} = 1$. Thus our notion of sufficiently regular functions coincides with the functions described in Remark 2.24.

We will directly give a direct (arguably more modern) proof of Weyl's law. This time we also get a reasonable error term.

Theorem 6.4 (Weyl law II). *Let* Γ *be a co-compact discrete subgroup of* $SL_2(\mathbb{R})$*. Then we have*

$$N_\Gamma(T) = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi} T + O_\Gamma(\sqrt{T}).$$

In the proof we will assume that Γ has no elliptic elements. This assumption is for convenience only. It is left as an exercise for the reader to handle the general case.

Proof. We start with some preparations. Choose $g \in \mathcal{C}_c^\infty(\mathbb{R})$ and let $h = \widehat{g}$. We assume that the support of g is contained in $(-l_\Gamma, l_\Gamma)$ and that $g(0) = \widehat{h}(0) = 1$. That such a function exists is easy to verify.

For a parameter T we define

$$h_T(z) = h(T - z) + h(T + z).$$

We have

$$\widehat{h}_T(r) = e^{-iTr} g(r) + e^{iTr} g(-r).$$

Since the support of g is contained in the interval $(-l_\Gamma, l_\Gamma)$ and we are assuming that there are no elliptic elements the trace formula reads

$$\sum_{j=-\infty}^{\infty} h(T - t_j) = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{2\pi} \int_{\mathbb{R}} h(T - t) t \tanh(\pi t) dt. \tag{51}$$

Here we have doubled the spectrum by setting $t_{-j} = -t_j$.

[21] More precisely we will use the Hardy-Littlewood Tauberian theorem, which can be formulated as follows. Let $F \colon [0, \infty) \to \mathbb{R}$ be a real valued function and define $w(s) = \int_0^\infty e^{-st} dF(t)$. Then, for $\rho > 0$ and a constant C, we have $w(s) \sim Cs^{-\rho}$ as $s \to 0$ if and only if $F(t) \sim \frac{C}{\Gamma(\rho+1)} t^\rho$ as $t \to \infty$.

We now additionally assume that $h(t) \geq 0$ for $t \in \mathbb{R}$ and $h(t) > 0$ for $t \in [-a, a]$.[22] We now observe that

$$\sharp\{j \in \mathbb{Z} \colon |t_j - \mu \leq a,\, t_j \in \mathbb{R}\} \cdot \min_{u \in [-a,a]} h(u) \leq \sum_{j \in \mathbb{Z}} h(\mu - t_j).$$

We estimate the right hand side using (51). One easily sees that

$$\sum_{j \in \mathbb{Z}} h(\mu - t_j) \ll_h 1 + |\mu|.$$

Since there are only finitely many eigenvalues with $t_j \notin \mathbb{R}$ we conclude that

$$\sharp\{j \in \mathbb{Z} \colon |t_j - \mu| \leq a\} \ll_{a,\Gamma} 1 + |\mu| \tag{52}$$

After having established this intermediate counting result we return to the main argument. Integrating (51) over $T \in [-\lambda, \lambda]$ yields

$$\int_{-\lambda}^{\lambda} \sum_{j=-\infty}^{\infty} h(T - t_j) dT = \frac{\mathrm{Vol}(\Gamma \backslash \mathbb{H})}{2\pi} \lambda^2 + O_\Gamma(\lambda). \tag{53}$$

To see this we write $p(t) = t \tanh(\pi t)$ and compute

$$\begin{aligned}\int_{-\lambda}^{\lambda} \int_{\mathbb{R}} h(T-t)p(t)dtdT = \int_{-\lambda}^{\lambda} \int_{\mathbb{R}} h(T-r)dTp(t)dt + \int_{-\lambda}^{\lambda} \int_{\mathbb{R}\backslash[-\lambda,\lambda]} h(T-t)p(t)dtdT \\ - \int_{-\lambda}^{\lambda} \int_{\mathbb{R}\backslash[-\lambda,\lambda]} h(T-t)dTp(t)dt.\end{aligned}$$

The first integral is easily seen to give the main term:

$$\int_{-\lambda}^{\lambda} \underbrace{\int_{\mathbb{R}} h(T-r)dT}_{=\widehat{h}(0)=1} p(t)dt = \int_{-\lambda}^{\lambda} p(t)dt = \lambda^2 + O(\lambda).$$

The remaining two integrals are easily handled using the rapid decay of h and we obtain

$$\int_{-\lambda}^{\lambda} \int_{\mathbb{R}} h(T-t)p(t)dtdT = \lambda^2 + O(\lambda).$$

The identity (53) is a direct consequence of these considerations.

Further, we claim that

$$\sum_{|t_j| \leq \lambda} \left| \int_{\mathbb{R}\backslash[-\lambda,\lambda]} h(T - t_j) dT \right| \ll \lambda \tag{54}$$

[22]Such a choice is possible by choosing $h_0 \in \mathcal{S}(\mathbb{R})$ such that g_0 has the desired properties. The desired h is then obtained by appropriately re-scaling $h_0 \cdot \overline{h_0}$.

and

$$\sum_{|t_j|>\lambda} \left| \int_{-\lambda}^{\lambda} h(T-t_j)dT \right| \ll \lambda \tag{55}$$

These are again shown by using the rapid decay of h together with (52). Indeed, we can estimate

$$\begin{aligned}\sum_{|t_j|\leq\lambda} \left| \int_{\lambda}^{\infty} h(T-t_j)dT \right| &\ll \sum_{|t_j|\leq\lambda} \int_{\lambda-t_j}^{\infty} |h(t)|dt \ll \sum_{|t_j|\leq\lambda} (1+\lambda-t_j)^{-5} \\ &\ll \sum_{-\lambda\leq k<\lambda} \frac{\sharp\{j\in\mathbb{Z}\colon |t_j-k|\leq 1\}}{(1+\lambda-k)^5} \ll \sum_{-\lambda\leq k<\lambda} \frac{1+|k|}{(1+\lambda-k)^5} \ll \lambda.\end{aligned}$$

The lower piece of the integral (i.e. from $-\lambda$ to $-\infty$) is handled similarly yielding (54). The proof for (55) is analogous.

We are now ready to finish the proof. We compute

$$\begin{aligned}\sharp\{j\in\mathbb{Z}\colon |t_j|\leq\lambda\} &= \sum_{|t_j|\leq\lambda} \widehat{h}(0) = \sum_{|t_j|\leq\lambda} \int_{\mathbb{R}} h(T-t_j)dT \\ &= \int_{-\lambda}^{\lambda} \sum_{j=-\infty}^{\infty} h(T-t_j)dT + \sum_{|t_j|\leq\lambda} \int_{\mathbb{R}\setminus[-\lambda,\lambda]} h(T-t_j)dT + \sum_{|t_j|>\lambda} \int_{-\lambda}^{\lambda} h(T-t_j)dT.\end{aligned}$$

This is estimated using (53), (54) and (55). We arrive at

$$\sharp\{j\in\mathbb{Z}\colon |t_j|\leq\lambda\} = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{2\pi}\lambda^2 + O_\Gamma(\lambda).$$

Converting this count to $N_\Gamma(T)$ concludes the proof. □

Remark 6.5. One can even show that

$$N_\Gamma(T) = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi}T + o_\Gamma(\sqrt{T}).$$

However, the saving is only logarithmic. Note that such a result is true in much greater generality. Indeed one only needs mild dynamical conditions to exclude surfaces like the sphere. We refer to [Iv] for a more exhaustive discussion.

6.2. **The prime geodesic theorem.** It will make sense to introduce the counting function

$$\widetilde{\pi}_{\Gamma\backslash\mathbb{H}}(T) = \sharp\{\text{closed geodesic of length } \leq T\}.$$

Here we are counting all closed geodesics not just primitive ones.

Lemma 6.6. *We have*

$$\widetilde{\pi}_{\Gamma\backslash\mathbb{H}}(T) = \pi_{\Gamma\backslash\mathbb{H}}(T) + O_\Gamma(e^{T/2}).$$

Proof. We simply observe that every geodesic is obtained as iterations of primitive geodesics. Thus we have

$$\widetilde{\pi}_{\Gamma\backslash\mathbb{H}}(T) = \sum_{1\leq k\leq T/l_\Gamma} \pi_{\Gamma\backslash\mathbb{H}}(T/k) = \pi_{\Gamma\backslash\mathbb{H}}(T) + O(\widetilde{\pi}_{\Gamma\backslash\mathbb{H}}(T/2)).$$

We are done after using the trivial bound

$$\widetilde{\pi}_{\Gamma\backslash\mathbb{H}}(T/2) \ll e^{\frac{T}{2}}$$

from Proposition 3.20. □

We will need too following result concerning the eigenvalues.

Lemma 6.7. *As $T\to\infty$ we have*

$$\sum_{j=1}^{\infty} e^{-\lambda_j T} \ll_\Gamma e^{-\lambda_1 T}.$$

Proof. For a parameter U we can write

$$\sum_{j=1}^{\infty} e^{-\lambda_j T} = \sum_{0<\lambda_j<U} e^{-\lambda_j T} + \underbrace{\int_U^\infty e^{-Tx} dN_\Gamma(x)}_{\ll_\Gamma U e^{-UT}}.$$

We are done after choosing $U=\lambda_1$. □

Proposition 6.8. *We have $\pi_{\Gamma\backslash\mathbb{H}}(T) \ll \frac{e^T}{\sqrt{T}}$.*

This improves the trivial bound from Proposition 3.20, but is still not optimal. Note that one should read this as

$$\pi_{\Gamma\backslash\mathbb{H}}(T) \ll \frac{e^T}{\sqrt{\log(e^T)}}.$$

Proof. We start by reading the trace formula for

$$h(t) = e^{-T(t^2+\frac{1}{4})} \text{ and } g(u) = \frac{1}{4\pi T}e^{-\frac{u^2}{4T}-\frac{T}{4}}$$

backwards. This leads to

$$\frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{2\pi}\int_0^\infty te^{-(t^2+\frac{1}{4})T}\tanh(\pi t)dt + \frac{1}{2}\sum_{l\in\mathcal{L}_{\Gamma\backslash\mathbb{H}}} l\sum_{k=1}^{\infty}\frac{1}{\sqrt{4\pi T}}e^{-\frac{k^2l^2}{4T}-\frac{T}{4}}\sinh(kl/2)^{-1}$$
$$+\text{ Elliptic Terms} = 1 + O(e^{-\lambda_1 T}).$$

The identity contribution as well as the elliptic terms are easily seen to be $O_\Gamma(T^{\frac{1}{2}}e^{-\frac{T}{4}})$. Thus we find that

$$\frac{1}{2}\sum_{l\in\mathcal{L}_{\Gamma\backslash\mathbb{H}}} l\sum_{k=1}^{\infty}\frac{1}{\sqrt{4\pi T}}e^{-\frac{k^2l^2}{4T}-\frac{T}{4}}\sinh(kl/2)^{-1} = 1+O(e^{-aT}),$$

for $0 < a < \min(\lambda_1, \frac{1}{4})$. Since at the moment we are only interest in upper bounds we only record the resulting estimate

$$\frac{1}{\sqrt{T}} \sum_{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}} l e^{-\frac{l^2}{4T} - l/2 - \frac{T}{4}} \ll_\Gamma 1.$$

Here we have estimated $\sinh(kl/2) \leq \frac{1}{2} e^{kl/2}$ and the terms for $k > 1$ are absorbed in the error. This is easily rewritten as

$$\int_0^\infty u e^{-(u+T)^2/4T} d\pi_{\Gamma\backslash\mathbb{H}}(u) \ll T^{\frac{1}{2}}.$$

By partially integrating the Riemann-Stieljes integral we find

$$\int_0^\infty \pi_{\Gamma\backslash\mathbb{H}}(u) e^{-(T+u)^2/4T} \left(\frac{u^2}{2T} + \frac{u}{2} - 1 \right) du = \int_0^\infty u e^{-(u+T)^2/4T} d\pi_{\Gamma\backslash\mathbb{H}}(u) \ll T^{\frac{1}{2}}.$$

We can now estimate

$$T e^{-T} \pi_{\Gamma\backslash\mathbb{H}}(T) \ll \int_T^{T+1} \pi_{\Gamma\backslash\mathbb{H}}(u) e^{-(T+u)^2/4T} \left(\frac{u^2}{2T} + \frac{u}{2} - 1 \right) du \ll T^{\frac{1}{2}}.$$

The claimed upper bound is immediate. □

Unfortunately it turns out that the test function $h(t) = e^{-T(t^2 + \frac{1}{4})}$ is not sufficient to derive the prime geodesic theorem. Nonetheless, we can still push our estimate to obtain some useful results.

First, we note that in the above computations we have often passed between sums over all hyperbolic conjugacy classes (i.e. closed geodesics) as they appear in the trace formula and the contribution from primitive conjugacy classes. We record a precise version of this argument in the following lemma.

Lemma 6.9. *For $a = \min(l_\Gamma, \log(100))$ we have*

$$\frac{e^{-\frac{T}{4}}}{2\sqrt{4\pi T}} \sum_{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}} l \sum_{k=1}^\infty e^{-\frac{k^2 l^2}{4T}} \sinh(kl/2)^{-1} = \frac{e^{-\frac{T}{4}}}{\sqrt{4\pi T}} \sum_{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}} \frac{l}{e^{l/2}} e^{-\frac{l^2}{4T}} + O\left((T^{\frac{1}{2}} + T^{-\frac{1}{2}}) e^{-a^2/4T - \frac{T}{4}} \right).$$

Furthermore, we have

$$\text{Elliptic Terms} \ll T^{\frac{1}{2}} e^{-\frac{T}{4}}.$$

This allows us to write

$$\sum_{j=0}^\infty e^{-\lambda_j T} = \frac{e^{-\frac{T}{4}}}{\sqrt{4\pi T}} \int_0^\infty \frac{u}{e^{u/2}} e^{-\frac{u^2}{4T}} d\pi_{\Gamma\backslash\mathbb{H}}(u) + O(T^{\frac{1}{2}} e^{-\frac{T}{4}}). \tag{56}$$

Using Lemma 6.7 we can replace the right hand side by $1 + O(e^{-\eta T})$ for any $0 < \eta < \min(\lambda_1, \frac{1}{4})$. The so created O-term obviously absorbs the error term on the right hand side. Integrating everything against a non-negative function $f\colon [1,\infty) \to \mathbb{R}$ we obtain

$$\begin{aligned}\int_1^\infty f(T)(1+O(e^{-\eta T}))dT &= \int_1^\infty \int_0^\infty \frac{e^{-\frac{T}{4}}}{\sqrt{4\pi T}} \frac{u}{e^{u/2}} e^{-\frac{u^2}{4T}} d\pi_{\Gamma\backslash\mathbb{H}}(u) f(T) dT \\ &= \int_0^\infty \frac{u}{e^{u/2}} \int_1^\infty \frac{e^{-\frac{T}{4}}}{\sqrt{4\pi T}} e^{-\frac{u^2}{4T}} f(T) dT d\pi_{\Gamma\backslash\mathbb{H}}(u)\end{aligned}$$

We take $f(T) = e^{-\beta T}$ for $\beta = \alpha - \frac{1}{4}$. Note that we have the tail bound

$$\int_0^\infty u e^{-\frac{u}{2}} \int_0^1 \frac{e^{-\alpha T}}{\sqrt{4\pi T}} e^{-u^2/4T} dT d\pi_{\Gamma\backslash\mathbb{H}}(u) \ll 1.$$

This leads us to

$$\frac{1}{\beta} + O(1) = \int_0^\infty u e^{-\frac{u}{2}} \int_0^\infty \frac{e^{-\alpha T}}{\sqrt{4\pi T}} e^{-u^2/4T} dT d\pi_{\Gamma\backslash\mathbb{H}}(u).$$

evaluating the inner integral gives

$$\frac{1}{\beta} + O(1) = \frac{1}{2\sqrt{\alpha}} \int_0^\infty u e^{-\frac{u}{2} - u\sqrt{\alpha}} d\pi_{\Gamma\backslash\mathbb{H}}(u).$$

We note that

$$\frac{2\sqrt{\beta+1/4}}{\beta} = \frac{1}{\beta} + O(1) \text{ and } \frac{1}{2} + \sqrt{\beta + \frac{1}{4}} = 1 + \beta + O(\beta^2).$$

Thus, as $t \to 0$ we arrive at

$$\int_0^\infty e^{-tu} u e^{-u} d\pi_{\Gamma\backslash\mathbb{H}}(u) = \frac{1}{t} + O(1). \tag{57}$$

We record the following consequences.

Proposition 6.10. *We have*

(1) *As $t \to 0$ we have*

$$\sum_{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}} l \cdot e^{-(1+t)l} = \frac{1}{t} + O(1).$$

(2) *As $X \to \infty$ we have*

$$\sum_{\substack{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}, \\ |l| \le X}} l e^{-l} = X(1+o(1)).$$

(3) *As $X \to \infty$ we have*

$$\sum_{\substack{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}, \\ |l| \le X}} e^{-l} = \log(X)(1+o(1)).$$

Proof. The first point is simply a reformulation of (57). The second identity is a direct consequence of a classical Tauberian Theorem. Finally, the third claim follows from the second one by partial summation. □

We conclude here by giving some (essentially trivial) lower bounds.

Proposition 6.11. *For any* $0 < \alpha < \frac{1}{4}$ *we have*

$$\liminf_{t\to\infty} \pi_{\Gamma\backslash\mathbb{H}}(t)e^{-\alpha t} = \infty.$$

Proof. We start from (56) together with Lemma 6.7. Truncating the integral at $S = T + 2\sqrt{T\log(T)}$ yields

$$\frac{e^{-T/4}}{\sqrt{4\pi T}} \int_0^S ue^{-\frac{u}{2}}e^{-u^2/4T} d\pi_{\Gamma\backslash\mathbb{H}}(u) = 1 + o(1).$$

As before we can rewrite this as

$$\begin{aligned} 1 + o(1) = \frac{e^{-T/4}}{\sqrt{4\pi T}} \int_0^S \pi_{\Gamma\backslash\mathbb{H}}(u)e^{-\frac{u}{2}-\frac{u^2}{4T}} \left(\frac{u^2}{2T} + \frac{u}{2} - 1\right) du \\ \leq \pi_{\Gamma\backslash\mathbb{H}}(S)\frac{e^{-T/4}}{\sqrt{4\pi T}} \int_0^S e^{-\frac{u}{2}-\frac{u^2}{4T}} T(1 + o(1))du. \end{aligned}$$

The integral can be evaluated explicitly. We obtain

$$\pi(S) \geq (1 + o(1))\sqrt{\pi}\frac{e^{\frac{T}{4}}}{\sqrt{T}}.$$

Since $S \asymp T$ we are done. □

Our so far *failed attempt* to prove the prime geodesic theorem leads us naturally to the definition of the following Dirichlet series.[23] For $\mathrm{Re}(s) > 1$ we set

$$E(s) = 2 \sum_{\substack{l\in\mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ l>0}} \frac{l}{e^{sl}}.$$

According to Proposition 6.10 this is absolutely convergence for $\mathrm{Re}(s) > 1$. Our goal is to find a suitable analytic continuation of $E(s)$.

Remark 6.12. We would like to choose $g(u) = e^{-\alpha|u|}$ in the trace formula. This would lead to

$$\text{Hyperbolic Terms} \sim E(\frac{1}{2} + \alpha).$$

Thus in order to obtain new analytic information we need to take α with $\mathrm{Re}(\alpha) \leq \frac{1}{2}$.

[23] Here I am assuming that we are classically trained analytic number theorists.

However, if we compute

$$h(t) = \int_{\mathbb{R}} e^{-\alpha|u|+itu} du = \frac{1}{\alpha + it} + \frac{1}{\alpha - it} = \frac{2\alpha}{\alpha^2 + t^2}.$$

From this we can read of the location of the poles and observe that they occur at $\mathrm{Im}(t) = \pm \mathrm{Re}(\alpha)$. Since we need h to be sufficiently regular, we can not use this particular test function to obtain any new analytic information.

We will use the test function

$$g(u) = \frac{1}{2\alpha} e^{-\alpha|u|} - \frac{1}{2\beta} e^{-\beta|u|}, \tag{58}$$

for $\frac{1}{2} < \mathrm{Re}(\alpha) < \mathrm{Re}(\beta)$. One computes that

$$h(t) = \frac{1}{t^2 + \alpha^2} - \frac{1}{t^2 + \beta^2} = \frac{\beta^2 - \alpha^2}{(t^2 + \alpha^2)(t^2 + \beta^2)}.$$

We see with bare eyes that this function is analytic for $|\,\mathrm{Im}(t)| < \mathrm{Re}(\alpha)$. Furthermore, we have

$$h(t) \ll (1 + |\,\mathrm{Re}(t)|)^{-4}$$

in an appropriate strip $\mathrm{Im}(t) \leq \frac{1}{2} + \epsilon$, where $\epsilon > 0$ depends on α. From now on we will fix $\beta \geq 2$.

The trace formula for this choice of h reads

$$\begin{aligned}
\sum_{j=0}^{\infty} \left(\frac{1}{t_j^2 + \alpha^2} - \frac{1}{t_j^2 + \beta^2} \right) = {} & \frac{\mathrm{Vol}(\Gamma \backslash \mathbb{H})}{4\pi} \int_{-\infty}^{\infty} t \left(\frac{1}{t^2 + \alpha^2} - \frac{1}{t^2 + \beta^2} \right) \tanh(\pi t) dt \\
& + \text{Elliptic Terms} \\
& + \frac{1}{2\alpha} \sum_{\substack{l \in \mathcal{L}_{\Gamma \backslash \mathbb{H}}, \\ l > 0}} \sum_{k=1}^{\infty} \frac{l}{\sinh(kl/2)} e^{-\alpha l} \\
& - \frac{1}{2\beta} \sum_{\substack{l \in \mathcal{L}_{\Gamma \backslash \mathbb{H}}, \\ l > 0}} \sum_{k=1}^{\infty} \frac{l}{\sinh(kl/2)} e^{-\beta l}.
\end{aligned}$$

We will first rewrite the hyperbolic terms. To simplify notation we first define

$$c_1(\beta) = \frac{1}{2\beta} \sum_{\substack{l \in \mathcal{L}_{\Gamma \backslash \mathbb{H}}, \\ l > 0}} \sum_{k=1}^{\infty} \frac{l}{\sinh(kl/2)} e^{-\beta l}. \tag{59}$$

Next we write $\alpha = s - \frac{1}{2}$ for $\mathrm{Re}(s) > 1$. We compute

$$e^{-(s-\frac{1}{2})lk} \sinh(kl/2)^{-1} = \frac{2}{(e^{\frac{kl}{2}} - e^{-\frac{kl}{2}}) e^{kl(s-\frac{1}{2})}} = \frac{2}{e^{kls}} + \frac{2}{e^{kls}(e^{kl} - 1)}.$$

This leads us to the definition of the Dirichlet series

$$A_1(s) = \sum_{\substack{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ l>0}} \sum_{k=1}^{\infty} \frac{l}{e^{kls}(e^{kl}-1)} \text{ for } \operatorname{Re}(s) > 0. \tag{60}$$

Note that this is absolutely convergent in the region of definition. Furthermore, it is easy to see that $A_1(s)$ is uniformly bounded on each half plane $\operatorname{Re}(s) \geq \epsilon > 0$.

Next we observe that

$$\sum_{k=2}^{\infty} e^{-kls} = \frac{e^{-2ls}}{1-e^{-ls}}.$$

Here we have simply summed the geometric series. This leads us to set

$$A_2(s) = \sum_{\substack{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ l>0}} l \cdot \frac{e^{-2ls}}{1-e^{-ls}}, \text{ for } \operatorname{Re}(s) > \frac{1}{2}. \tag{61}$$

Note that the series converges absolutely and the resulting function is uniformly bounded in half planes of the form $\operatorname{Re}(s) \geq \frac{1}{2} + \epsilon > \frac{1}{2}$.

All together we obtain that

$$\begin{aligned}\text{Hyperbolic Terms} &= \frac{1}{2\alpha} \sum_{\substack{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ l>0}} \sum_{k=1}^{\infty} \frac{l}{\sinh(kl/2)} e^{-\alpha l} - \frac{1}{2\beta} \sum_{\substack{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ l>0}} \sum_{k=1}^{\infty} \frac{l}{\sinh(kl/2)} e^{-\beta l} \\ &= \frac{1}{2s-1} E(s) + \frac{2}{2s-1} A_2(s) + \frac{2}{2s-1} A_1(s) - c(\beta).\end{aligned}$$

Next we compute the identity contribution. This is done with a trick. Indeed, we recall the automorphic resolvent operator $R_{\Gamma,s}$ and note that

$$R_{\Gamma,s}\phi_j = \frac{1}{(t_j^2 + (s-\frac{1}{2})^2)} \cdot \phi_j.$$

Therefore, we recognize the spectral side as the trace of $R_{\Gamma,s} - R_{\Gamma,\beta+\frac{1}{2}}$. In particular, the corresponding integral kernel is the automorphic Green function. We conclude that

$$k(u) = G_s(u) - G_{\beta+\frac{1}{2}}(u).$$

We can now compute the identity contribution from Lemma 5.4 as

$$\text{Identity Term} = \operatorname{Vol}(\Gamma\backslash\mathbb{H}) k(0) = \operatorname{Vol}(\Gamma\backslash\mathbb{H}) \lim_{u\to 0} (G_s(u) - G_{\beta+\frac{1}{2}}(u))$$

In view of (24) we arrive at

$$\text{Identity Term} = \frac{\operatorname{Vol}(\Gamma\backslash\mathbb{H})}{2\pi} (\psi(\beta + \frac{1}{2}) - \psi(s)).$$

Recall that ψ is the digamma function satisfying

$$\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)} = -\gamma - \sum_{n=0}^{\infty} \left(\frac{1}{n+z} - \frac{1}{n+1} \right). \tag{62}$$

Remark 6.13. This contribution is also computed in [Iw, (10.24)]. However, the formula in loc. cit. seems to have a wrong power of 2. Since we were unable to locate where this inconsistency comes from we will give an alternative evaluation of the identity contribution.

We start from

$$\text{Identity Term} = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi} \int_{\mathbb{R}} t \cdot \left(\frac{1}{t^2+(s-\frac{1}{2})^2} - \frac{1}{t^2+\beta^2} \right) \tanh(\pi t) dt.$$

Recall that the poles of $\tanh(\pi t)$ is odd with period i. Furthermore, its poles are simple and located at $(k+\frac{1}{2})i$ with $k \in \mathbb{Z}$. These poles all have residue $\frac{1}{\pi}$. Let $K \in \mathbb{N}$ be a large parameter and shift the contour to the horizontal line $\mathbb{R} + iK$. By the residue theorem we obtain

$$\begin{aligned}
&\int_{\mathbb{R}} t \cdot \left(\frac{1}{t^2+(s-\frac{1}{2})^2} - \frac{1}{t^2+\beta^2} \right) \tanh(\pi t) dt \\
&\quad = 2\pi i \left(\frac{1}{2} \tanh(\pi i (s - \frac{1}{2})) - \frac{1}{2} \tanh(\pi i \beta) \right) \\
&\qquad + 2\pi i \sum_{0 \leq k < K} \frac{i(k+1/2)}{\pi} \left(\frac{1}{(s-\frac{1}{2})^2 - (k+\frac{1}{2})^2} - \frac{1}{\beta^2 - (k+\frac{1}{2})^2} \right) \\
&\qquad + \int_{-\infty+iK}^{\infty+iK} t \cdot \left(\frac{1}{t^2+(s-\frac{1}{2})^2} - \frac{1}{t^2+\beta^2} \right) \tanh(\pi t) dt.
\end{aligned}$$

Here we are picking up the poles at $t = i(s-\frac{1}{2})$ and $t = i\beta$, which we assume to lie in the upper half plane as well as the poles of $\tanh(\pi t)$. We check that the remaining integral tends to 0 as $K \to \infty$. Thus we end up with

$$\begin{aligned}
\text{Identity Term} = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi} \bigg(& \pi \tan(\pi\beta) - \pi \tan(\pi(s - \frac{1}{2})) \\
& - \sum_{k=0}^{\infty} \left(\frac{2k+1}{(s-\frac{1}{2})^2 - (k+\frac{1}{2})^2} - \frac{2k+1}{\beta^2 - (k+\frac{1}{2})^2} \right) \bigg).
\end{aligned}$$

We have to play a bit with the sum. First one checks that

$$\frac{2k+1}{(s-\frac{1}{2})^2 - (k+\frac{1}{2})^2} - \frac{2k+1}{\beta^2 - (k+\frac{1}{2})^2} = \frac{1}{s-k-1} - \frac{1}{s+k} - \frac{1}{\beta - k - \frac{1}{2}} + \frac{1}{\beta + k + \frac{1}{2}}.$$

At this point we recall that

$$\pi \tan(\pi z) = -\sum_{k=0}^{\infty} \left(\frac{1}{z-k-\frac{1}{2}} + \frac{1}{z+k+\frac{1}{2}} \right).$$

Inserting this above yields

$$\text{Identity Term} = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{2\pi}\left(\sum_{k=0}^{\infty}\left(\frac{1}{s+k} - \frac{1}{\beta+k+\frac{1}{2}}\right)\right).$$

Comparing the k-sum to (62) allows us to conclude that

$$\text{Identity Term} = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{2\pi}(\Psi(\beta+\frac{1}{2}) - \Psi(s))$$

as desired.

The next step is to analyze the elliptic terms. We first write the elliptic contribution using Lemma 5.6 as

$$\begin{aligned}\text{Elliptic Terms} = \frac{1}{2s-1}\sum_{\substack{\{\gamma_0\}\\ \text{primitive}\\ \text{elliptic}}}\frac{1}{m(\gamma_0)}\sum_{k=1}^{m(\gamma_0)-1}\int_0^\infty \frac{e^{-(s-\frac{1}{2})r}\cosh(r/2)}{\cosh(r)-\cos(2\pi k/m(\gamma_0))}dr \\ - \frac{1}{2\beta}\sum_{\substack{\{\gamma_0\}\\ \text{primitive}\\ \text{elliptic}}}\frac{1}{m(\gamma_0)}\sum_{k=1}^{m(\gamma_0)-1}\int_0^\infty \frac{e^{-\beta r}\cosh(r/2)}{\cosh(r)-\cos(2\pi k/m(\gamma_0))}dr.\end{aligned}$$

We define

$$c_2(\beta) = \frac{1}{2\beta}\sum_{\substack{\{\gamma_0\}\\ \text{primitive}\\ \text{elliptic}}}\frac{1}{m(\gamma_0)}\sum_{k=1}^{m(\gamma_0)-1}\int_0^\infty \frac{e^{-\beta r}\cosh(r/2)}{\cosh(r)-\cos(2\pi k/m(\gamma_0))}dr. \tag{63}$$

Computing the remaining integral is a bit cumbersome. We start by recalling the identity

$$\int_0^\infty e^{-(\mu\pm\frac{1}{2})x}(\cosh(x)-\cos(t))^{-1}dx = \frac{2}{\sin(t)}\sum_{k=1}^{\infty}\frac{\sin(kt)}{\mu\pm\frac{1}{2}+k} \text{ for } \mathrm{Re}(\mu) > -\frac{1}{2} \text{ and } t\notin 2\pi\mathbb{Z}.$$

Summing over the choice of sign $\pm$ leads us to the formula

$$\int_0^\infty e^{-\mu r}\frac{\cosh(r/2)}{\cosh(r)-\cos(2\alpha)}dr = \frac{1}{\sin(\alpha)}\sum_{k=0}^{\infty}\frac{\sin((2k+1)\alpha)}{k+\mu+\frac{1}{2}}.$$

We thus obtain

$$\begin{aligned}&\frac{1}{(2s-1)m(\gamma_0)}\int_0^\infty \frac{e^{-(s-\frac{1}{2})r}\cosh(r/2)}{\cosh(r)-\cos(2\pi k/m(\gamma_0))}dr \\ &\quad = ((2s-1)m(\gamma_0)\sin(2\pi k/m(\gamma_0)))^{-1}\sum_{l=0}^{\infty}(s+l)^{-1}\sin\left((2l+1)\frac{\pi k}{m(\gamma_0)}\right).\end{aligned}$$

This is already a good expression, but we can do better. We do so by reworking the l-sum. First, we write

$$\sum_{l=0}^{\infty}(s+l)^{-1}\sin\left((2l+1)\frac{\pi k}{m(\gamma_0)}\right)$$
$$=\sum_{0\le l<m(\gamma_0)}\sin\left((2l+1)\frac{\pi k}{m(\gamma_0)}\right)\sum_{n=0}^{\infty}\left((s+l+m(\gamma_0)n)^{-1}-(m(\gamma_0)+m(\gamma_0)n)^{-1}\right).$$

Here we have simply split the sum into residue classes modulo $m(\gamma_0)$ and included the convergence terms $m(\gamma_0)+m(\gamma_0)n$ using character orthogonality. We recognize this as

$$\sum_{l=0}^{\infty}(s+l)^{-1}\sin\left((2l+1)\frac{\pi k}{m(\gamma_0)}\right)=-\frac{1}{m(\gamma_0)}\sum_{0\le l<m(\gamma_0)}\psi\left(\frac{s+l}{m(\gamma_0)}\right)\sin\left((2l+1)\frac{\pi k}{m(\gamma_0)}\right).$$

Inserting this above allows us to express the elliptic contribution as

$$\text{Elliptic Terms}+c_2(\beta)$$
$$=\frac{-1}{2s-1}\sum_{\substack{\{\gamma_0\}\\ \text{primitive}\\ \text{elliptic}}}\frac{1}{m(\gamma_0)^2}\sum_{k=1}^{m(\gamma_0)-1}\sum_{0\le l<m(\gamma_0)}\psi\left(\frac{s+l}{m(\gamma_0)}\right)\frac{\sin\left((2l+1)\frac{\pi k}{m(\gamma_0)}\right)}{\sin(2\pi k/m(\gamma_0))}.$$

This can be simplified by observing that

$$\frac{\sin\left((2l+1)\frac{\pi k}{m(\gamma_0)}\right)}{\sin(2\pi k/m(\gamma_0))}=\sum_{|n|\le l}e\left(\frac{kn}{m(\gamma_0)}\right),$$

so that

$$\sum_{0<k<m(\gamma_0)}\frac{\sin\left((2l+1)\frac{\pi k}{m(\gamma_0)}\right)}{\sin(2\pi k/m(\gamma_0))}=m(\gamma_0)-2l-1.$$

We arrive at

$$\text{Elliptic Terms}+c_2(\beta)$$
$$=\frac{1}{2s-1}\sum_{\substack{\{\gamma_0\}\\ \text{primitive}\\ \text{elliptic}}}\frac{1}{m(\gamma_0)}\sum_{l=0}^{m(\gamma_0)-1}\left(\frac{2l+1}{m(\gamma_0)}-1\right)\psi\left(\frac{s+l}{m(\gamma_0)}\right).$$

All together we obtain the following result.

Proposition 6.14 (Resolvent Trace Formula I)**.** *For s with $\mathrm{Re}(s) > 1$ and $\beta \geq 2$ we have*

$$\sum_{j=0}^{\infty}\left(\frac{1}{t_j^2+(s-\frac{1}{2})^2}-\frac{1}{t_j^2+\beta^2}\right)=\frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{2\pi}\left(\psi\left(\beta+\frac{1}{2}\right)-\psi(s)\right)$$

$$+\frac{1}{2s-1}\sum_{\substack{\{\gamma_0\}\\ \textit{primitive}\\ \textit{elliptic}}}\frac{1}{m(\gamma_0)}\sum_{l=0}^{m(\gamma_0)-1}\left(\frac{2l+1}{m(\gamma_0)}-1\right)\psi\left(\frac{s+l}{m(\gamma_0)}\right)$$

$$+\frac{1}{2s-1}E(s)+\frac{2}{2s-1}A_2(s)+\frac{2}{2s-1}A_1(s)-c_1(\beta)-c_2(\beta).$$

The Dirichlet series $A_1(s)$ and $A_2(s)$ are defined in (60) *and* (61) *respectively and $c_1(\beta), c_2(\beta)$ are given in* (59), (63).

On the spectral side a special role is played by potentially small eigenvalues. This motivates the following definition.

Definition 6.1. We define $M_\Gamma \in \mathbb{Z}_{\geq 0}$ so that

$$\lambda_{M_\Gamma} < \frac{1}{4} \leq \lambda_{M_{\Gamma+1}}.$$

For the eigenvalues $\lambda_0, \ldots, \lambda_{M_\Gamma}$ the corresponding parameters t_j are purely imaginary and we can assume that their imaginary part is negative. We set

$$s_j = \frac{1}{2} + it_j \text{ and } s_j^\vee = \frac{1}{2} - it_j.$$

Thus we have $s_j \in (\frac{1}{2}, 1]$ and $s_j^\vee \in [0, \frac{1}{2})$ In particular, $s_0 = 1$ and $s_0^\vee = 0$.

Note that we can write

$$\frac{1}{t_j^2+(s-\frac{1}{2})^2}=\frac{1}{2s-1}\left(\frac{1}{s-s_j}+\frac{1}{s-s_j^\vee}\right).$$

We arrive at the following theorem.

Theorem 6.15. *The function $E(s)$ has a meromorphic extension to $\mathrm{Re}(s) > \frac{1}{2}$ with simple poles of residue 1 at the points $s_0, \ldots, s_{M_\Gamma}$.*

Proof. According to our discussion above the trace formula allows us to write

$$E(s)=\sum_{j=0}^{M_\Gamma}\left(\frac{1}{s-s_j}+\frac{1}{s-s_j^\vee}\right)+\Psi(s)$$

for $\mathrm{Re}(s) > 1$ and some function $\Psi(s)$. We have seen that $\Psi(s)$, which is explicitly given by the trace formula, is holomorphic for $\mathrm{Re}(s) > \frac{1}{2}$. The result is a consequence of the principle of analytic continuation. □

We also need some control on the growth of $E(s)$. This is supplied by the following result.

Theorem 6.16. *For $\epsilon > 0$ sufficiently small and $t \in \mathbb{R}$ sufficiently far away from 0 (i.e. $|t| \geq 1000$) we have*

$$E(\sigma + it) \ll \frac{|t|^2}{\epsilon}, \ \textit{for } \frac{1}{2} + \epsilon \leq \sigma \leq 2.$$

Proof. We look at the resolvent trace formula. First, a version of Stirling's formula for the digamma function tells us that the contribution of the identity and the elliptic terms is $O_\Gamma(\log(|t|))$ in the region under consideration. Similarly, since $\beta \geq 2$ is considered fixed we have $c_1(\beta) + c_2(\beta) \ll 1$. Furthermore, by Proposition 6.10 we have

$$A_1(\sigma + it) \ll 1 \text{ and } A_2(\sigma + it) \ll \frac{1}{\epsilon}.$$

Thus, we obtain

$$\frac{E(s)}{2s-1} = \sum_{j=0}^{\infty} \left(\frac{1}{t_j^2 + (s - \frac{1}{2})^2} - \frac{1}{t_j^2 + \beta^2} \right) + O_\Gamma \left(\frac{1}{|t|\epsilon} + \log(|t|) \right)$$

for $s = \sigma + it$.

We have to estimate the spectral contribution. First, note that since t is sufficiently large we can trivially bound the contribution from the eigenvalues $\lambda_0, \ldots, \lambda_{M_\Gamma}$. For the record:

$$\sum_{j=0}^{M_\Gamma} \left(\frac{1}{t_j^2 + (s - \frac{1}{2})^2} - \frac{1}{t_j^2 + \beta^2} \right) \ll_\Gamma 1.$$

Thus, we only need to carefully treat the piece of the sum where the $t_j's$ are real. Let us temporarily write

$$\widetilde{N}_\Gamma(y) = \sharp\{j > M_\Gamma \colon 0 \leq t_j \leq y\}.$$

We first handle the contribution from $0 \leq t_j \leq 2|t|$ by estimating

$$\begin{aligned}
&\sum_{0 \leq t_j \leq 2|t|} \left| \frac{1}{t_j^2 + (s - \frac{1}{2})^2} - \frac{1}{t_j^2 + \beta^2} \right| \\
&\quad \leq \sum_{0 \leq t_j \leq 2|t|} \frac{1}{|(\sigma - \frac{1}{2}) + i(t - t_j)| \cdot |(\sigma - \frac{1}{2}) + i(t + t_j)|} + \sum_{0 \leq t_j \leq 2|t|} \frac{1}{t_j^2 + 1} \\
&\quad \leq \sum_{0 \leq t_j \leq 2|t|} \frac{1}{\epsilon(|t| + t_j)} + \sum_{0 \leq t_j \leq 2|t|} \frac{1}{t_j^2 + 1} \\
&\quad \ll_\Gamma 1 + \frac{\widetilde{N}_\gamma(2|t|)}{\epsilon|t|} + \int_1^{2|t|} x^{-2} d\widetilde{N}_\Gamma(x) \ll \frac{|t|}{\epsilon}.
\end{aligned}$$

In the last step we have used that $\widetilde{N}_\Gamma(y) \ll_\Gamma y^2$ via the Weyl law. On the other hand we have

$$\begin{aligned}
&\sum_{2|t|<t_j} \left| \frac{1}{t_j^2+(s-\frac{1}{2})^2} - \frac{1}{t_j^2+\beta^2} \right| \\
&\qquad = \sum_{2|t|<t_j} \left| \frac{\beta^2-(\sigma-\frac{1}{2}+it)^2}{(t_j^2+(\sigma-\frac{1}{2}+it)^2)(t_j^2+\beta^2)} \right| \\
&\qquad \ll \sum_{2|t|<t_j} \frac{T^2}{(t_j+t)(t_j-t)t_j^2} \ll \sum_{2|t|<t_j} \frac{T^2}{t_j^4} \\
&\qquad \ll T^2 \int_{2|t|}^{\infty} x^{-4} d\widetilde{N}_\Gamma(x) \ll 1.
\end{aligned}$$

The final estimate again follows from Weyl's law. Combining all the bounds completes the proof. $\square$

Using the Phragmén-Lindelöf principle we arrive at the following result.

Theorem 6.17. *For $s=\sigma+it$ with $\sigma \geq \frac{1}{2}+\epsilon$ and $|t| \geq 1000$, we have*

$$E(s) \ll_{\Gamma,\epsilon} |t|^{4\cdot\max(0,1+\epsilon-\sigma)}.$$

Proof. The result is an interpolation between the bounds

$$E(s) \ll_\epsilon |t|^2 \text{ for } \frac{1}{2}+\epsilon \leq \sigma \leq 1+\epsilon$$

and

$$E(s) \ll_\epsilon 1 \text{ for } 1+\epsilon \leq \sigma.$$

$\square$

We define

$$\theta(x) = 2 \sum_{\substack{\log(L)\in\mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ 1<L\leq x}} \log(L).$$

and

$$\theta_1(x) = \int_1^x \theta(t)dt.$$

Note that $\theta(x)=\theta_1(x)=0$ for $x<e^{l_\Gamma}$. The exponential re-scaling will make it easier to refer to tools from classical analytic number theory. In particular,

$$E(s) = 2 \sum_{\substack{\log(L)\in\mathcal{L},\\ 1<L}} \frac{\log(L)}{L^s}$$

reads (essentially) like a classical Dirichlet series. We first proof the following result.

Proposition 6.18. *For each $\epsilon > 0$ we have*

$$\theta_1(x) = \sum_{j=0}^{M_\Gamma} \frac{t^{1+s_j}}{s_j(s_j+1)} + O_{\Gamma,\epsilon}(t^{\frac{7}{4}+\epsilon}).$$

Proof. Our starting point is the formula

$$\theta_1(x) = \frac{1}{2\pi i}\int_{(2)} \frac{x^{1+s}}{s(s+1)} E(s) ds + O(1).$$

This is a version of Perron's formula. Let $\eta > 0$ be such that $\frac{3}{4} + \eta \neq s_j$. We can shift the line of integration to $\mathrm{Re}(s) = \frac{3}{4} + \eta$ picking up the residues at the poles we are crossing. We arrive at

$$\theta_1(x) = \sum_{\frac{3}{4}+\eta \le s_j \le 1} \frac{x^{1+s_j}}{s_j(s_j+1)} + \frac{1}{2\pi i}\int_{(\frac{3}{4}+\eta)} \frac{x^{s+1}}{s(s+1)} E(s) ds + O(1).$$

By Theorem 6.17 (with $\epsilon = \frac{3}{4}\eta$) we have

$$E(s) \ll (1 + |\operatorname{Im}(s)|)^{1-\eta}$$

on the line of integration $\mathrm{Re}(s) = \frac{3}{4} + \eta$. Estimating the remaining integral trivially and choosing η sufficiently small gives the desired error term. □

Theorem 6.19 (Prime Geodesic Theorem I)**.** *We have*

$$\pi_{\Gamma\backslash\mathbb{H}}(t) = li(e^t) + \sum_{j=1}^{M_\Gamma} li(e^{s_j t}) + O(e^{(\frac{7}{8}+\epsilon)t}),$$

where $li(x) = \int_2^x \log(t)^{-1} dt$.

Proof. We start by proving an asymptotic formula for θ. Note that we can write the formula from Proposition 6.18 as

$$\int_0^t \theta(x) dx = \sum_{j=0}^{M_\Gamma} \int_0^t \frac{x^{s_j}}{s_j} dx + O(t^{\frac{7}{4}+\epsilon}).$$

For $0 < h < t/2$ we have

$$\begin{aligned}\theta(t) \le \frac{1}{h}\int_t^{t+h} \theta(x) dx &= \sum_{j=0}^{M_\Gamma} \frac{1}{h}\int_t^{t+h} \frac{x^{s_j}}{s_j} dx + O(\frac{1}{h} e^{(\frac{7}{4}+\epsilon)t}) \\ &\le \sum_{j=0}^{M_\Gamma} \frac{(t+h)^{s_j}}{s_j} + O(\frac{1}{h} t^{\frac{7}{4}+\epsilon}).\end{aligned}$$

Here we have used the mean value theorem for integrals. We now use the Taylor expansion

$$(t+h)^{s_j} = t^{s_j} + O(h).$$

Note that this is sharp for $j = 0$, when $s_j = 1$. Thus we have obtained

$$\theta(t) \leq \sum_{j=0}^{M_\Gamma} \frac{t^{s_j}}{s_j} + O\left(h + \frac{1}{h} t^{\frac{7}{4}+\epsilon}\right).$$

One similarly obtains a lower bound of the same shape. Thus, after choosing $h = t^{\frac{7}{8}}$, we arrive at

$$\theta(t) = \sum_{j=0}^{M_\Gamma} \frac{t^{s_j}}{s_j} + O\left(t^{\frac{7}{8}+\epsilon}\right).$$

To conclude the proof we note that

$$\pi_{\Gamma\backslash\mathbb{H}}(\log(t)) = \int_2^t \frac{1}{\log(x)} d\theta(x) + O(1).$$

The desired result follows from rather elementary computations. $\square$

Remark 6.20. With more work it is even possible to show that

$$\pi_{\Gamma\backslash\mathbb{H}}(t) = \mathrm{li}(e^t) + \sum_{j=1}^{M_\Gamma} \mathrm{li}(e^{s_j t}) + O_\Gamma\left(e^{\frac{3}{4}t}\sqrt{t}\right).$$

We note the presence of the *exceptional* eigenvalues $\lambda_1, \ldots, \lambda_{M_\Gamma}$ as secondary terms. With our current technology the bottleneck is that our bound for $E(s)$ close to the line $\mathrm{Re}(s) = \frac{1}{2}$ is too weak. This does not allow us to shift the contour beyond $\mathrm{Re}(s) = \frac{3}{4} + \eta$ in the proof Proposition 6.18.

6.3. **Selberg's zeta function.** In the previous section we have used analytic information on the Dirichlet series $E(s)$ to prove the prime geodesic theorem. It however turns out that it is beneficial to study a slightly different generating function.

Recall that the full hyperbolic contribution to the resolvent trace formula was given by

$$\sum_{\substack{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}, \\ l>0}} \sum_{k=1}^{\infty} \frac{l}{\sinh(kl/2)} e^{-(s-\frac{1}{2})l} = 2 \sum_{\substack{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}}, \\ l>0}} \sum_{k=1}^{\infty} \frac{l}{e^{kls} - e^{kl(s-1)}}.$$

We will rearrange this as follows

$$
\begin{aligned}
2\sum_{\substack{l\in\mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ l>0}}\sum_{k=1}^{\infty}\frac{l}{e^{kls}-e^{kl(s-1)}} &= 2\sum_{\substack{l\in\mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ l>0}} l\sum_{k=1}^{\infty}\frac{e^{-kls}}{1-e^{-kl}} \\
&= 2\sum_{\substack{l\in\mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ l>0}} l\sum_{k=1}^{\infty}\sum_{n=0}^{\infty}e^{-kls}e^{-kln} \\
&= 2\sum_{\substack{l\in\mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ l>0}} l\sum_{n=0}^{\infty}\frac{e^{-sl-nl}}{1-e^{-sl-nl}} \\
&= 2\sum_{\substack{l\in\mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ l>0}}\sum_{n=0}^{\infty}\frac{d}{ds}\left[\log(1-e^{-sl-nl})\right] \\
&= \frac{d}{ds}\log\left[\prod_{l\in\mathcal{L}_{\Gamma\backslash\mathbb{H}}}\prod_{n=0}^{\infty}(1-e^{-(s+k)|l|})\right].
\end{aligned}
$$

Recall the following definition.

Definition 6.2. For $\mathrm{Re}(s)>1$ we define the Selberg Zeta function of Γ by

$$
Z_\Gamma(s)=\prod_{l\in\mathcal{L}_{\Gamma\backslash\mathbb{H}}}\prod_{n=0}^{\infty}(1-e^{-(s+k)|l|}).
$$

Thus, our computation above shows that

$$
\frac{Z_\Gamma'(s)}{Z(s)}=\sum_{\substack{l\in\mathcal{L}_{\Gamma\backslash\mathbb{H}},\\ l>0}}\sum_{k=1}^{\infty}\frac{l}{\sinh(kl/2)}e^{-(s-\frac{1}{2})l}=E(s)+2A_1(s)+2A_2(s),
$$

where $A_1(s)$ and $A_2(s)$ were defined in (60) and (61) respectively. We can thus write the Resolvent trace formula in the following much nicer form.

Theorem 6.21 (Resolvent trace formula II)**.** *Let $\Gamma \subseteq SL_2(\mathbb{R})$ be a discrete co-compact subgroup. For $s, z \in \mathbb{C}$ with $1 < \mathrm{Re}(s) < \mathrm{Re}(z)$ we have*

$$\begin{aligned}\sum_{j=0}^{\infty}&\left(\frac{1}{t_j^2+(s-\frac{1}{2})^2}-\frac{1}{t_j^2+(z-\frac{1}{2})^2}\right)\\ &= \frac{1}{2s-1}\frac{Z_\Gamma'(s)}{Z_\Gamma(s)}-\frac{1}{2z-1}\frac{Z_\Gamma'(z)}{Z_\Gamma(z)}+\chi(\Gamma)\cdot(\psi(z)-\psi(s))\\ &\qquad+\sum_{\substack{\{\gamma_0\}\\ \textit{primitive}\\ \textit{elliptic}}}\left(\frac{1}{2s-1}R_{m(\gamma_0)}(s)-\frac{1}{2z-1}R_{m(\gamma_0)}(z)\right).\end{aligned}$$

for

$$R_m(s) = m^{-2}\sum_{0\le l<m}(2s+2k-m)\psi\left(\frac{s+l}{m}\right) \quad \textit{and} \quad \chi(\Gamma)=\frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{2\pi}+\sum_{\substack{\{\gamma_0\},\\ \textit{prim. ell.}}}\frac{1}{m(\gamma_0)}.$$

Proof. This follows by rearranging the expression in Proposition 6.14 with $\beta = z-\frac{1}{2}$. Note that the terms $c_1(\beta)$ and $c_2(\beta)$, which previously were simply constants, are computed as their $\alpha = s-\frac{1}{2}$ counterparts. We will only slightly rearrange the elliptic contribution, which currently reads

$$\sum_{\substack{\{\gamma_0\}\\ \text{primitive}\\ \text{elliptic}}}\frac{1}{m(\gamma_0)}\sum_{l=0}^{m(\gamma_0)-1}\left(\frac{2l+1}{m(\gamma_0)}-1\right)\left(\frac{1}{2s-1}\psi\left(\frac{s+l}{m(\gamma_0)}\right)-\frac{1}{2z-1}\psi\left(\frac{z+l}{m(\gamma_0)}\right)\right).$$

To do so we recall the curious identities

$$\frac{1}{m}\sum_{0\le l<m}\psi\left(\frac{s+l}{m}\right)=\psi(s)-\log(m).$$

and

$$\frac{2l+1}{m}-1=\frac{2s+2l-m}{m}-\frac{2s-1}{m}.$$

With this in mind we define

$$R_m(s) = m^{-2}\sum_{0\le l<m}(2s+2k-m)\psi\left(\frac{s+l}{m}\right). \tag{64}$$

The upshot is that we can write

$$\frac{1}{m(\gamma_0)}\sum_{l=0}^{m(\gamma_0)-1}\left(\frac{2l+1}{m(\gamma_0)}-1\right)\frac{1}{2s-1}\psi\left(\frac{s+l}{m(\gamma_0)}\right)$$
$$=\frac{1}{m(\gamma_0)}(\log(m(\gamma_0))-\psi(s))+\frac{1}{2s-1}R_{m(\gamma_0)}(s).$$

Inserting this for the elliptic contribution gives the desired result. □

Remark 6.22. The functions $R_m(z)$ are meromorphic functions on the complex plane with simple poles at the non-negative integers $-d$ (i.e. $d \in \mathbb{Z}_{\geq 0}$). The residues are positive integers. Furthermore, by the Gauß- Bonnet formula given in (39), we have

$$\chi(\Gamma)=2g-2+\sharp\{\text{prim. ell. conj. cl.}\}\in\mathbb{N}.$$

Since $\frac{Z_\Gamma'(s)}{Z_\Gamma(s)}=E(s)+2A_1(s)+2A_2(s)$, we can directly infer the following properties:

- The function $\frac{Z_\Gamma'(s)}{Z_\Gamma(s)}$ has a meromorphic continuation to the half plane $\mathrm{Re}(s) > \frac{1}{2}$ with poles at $s_0,\dots,s_{M_\Gamma}$ with residue 1. (This follows from Theorem 6.15.)
- For $\frac{1}{2}+\epsilon\leq\sigma\leq 1+\epsilon$ and $|t|\geq 2024$ we have $\frac{Z_\Gamma'(\sigma+it)}{Z_\Gamma(\sigma+it)}\ll_\Gamma \frac{|t|^2}{\epsilon}$. (This is Theorem 6.16.)
- For $\frac{1}{2}+\epsilon\leq\sigma\leq 1+\epsilon$ and $|t|\geq 2024$ we have $\frac{Z_\Gamma'(\sigma+it)}{Z_\Gamma(\sigma+it)}\ll_{\Gamma,\epsilon} |t|^{4(1-\sigma)+\epsilon}$. (This is essentially Theorem 6.17.)

However, we can obtain much more information now.

Theorem 6.23. *Let $\Gamma\subseteq \mathrm{SL}_2(\mathbb{R})$ be discrete and co-compact. The Selberg Zeta function $Z_\Gamma(s)$ defined in Definition 6.2 for $\mathrm{Re}(s)>1$ has an analytic continuation to the complex plane and the identity given in Theorem 6.21 remains true for all $s\in\mathbb{C}$. Furthermore, we have the functional equation*

$$\frac{Z_\Gamma'(s)}{Z_\Gamma(s)}+\frac{Z_\Gamma'(1-s)}{Z_\Gamma(1-s)}=-(2s-1)\chi(\Gamma)\cdot\pi\cot(\pi s)-\sum_{\substack{\{\gamma_0\}\\ \text{primitive}\\ \text{elliptic}}}(R_{m(\gamma_0)}(s)+R_{m(\gamma_0)}(1-s)).$$

Proof. Using Theorem 6.21 we can meromorphically continue $\frac{Z_\Gamma'(s)}{Z(s)}$ to all $\mathbb{C}$. Looking at the poles allows shows holomorphicity of $Z(s)$.

To see the functional equation we look at the resolvent trace formula with $z = 1-s$. Note that for this choice the spectral side vanishes. Furthermore, we recall the identity

$$\psi(1-s)-\psi(s)=\pi\cot(\pi s).$$

The resolvent trace formula therefore reads

$$\frac{1}{2s-1}\left(\frac{Z_\Gamma'(s)}{Z_\Gamma(s)}+\frac{Z_\Gamma'(1-s)}{Z_\Gamma(1-s)}\right)$$
$$=-\chi(\Gamma)\cdot\pi\cot(\pi s)-\frac{1}{2s-1}\sum_{\substack{\{\gamma_0\}\\ \text{primitive}\\ \text{elliptic}}}(R_{m(\gamma_0)}(s)+R_{m(\gamma_0)}(1-s)),$$

which is the desired formula. □

Remark 6.24. One can reformulate the functional equation in terms of $Z_\Gamma(s)$ with auxiliary factors involving the Barnes G-function.

6.4. **Odds and ends.** We will now sketch some interesting results that can be derived by refining the analysis carried out in the previous section. Doing so we will pick some interesting results from [He76, Chapter Two]. Note Hejhal's approach is strongly motivated by classical analytic number theory. This means that the results are usually derived from a careful analysis of the Selberg-Zeta function, which takes the place of the Riemann zeta function.

Throughout we will assume that Γ has no elliptic elements. This assumption is only for technical convenience and in accordance with the set-up from [He76].

We start by defining

$$\Lambda(L)=\frac{\log(L)}{1-L^{-1}}.$$

This function is defined such that

$$\frac{Z_\Gamma'(s)}{Z_\Gamma(s)}=2\sum_{\substack{L>1,\\ \log(L)\in\mathcal{L}_{\Gamma\backslash\mathbb{H}}}}\frac{\Lambda(L)}{L^s}\ \text{for}\ \operatorname{Re}(s)>1.$$

We define the counting functions

$$\psi(x)=2\sum_{\substack{1<L\leq x,\\ \log(L)\in\mathcal{L}_{\Gamma\backslash\mathbb{H}}}}\Lambda(L)\ \text{and}\ \psi_1(x)=\int_1^x\psi(t)dt. \tag{65}$$

Note that these functions are closely related to the counting functions $\theta(x)$ and $\theta_1(x)$ that we encountered earlier.

Theorem 6.25 (Explicit formula)**.** *There are number $\alpha_0,\alpha_1,\beta_0$ and β_1 depending only on Γ such that we have*

$$\psi_1(x)=\alpha_0x+\beta_0x\log(x)+\alpha_1+\beta_1\log(x)+\chi(\Gamma)\sum_{k=2}^{\infty}\frac{2k+1}{k(k-1)}x^{1-k}$$
$$+\sum_{j=1}^{M_\Gamma}\left(\frac{x^{1+s_j}}{s_j(1+s_j)}+\frac{x^{1+s_j^\vee}}{s_j^\vee(1+s_j^\vee)}\right)+\sum_{t_j\geq0}\frac{x^{1+s_j}}{s_j(1+s_j)}+\sum_{t_j\geq0}\frac{x^{1+s_j^\vee}}{s_j^\vee(1+s_j^\vee)}.$$

The t_j-sums converge uniformly for x in a fixed interval, but not absolutely.

Proof. This is [He76, Theorem 5.12]. We will only sketch the idea of the proof and skip the details. One starts from the formula

$$\psi_1(x) = \frac{1}{2\pi i}\int_{(2)} \frac{x^{s+1}}{s(s+1)} \cdot \frac{Z_\Gamma'(s)}{Z_\Gamma(s)} ds.$$

One now carefully shifts the contour to $-A - i\mathbb{R}$ with $A \to \infty$. Doing so one picks up the residues at all poles of $\frac{Z_\Gamma'(s)}{Z_\Gamma(s)}$. These come from the negative integers as well as from the eigenvalues of Δ. The numbers $\alpha_0, \alpha_1, \beta_0, \beta_1$ are arise from the poles at $s = 0$ and $s = -1$, whose treatment requires some special care. □

As a corollary of the proof of this formula one obtains the following strengthened error term in the prime geodesic theorem. For more details we refer to [He76, Theorem 5.14].

Theorem 6.26 (Prime Geodesic Theorem II)**.** *We have*

$$\pi_{\Gamma\backslash\mathbb{H}}(t) = li(e^t) + \sum_{j=1}^{M_\Gamma} li(e^{s_j t}) + O(e^{(\frac{3}{4}+\epsilon)t}).$$

We will now revisit Weyl's law. First note that for $\mathrm{Re}(s) > 1$ the Selberg zeta function $Z_\Gamma(s)$ is non-zero, so that we have no problem in defining $\log(Z_\Gamma(s))$. We then analytically continue $\log(Z_\Gamma(s))$ to the plane $\mathrm{Re}(s) \geq \frac{1}{2}$ where we remove the poles.

Recall that we had studied

$$N_\Gamma(x) = \sharp\{j \geq 0\colon \lambda_j \leq x\}.$$

It is convenient to introduce the variant

$$\widetilde{N}_\Gamma(x) = \sharp\{j > M_\Gamma\colon t_j \leq x\}.$$

of this counting function. It turns out that this is closely related to the argument of the Selberg-zeta function.

Theorem 6.27 (Theorem 7.1, [He76])**.** *Let $T \geq 1$ be a parameter with $T \neq t_j$ for all $j > M_\Gamma$. Then we have*

$$\widetilde{N}_\Gamma(T) = \frac{\mathrm{Vol}(\Gamma\backslash\mathbb{H})}{4\pi}T^2 + S(T) + E(T),$$

where

$$S(T) = \frac{1}{\pi}\arg\left(Z_\Gamma(\frac{1}{2}+iT)\right) \text{ and}$$

$$E(T) = \frac{Vol(\Gamma\backslash\mathbb{H})}{2\pi}\int_0^T t[\tanh(\pi t) - 1]dt - (M_\Gamma + 1).$$

Proof. We consider the rectangle

$$R_A(T) = [1 - A, A] \times i[-T \times T] \text{ for } 1 < A < 2.$$

Integrating over the boundary of this rectangle gives

$$2\widetilde{N}_\Gamma(T) + (M_\Gamma + 1) + M_\Gamma + (2g - 1) = \int_{\partial R_A(T)} \frac{Z'_\Gamma(s)}{Z_\Gamma(s)} ds.$$

This is a simple consequence of the residue theorem. Note that, since we are assuming that there are no elliptic elements the formula is particularly nice. In particular we have $2g - 2 = \chi(\Gamma)$. We write the functional equation as

$$\frac{Z'_\Gamma(s)}{Z_\Gamma(s)} + \frac{Z'_\Gamma(1-s)}{Z_\Gamma(1-s)} = -\operatorname{Vol}(\Gamma\backslash\mathbb{H})(s - \frac{1}{2})\cot(\pi s) = \frac{B'(s)}{B(s)}.$$

Let $\partial_{\text{right}} R_A(T)$ denote the part of $\partial R_A(T)$ lying on the right hand side of the vertical line $(\frac{1}{2})$. The functional equation allows us to write

$$2\widetilde{N}_\Gamma(T) + 2M_\Gamma + 2g = \frac{1}{\pi i}\int_{\partial_{\text{right}} R_A(T)} \frac{Z'_\Gamma(s)}{Z_\Gamma(s)} ds - \frac{1}{2\pi i}\int_{\partial_{\text{right}} R_A(T)} \frac{B'(s)}{B(s)} ds.$$

We define

$$\frac{\operatorname{Vol}(\Gamma\backslash\mathbb{H})}{4\pi} T^2 + E(T) = -\frac{1}{4\pi i}\int_{\partial_{\text{right}} R_A(T)} \frac{B'(s)}{B(s)} ds - M_\Gamma - g \tag{66}$$

The function $B(s)$ is explicit enough, so that the desired formula for $E(T)$ is easily derived. We are left with

$$S(T) = \frac{1}{2\pi i}\int_{\partial_{\text{right}} R_A(T)} \frac{Z'_\Gamma(s)}{Z_\Gamma(s)} ds.$$

To evaluate this we note that the contour $\partial_{\text{right}} R_A(T)$ is in the part of the domain where $\log(Z_\Gamma(s))$ is defined and analytic. One computes the desired formula for $S(T)$ after recalling that $\frac{Z'_\Gamma(s)}{Z_\Gamma(s)} = \frac{d}{ds}\log(Z_\Gamma(s))$. This concludes our sketch of this proof. □

Using standard techniques one can show that $E(T) = O(1)$ and $S(T) = O(T)$. This gives an alternative proof of Theorem 6.4. A refined analysis of $S(T)$, see [He76, Theorem 8.1], will lead to

Theorem 6.28 (Weyl law III)**.** *We have*

$$N_\Gamma(T) = \frac{\operatorname{Vol}(\Gamma\backslash\mathbb{H})}{4\pi} T + O_\Gamma\left(\frac{\sqrt{T}}{\log(T)}\right).$$

We will conclude this section by giving an example for co-compact Γ (without elliptic elements). To do so we fix a prime $p \equiv 1 \bmod 4$ and a positive integer a, which is not a quadratic residue modulo p (i.e. $a \not\equiv x^2 \bmod p$). We define

$$\Gamma(a,p) = \left\{ \begin{pmatrix} y_0 + y_1\sqrt{a} & (y_2 + y_3\sqrt{a})\sqrt{p} \\ (y_2 - y_3\sqrt{a})\sqrt{p} & y_0 - y_1\sqrt{a} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}) \colon y_0, y_1, y_2, y_3 \in \mathbb{Z} \right\}.$$

If $\gamma \in \Gamma(a,p)$, then $\mathrm{Tr}(\gamma) = 2y_0$. Thus, potential elliptic elements must have $y_0 = 0$. In this case we write $\det(\gamma) = 1$ as

$$1 + y_1^2 a = -p(y_2^2 - y_3^2 a).$$

In particular we obtain

$$y_1^2 a \equiv -1 \bmod p.$$

Since $p \equiv 1 \bmod 4$ we have that -1 is a quadratic residue modulo 4. Thus, the above congruence implies that a is a quadratic residue modulo p. This contradicts our assumption and we conclude that there are no elliptic elements.

Similarly, if γ is parabolic, then we must have $y_0 = \pm 1$. Using the determinant equation we arrive at

$$y_1^2 a = -p(y_2^2 - y_3^2 a).$$

This implies that $p \mid y_1$, so that

$$y_2^2 - y_3^2 a \equiv 0 \bmod p.$$

Using the fact that a is no quadratic residue modulo p we arrive at $y_2 = y_3 = 0$, so that ± 1 is the only element with trace ± 2.

We can now summarize properties of $\Gamma(a,p)$:

- $\Gamma(a,p)$ is a non-commutative subgroup of $\mathrm{SL}_2(\mathbb{R})$ with only hyperbolic elements. In particular, by Theorem 3.11, we deduce that $\Gamma(a,p)$ is discrete.
- One can show that $\Gamma(a,p)$ has finite co-volume. This is not trivial and we use this as a fact. As a consequence we see that $\Gamma(a,p)$ is co-compact. This is due to Theorem 3.10.

We have thus seen a concrete example of a co-compact Fuchsian group with no elliptic elements.

Remark 6.29. The construction of $\Gamma(a,p)$ might appear to come put of thin air. However it is quite natural. Indeed it is obtained from an order in the (rational) quaternion division algebra $\mathcal{H}(\mathbb{Q}) = \mathbb{Q} + \mathbb{Q}i + \mathbb{Q}j + \mathbb{Q}k$ where $i^2 = a$ and $j^2 = p$. This connection can also be used to establish the fact that $\Gamma(a,p)$ has finite co-volume.

7. Spectral theory for the modular curve

In Section 4 we have developed the spectral theory for compact quotients $\Gamma\backslash\mathbb{H}$. We now turn our attention to the case of non-compact quotients with finite volume, where the theory is very different. Instead of considering general Fuchsian groups we will focus on $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. This group has many special features which simplify the theory considerably. However, we still encounter (essentially) all features that one finds for general non-compact quotients of finite co-volume.

Throughout this section we write

$$\Gamma = \mathrm{SL}_2(\mathbb{Z}) \text{ and } \overline{\Gamma} = \mathrm{PSL}_2(\mathbb{Z}).$$

Once and for all we fix the standard fundamental domain

$$\mathcal{F} = \{z = x + iy \in \mathbb{H} \colon |x| \leq \frac{1}{2},\ |z| \geq 1\}.$$

We have a parabolic element given by

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = n(1).$$

It stabilizes ∞, which is a vertex of our fundamental domain. We have

$$\Gamma_\infty = \{\pm T^m \colon m \in \mathbb{Z}\} \text{ and } \overline{\Gamma}_\infty = \langle T \rangle.$$

The set of parabolic fixed points is $\mathbb{Q} \cup \{\infty\}$, but all these are Γ-equivalent. Thus, the modular curve $\Gamma\backslash\mathbb{H}$ has only one cusp $\mathfrak{a}$. We choose ∞ as a canonical representative for the cusp $\mathfrak{a}$.

Recall that our goal is to obtain a suitable spectral decomposition of $L^2(\Gamma\backslash\mathbb{H})$ for Δ. Given $f \in L^2(\Gamma\backslash\mathbb{H})$ we observe that

$$f(z+1) = f(T.z) = f(z).$$

Thus, such a function f is one-periodic in the x-direction, so that we obtain a Fourier expansion of the form

$$f(x+iy) = \sum_{n \in \mathbb{Z}} f_n(y) e(nx) \text{ for } f_n(y) = \int_0^1 f(iy + x) e(-nx) dx.$$

If f is smooth this series converges absolutely and uniformly on compacta. The following general result will turn out to be useful.

Proposition 7.1. *Let $f \colon \mathbb{H} \to \mathbb{C}$ be a function such that*

(1) *$f(z+m) = f(z)$ for all $m \in \mathbb{Z}$;*
(2) *$\Delta f = \lambda f$ for $\lambda = s(1-s)$; and*
(3) *$f(z) = o(e^{2\pi y})$ as $y \to \infty$.*

Then we can write f as

$$f(z) = f_0(y) + \sum_{n \neq 0} a_f(n) \cdot W_s(nz)$$

for

$$W_s(nz) = 2\sqrt{|n|y} K_{s-\frac{1}{2}}(2\pi|n|y) e(nx) \text{ and}$$
$$f_0(y) = \frac{A_f}{2}(y^s + y^{1-s}) + \frac{B_f}{2s-1}(y^s - y^{1-s}).$$

Proof. Due to (1) we have a Fourier expansion for f. For $n \neq 0$ we make the Ansatz $f_n(y) = g(2\pi n y)$. This allows us to write (2) as

$$g''(y) + (\frac{\lambda}{y} - 1)g(y) = 0.$$

This second order ODE is well known to have the following two linear independent solutions

$$\sqrt{2\pi y}K_{s-\frac{1}{2}}(y) \sim e^{-y} \text{ and } \sqrt{2\pi y}I_{s-\frac{1}{2}}(y) \sim e^{y}.$$

The second solution is excluded by the growth condition (3). This shows that $f_n(y)$ must have the predicted shape (as a function of y). The zeroth coefficient is determined similarly. □

We will now turn towards the construction of certain Γ-invariant functions. As in earlier situations we do so using an averaging argument. However, the presence of the cusp allows us a slightly different construction. Indeed, given $\psi\colon \mathbb{R}_{>0} \to \mathbb{C}$ we define

$$E_\infty(z|\psi) = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \psi(\mathrm{Im}(\gamma.z)) = \frac{1}{2}\sum_{\substack{c,d\in\mathbb{Z},\\ (c,d)=1}} \psi\left(\frac{\mathrm{Im}(y)}{|cz+d|^2}\right).$$

The second expression allows us to deduce that the sum converges absolutely as soon as

$$\psi(y) \ll y\log(y)^{-2} \text{ as } y \to 0.$$

Before further studying $E_\infty(z|\psi)$ we make the following important definition.

Definition 7.1 (Eisenstein series)**.** Given $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$ we define the Eisenstein series as

$$E(z,s) = E_\infty(z|(\cdot)^s) = \mathrm{Im}(z)^s \cdot \frac{1}{2}\sum_{\substack{c,d\in\mathbb{Z},\\ (c,d)=1}} |cz+d|^{-2s}.$$

We make the following two important observations

- The function $E(\cdot,s)$ is Γ-invariant and satisfies $\Delta E(\cdot,s) = \lambda E(\cdot,s)$ for $\lambda = s(1-s)$.
- $E(\cdot,s)$ is not in $L^2(\Gamma\backslash\mathbb{H})$.

Our goal is to compute the Fourier expansion of $E(z,s)$. To do so we have to recall the Riemann zeta function:

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s} = \prod_p (1-p^{-s})^{-1} \text{ for } \mathrm{Re}(s) > 1.$$

The product is taken over all prime numbers p and is the so called Euler product of $\zeta(s)$. More generally, given coefficients $a(n)$ with $a(nm) = a(n)a(m)$ for $(n,m) = 1$

we have

$$\sum_{n=1}^{\infty} a(n)n^{-s} = \prod_p \left(\sum_{k=0}^{\infty} a(p^k)p^{-ks} \right)$$

whenever the sum is absolutely convergent. This is essentially equivalent to the fundamental theorem of arithmetic. We will use these Euler product for several computations in the proof of the following important theorem.

Theorem 7.2. *For* $\mathrm{Re}(s) > 1$ *we have*

$$E(z,s) = \mathrm{Im}(z)^s + \varphi(s)\,\mathrm{Im}(z)^{1-s} + \frac{1}{\zeta^*(2s)} \sum_{m \neq 0} |m|^{-\frac{1}{2}} \eta_s(m) W_s(mz).$$

for

$$\zeta^*(s) = \pi^{-\frac{s}{2}}\Gamma(s/2)\zeta(s),\ \varphi(s) = \frac{\zeta^*(2s-1)}{\zeta^*(2s)} \text{ and}$$

$$\eta_s(n) = \sum_{ab=n} \left(\frac{a}{b}\right)^s.$$

Proof. We start from the expression

$$E(z,s) = \frac{1}{2} \sum_{\substack{c,d\in\mathbb{Z},\\ (c,d)=1}} |cz+d|^{-2s}.$$

We first sum over c and observe that if $c = 0$ we can only have $d = \pm 1$. For $c \neq 0$ we write d as $d + ct$ for $t \in \mathbb{Z}$ and d mod c with $(c,d) = 1$. We obtain

$$E(x+iy,s) = y^s + y^s \sum_{c>0} \sum_{\substack{d \bmod c\\ (c,d)=1}} \sum_{t\in\mathbb{Z}} (c^2y^2 + (cx+d+ct)^2)^{-s}.$$

Applying Poisson summation in the t-sum yields

$$E(x+iy,s) = y^s + y^s \sum_{c>0} \sum_{\substack{d \bmod c\\ (c,d)=1}} \sum_{m\in\mathbb{Z}} I_m(c,d,x,y,s),$$

for

$$I_m(c,d,x,y,s) = \int_{\mathbb{R}} (c^2y^2 + (cx+d+ct)^2)^{-s} e(-mt)dt.$$

After some simple changes of variables we obtain

$$I_m(c,d,x,y,s) = c^{-2s}y^{1-2s}e(mx + m\frac{d}{c}) \int_{\mathbb{R}} (1+x^2)^{-s} e(-myt)dt.$$

If $m = 0$, we easily recognize this integral as

$$I_0(c,d,x,y,s) = y^{1-2s}c^{-2s}\sqrt{\pi}\frac{\Gamma(s-1/2)}{\Gamma(s)}.$$

For $m \neq 0$ we further write this as

$$I_m(c,d,x,y,s) = 2c^{-2s}y^{1-2s}e(mx+m\frac{d}{c})\int_0^\infty (1+x^2)^{-s}\cos(2\pi myt)dt.$$

The remaining integral is the Basset-integral for the K-Bessel function. Indeed we find that

$$\begin{aligned}
I_m(c,d,x,y,s) &= 2c^{-2s}y^{1-2s}e(mx+m\frac{d}{c})\sqrt{\pi}\Gamma(s)^{-1}\left(\frac{2\pi|m|y}{2}\right)^{2-\frac{1}{2}} K_{s-\frac{1}{2}}(2\pi|m|y) \\
&= e(m\frac{d}{c})c^{-2s}\frac{\pi^s|m|^{s-1}}{\Gamma(s)}W_s(m(x+iy)).
\end{aligned}$$

Inserting this in our expression for $E(x+iy,s)$ leaves us with

$$\begin{aligned}
E(x+iy,s) = y^s + y^{1-s}\sqrt{\pi}\frac{\Gamma(s-1/2)}{\Gamma(s)}\sum_{c>0}\sum_{\substack{d \bmod c\\(c,d)=1}} c^{-2s} \\
+ \frac{\pi^s}{\Gamma(s)}\sum_{m\neq 0}|m|^{s-1}\sum_{c>0}\sum_{\substack{d \bmod c\\(c,d)=1}} c^{-2s}e(m\frac{d}{c})W_s(m(x+iy)).
\end{aligned}$$

Our remaining job is to compute the c-sum.

We first note that

$$\varphi(c) = \sum_{\substack{d \bmod c\\(c,d)=1}} = \sharp(\mathbb{Z}/c\mathbb{Z})^\times. \tag{67}$$

This is Euler's phi-function. By the Chinese Remainder Theorem we obtain that

$$\varphi(c) = \varphi(c_1)\varphi(c_2) \text{ for } c = c_1c_2 \text{ with } (c_1,c_2)=1.$$

One easily checks that $\varphi(p^k) = (1-1/p)p^k$ for prime powers p^k with $k>0$. Thus we can write

$$\begin{aligned}
\sum_{c>0}\sum_{\substack{d \bmod c\\(c,d)=1}} c^{-2s} = \sum_{c>0}\varphi(c)c^{-2s} &= \prod_p\left(\sum_{k=0}^\infty \varphi(p^k)p^{-2ks}\right) \\
&= \prod_p\left(1+(1-1/p)\sum_{k=1}^\infty p^{-(2s-1)k}\right) \\
&= \prod_p\left(1+(1-1/p)p^{-(2s-1)k}\frac{1}{1-p^{-(2s-1)}}\right) \\
&= \prod_p\frac{1-p^{-2s}}{1-p^{-(2s-1)}} = \frac{\zeta(2s-1)}{\zeta(2s)}.
\end{aligned}$$

Thus we have obtained

$$E(x+iy,s) = y^s + \frac{\zeta^*(2s-1)}{\zeta^*(2s)}y^{1-s} \\ + \frac{\pi^s}{\Gamma(s)}\sum_{m\neq 0}|m|^{s-1}\sum_{c>0}c^{-2s}\sum_{\substack{d \bmod c\\(c,d)=1}} e(m\frac{d}{c})W_s(m(x+iy)).$$

To compute the final sum we define

$$R_c(m) = \sum_{\substack{d \bmod c\\(c,d)=1}} e(m\frac{d}{c}).$$

These are so called Ramanujan sums. Again one can use the Chinese Remainder Theorem to show that $R_c(m) = R_{c_1}(m)\cdot R_{c_2}(m)$ for $c = c_1c_2$ with $(c_1,c_2) = 1$. We obtain

$$\sum_{c>0}c^{-2s}\sum_{\substack{d \bmod c\\(c,d)=1}} e(m\frac{d}{c}) = \sum_{c>0}c^{-2s}R_c(m) = \prod_p\left(\sum_{k=0}^{\infty}R_{p^k}(m)p^{-2sk}\right).$$

We thus need to understand the Ramanujan sums on prime powers. To evaluate this we let $v_p(m)$ denote the exact power of p that divides m (i.e. $p^{v_p(m)} \mid m$ and $p^{v_p(m)+1} \nmid m$). We make the following observations:

- If $0 \le k \le v_p(m)$ (i.e. $p^k \mid m$), then we have $R_{p^k}(m) = \varphi(p^k)$.
- If $v_p(m) \le k$ (i.e. $(p^k, m) = p^{v_p(m)}$), then we have $R_{p^k}(m) = p^{v_p(m)}R_{p^{k-v_p(m)}}(m)$.
- If $v_p(m) = 0$ and $k = 1$, then

$$R_p(m) = \sum_{d \bmod p} e\left(m\frac{d}{p}\right) - 1 = -1 \tag{68}$$

  by character orthogonality.
- If $v_p(m) = 0$ and $k > 1$, then

$$R_{p^k}(m) = \sum_{d \bmod p^k} e\left(m\frac{d}{p^k}\right) - \sum_{d \bmod p^{k-1}} e\left(m\frac{d}{p^{k-1}}\right) = 0.$$

These facts lead to a complete evaluation of $R_c(m)$. We can now compute the Euler factors

$$\sum_{k=0}^{\infty} R_{p^k}(m)p^{-2sk} = 1 + \sum_{k=1}^{v_p(m)} p^{-2ks}\varphi(p^k) - p^{-2s(v_p(m)+1)+v_p(m)}$$
$$= 1 + (1-1/p)p^{-(2s-1)}\frac{1-p^{-v_p(m)(2s-1)}}{1-p^{-(2s-1)}} - p^{-2s(v_p(m)+1)+v_p(m)}$$
$$= (1-p^{-2s})\frac{1-p^{-(2s-1)(v_p(m)+1)}}{1-p^{-(2s-1)}} = (1-p^{-2s})\sum_{d|p^{v_p(m)}} d^{1-2s}.$$

We conclude that

$$\sum_{c>0} c^{-2s} \sum_{\substack{d \bmod c\\(c,d)=1}} e(m\frac{d}{c}) = \zeta(2s)^{-1}\sum_{d|m} d^{1-2s}.$$

Inserting this above yields

$$E(x+iy,s) = y^s + \frac{\zeta^*(2s-1)}{\zeta^*(2s)}y^{1-s} + \frac{1}{\zeta^*(2s)}\sum_{m\neq 0}|m|^{s-1}\sum_{d|m} d^{1-2s}W_s(m(x+iy)).$$

This completes the proof. $\square$

Remark 7.3. A similar computation shows that the constant term of $E(\cdot|\psi) \in \mathfrak{E}(\Gamma\backslash\mathbb{H})$ is given by

$$\psi(y) + \sum_{c>0}\varphi(c)\int_{\mathbb{R}}\psi\left(\frac{y}{c^2(t^2+y^2)}\right)dt.$$

We recall some sets about the completed Riemann zeta function $\zeta^*(s)$:

- The function $\zeta^*(s)$ has a meromrphic continuation to $s \in \mathbb{C}$ with a simple pole at $s = 1$ and functional equation
$$\zeta^*(s) = \zeta^*(1-s).$$
We also record that $\mathrm{Res}_{s=1}\,\zeta^*(s) = 1$.
- $\zeta^*(s)$ has obviously no zeros in the (open) half plane $\mathrm{Re}(s) > 1$. It is even known that $\zeta^*(s) \neq 0$ for $\mathrm{Re}(s) \geq 1$. The latter is a relatively deep result, which is equivalent to the prime number theorem.

These facts allows us to draw the following conclusion from the Fourier expansion of $E(z,s)$.

Corollary 7.4. *The function $s \mapsto E(\cdot,s)$ has a meromorphic continuation to $s \in \mathbb{C}$. In the half plane $\mathrm{Re}(s) \geq \frac{1}{2}$ there is exactly one simple pole at $s = 1$ with*

$$\mathrm{Res}_{s=1}\,E(z,s) = \frac{3}{\pi}.$$

Furthermore, we have the functional equation

$$\zeta^*(2s)E(z,s) = \zeta^*(2(1-s))E(z,1-s).$$

Proof. The meromorphic continuation follows directly from the analytic properties of $\zeta^*(s)$, $\eta_s(n)$ and $K_s(y)$. To verify the functional equation we only note that $W_s(nz) = W_{-s}(nz)$ and $\eta_s(n) = \eta_{-s}(n)$. Finally, we need to compute the residuum

$$\mathrm{Res}_{s=1} E(z,s) = \frac{\pi}{\zeta(2)} \mathrm{Res}_{s=1} \zeta^*(2s-1) = \frac{3}{\pi}.$$

This completes the proof. □

We return to our study of the space $L^2(\Gamma\backslash\mathbb{H})$. We define the following subspace

$$\mathfrak{E}(\Gamma\backslash\mathbb{H}) = \{E_\infty(z|\psi) \colon \psi \in \mathcal{C}_c^\infty(\mathbb{R}_{>0})\}.$$

This is the space of so called *incomplete Poincaré series.* Recall that $\mathcal{C}_b^\infty(\Gamma\backslash\mathbb{H})$ is the original domain of Δ before taking the self-adjoint extension. We have the inclusion

$$\mathfrak{E}(\Gamma\backslash\mathbb{H}) \subseteq \mathcal{C}_b^\infty(\Gamma\backslash\mathbb{H}) \subseteq L^2(\Gamma\backslash\mathbb{H}).$$

We further denote the orthogonal complement of $\mathfrak{E}(\Gamma\backslash\mathbb{H})$ in $L^2(\Gamma\backslash\mathbb{H})$ by $L_0^2(\Gamma\backslash\mathbb{H})$. More precisely

$$L_0^2(\Gamma\backslash\mathbb{H}) = \{f \in L^2(\Gamma\backslash\mathbb{H}) \colon \langle f, E(\cdot|\psi)\rangle = 0 \text{ for all } \psi \in \mathcal{C}_c^\infty(\mathbb{R}_{>0})\}.$$

We have the decomposition

$$L^2(\Gamma\backslash\mathbb{H}) = L_0^2(\Gamma\backslash\mathbb{H}) \oplus \overline{\mathfrak{E}(\Gamma\backslash\mathbb{H})}.$$

We define the space of *cusp forms* as

$$\mathfrak{C}(\Gamma\backslash\mathbb{H}) = L_0^2(\Gamma\backslash\mathbb{H}) \cap \mathcal{C}_b^\infty(\Gamma\backslash\mathbb{H}).$$

Note that the Laplace-Beltrami operator Δ maps $\mathfrak{E}(\Gamma\backslash\mathbb{H})$ (resp. $\mathfrak{C}(\Gamma\backslash\mathbb{H})$) into itself.

Remark 7.5. At this point we do not know if $\mathfrak{C}(\Gamma\backslash\mathbb{H})$ is non-zero. The next result will show that the constant function is not in this space.

Proposition 7.6. *We have $f \in L_0^2(\Gamma\backslash\mathbb{H})$ if and only it $f_0(y) = 0$ for almost all $y \in \mathbb{R}_{>0}$.*

Proof. We will simply compute the inner product

$$\begin{aligned}\langle f, E_\infty(\cdot|\psi)\rangle &= \int_{\mathcal{F}} f(z) \sum_{\Gamma_\infty\backslash\Gamma} \overline{\psi(\mathrm{Im}(\gamma z))} d\mu(z) \\ &= \int_{\Gamma_\infty\backslash\mathbb{H}} f(z)\overline{\psi(\mathrm{Im}(z))} d\mu(z) \\ &= \int_0^\infty \left(\int_0^1 f(x+iy)dx\right) \overline{\psi(y)} y^{-2} dy \\ &= \int_0^\infty f_0(y)\overline{\psi(y)} y^{-2} dy.\end{aligned}$$

We see that $f_0(y) = 0$ implies that $\langle f, E_\infty(\cdot|\psi)\rangle = 0$ for all ψ. On the other hand, if $f \in L^2_0(\Gamma\backslash\mathbb{H})$, then

$$\int_0^\infty f_0(y)\overline{\psi(y)} y^{-2} dy = 0 \text{ for all } \psi \in \mathcal{C}_c^\infty(\mathbb{R}_{>0}). \tag{69}$$

It is a standard fact that this implies $f_0(y) = 0$ for almost all y. □

Corollary 7.7. *Let $f \in \mathfrak{C}(\Gamma\backslash\mathbb{H})$ be such that $\Delta f = s(1-s)f$. Then we have*

$$f(z) = \sum_{n\neq 0} a_f(n) W_s(nz).$$

To analyze the space $\mathfrak{C}_{\Gamma\backslash\mathbb{H}}$ we will use invariant integral operators

$$[T_k f](z) = \int_{\mathcal{F}} k_\Gamma(z,w) f(w) d\mu(w),$$

for $k(u) \in \mathcal{C}_c^\infty(\mathbb{R}_{\geq 0})$ and

$$k_\gamma(z,w) = \sum_{\gamma\in\overline{\Gamma}} k(z,\gamma w).$$

In contrast to the compact case the automorphic kernel $k_\Gamma(z,w)$ is not bounded on $\mathcal{F}\times\mathcal{F}$. The reason is that points close to the cusp are approximately fixed under the action of Γ_∞. Thus, even if k has very small support the γ-sum in the definition of k_Γ can have arbitrary many non-trivial terms.

Lemma 7.8. *Invariant integral operators T_k map the space $\mathfrak{C}(\Gamma\backslash\mathbb{H})$ into itself.*

Proof. Let $g = T_k f$. We simply compute the constant term

$$\begin{aligned}g_0(\mathrm{Im}(z)) &= \int_0^1 g(z+t)dt = \int_0^1 \int_{\mathbb{H}} k(n(t)z,w) f(w) d\mu(w) dt \\ &= \int_{\mathbb{H}} k(z,w) \int_0^1 f(n(t)w) dt d\mu(w) = \int_{\mathbb{H}} k(z,w) f_p(\mathrm{Im}(w)) d\mu(w) = 0.\end{aligned}$$

□

Definition 7.2 (Principal Part)**.** We define the principal part of k_Γ as

$$H_{\mathfrak{k}}(z,w) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \int_{\mathbb{R}} k(z, n(t)\gamma w) dt.$$

It is easy to see that the principal is a well defined function which is Γ-invariant in the second variable.

Lemma 7.9. *We have*

$$H_k(z,w) \ll_k 1 + \operatorname{Im}(z).$$

Proof. Without loss of generality we can assume that $w \in \mathcal{F}$. We start from the definition

$$H_k(z,w) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \int_{\mathbb{R}} k(z, t + \gamma w) dt.$$

Suppose that the support of $k(u)$ is in $[0, A]$. Then the t-integral is supported on the set of t's with

$$|z - t - \gamma w|^2 \leq 4A \operatorname{Im}(z) \operatorname{Im}(\gamma w).$$

By looking at the imaginary part of $z - t - \gamma w$ we observe that we must have $\operatorname{Im}(z) \asymp \operatorname{Im}(\gamma w)$ in order for this set to be non-empty. If this is the case, then t lies in an interval of length $\asymp \operatorname{Im}(z)$. This leads to the bound

$$H_k(z,w) \ll_k \operatorname{Im}(z) \cdot \sharp\{\gamma \in \Gamma_\infty \backslash \Gamma \colon \operatorname{Im}(\gamma w) \asymp \operatorname{Im}(z)\}.$$

To finish the proof we need to count the possible choices for γ. Since elements in $\Gamma_\infty \backslash \Gamma$ are parametrized by their bottom row, we can instead estimate

$$\sharp\{(c,d) \in \mathbb{Z}^2 \colon (c,d) = 1 \text{ and } \frac{\operatorname{Im}(w)}{|cz+d|^2} \asymp \operatorname{Im}(z)\}.$$

Since $w \in \mathcal{F}$ we must have $|cz + d| \geq 1$. We obtain the following inequalities

$$\begin{aligned} \operatorname{Im}(z) &\ll \operatorname{Im}(w), \\ c &\ll (\operatorname{Im}(w) \operatorname{Im}(z))^{-\frac{1}{2}} \text{ and} \\ |c \operatorname{Re}(z) + d| &\ll \left(\frac{\operatorname{Im}(w)}{\operatorname{Im}(z)} \right)^{\frac{1}{2}}. \end{aligned}$$

If $c = 0$, there is only one choice for d. On the other hand, for $c \neq 0$ we have $\ll (\operatorname{Im}(w) \operatorname{Im}(z))^{\frac{1}{2}}$ choices. Finally, if $c \neq 0$ is fixed, then we have $\ll 1 + \left(\frac{\operatorname{Im}(w)}{\operatorname{Im}(z)}\right)^{-\frac{1}{2}}$ choices for d. Thus we have

$$\begin{aligned} &\sharp\{(c,d) \in \mathbb{Z}^2 \colon (c,d) = 1 \text{ and } \frac{\operatorname{Im}(w)}{|cz+d|^2} \asymp \operatorname{Im}(z)\} \\ &\quad \ll 1 + (\operatorname{Im}(w) \operatorname{Im}(z))^{-\frac{1}{2}} \left(1 + \left(\frac{\operatorname{Im}(w)}{\operatorname{Im}(z)} \right)^{\frac{1}{2}} \right) \ll 1 + \operatorname{Im}(z)^{-1}. \end{aligned}$$

Plugging this into our estimate for $H_k(z,w)$ completes the proof. $\square$

Lemma 7.10. *For $f \in \mathfrak{C}(\Gamma\backslash\mathbb{H})$ we have*

$$\langle H_k(z,\cdot), f\rangle = 0.$$

Proof. We simply compute the inner product using the unfolding trick:

$$\begin{aligned}\langle H_k(z,\cdot), f\rangle &= \int_{\Gamma_\infty\backslash\mathbb{H}}\int_{\mathbb{R}} k(z,n(t)w)dt\overline{f(w)}d\mu(w)\\ &= \int_0^\infty\int_{\mathbb{R}} k(z,t+iv)dt\int_0^1 f(iv+x)dxv^{-2}dv\\ &= \int_0^\infty\int_{\mathbb{R}} k(z,t+iv)dtf_0(v)v^{-2}dv = 0.\end{aligned}$$

$\square$

At this point we define

$$\widehat{k}_\Gamma = k_\Gamma - H_\Gamma.$$

We associate the integral operator $\widehat{T}_k$ with kernel $\widehat{k}_\Gamma$ acting on functions $f\colon \mathcal{F} \to \mathbb{C}$. Recall that $\mathcal{F}$ is considered fixed. The definition and the previous lemma directly imply the following.

Corollary 7.11. *For $f \in \mathfrak{C}(\Gamma\backslash\mathbb{H})$ we have*

$$T_k f = \widehat{T}_k f.$$

We are now ready to establish the following important result.

Proposition 7.12. *The kernel $\widehat{k}_\Gamma$ is bounded on $\mathcal{F}\times\mathcal{F}$.*

Proof. Let $z, w \in \mathcal{F}$. We start by observing that

$$k_\gamma(z,w) = \sum_{\gamma \text{ parabolic}} k(z,\gamma w) + O(1).$$

This is true since for non-parabolic motions the points z and $\gamma.w$ are separated by large distances for almost all γ. Similarly, using Lemma 7.9, we see that

$$H_k(z,w) = \int_{\mathbb{R}} k(z,n(t)w)dt + O(1).$$

We arrive at

$$\widehat{k}_\gamma(z,w) = \underbrace{\sum_{\gamma\in\Gamma_\infty} k(z,\gamma w) - \int_{\mathbb{R}} k(z,n(t)w)dw}_{=J(z,w)} + O(1).$$

We will show that $J(z,w)$ is bounded on $\mathbb{H}\times\mathbb{H}$. To do so we will use the Euler-MacLaurin summation formula

$$\sum_{b\in\mathbb{Z}} F(b) = \int_{\mathbb{R}} F(t)dt + \int_{\mathbb{R}} \psi(t)dF(t),$$

where $\psi(t) = t - [t] - \frac{1}{2}$. This leads to

$$J(z,w) = \sum_{b\in\mathbb{Z}} k(z,w+b) - \int_{\mathbb{R}} k(z,w+t)dt = \int_{\mathbb{R}} \psi(t)dk(z,w+t).$$

We conclude the proof by observing that

$$\int_{\mathbb{R}} \psi(t)dk(z,w+t) \ll \int_0^\infty |k'(u)|du \ll 1.$$

□

The upshot of this proposition is that the operator $\widehat{T}_k$ is Hilbert-Schmidt.

Remark 7.13. So far we have assumed that the kernel $k(u)$ are smooth and compactly supported. However, we will want to work with more general kernels. Thus it is useful to note that our results remain true as long as the kernels satisfy

$$k(u), k'(u) \ll (u+1)^{-2}. \tag{70}$$

This can be seen by a standard approximation argument.

Theorem 7.14. *The Laplace operator Δ has pure point spectrum in $L^2_0(\Gamma\backslash\mathbb{H})$ and the eigenspaces have finite dimension. There is a complete orthonormal system $\{\phi_j\}$ of eigenfunctions such that*

$$f(z) = \sum_j \langle f, \phi_j\rangle \phi_j(z) \text{ for all } f \in L^2_0(\Gamma\backslash\mathbb{H}).$$

If $f \in \mathfrak{C}(\Gamma\backslash\mathbb{H})$, then the series converges absolutely and uniformly on compacta.

Proof. As in the compact case we proof this result by studying the resolvent operator. To do so we fix $a > s \geq 2$ and construct the (invariant integral) operator

$$L = R_s - R_a = (s(1-s) - a(1-a))R_sR_a.$$

The latter identity is known as Hilbert's formula. If $k(u) = G_a(u) - G_s(u)$, then $L = T_k$. Note that, since the singularity is canceled out, the function k satisfies (70). We also recall that $R_s = (\Delta + s(1-s))^{-1}$ has dense range in $L^2(\Gamma\backslash\mathbb{H})$.[24] This implies that also L has dense range. Now the operator $\widehat{L} = \widehat{T}_k$ is bounded on $L^2(\mathcal{F})$. We claim that $\widehat{L}$ has still dense range in $\mathfrak{C}(\Gamma\backslash\mathbb{H})$. To see this we put

$$g = (s(1-s) - a(1-a))^{-1}(\Delta + a(1-a))(\Delta + s(1-s))f$$

[24]This follows from Lemma 4.3, which is easily seen to remain true for non-compact quotients.

We have $Lg = f$. Furthermore, if $f \in \mathfrak{C}(\Gamma\backslash\mathbb{H})$, then $g \in \mathfrak{C}(\Gamma\backslash\mathbb{H})$. Our claim follows because $\widehat{L}g = Lg = f$.

Thus $\widehat{L}$ is a compact (even Hilbert-Schmidt) operator with dense range. By the corresponding spectral theorem we conclude that $\widehat{L}$ has discrete spectrum on $L^2_0(\Gamma\backslash\mathbb{H})$ and each element $f \in L^2_0(\Gamma\backslash\mathbb{H})$ can be expanded in eigenfunctions of $\widehat{L}$. We are done, since when restricted to $\mathfrak{C}(\Gamma\backslash\mathbb{H})$ the operators $\widehat{L}$ and L agree. $\square$

Let $J_0 \subseteq \mathbb{N}$ denote an index set and put $J = J_0 \cup \{0\}$. We index the eigenfunctions (resp. eigenvalues) contributing to the spectrum of $L^2_0(\Gamma\backslash\mathbb{H})$ by $\{\phi_j : j \in J_0\}$ (resp. $(\lambda_j)_{j\in J_0}$). We also put $\lambda_0 = 0$ and $\phi_0(z) = \sqrt{3/\pi}$. Note that ϕ_0 is the L^2-normalized constant function.

Recall that the Mellin transform is given by

$$[\mathfrak{M}\psi](s) = \int_0^\infty \psi(y)y^s\frac{dy}{y}.$$

If ψ is smooth and compactly supported, then $\mathfrak{M}\psi$ is entire. We have the inversion formula

$$\psi(y) = \frac{1}{2\pi i}\int_{(\sigma)}[\mathfrak{M}\psi](s)y^{-s}ds,$$

for a suitable vertical line (σ).

Our remaining task will be to derive a spectral decomposition for $\mathfrak{E}(\Gamma\backslash\mathbb{H})$. We start by taking $E_\infty(\cdot|\psi)$ in this space and computing

$$\begin{aligned}\langle E_\infty(\cdot|\psi), \phi_0\rangle \cdot \phi_0(z) &= \frac{3}{\pi}\int_{\Gamma\backslash\mathbb{H}} E_\infty(z|\psi)d\mu(z)\\ &= \frac{3}{\pi}\int_{\Gamma_\infty\backslash\mathbb{H}}\psi(\mathrm{Im}(z))d\mu(z) \quad = \frac{3}{\pi}\int_0^\infty \psi(y)y^{-2}dy = \frac{3}{\pi}[\mathfrak{M}\psi](-1).\end{aligned}$$

This is already quite interesting.

We continue our study of incomplete Poincaré series by writing

$$\begin{aligned}E_\infty(z|\psi) &= \sum_{\gamma\in\Gamma_\infty\backslash\Gamma}\frac{1}{2\pi i}\int_{(2)}[\mathfrak{M}\psi](-s)\,\mathrm{Im}(\gamma z)^s ds\\ &= \frac{1}{2\pi i}\int_{(2)}[\mathfrak{M}\psi](-s)E(z,s)ds.\end{aligned}$$

Recall that $E(z,s)$ is meromorphic in the half plane $\mathrm{Re}(s) \geq \frac{1}{2}$ with a single simple pole at $s = 1$. we are thus invited to shift the contour and obtain

$$\begin{aligned}E_\infty(z|\psi) &= \frac{3}{\pi}[\mathfrak{M}\psi](-1) + \frac{1}{2\pi i}\int_{(\frac{1}{2})}[\mathfrak{M}\psi](s)E(z,s)ds\\ &= \langle E_\infty(\cdot|\psi), \phi_0\rangle \cdot \phi_0(z) + \frac{1}{2\pi i}\int_{(\frac{1}{2})}[\mathfrak{M}\psi](s)E(z,s)ds.\end{aligned}$$

We want to rewrite the final integral slightly. To do so we recall that the Eisenstein series satisfies the functional equation

$$E(z,1-s) = \underbrace{\frac{\zeta^*(2s)}{\zeta^*(2(1-s))}}_{=\varphi(1-s)} E(z,s) = \varphi(1-s)\,\mathrm{Im}(z)^s + \mathrm{Im}(z)^{1-s} + \dots.$$

We also note that we have

$$\overline{E(z,s)} = E(z,\overline{s}) = E(z,1-s) = \varphi(1-s)E(z,s) \text{ on the line } \mathrm{Re}(s) = \frac{1}{2}.$$

In particular, we must have $\varphi(\frac{1}{2}+it) \in S^1$. This turns out to be an important observation.

Now we compute

$$\begin{aligned}
\langle E_\infty(\cdot|\psi), E(\cdot,s)\rangle &= \int_{\Gamma\backslash\mathbb{H}} E_\infty(\cdot|\psi)\overline{E(z,s)}d\mu(z) \\
&= \int_{\Gamma\backslash\mathbb{H}} E_\infty(\cdot|\psi)E(z,1-s)d\mu(z) \\
&= \int_0^\infty y^{-2}\psi(y)[y^{1-s} + \varphi(1-s)y^s]dy \\
&= [\mathfrak{M}\psi](-s) + \varphi(1-s)[\mathfrak{M}\psi](s-1),
\end{aligned}$$

for $\mathrm{Re}(s) = \frac{1}{2}$. In the last step we have unfolded the integral as in the proof of Proposition 7.6. Multiplying the last expression with $E(z,s)$ and integrating over $\mathrm{Re}(s) = \frac{1}{2}$ yields

$$\begin{aligned}
\int_{(\frac{1}{2})} \langle E_\infty(\cdot|\psi), E(\cdot,s)\rangle \cdot E(z,s)ds &= \int_{(\frac{1}{2})} [\mathfrak{M}\psi](-s)E(z,s)ds \\
&\qquad + \int_{(\frac{1}{2})} \varphi(1-s)[\mathfrak{M}\psi](s-1)E(z,s)ds \\
&= \int_{(\frac{1}{2})} [\mathfrak{M}\psi](-s)E(z,s)ds \\
&\qquad + \int_{(\frac{1}{2})} [\mathfrak{M}\psi](s-1)E(z,1-s)ds \\
&= 2\int_{(\frac{1}{2})} [\mathfrak{M}\psi](-s)E(z,s)ds.
\end{aligned}$$

Putting everything together gives us the decomposition

$$E_\infty(z|\psi) = \langle E_\infty(\cdot|\psi), \phi_0\rangle \cdot \phi_0(z) + \frac{1}{4\pi}\int_{\mathbb{R}} \langle E_\infty(\cdot|\psi), E(\cdot, \frac{1}{2}+it)\rangle \cdot E(z, \frac{1}{2}+it)dt \quad (71)$$

This already looks like a spectral decomposition. We only need to understand the nature of the t-integral.

We define

$$\mathfrak{R}(\Gamma\backslash\mathbb{H}) = \mathbb{C}\phi_0.$$

This space is orthogonal to $L_0^2(\Gamma\backslash\mathbb{H})$ and $\{\phi_j\}_{j\in J}$ is a complete orthonormal system for

$$L^2_{\text{disc}}(\Gamma\backslash\mathbb{H}) = \mathfrak{R}(\Gamma\backslash\mathbb{H}) \oplus L_0^2(\Gamma\backslash\mathbb{H})$$

consisting of eigenfunctions of Δ. It turns out that the spectrum of Δ on the complement of $L^2_{\text{disc}}(\Gamma\backslash\mathbb{H})$ is absolutely continuous. To describe it we introduce the so called Eisenstein transform

$$[E_\infty f](z) = \frac{1}{4\pi}\int_0^\infty f(t)E(z,\frac{1}{2}+it)dt,$$

for $f \in \mathcal{C}_c^\infty(\mathbb{R}_{>0})$.

Lemma 7.15. *For $f \in \mathcal{C}_c^\infty(\mathbb{R}_{>0})$ we have $E_\infty f \in L^2(\Gamma\backslash\mathbb{H})$*

Proof. Using the Fourier expansion of $E(\cdot, \frac{1}{2}+it)$ one can show that

$$E(x+iy,\frac{1}{2}+it) = y^{\frac{1}{2}+it} + \varphi(\frac{1}{2}+it)y^{\frac{1}{2}-t} + O(e^{-2\pi y})$$

for $x+iy \in \mathcal{F}$. To see this one estimates the contribution of the non-constant Fourier modes trivially using the exponential decay of the K-Bessel function. The trick is to use partial integration to see that

$$\int_0^\infty f(t)y^{\frac{1}{2}+it}dt \ll_f y^{\frac{1}{2}}\log(y)^{-1}.$$

With this at hand we estimate

$$\|E_\infty f\|^2 \ll_f \int_{\frac{\sqrt{3}}{2}}^\infty y^{-1}\log(y)^{-2}dy \ll_f 1.$$

This gives the desired result. $\square$

Now we equip $\mathcal{C}_c^\infty(\mathbb{R}_{>0})$ with the inner product

$$\langle f,g\rangle = \frac{1}{2\pi}\int_0^\infty f(y)\overline{g(y)}dy,$$

so that $\mathcal{C}_c^\infty(\mathbb{R}_{>0}) \subseteq L^2(\mathbb{R}_{>0})$. Our goal is to show that E_∞ extends to an isometry from $L^2(\mathbb{R}_{>0})$ into $L^2(\Gamma\backslash\mathbb{H})$. This is achieved in Proposition 7.17 below. We need some preparations.

Definition 7.3 (Truncated Eisenstein series)**.** For a large parameter $Y > 1$ we define

$$E^Y(z,s) = \begin{cases} E(z,s) - \operatorname{Im}(z)^s - \varphi(s)\operatorname{Im}(z)^{1-s} & \text{if } \operatorname{Im}(z) > Y, \\ E(z,s) & \text{if } \operatorname{Im}(z) \leq Y, \end{cases}$$

for $z \in \mathcal{F}$. We extend $E^Y(z,s)$ to a Γ-invariant function in the obvious way.

An easy computation shows that $E^Y(\cdot, s) \in L^2(\Gamma\backslash\mathbb{H})$ as long as we are staying away from the poles.

Lemma 7.16 (Maaß-Selberg Relations)**.** *For $s_1, s_2 \in \mathbb{C}$ and $Y > 1$ we have*

$$\begin{aligned}&\langle E^Y(\cdot, s_1), E^Y(\cdot, s_2)\rangle \\ &\quad = \frac{Y^{s_1+\overline{s_2}-1}}{s_1+\overline{s_2}-1} + \varphi(s_1)\frac{Y^{\overline{s_2}-s_1}}{\overline{s_2}-s_1} + \overline{\varphi(s_2)}\frac{Y^{s_1-\overline{s_2}}}{s_1-\overline{s_2}} + \varphi(s_1)\overline{\varphi(s_2)}\frac{Y^{1-s_1-\overline{s_2}}}{1-s_1-\overline{s_2}}.\end{aligned}$$

Proof. We first compute that inner product for $\mathrm{Re}(s_1) > \mathrm{Re}(s_2) > 1$ The general statement follows by analytic continuation.

First, a simple computation shows that

$$\langle E^Y(\cdot, s_1), E^Y(\cdot, s_2)\rangle = \langle E^Y(\cdot, s_1), E(\cdot, s_2)\rangle.$$

Indeed, we have

$$\begin{aligned}\langle f, E(\cdot, s_2) - E^Y(\cdot, s_2)\rangle &= \int_0^1 \int_Y^\infty f(x+iy)\cdot \overline{y^{s_2} + \varphi(s_2)y^{1-s_2}}y^{-2}dydx \\ &= \int_Y^\infty f_0(y)\cdot y^{-2}dy.\end{aligned}$$

However, if $f = E^Y(\cdot, s_2)$, then we observe that $f_0|_{(Y,\infty)} = 0$ and our claim is verified.

Now we observe that $E^Y(\cdot, s_1) \in L^2(\Gamma\backslash\mathbb{H})$. We can thus write it as

$$E^Y(\cdot, s_1) = \phi + E(\cdot|\psi)$$

for $\phi \in L^2_0(\Gamma\backslash\mathbb{H})$ and $E(\cdot|\psi) \in \overline{\mathfrak{E}(\Gamma\backslash\mathbb{H})}$. Our standard unfolding trick (using that $\mathrm{Re}(s_2) > 1$) shows that $\langle \phi, E(\cdot, s_2)\rangle = 0$. We conclude that.

$$\langle E^Y(\cdot, s_1), E(\cdot, s_2)\rangle = \langle E(\cdot|\psi), E(\cdot, s_2)\rangle.$$

We can now determine ψ explicitly. Indeed, we only have to choose ψ so that its constant term matches the one of $E^Y(\cdot, s_1)$. One finds that

$$\psi(y) = \begin{cases} y^{s_1} & \text{if } y < Y \\ -\varphi(s_1)y^{1-s_1} & \text{if } y \geq Y. \end{cases}$$

With this at hand we unfold again to find

$$\begin{aligned}\langle E^Y(\cdot, s_1), E(\cdot, s_2)\rangle &= \int_0^\infty y^{-2}\psi(y)\overline{[y^{s_2} + \varphi(s_2)y^{1-s_2}]}dy \\ &= \int_0^Y y^{\overline{s_2}-s_1-2}dy + \overline{\varphi(s_2)}\int_0^Y y^{s_1-\overline{s_2}-1}dy \\ &\quad - \varphi(s_1)\int_Y^\infty y^{\overline{s_2}-s_1-1}dy - \varphi(s_1)\overline{\varphi(s_2)}\int_Y^\infty y^{-s_1-\overline{s_2}}dy.\end{aligned}$$

Evaluating the y-integrals gives the desired result. $\square$

We are now ready to establish the main technical result.

Proposition 7.17. *For $f, g \in \mathcal{C}_c^\infty(\mathbb{R}_{>0})$ we have*

$$\langle E_\infty f, E_\infty g\rangle = \langle f, g\rangle.$$

Proof. We consider the truncated Eisenstein transform

$$E_\infty^Y f = \frac{1}{4\pi}\int_0^\infty f(r)E^Y(\cdot, \frac{1}{2}+ir)dr.$$

We compute

$$\begin{aligned}\|(E_\infty - E_\infty Y)f\|^2 &= \int_Y^\infty \left|\int_0^\infty [y^{\frac{1}{2}+ir} + \varphi(\frac{1}{2}+ir)y^{\frac{1}{2}-ir}]f(r)dr\right|^2 y^{-2}dy\\ &\ll_f \int_Y^\infty y^{-1}\log(y){-2}dy \ll \log(Y)^{-1}.\end{aligned}$$

Here we have used partial integration in the r-integral to obtain the extra factor of $\log(y)^{-1}$. Using Cauchy-Schwarz we conclude that

$$\langle E_\infty f, E_\infty g\rangle = \langle E_\infty^Y f, E_\infty^Y g\rangle + O_{f,g}(\log(Y)^{\frac{1}{2}}).$$

Since $E_Y(\cdot, \frac{1}{2}+ir)$ is square integrable we can use Fubini to obtain

$$\langle E_\infty^Y f, E_\infty^Y g\rangle = \frac{1}{(4\pi)^2}\int_0^\infty\int_0^\infty f(r')\overline{g(r)}\langle E^Y(\cdot,\frac{1}{2}+ir'), E^Y(\cdot,\frac{1}{2}+ir)\rangle drdr'.$$

Here we insert the Maaß-Selberg relations to obtain

$$\begin{aligned}\langle E_\infty^Y f, E_\infty^Y g\rangle = \frac{1}{(4\pi)^2}\int_0^\infty\int_0^\infty f(r')\overline{g(r)}\Big[&\varphi(\frac{1}{2}+ir')\frac{Y^{-i(r'+r)}}{-i(r'+r)} + \overline{\varphi(\frac{1}{2}+ir)}\frac{Y^{i(r'+r)}}{i(r'+r)}\\ &\varphi(\frac{1}{2}+ir')\overline{\varphi(\frac{1}{2}+ir)}\frac{Y^{i(r-r')}}{i(r-r')} + \frac{Y^{i(r'-r)}}{i(r'-r)}\Big]drdr'.\end{aligned}$$

The terms with $r+r'$ in the denominator have no pole in the support of $f(r')\overline{g(r)}$ and one can estimate their contribution using the by now familiar integration by parts argument. Further we note that

$$\varphi(\frac{1}{2}+ir')\overline{\varphi(\frac{1}{2}+ir)} - 1$$

has a zero at $r = r'$, which can be used to kill the pole coming from $(r-r')^{-1}$. This allows us to write the integral as

$$\begin{aligned}\langle E_\infty^Y f, E_\infty^Y g\rangle &= \frac{1}{(4\pi)^2}\int_0^\infty\int_0^\infty f(r')\overline{g(r)}\left(\frac{Y^{i(r'-r)}}{i(r'-r)} + \frac{Y^{i(r-r')}}{i(r-r')}\right)drdr' + O_{f,g}(\log(Y)^{-1})\\ &= \frac{1}{8\pi^2}\int_0^\infty\int_0^\infty f(r')\overline{g(r)}\frac{\sin((r'-r)\log(Y))}{r'-r}drdr' + O_{f,g}(\log(Y)^{-1}).\end{aligned}$$

Now, as $Y \to \infty$, the function $\frac{\sin(u \log(Y))}{u}$ approximates the distribution $\pi \cdot \delta_0$. This can be made precise using standard Fourier analysis and we omit the argument. This shows that

$$\langle E_\infty^Y f, E_\infty^Y g\rangle = \frac{1}{8\pi^2}\int_0^\infty f(r')\overline{g(r')}dr + o_{f,g}(1). \tag{72}$$

Taking the limit $Y \to \infty$ completes the proof. □

As indicated above this allows us to extend E_∞ to an isometry from $L^2(\mathbb{R}_{>0})$ to $L^2(\Gamma\backslash\mathbb{H})$. We denote the image by $L_E(\Gamma\backslash\mathbb{H})$.[25] A key observation is that

$$\Delta E_\infty f = E_\infty M f$$

where M is the multiplication operator

$$[Mf](r) = (r^2 + \frac{1}{4})f(r).$$

The spectrum of the multiplication operator is easily determined and we obtain the the spectrum of Δ in $L^2_E(\Gamma\backslash\mathbb{H})$ is $[\frac{1}{4}, \infty)$ and is absolutely continuous. It is also immediate that $\mathfrak{R}(\Gamma\backslash\mathbb{H})$ is orthogonal to $L^2_E(\Gamma\backslash\mathbb{H})$. In view of (71) we obtain the decomposition

$$\overline{\mathfrak{E}(\Gamma\backslash\mathbb{H})} = \mathfrak{R}(\Gamma\backslash\mathbb{H}) \oplus L^2_E(\Gamma\backslash\mathbb{H}).$$

In total we have decomposed

$$L^2(\Gamma\backslash\mathbb{H}) = L^2_{\text{disc}}(\Gamma\backslash\mathbb{H}) \oplus L^2_E(\Gamma\backslash\mathbb{H})$$

into Δ-invariant spaces and we have understood the spectrum of Δ on each piece. As a result we obtain the following theorem.

Theorem 7.18 (Spectral decomposition for $\mathrm{SL}_2(\mathbb{Z})$)**.** *For $f \in L^2(\Gamma\backslash\mathbb{H})$ we have*

$$f(z) = \sum_{j\in J}\langle f, \phi_j\rangle \cdot \phi_j(z) + \frac{1}{4}\int_{-\infty}^\infty \langle f, E(\cdot, \frac{1}{2} + ir)\rangle \cdot E(z, \frac{1}{2} + it)dt.$$

In general this expansion converges in norm. However, if $f \in \mathcal{C}_b^\infty(\Gamma\backslash\mathbb{H})$, then it converges absolutely and uniformly on compacta.

This can be used to obtain the following spectral expansion of an automorphic kernel.

Theorem 7.19. *Let $k(u)$ be a function such that its Selberg-Harish–Chandra transform h is sufficiently regular. Then we have*

$$k_\Gamma(z, w) = \sum_{j\in J} h(t_j)\phi_j(z)\overline{\phi_j(w)} + \frac{1}{4\pi}\int_{\mathbb{R}} h(t)E(z, \frac{1}{2} + it)\overline{E(w, \frac{1}{2} + it)}dt. \tag{73}$$

This expansion converges absolutely and uniformly on compacta.

We conclude this section by proving the following theorem.

[25]This is not standard notation.

Theorem 7.20. *Suppose λ is an eigenvalue of Δ on the space $L_0^2(\Gamma\backslash\mathbb{H})$, then we have*
$$\lambda \geq \frac{3}{2}\pi^2.$$
In particular, the spectrum of Δ acting on $L^2(\Gamma\backslash\mathbb{H})$ is $\{0\} \cup [\frac{1}{4}, \infty)$.

Proof. Let $\phi \in L_0^2(\Gamma\backslash\mathbb{H})$ be an L^2-normalized eigenfunction of Δ with eigenvalue λ. We start by recalling that
$$\lambda = \langle \Delta\phi, \phi\rangle = \int_{\mathcal{F}} |\operatorname{Im}(z)\nabla\phi(z)|^2 d\mu(z).$$
Now we observe that the union of two fundamental domains $\mathcal{F} \cup S.\mathcal{F}$ contains the strip
$$\{z \in \mathbb{H}\colon \operatorname{Im}(z) \geq \frac{\sqrt{3}}{2} \text{ and } |\operatorname{Re}(z)| \leq \frac{1}{2}\}.$$
We can this estimate
$$2\lambda \geq \int_{\frac{\sqrt{3}}{2}}^{\infty}\int_{-\frac{1}{2}}^{\frac{1}{2}} |\nabla\phi(x+iy)|^2 dxdy. \tag{74}$$
We now write the Fourier expansion of ϕ as
$$\phi(z) = \sum_{n\neq 0} a_n(y)e(nx).$$
The gradient is then given by
$$\nabla\phi(z) = \left(2\pi i \sum_{n\neq 0} n\cdot a_n(y)e(nx), \sum_{n\neq 0} a_n'(y)e(nx)\right).$$
Thus we can estimate the norm of this from below by
$$|\nabla\phi(z)|^2 \geq 4\pi^2 \left|\sum_{n\neq 0} n\cdot a_n(y)e(nx)\right|^2 = 4\pi^2 \sum_{n,m\neq 0} nm\cdot a_n(y)\overline{a_m(y)}e((n-m)x).$$
Inserting this in (74) and executing the x-integral yields
$$\begin{aligned}
2\lambda &\geq 4\pi^2 \int_{\frac{\sqrt{3}}{2}}^{\infty} \sum_{n\neq 0} |n\cdot a_n(y)|^2 dy \\
&\geq 3\pi^2 \int_{\frac{\sqrt{3}}{2}}^{\infty} \sum_{n\neq 0} |a_n(y)|^2 y^{-2} dy \\
&=\geq 3\pi^2 \int_{\frac{\sqrt{3}}{2}}^{\infty}\int_{-\frac{1}{2}}^{\frac{1}{2}} |\phi(x+iy)|^2 y^{-2} dxdy \\
&\geq 3\pi^2 \int_{\mathcal{F}} |\phi(z)|^2 d\mu(z) \geq 3\pi^2.
\end{aligned}$$

This completes the proof. □

Note that this argument is very wasteful. One can numerically compute that

$$\lambda_1 \approx 91,14\ldots.$$

8. Selberg's trace formula for the modular curve

We stick to the important arithmetic case $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. We would like to follow the computation of the trace of an (invariant) integral operator T_k as in Section 5. However, doing so quickly leads to convergence issues. To circumvent these we introduce the truncated fundamental domain

$$\mathcal{F}(Y) = \{x+iy \in \mathbb{H} \colon |x| \leq \frac{1}{2},\ y \leq Y \text{ and } |x+iy| \geq 1\}.$$

We also write

$$\mathcal{P}(Y) = \{x+iy \in \mathbb{H} \colon |x| \leq \frac{1}{2} \text{ and } y > Y\},$$

so that $\mathcal{F}(Y) = \mathcal{F} \setminus \mathcal{P}(Y)$.

The truncated trace is defined as

$$\mathrm{Tr}^Y T_k = \int_{\mathcal{F}(Y)} k_\Gamma(z,z)d\mu(z).$$

Note that if T_k would be trace class, then this would approach the usual trace as $Y \to \infty$. However, in our current set-up T_k fails to be trace class and we will only be able to establish an asymptotic formula as $Y \to \infty$.

Throughout this section we will make the following additional assumptions on h and k:

- $0 \leq k(u) \ll (u+1)^{-s}$ for $s > 1$,
- $0 \leq h(t) \ll (|t|+1)^{-4}$ and
- $0 \leq g(x) \ll e^{-|x|/2}$.

This will greatly simplify certain parts of the analysis.

8.1. **The spectral trace.** We start by computing the spectral trace. More precisely, we spectrally expand k_Γ using Theorem 7.19. Note that for $z \in \mathcal{F}(Y)$ we have $E(z,s) = E^Y(z,s)$. Thus, we can write

$$\mathrm{Tr}^Y T_k = \sum_{j\in J} h(t_j) \int_{\mathcal{F}(Y)} |\phi_j(z)|^2 d\mu(z) + \frac{1}{4\pi}\int_{\mathbb{R}} h(t) \int_{\mathcal{F}(Y)} |E^Y(z, \frac{1}{2}+it)|^2 d\mu(z)dt.$$

For $j \in J$ we have

$$\int_{\mathcal{F}(Y)} |\phi_j(z)|^2 d\mu(z) = 1 - \int_{\mathcal{P}(Y)} |\phi_j(z)|^2 d\mu(z).$$

For $j = 0$ this simply is

$$\int_{\mathcal{F}(Y)} |\phi_0(z)|^2 d\mu(z) = 1 - \frac{3}{\pi}\mathrm{Vol}(\mathcal{P}(Y)) = 1 + O(Y^{-1}).$$

On the other hand, for $j \in J_0$ (i.e. ϕ_j is a cusp form) we have the following result.

Lemma 8.1. *Let ϕ_j be a cusp form and $Y > 1$ sufficiently large. Then we have*

$$\int_{\mathcal{P}(Y)} |\phi_j(z)|^2 d\mu(z) \ll_\epsilon |s_j|^{\frac{4}{3}+\epsilon} \cdot Y^{-1}.$$

Proof. We will show that $\phi_j(z) \ll |s_j|^{2/3+\epsilon}$ for $y \geq 1$. The desired result follows after integrating this bound from Y to ∞ (against $y^{-2}dy$).

The desired estimate is derived from the Fourier expansion of ϕ_j given in Proposition 7.1. The first step will be to prove suitable (L^2)-estimates for the Fourier coefficients. To deduce these we use Parseval's identity to obtain

$$\sum_{n \neq 0} |a_{\phi_j}(n) W_{s_j}(iny)|^2 = \int_0^1 |\phi_j(x+iy)|^2 dx.$$

Integrating this over the interval (X, ∞) yields

$$\sum_{n \neq 0} |a_{\phi_j}(n)|^2 \int_X^\infty |W_{s_j}(iny)|^2 y^{-2} dy = \int_0^1 \int_X^\infty |\phi_j(x+iy)|^2 \frac{dxdy}{y^2}.$$

Here we think of $X > 0$ being small, so that we can estimate the right hand side by the number of translates of $\mathcal{F}$ that has a non-zero intersection with the strip $\{x + iy \colon x \in [-1/2, 1/2],\, y \geq X\}$. An easy count (see the proof of Lemma 7.9 for similar estimates) shows that

$$\int_0^1 \int_X^\infty |\phi_j(x+iy)|^2 \frac{dxdy}{y^2} \ll 1 + X^{-1}.$$

Now, from the bound

$$\int_{|s|/2}^\infty K_{s-1/2}(y)^2 y^{-1} dy \gg |s|^{-1} e^{-\pi|s|}$$

we deduce that

$$\int_X^\infty W_s(iny)^2 y^{-2} dy \gg |n| \cdot |s|^{-1} e^{-\pi|s|}$$

as long as $4\pi|n|X \leq |s|$. After choosing $4\pi N X = |s_j|$ we can estimate

$$\sum_{0 \neq |n| \leq N} |n| \cdot |a_{\phi_j}(n)|^2 \ll |s_j| \cdot e^{\pi|s_j|} \sum_{n \neq 0} |a_{\phi_j}(n)|^2 \int_X^\infty |W_{s_j}(iny)|^2 y^{-2} dy \tag{75}$$

Using the above bounds to majorize the right hand side we get the important estimate

$$\sum_{0 \neq |n| \leq N} |n| \cdot |a_{\phi_j}(n)|^2 \ll (N + |s_j|) e^{\pi|s_j|}. \tag{76}$$

A trivial consequence is

$$a_{\phi_j}(n) \ll e^{\pi|s_j|/2} \cdot \begin{cases} \left(\frac{|s_j|}{n}\right)^{\frac{1}{2}} & \text{if } n \le |s_j|, \\ 1 & \text{if } n \ge |s_j|. \end{cases}$$

At this point we record the important uniform estimate

$$\frac{\sqrt{|x|}}{|\Gamma(s_j)|} K_{s_j-\frac{1}{2}}(2\pi|x|) \ll \begin{cases} |s_j|^{\frac{1}{6}} & \text{if } |x| \le |s_j|, \\ e^{-\pi|x|} & \text{if } |x| > |s_j|. \end{cases} \tag{77}$$

We are now able to estimate ϕ_j for large $y > 1$ using its Fourier expansion. To do so we recall the definition of W_s and Stirling's formula. We then use the estimates above and Cauchy-Schwarz to bound[26]

$$\begin{aligned} \phi_j(x+iy) \ll & \left(\sum_{0 \neq |n| \le |s_j|/y} |n| \cdot |a_{\phi_j}(n)|^2 e^{-\pi|s_j|} \right)^{\frac{1}{2}} \left(\sum_{0 \neq |n| \le |s_j|/y} |n|^{-1} \cdot |W_{s_j}(ny)|^2 e^{\pi|s_j|} \right)^{\frac{1}{2}} \\ & + |s_j|^{\frac{1}{2}} \sum_{|n| > |s_j|/y} e^{-\pi|n|y} \\ \ll_\epsilon & \ |s_j|^{2/3+\epsilon} + |s_j|^{\frac{1}{2}} y^{-1}. \end{aligned}$$

This gives the desired result on the size of ϕ_j. □

Now recall that by our assumption on h we have

$$\sum_{j \in J} h(t_j)|t_j|^{\frac{4}{3}+\epsilon} \ll 1.$$

We conclude that

$$\sum_{j \in J} h(t_j) \int_{\mathcal{F}(Y)} |\phi_j(z)|^2 d\mu(z) = \sum_{j \in J} h(t_j) + O(Y^{-1}).$$

For the Eisenstein series we use the Maaß-Selberg relations to compute

$$\begin{aligned} \int_{\mathcal{F}} |E^Y(z, \sigma+it)|^2 d\mu(z) = & \frac{1}{2\sigma-1} \left(Y^{2\sigma-1} - \varphi(\sigma+it)\varphi(\sigma-it) Y^{1-2\sigma} \right) \\ & + \frac{1}{2it} \left(\varphi(\sigma-it) Y^{2it} - \varphi(\sigma+it) Y^{-2it} \right). \end{aligned}$$

In order to take the limit $\sigma \to \frac{1}{2}$ we record the approximation

$$Y^{\pm(2\sigma-1)} = 1 \pm (2\sigma-1)\log(Y) + \dots .$$

Furthermore one obtains

$$\varphi(\sigma+it) = \varphi(\sigma+it) + (\sigma - \frac{1}{2})\varphi'(\sigma+it) + \dots$$

[26] We are very wasteful here, but the result will be sufficient for our purposes.

and using the functional equation also

$$\varphi(\sigma+it)\varphi(\sigma-it) = 1 + (2\sigma-1)\frac{\varphi'(\sigma+it)}{\varphi(\sigma+it)} + \cdots .$$

This gives

$$\int_{\mathcal{F}} |E^Y(z,\sigma+it)|^2 d\mu(z) = \frac{1}{2it}\left(\varphi(\sigma-it)Y^{2it} - \varphi(\sigma+it)Y^{-2it}\right) + 2\log(Y) - \frac{\varphi'(\frac{1}{2}+it)}{\varphi(\frac{1}{2}+it)}.$$

Taking the integral over $\mathbb{R}$ against the test function h yields

$$\frac{1}{4\pi}\int_{-\infty}^{\infty} h(t)\int_{\mathcal{F}} |E^Y(z,\sigma+it)|^2 d\mu(z)dt = g(0)\log(Y) - \int_{-\infty}^{\infty} h(t)\frac{\varphi'(\sigma+it)}{\varphi(\sigma+it)}dt + \frac{1}{8\pi i}\int_{-\infty}^{\infty} h(t)\left(\varphi(\sigma-it)Y^{2it} - \varphi(\sigma+it)Y^{-2it}\right)dt.$$

The remaining integral can be computed by rewriting it as

$$\frac{1}{8\pi i}\int_{-\infty}^{\infty} h(t)\left(\varphi(\sigma-it)Y^{2it} - \varphi(\sigma+it)Y^{-2it}\right)dt = \frac{1}{4\pi i}\int_{-\infty}^{\infty} t^{-1}h(t)\left(\varphi(\frac{1}{2}-it)Y^{2it} - \varphi(\frac{1}{2})\right)dt.$$

We move the line of integration to $\mathrm{Im}(\cdot) = \delta$ and obtain

$$\frac{1}{8\pi i}\int_{-\infty}^{\infty} h(t)\left(\varphi(\sigma-it)Y^{2it} - \varphi(\sigma+it)Y^{-2it}\right)dt = -\frac{\varphi(\frac{1}{2})}{4\pi i}\int_{\mathrm{Im}(t)=\delta} t^{-1}h(t)dt + O(Y^{-2\delta}).$$

The t-integral can be evaluated by moving the contour to $\mathrm{Im}(t) = -\delta$ picking up the pole at 0. Using symmetry of h we arrive at

$$\frac{1}{8\pi i}\int_{-\infty}^{\infty} h(t)\left(\varphi(\sigma-it)Y^{2it} - \varphi(\sigma+it)Y^{-2it}\right)dt = \frac{1}{4}\varphi(\frac{1}{2})h(0). \tag{78}$$

We still need to control the integral of the truncated Eisenstein series over $\mathcal{P}(Y)$. This is again done using the Fourier expansion. We prove the following estimate.

Lemma 8.2. *Under our current assumptions we have*

$$\frac{1}{4\pi}\int_{\mathbb{R}} h(t)\int_{\mathcal{P}(Y)} |E^Y(z,\frac{1}{2}+it)|^2 d\mu(z) \ll Y^{-2}.$$

Proof. This is also proved using the Fourier expansion. It is important to recall that on $\mathcal{P}(Y)$ the constant term of E^Y vanishes. Thus, we have

$$E^Y(z,s) = \frac{1}{\zeta^*(2s)}\sum_{m\neq 0} |m|^{-\frac{1}{2}}\eta_s(m)W_s(mz) \text{ for } z \in \mathcal{P}(Y).$$

We have the divisor bound $\eta_s(m) \ll |m|^{\mathrm{Re}(s)+\epsilon}$ and we recall (77). Note that the Γ-factor normalizing the Bessel function in (77) comes from the completed zeta function $\zeta^*(2s)$. In particular, we estimate

$$\begin{aligned}\int_{\mathcal{P}(Y)} |E^Y(z, \frac{1}{2}+it)|^2 d\mu(z) &= \frac{1}{|\zeta^*(1+it)|^2}\int_Y^\infty \sum_{m\neq 0} |m|^{-1}|\eta_s(m)W_s(miy)|^2\frac{dy}{y^2} \\ &\ll \frac{1}{|\zeta(1+it)|^2}\int_Y^\infty \left(\sum_{m=1}^{t/y} m^\epsilon t^{\frac{1}{3}} + \sum_{m>t/y} m^\epsilon e^{-2my}\right)\frac{dy}{y^2} \\ &\ll \frac{|t|^{\frac{4}{3}+\epsilon}}{|\zeta(1+i2t)|^2}Y^{-2}.\end{aligned}$$

Using suitable (lower) bounds for $\zeta(1+2it)$ allows us to integrate this against $h(t)$ and complete the proof. □

Putting everything together yields the following important proposition.

Proposition 8.3. *Under our working assumptions we have*

$$\mathrm{Tr}^Y T_k = g(0)\log(Y) + \sum_{j\in J} h(t_j) - \frac{1}{4\pi}\int_{-\infty}^\infty h(t)\frac{\varphi'(\frac{1}{2}+it)}{\varphi(\frac{1}{2}+it)}dt + \frac{h(0)}{4}\varphi(\frac{1}{2}) + O_{h,\delta}(Y^{-\delta})$$

for some small $\delta > 0$ and as $Y \to \infty$.

8.2. **Conjugacy classes for $\mathbf{SL}_2(\mathbb{Z})$.** Before we turn towards the geometric side let us compute the conjugacy classes for $\overline{\Gamma}$. First, the two parabolic conjugacy classes are $\{T^{\pm 1}\}$. We deduce that a complete list of parabolic conjugacy classes is given by $\{T^m\}$ where $0 \neq m \in \mathbb{Z}$.

The elliptic classes can be computed as follows. Let $\gamma \in \Gamma$ be elliptic. Because the entries are integers we must have $\mathrm{tr}(\gamma) \in \{0, \pm 1\}$. We know that each elliptic conjugacy class corresponds to precisely one fixed point in $\mathcal{F}$. For $\mathrm{tr}(\gamma) = 0$ this leads to the condition[27]

$$\frac{a}{c} + i\frac{1}{c} \in \mathcal{F}.$$

After recalling the definition of $\mathcal{F}$ we obtain that $|a/c| \leq \frac{1}{2}$ and $\frac{a^2+1}{c^2} \geq 1$. We conclude that $a = 0$ and $c = \pm 1$. Therefore, we have found exactly one elliptic conjugacy class in $\overline{\Gamma}$ with trace 0. It is given by $\{S\}$ where

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

has order two in $\overline{\Gamma}$.

[27]The case $c = 0$ is easily ruled out.

Next we look at the case $\mathrm{tr}(\gamma) = 1$. (Since we are working in $\overline{\Gamma}$ this covers also the case $\mathrm{tr}(\gamma) = -1$.) Here we obtain the condition

$$\frac{2a-1}{2c} + i\frac{\sqrt{3}}{2c} \in \mathcal{F}.$$

A solution is $c = a = 1$ and $d = 0$. The determinant condition then implies $b = -1$. We conclude that

$$\gamma = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \in \Gamma$$

is elliptic and fixes $\frac{1}{2} + i\frac{\sqrt{3}}{2}$. One computes $\gamma^3 = -I_2$, so that γ has order 3.

The most interesting conjugacy classes are the hyperbolic ones. As we will see below these classes carry very special arithmetic information. Indeed, for $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ we can relate them to certain class groups in real quadratic fields. We will sketch this correspondence following [He83, Chapter Eleven, Section 2].

We start by defining a map

$$W\colon \mathrm{GL}_2(\mathbb{C}) \to \mathrm{GL}_3(\mathbb{C}), \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{pmatrix} a^2 & ab & b^2 \\ 2ac & ad+bc & 2bd \\ c^2 & cd & d^2 \end{pmatrix}.$$

The following properties can be verified by direct computation:

(1) The map W is a group homomorphism;
(2) $\det(W(T)) = \det(T)^3$ for $T \in \mathrm{GL}_2(\mathbb{C})$;
(3) For the matrix

$$D = \begin{pmatrix} 0 & 0 & -2 \\ 0 & 1 & 0 \\ -2 & 0 & 0 \end{pmatrix}$$

we have $W(T)^t D W(T) = \det(T)^2 \cdot S$ for $T \in \mathrm{GL}_2(\mathbb{C})$.
(4) If $e_2(0,1,0)^t$, then $W(T)e_2 = e_2$ for $T \in SL_2(\mathbb{R})$ if and only if T is diagonal.
(5) The hyperboloid

$$\{\mathbf{x} \in \mathbb{R}^3 \colon x_2^2 - 4x_1x_3 = 1\} \subseteq \mathbb{R}^3$$

can we realized as the set $W(\mathrm{SL}_2(\mathbb{R})) \cdot e_2$.

A binary quadratic form is a homogeneous polynomial in two variables. We write

$$q\left[\begin{pmatrix} x \\ y \end{pmatrix}\right] = n_1x^2 + n_2xy + n_3y^2.$$

We note that we can identify a (general) binary quadratic form q with the vector $\mathbf{n} = (n_1, n_2, n_3)^t$. However, we have a natural action of $\mathrm{SL}_2(\mathbb{R})$ on these forms. Indeed we set

$$[g.q]\left[\begin{pmatrix} x \\ y \end{pmatrix}\right] = q\left[g\begin{pmatrix} x \\ y \end{pmatrix}\right].$$

Lemma 8.4. *Let q be a binary quadratic form associated to $\mathbf{n} \in \mathbb{R}^3$ and let $g \in \mathrm{SL}_2(\mathbb{R})$ be a matrix. Further, denote the vector associated to $q.q$ by $\mathbf{m} \in \mathbb{R}^3$. Then we have*

$$\mathbf{m} = W(g^t)\mathbf{n}.$$

Proof. This is a direct computation. □

The discriminant of a quadratic form is defined by

$$\mathrm{disc}(q) = n_2^2 - 4n_1n_3.$$

We will now turn our attention to integral quadratic forms. These are those with integral coefficients (i.e. their associated vector $\mathbf{n}$ is in $\mathbb{Z}^3$). For $d \in \mathbb{N}$ with $d \equiv 0, 1 \bmod 4$ and non-square we define

$$\mathcal{C}(d) = \{\mathbf{n} \in \mathbb{Z}^3 \colon n_2^2 - 4n_1n_3 = d\}.$$

Further, put

$$\mathcal{C} = \bigcup_d \mathcal{C}(d),$$

where the union is taken over all d satisfying the conditions above. Furthermore we call $\mathbf{n} \in \mathbb{Z}^3$ as well as the associated quadratic form primitive if $(n_1, n_2, n_3) = 1$.

The group $W(\Gamma)$ induces an obvious equivalence relation on $\mathcal{C}(d)$, which respects primitive forms. Let $h(d)$ denote the class number of primitive quadratic corms. It is easy to see that the number of all equivalence classes in $\mathcal{C}(d)$ is given by

$$\sum_{\substack{g^2 \mid d, \\ g>0}} h(d/g^2).$$

Lemma 8.5. *Let $T = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R})$ be hyperbolic. Then*

$$\{\mathbf{x} \in \mathbb{R}^3 \colon W(T)\mathbf{x} = \mathbf{x}\} = \mathbb{R} \cdot \begin{pmatrix} b \\ d-a \\ -c \end{pmatrix}.$$

Proof. The inclusion of the right hand side in the left hand side can be directly verified. Next we note that since T is hyperbolic we can write $T = g \cdot a(\lambda) \cdot g$ with $\lambda > 1$. In particular, the vector $W(T)$ has eigenvalues λ^2, 1 and λ^{-2}. We conclude that the space on the left hand side is one dimensional. This completes the proof. □

Lemma 8.6. *Let $\mathbf{n} \in \mathcal{C}(d)$ be primitive and set $\epsilon_d = \frac{\alpha_d + \beta_d\sqrt{d}}{2}$ where α_d, β_d is the minimal solution to Pell's equation $x^2 - dy^2 = 4$. We define*

$$P_{\mathbf{n}} = \begin{pmatrix} \frac{\alpha_d - \beta_d n_2}{2} & \beta n_1 \\ -\beta n_3 & \frac{\alpha_d + \beta_d n_2}{2} \end{pmatrix} \in \mathrm{SL}_2(\mathbb{R}).$$

Then

$$\Gamma_{\mathbf{n}} = \{T \in \mathrm{PSL}_2(\mathbb{Z})\colon W(T)\mathbf{n} = \mathbf{n}\} = \langle P_{\mathbf{n}} \rangle$$

and $P_{\mathbf{n}}$ *is primitive hyperbolic.*

Proof. It is a classical result from number theory that the matrix $P_{\mathbf{n}}$ generates the stabilizer of $\mathbf{n}$. To see that $P_{\mathbf{n}}$ is primitive and hyperbolic we observe that $\Gamma_{\mathbf{n}}$ is a discrete subgroup of a group isomorphic to $\mathbb{R}^{\times}$. In particular, $\Gamma_{\mathbf{n}}$ is cyclic and its generator must be primitive by default. □

Given $T \in \mathrm{PSL}_2(\mathbb{Z})$ primitive hyperbolic we define

$$\mathbf{n}(T) = \frac{\mathrm{sgn}(a+d)}{(b, d-a, c)} \begin{pmatrix} b \\ d-a \\ -c \end{pmatrix}.$$

Lemma 8.7. *For* $\gamma_1, \gamma_2 \in \mathrm{PSL}_2(\mathbb{Z})$ *primitive hyperbolic we have*

(1) $\mathbf{n}(\gamma_1) \in \mathcal{C}$ *primitive;*
(2) $\overline{\Gamma}_{\mathbf{n}}(\gamma_1) = \langle \gamma_1 \rangle$*;*
(3) $\mathbf{n}(g\gamma_1 g^{-1}) = W(g)\mathbf{n}(\gamma)$ *for* $g \in \mathrm{PSL}_2(\mathbb{Z})$*;*
(4) $\mathbf{n}(\gamma_1^{-1}) = -\mathbf{n}(\gamma_1)$*;*
(5) $\mathbf{n}(\gamma_1) = \mathbf{n}(\gamma_2)$ *if and only if* $\gamma_1 = \gamma_2$*; and*
(6) $\mathbf{n}(P_{\mathbf{n}}) = \mathbf{n}$ *for primitive* $\mathbf{n} \in \mathcal{C}$.

Proof. These properties are all very straight forward to verify. □

Corollary 8.8. *There is a one to one correspondence between primitive hyperbolic conjugacy classes in* $\mathrm{PSL}_2(\mathbb{Z})$ *and inequivalent primitive elements in* $\mathcal{N}$.

Proof. By the lemma above the map $\{\gamma\} \mapsto W(\Gamma)\mathbf{n}(\gamma)$ is well defined and gives the desired correspondence. □

Suppose γ is a primitive hyperbolic matrix with $l = d(\gamma) > 0$. Then we have

$$\alpha_d = \mathrm{tr}(\gamma) = e^{l/2} + e^{-l/2},$$

where $\mathbf{n}(\gamma) \in \mathcal{C}(d)$. Solving this equation leads us to

$$e^l = \epsilon_d^2.$$

Together with the correspondence above we deduce that

$$Z_\Gamma(s) = \prod_{k=0}^{\infty} \prod_{d} [1 - \epsilon_d^{-2(s+k)}]^{h(d)}.$$

Similarly we have

$$\frac{Z'_\Gamma}{Z_\Gamma}(s) = 2\sum_{k=1}^{\infty} \sum_{d} \frac{h(d)\log(\epsilon_d)}{\epsilon_d^{2ls} - \epsilon_d^{2l(s-1)}}.$$

We conclude that for the lattice $\mathrm{SL}_2(\mathbb{Z})$ the Selberg Zeta Function is very arithmetic in nature. These arithmetic features can be used to establish strong forms of the prime geodesic theorem using tools from analytic number theory. We refer to [SY] for more information.

8.3. **Computing the orbital integrals.** After this interlude we return to the study of orbital integrals. The main task is to handle the parabolic conjugacy classes. Thus we can assume that $\gamma = T^k$ for $k \in \mathbb{N}$. We need to compute

$$I_Y(\gamma) = \int_{\mathcal{D}_\gamma(Y)} k(\gamma.z, z) d\mu(z),$$

where

$$\mathcal{D}_\gamma(Y) = \bigcup_{T \in \overline{\Gamma}_\gamma \backslash \overline{\Gamma}} T.\mathcal{F}(Y).$$

Since $\gamma = T^k$, we can understand the set $\mathcal{D}_\gamma(Y)$ as follows. We first define the part of the upper half plane with cuspidal zones removed

$$\mathbb{H}(Y) = \Gamma.\mathcal{F}(Y).$$

This is left invariant by $\Gamma_\gamma = \Gamma_\infty$. We thus obtain

$$I_Y(\gamma) = \int_{\overline{\Gamma}_\infty \backslash \mathbb{H}(Y)} k(z + k, z) d\mu(z).$$

We can find a set of representatives for the quotient $\overline{\Gamma}_\infty \backslash \mathbb{H}(Y)$ that contains $\{x + iy \colon 0 < x \leq 1 \text{ and } Y^{-1} < y \leq Y\}$ and is contained in $\{x + iy \colon 0 < x \leq 1 \text{ and } y \leq Y\}$. We thus have

$$\int_0^1 \int_{Y^{-1}}^Y k(z + k, z) d\mu(z) \leq I_Y(\gamma) \leq \int_0^1 \int_0^Y k(z + k, z) d\mu(z).$$

We can write the upper bound as

$$\begin{aligned}
\int_0^1 \int_0^Y k(z + k, z) d\mu(z) &= \int_0^Y k\left(\left(\frac{k}{2y}\right)^2\right) y^{-2} dy \\
&= |k|^{-1} \int_{(k/2Y)^2}^\infty k(u) u^{-\frac{1}{2}} du.
\end{aligned}$$

This will be summed over $k \in \mathbb{Z} \setminus \{0\}$ and we obtain

$$\begin{aligned}
\sum_{0 \neq k \in \mathbb{Z}} I_Y(\{T^k\}) &\leq 2 \sum_{k \in \mathbb{N}} \frac{1}{k} \int_{(k/2Y)^2}^\infty k(u) u^{-\frac{1}{2}} du \\
&= 2 \int_{(1/2Y)^2}^\infty k(u) u^{-\frac{1}{2}} \left(\sum_{1 \leq l < 2Y\sqrt{u}} l^{-1} \right) du \\
&= 2 \int_{(1/2Y)^2}^\infty k(u) u^{-\frac{1}{2}} \left(\log(2Y\sqrt{u}) + \gamma + O(u^{-\frac{1}{2}} Y^{-1}) \right) du.
\end{aligned}$$

The integral over the error is easily bound by $\ll Y^{-1}\log(Y)$. We conclude that

$$\sum_{0\neq k\in\mathbb{Z}} I_Y(\{T^k\}) \leq 2\int_0^\infty k(u)u^{-\frac{1}{2}}\left(\log(2Y\sqrt{u})+\gamma\right)du + O(Y^{-1}\log(Y)).$$

The lower bound can be written as

$$I_Y(\{T^k\}) \geq \int_0^1\int_0^Y k(z+k,z)d\mu(z) - \int_0^1\int_0^{Y^{-1}} k(z+k,z)d\mu(z).$$

Proceeding as above allows us to obtain a lower bound with the same asymptotic behavior. (Here we can treat the short integral $\int_0^{Y^{-1}}$) essentially trivially.) We arrive at the following preliminary result.

Lemma 8.9. *We have*

$$\sum_{k\in\mathbb{Z}} I_Y(\{T^k\}) = 2\int_0^\infty k(u)u^{-\frac{1}{2}}\left(\log(2Y\sqrt{u})+\gamma\right)du + O(Y^{-1}\log(Y)).$$

The integrals can be further computed and one obtains

Lemma 8.10. *We have*

$$\sum_{k\in\mathbb{Z}} I_Y(\{T^k\}) = g(0)\log(Y)$$
$$- g(0)\log(2) + \frac{1}{4}h(0) - \frac{1}{2\pi}\int_{\mathbb{R}} h(t)\psi(1+it)dt + O(Y^{-1}\log(Y)).$$

Proof. We only have to compute the u-integral. We first note that

$$\int_0^\infty k(u)u^{-\frac{1}{2}}du = q(0) = \frac{1}{2}g(0).$$

This immediately gives

$$2\int_0^\infty k(u)u^{-\frac{1}{2}}\left(\log(2Y\sqrt{u})+\gamma\right)du = g(0)\log(Y)+g(0)\log(2)+g(0)\gamma+\int_0^\infty k(u)u^{-\frac{1}{2}}\log(u)du.$$

We compute the remaining integral as follows. By inserting (35) and exchanging the integrals we obtain

$$\begin{aligned}\int_0^\infty k(u)u^{-\frac{1}{2}}\log(u)du &= -\frac{1}{\pi}\int_0^\infty\int_0^v \frac{\log(u)}{\sqrt{u(v-u)}}dudq(v)\\ &= -\frac{1}{\pi}\int_0^\infty\int_0^1 \frac{\log(uv)}{\sqrt{u(1-u)}}dudq(v)\\ &= \frac{1}{\pi}q(0)\int_0^1\frac{\log(u)}{\sqrt{u(1-u)}}du - \frac{1}{\pi}\int_0^1\frac{1}{\sqrt{u(1-u)}}du\int_0^\infty\log(v)dq(v)\\ &= -2q(0)\log(2) - \int_0^\infty \log(v)dq(v).\end{aligned}$$

After a change of variables we obtain

$$2\int_0^\infty k(u)u^{-\frac{1}{2}}\left(\log(2Y\sqrt{u})+\gamma\right)du = g(0)\log(Y)+g(0)\gamma-\int_0^\infty \log(\sinh(r/2))dg(r).$$

In order to replace g by h we recall

$$g'(r) = -\frac{1}{2\pi i}\int_{\mathrm{Im}(\cdot)=\epsilon} e^{irt}h(t)tdt$$

and

$$\int_0^\infty \log(\sinh(r/2))de^{-\nu r} = \gamma+\log(2)-\frac{1}{2\nu}+\psi(1+\nu).$$

The latter is a well known Laplace transform. Combining these two facts yields

$$\int_0^\infty \log(\sinh(r/2))dg(r) = (\gamma+\log(2))g(0)-\frac{1}{4}h(0)+\frac{1}{2\pi}\int_{\mathbb{R}} h(t)\psi(1+it)dt.$$

Note that the term containing $h(0)$ arises through the same trick that allowed us to arrive at (78). This completes the proof. ☐

8.4. **The final trace formula.** We are mow ready to collect all the pieces together and prove a version of Selberg's trace formula for $\mathrm{SL}_2(\mathbb{Z})$.

Theorem 8.11 (Selberg's trace formula for $\mathrm{SL}_2(\mathbb{Z})$)**.** *Suppose h satisfies the regularity properties from Remark 2.24. We make the following additional assumptions:*

- $0 \leq k(u) \ll (u+1)^{-s}$ *for* $s>1$,
- $0 \leq h(t) \ll (|t|+1)^{-4}$ *and*
- $0 \leq g(x) \ll e^{-|x|/2}$.

Then we have

$$\begin{aligned}\sum_{j\in J} h(t_j) &- \frac{1}{4\pi}\int_{\mathbb{R}} h(t)\frac{\varphi'}{\varphi}(\frac{1}{2}+it)dt \\ &= \frac{1}{12}\int_{\mathbb{R}} h(t)t\tanh(\pi t)dt - \frac{1}{2\pi}\int_{\mathbb{R}} h(t)\psi(1+it)dt \\ &+ \frac{h(0)}{4}(1-\varphi(\frac{1}{2})) - g(0)\log(2) \\ &+ \sum_d h(d)\log(\epsilon_d)\sum_{k=1}^\infty \frac{g(2k\log(\epsilon_d))}{\sinh(k\log(\epsilon_d))} \\ &+ \frac{1}{2}\int_0^\infty \frac{g(r)\cosh(r/2)}{\cosh(r)-\cos(\pi)}dr + \frac{1}{3}\sum_{k=1,2}\int_0^\infty \frac{g(r)\cosh(r/2)}{\cosh(r)-\cos(\frac{2\pi k}{3})}dr.\end{aligned}$$

Proof. By combining Proposition 8.3 and Proposition 8.10 we obtain

$$\begin{aligned}
&\mathrm{Tr}^Y T_k - \sum_{k\in\mathbb{Z}} I_Y(\{T^k\}) \\
&\quad = \sum_{j\in J} h(t_j) - \frac{1}{4\pi}\int_{-\infty}^{\infty} h(t)\frac{\varphi'(\frac{1}{2}+it)}{\varphi(\frac{1}{2}+it)}dt + \frac{h(0)}{4}\varphi(\frac{1}{2}) + g(0)\log(2) - \frac{1}{4}h(0) \\
&\qquad\qquad + \frac{1}{2\pi}\int_{\mathbb{R}} h(t)\psi(1+it)dt + O_{h,\delta}(Y^{-\delta}).
\end{aligned}$$

On the other hand we have

$$\mathrm{Tr}^Y T_k - \sum_{k\in\mathbb{Z}} I_Y(\{T^k\}) = \sum_{\substack{\{\gamma\} \\ \text{not parabolic}}} I_Y(\{\gamma\}) \leq \sum_{\substack{\{\gamma\} \\ \text{not parabolic}}} I(\{\gamma\}). \tag{79}$$

The last inequality holds by non-negativity of k. Recall that we have computed all non-parabolic orbital integrals in Lemma 5.4, 5.6 and 5.5. These computations were independent of the compactness of $\Gamma\backslash\mathbb{H}$.

Note that, due to the sufficient regularity of h, the right hand side of (79) converges absolutely. Since we further have $I_Y(\{\gamma\}) \to I(\{\gamma\})$ individually we can take the limit $Y \to \infty$. The formula as stated is then a direct consequence of our computations. □

As in the compact case one can relax the conditions on h by a suitable approximation argument. We will not pursue this here, because for all our applications the formula above is sufficient.

9. Application II

We now turn towards the standard applications. As in the co-compact case we will discuss the Weyl law, the Selberg zeta function and the Prime geodesic theorem. Even though some of the results remain valid more generally we will work exclusively with $\Gamma = \mathrm{SL}_2(\mathbb{Z})$.

9.1. **The Weyl law for** $\mathrm{SL}_2(\mathbb{Z})$**.**

Recall that the Weyl law is an asymptotic formula for the number of Δ-eigenvalues in a growing interval. It will be convenient to consider the following equivalent count

$$N_\Gamma(T) = \sharp\{j \in J \colon |t_j| \leq T\}.$$

Since the quotient $\Gamma\backslash\mathbb{H}$ is non-compact we will encounter the following artifact of the continuous spectrum

$$M_\Gamma(T) = \frac{1}{4\pi}\int_{-T}^{T}\frac{\varphi'}{\varphi}(\frac{1}{2}+it)dt.$$

Using the Bessel inequality (for a suitable kernel) one can show that

$$N_\Gamma(T), M_\Gamma(T) \ll T^2.$$

In order to upgrade these upper bounds to an asymptotic formula we consider the test pair

$$h(t) = e^{-\delta t^2} \text{ and } g(x) = \frac{1}{\sqrt{4\pi\delta}} e^{-\frac{x^2}{4\delta}},$$

for small $\delta > 0$. In the proof of Proposition 6.1 we have already seen that

$$\frac{1}{12}\int_{-\infty}^{\infty} e^{-\delta t^2} t \tanh(\pi t) dt = \frac{1}{12\delta} + O(1).$$

Furthermore, the hyperbolic and the elliptic contribution are uniformly bounded. The remaining contribution is

$$\begin{aligned}\frac{h(0)}{4}(1-\varphi(\frac{1}{2})) - g(0)\log(2) - \frac{1}{2\pi}\int_{\mathbb{R}} h(t)\psi(1+it)dt \\ = -\frac{\log(2)}{2\sqrt{\pi\delta}} - \frac{1}{2\pi}\int_{\mathbb{R}} e^{-\delta t^2}\psi(1+it)dt + O(1).\end{aligned}$$

To compute the integral we use that

$$\psi(1+it) + \psi(1-it) = \log(1+t^2) + O((1+t^2)^{-1}) = 2\log(t) + O((1+t^2)^{-1}).$$

This allows us to compute

$$\begin{aligned}\frac{1}{2\pi}\int_{\mathbb{R}} e^{-\delta t^2}\psi(1+it)dt &= \frac{1}{\pi}\int_0^{\infty} e^{-\delta t^2}\log(t)dt + O(1) \\ &= \frac{1}{4\pi\sqrt{\delta}}\int_0^{\infty} e^{-r}\log(r/\delta)r^{-\frac{1}{2}}dr + O(1) \\ &= \frac{1}{4\pi\sqrt{\delta}}(\Gamma'(\frac{1}{2}) - \Gamma(\frac{1}{2})\log(\delta)) + O(1) \\ &= \frac{1}{4\sqrt{\pi\delta}}(-\gamma - \log(4\delta)) + O(1).\end{aligned}$$

We conclude that

$$\frac{h(0)}{4}(1-\varphi(\frac{1}{2})) - g(0)\log(2) - \frac{1}{2\pi}\int_{\mathbb{R}} h(t)\psi(1+it)dt = \frac{\gamma}{4\sqrt{\pi\delta}} + \frac{\log(\delta)}{4\sqrt{\pi\delta}} + O(1).$$

We arrive at the following result.

Proposition 9.1. *As $\delta \to 0$ we have*

$$\sum_{j\in J} e^{-\delta t_j^2} - \frac{1}{4\pi}\int_{\mathbb{R}} e^{-\delta t^2}\frac{\varphi'}{\varphi}(\frac{1}{2}+it)dt = \frac{1}{12\delta} + \frac{\log(\delta)}{4\sqrt{\pi\delta}} + \frac{\gamma}{4\sqrt{\pi\delta}} + O(1).$$

Using a Tauberian theorem we obtain the following Weyl law:

Corollary 9.2. *We have*

$$N_\Gamma(T) - M_\Gamma(T) = \frac{T^2}{12}(1+o(1)).$$

Remark 9.3. With more effort one can show that

$$N_\Gamma(T) - M_\Gamma(T) = \frac{T^2}{12} - \frac{1}{\pi}T\log(T) + c_\Gamma T + O(T\log(T)^{-1}),$$

where c_Γ is a suitable constant. This is the correct generalization of Theorem 6.28 to the non-compact situation. A similar formula with slightly different constants is true for general co-finite groups Γ.

Note that, because we pick up N_Γ and M_Γ simultaneously this does not give us not a lot of information on N_Γ. However, in our case (i.e. $\Gamma = \mathrm{SL}_2(\mathbb{Z})$) we can recall that

$$\varphi(s) = \sqrt{\pi}\frac{\Gamma(s-\frac{1}{2})\zeta(2s-1)}{\Gamma(s)\zeta(2s)}.$$

This allows us to obtain useful information on $M_\Gamma(T)$. After expressing $\varphi(s)$ as a Hadamard canonical product, taking the logarithmic derivative and performing a contour shift in the definition of $M_\Gamma(T)$ we observe that

$$-M_\Gamma(T) = \sharp\{\mathrm{Im}(s_j) \leq T\colon s_j \text{ pole of } \varphi\} + O(T).$$

Since the poles of $\varphi(s)$ are essentially the zeros of $\zeta(2s)$, whose number in boxes is sufficiently well understood, we obtain

$$-M_\Gamma(T) \ll T\log(T).$$

This allows us to conclude that $\sharp J = \infty$. More precisely we have obtained the following.

Theorem 9.4 (Selberg 1954)**.** *We have*

$$N_\Gamma(T) = \frac{T^2}{12}(1+o(1)).$$

A similar result can be obtained more generally for congruence subgroups. However for arbitrary co-finite Γ the quantity $-M_\Gamma(T)$ can not be controlled this way. There are indeed many interesting questions connected to the existence of (many) cusp forms for general lattices. See [PS].

9.2. **The Selberg zeta function.** As in the compact case we can also choose the test function

$$h(t) = \frac{1}{t^2+\alpha^2} - \frac{1}{t^2+\beta^2} \text{ for } \alpha = s - \frac{1}{2} \text{ and } \beta = z - \frac{1}{2}.$$

If we assume $1 < s < z$ this function (and its transforms) satisfies the assumptions for our trace formula. We obtain the following formula.

Theorem 9.5 (Resolvent trace formula for $\mathrm{SL}_2(\mathbb{Z})$)**.** *For $1 < s < z$ we have*

$$\sum_{j\in J}\left[\frac{1}{(s-\frac{1}{2})^2+t_j^2}-\frac{1}{(z-\frac{1}{2})^2+t_j^2}\right]-\frac{1}{4\pi}\int_{\mathbb{R}}\left[\frac{1}{(s-\frac{1}{2})^2+t^2}-\frac{1}{(z-\frac{1}{2})^2+t^2}\right]\cdot\frac{\varphi'}{\varphi}(\frac{1}{2}+it)dt$$
$$=\frac{1}{2s-1}\frac{Z'_\Gamma}{Z_\Gamma}(s)-\frac{1}{2sz-1}\frac{Z'_\Gamma}{Z_\Gamma}(z)$$
$$+\psi(z)-\psi(s)+\left[\frac{1}{(2s-1)^2}-\frac{1}{(2z-1)^2}\right](1-\varphi(\frac{1}{2}))$$
$$+\frac{1}{2s-1}R_2(s)-\frac{1}{2z-1}R_2(z)+\frac{1}{2s-1}R_3(s)-\frac{1}{2z-1}R_3(z)$$
$$-\frac{1}{2s-1}(\psi(s+\frac{1}{2})+\log(2))+\frac{1}{2z-1}(\psi((z+\frac{1}{2}))+\log(2)),$$

where R_m is defined in (64)*.*

Proof. This is derived as Theorem 6.21. The additional terms arising from the non-compactness of $\Gamma\backslash\mathbb{H}$ are easily computed. Note that

$$\chi(\mathrm{SL}_2(\mathbb{Z}))=\frac{1}{6}+\frac{1}{3}+\frac{1}{2}=1.$$

□

This formula directly leads to a meromorphic continuation of $\frac{Z'_\Gamma}{Z_\Gamma}(s)$ to the complex plane. One checks that all poles are simple and have integral residues. This allows us to conclude the following.

Theorem 9.6. *The Selberg zeta function $Z_\Gamma(s)$ has a meromorphic continuation to $s\in\mathbb{C}$. Furthermore, the identity given in Theorem 9.5 holds for all $s,z\in\mathbb{C}$ and we have a functional equation of the form*

$$Z_\Gamma(s)=\Psi(s)Z_\Gamma(1-s),$$

where $\Psi(s)$ is a certain meromorphic function of order two.

Proof. After having seen the analytic continuation of $\frac{Z'_\Gamma}{Z_\Gamma}(s)$ we find that $Z_\Gamma(s)=Z_\Gamma(z)F(s)$ for

$$F(s)=\exp\left(\int_z^s\frac{Z'_\Gamma}{Z_\Gamma}(u)du\right).$$

That the formula in Theorem 9.5 remains valid for all $s,z\in\mathbb{C}$ follows by analytic continuation.

Finally, the functional equation is obtained by using $z=1-s$ in Theorem 9.5.

□

We end by listing some further properties of $Z_\Gamma(s)$ for $\Gamma=\mathrm{SL}_2(\mathbb{Z})$ that can be obtained from the resolvent trace formula:

- In the half plane $\mathrm{Re}(s) > \frac{1}{2}$ the function $Z_\Gamma(s)$ is holomorphic and has a zero at $s_0 = 1$;
- At $s = \frac{1}{2}$ there is a simple pole.
- There are so called topological zeros at $s = 0, -1, -2, \ldots$ and topological poles at $s = -\frac{1}{2}, -3/2, \ldots$. These arise form the terms involving the digamma function ψ.
- The remaining zeros are spectral and occur at so called resonances. These include zeros at s_j for $j \in J$.

9.3. **The prime geodesic theorem for the modular curve.** Recall that the prime geodesic theorem is about an asymptotic formula for the counting function

$$\pi_\Gamma(x) = \sharp\{l \in \mathcal{L}_{\Gamma\backslash\mathbb{H}} \colon |l| \leq x\}.$$

In this section we will prove the following theorem.

Theorem 9.7 (Prime geodesic theorem for $\mathrm{SL}_2(\mathbb{Z})$)**.** *For $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ we have*

$$\pi_\Gamma(x) = li(e^x) + O(e^{\frac{3}{4}x}).$$

This is a direct analogue of Theorem 6.26 for the con-compact quotient $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. Note that here we do not encounter secondary terms, because we have excluded eigenvalues of Δ in the interval $(0, \frac{1}{4})$.

We will give a proof of Theorem 9.7 using the trace formula. To do so we study the sum

$$H(T) = \sum_{\substack{\{\gamma_0\} \\ \text{prim. hyp.}}} l(\gamma_0) \sum_{1 \leq k \leq T/l(\gamma_0)} \frac{\cosh(kl(\gamma_0)/2)}{\sinh(kl(\gamma_0)/2)}.$$

The asymptotic behavior of the sum is as follows.

Proposition 9.8. *We have*

$$H(T) = e^T + O(e^{\frac{3T}{4}}).$$

Taking this result for granted allows us to deduce Theorem 9.7. Indeed, since

$$\frac{\cosh(x)}{\sinh(x)} = 1 + O(e^{-2x})$$

we can easily write

$$H(T) = \sum_{\substack{\{\gamma_0\} \\ \text{prim. hyp.} \\ |l(\gamma_0)| \leq T}} l(\gamma_0) + O(e^{\frac{T}{2}}).$$

Passing from the remaining sum to the counting function π_Γ is an exercise using partial integration.

Proof or Proposition 9.8. We define the function

$$g_T(x) = 2\cosh(x/2)\mathbb{1}_{[-T,T]}(x).$$

The hyperbolic contribution of Selberg's trace formula corresponding to g_T would precisely give $H(T)$. Unfortunately, the discontinuity of g_T forbids a direct use of it in the trace formula. To fix this we choose a (fixed) bump function $\varphi \in \mathcal{C}_c^\infty(\mathbb{R})$ satisfying

- φ is non-negative and even;
- φ is supported in the interval $[-1, 1]$;
- φ is L^1-normalized (i.e. $\int_{\mathbb{R}} \varphi(x)dx = 1$).

For a small parameter $\epsilon > 0$ we consider the re-scaled version

$$\varphi_\epsilon(x) = \frac{1}{\epsilon}\varphi(x/\epsilon).$$

Finally set

$$g_{T,\epsilon}(x) = [g_T * \varphi_\epsilon](x) = 2\int_{\mathbb{R}} \cosh((x-y)/2)\mathbb{1}_{[-T,T]}(x-y)\varphi_\epsilon(y)dy.$$

Note that $g_{T,\epsilon} \in \mathcal{C}_c^\infty(\mathbb{R})$. We define

$$H_\epsilon(T) = \frac{1}{2}\sum_{\substack{\{\gamma_0\}\\ \text{prim. hyp.}}} l(\gamma_0)\sum_{k=1}^{\infty}\frac{g_{T,\epsilon}(kl(\gamma_0))}{\sinh(kl(\gamma_0)/2)}.$$

Note that we have $H_\epsilon(T) \to H(T)$ as $\epsilon \to 0$. More important is the observation that

$$H_\epsilon(T-\epsilon) \leq H(T) \leq H_\epsilon(T+\epsilon). \tag{80}$$

We have

$$h_{T,\epsilon}(t) = \widehat{g}_{T,\epsilon}(t) = \widehat{g}_T(t)\cdot\widehat{\varphi}_\epsilon(t).$$

Before applying the trace formula we gather some estimates. First, we can compute

$$\widehat{g}_T(t) = \frac{2}{\frac{1}{2}+it}\sinh(T(\frac{1}{2}+it)) + \frac{2}{\frac{1}{2}-it}\sinh(T(\frac{1}{2}-it)).$$

For $t \in \mathbb{R}$ we obtain the useful estimate

$$\widehat{g}_T(t) \ll (1+|t|)^{-1}e^{T/2}.$$

On the other hand we have

$$\widehat{g}_T(\pm i/2) = e^T + 2T + e^{-T} = e^T + O(T).$$

We also compute that

$$\widehat{\varphi}_\epsilon(t) = \widehat{\varphi}(\epsilon t) \ll (1+|t\epsilon|)^{-2}$$

and

$$\widehat{\varphi}_\epsilon(i/2) = \widehat{\varphi}(i\epsilon/2) = 1 + O(\epsilon).$$

Combining this we obtain

$$h_{T,\epsilon}(i/2) = e^T + O(\epsilon e^T)$$

and

$$h_{T,\epsilon}(t) \ll e^{T/2}(1+|t|)^{-1}(1+|\epsilon t|)^{-2} \text{ for } t \in \mathbb{R}.$$

We now begin our analysis of the terms in the trace formula Theorem 8.11. The discrete contribution is bounded by

$$\sum_{j \in J} h_{T,\epsilon}(t_j) = e^T + O\left(\epsilon e^T + e^{T/2} \sum_{j \in J_0} (1+|t_j|)^{-1}(1+|\epsilon t_j|)^{-2}\right).$$

After splitting the remaining j-sum into the pieces $\{t_j \leq \frac{1}{\epsilon}\}$ and $\{t_j > \frac{1}{\epsilon}\}$ one uses Weyl's law to estimate

$$\sum_{j \in J} h_{T,\epsilon}(t_j) = e^T + O\left(\epsilon e^T + \frac{1}{\epsilon} e^{\frac{T}{2}}\right).$$

Using standard bounds for $\zeta(s)$ on the line $\mathrm{Re}(\cdot) = 1$ one can easily show that the contribution from

$$-\frac{1}{4\pi} \int_{\mathbb{R}} \frac{\varphi'}{\varphi}(\frac{1}{2} + it) h_{T,\epsilon}(t) dt$$

is absorbed in the error term. The remaining geometric terms can also be estimated directly using our bounds. All together the trace formula produces the asymptotic

$$H_\epsilon(T) = e^T + O(\epsilon e^T + \frac{1}{\epsilon} e^{\frac{T}{2}}).$$

Choosing $\epsilon = e^{-T/4}$ and using (80) completes the proof. $\square$

Using our description of the primitive hyperbolic conjugacy classes we can write

$$\pi_\Gamma(\log(x^2)) = \sum_{\substack{\{\gamma_0\} \\ \text{prim. hyp.} \\ |l(\gamma_0)| \leq \log(x^2)}} 1 = \sum_{\substack{d \equiv 0,1 \bmod 4, \\ \text{non-square} \\ \epsilon_d \leq x}} h(d).$$

We put

$$\mathcal{D}(x) = \{d \in \mathbb{N} \colon \epsilon_d \leq x,\ d \equiv 0, 1 \bmod 4, \text{ non-square}\}.$$

Then it was shown by Sarnak in [Sa, Proposition 4.1] that

$$\sharp \mathcal{D}(x) = \frac{35}{16} x + O_\epsilon(x^{\frac{2}{3}+\epsilon}).$$

We can thus write our prime geodesic theorem as

$$\frac{1}{\sharp \mathcal{D}(x)} \sum_{d \leq \mathcal{D}(x)} h(d) = \frac{16}{35} \cdot \frac{\mathrm{li}(x^2)}{x} + O_\epsilon(x^{\frac{2}{3}+\epsilon}).$$

It is a key feature of our summation, that we order by the size of the so called regulator or $\mathbb{Q}(\sqrt{d})$. That the ordering matters can be seen by the following asymptotic formula, which is due to Siegel:

$$\sum_{\substack{1\leq d\leq x\\ d\equiv 0,1 \bmod 4\\ \text{non-suqare}}} h(d)\log(\epsilon_d) = \frac{\pi^2}{18\zeta(3)}x^{\frac{3}{2}} + O(x\log(x)).$$

Remark 9.9. Note that for $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ the error term in the prime geodesic theorem can be improved. Such improvements usually rely on deep results from analytic number theory. We refer to the paper [SY] for example. The current (unconditional) record[28] is

$$\sum_{\substack{\{\gamma_0\}\\ \text{prim. hyp.}\\ |l(\gamma_0)|\leq T}} l(\gamma_0) = e^T + O_\epsilon(e^{(\frac{2}{3}+\epsilon)T})$$

due to I. Kaneko in 2024.

References

[Bo] D. Borthwick, *Spectral theory of infinite-area hyperbolic surfaces*, Progress in Mathematics, 256. Birkhäuser Boston, Inc., Boston, MA, 2007.

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

[He76] D. A. Hejhal, *The Selberg trace formula for* $\mathrm{PSL}(2,R)$. *Vol. I*, Lecture Notes in Mathematics, Vol. 548. Springer-Verlag, Berlin-New York, 1976.

[He83] D. A. Hejhal, *The Selberg trace formula for* $\mathrm{PSL}(2,\mathbf{R})$. *Vol. 2*, Lecture Notes in Mathematics, 1001. Springer-Verlag, Berlin, 1983.

[Hu] M. N. Huxley, *Integer points, exponential sums and the Riemann zeta function*, Number theory for the millennium, II (Urbana, IL, 2000), 275–290, A K Peters, Natick, MA, 2002.

[Iv] V. Ivrii, *100 years of Weyl's law*, Bull. Math. Sci. 6 (2016), no. 3, 379–452.

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

[IK] H. Iwaniec, E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, 53. American Mathematical Society, Providence, RI, 2004. xii+615 pp.

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

[Mü] W. Müller, *Weyl's law in the theory of automorphic forms*, Groups and analysis, 133–163, London Math. Soc. Lecture Note Ser., 354, Cambridge Univ. Press, Cambridge, 2008.

[vN] J. von Neumann, *Zum Haarschen Maß in topologischen Gruppen*, (German) Compositio Math. 1 (1935), 106–114.

[28]At the time of writing.

[PS] R. S. Phillips, P. Sarnak, *On cusp forms for co-finite subgroups of* $\mathrm{PSL}(2, \mathbf{R})$, Invent. Math. 80 (1985), no. 2, 339–364.

[Sa] P. Sarnak, *Class numbers of indefinite binary quadratic forms*, J. Number Theory 15 (1982), no. 2, 229–247.

[Se] A. Selberg, *Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series*, J. Indian Math. Soc. (N.S.) 20 (1956), 47–87.

[Si] C. L. Siegel,*Topics in complex function theory. Vol. II. Automorphic functions and abelian integrals*, Translated from the German by A. Shenitzer and M. Tretkoff. With a preface by Wilhelm Magnus. Reprint of the 1971 edition. Wiley Classics Library. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1988.

[SY] K. Soundararajan, M. P. Young, *The prime geodesic theorem*, J. Reine Angew. Math. 676 (2013), 105–120.

[Tr] H. Triebel, *Höhere Analysis*, Hochschulbücher für Mathematik, Band 76. VEB Deutscher Verlag der Wissenschaften, Berlin, 1972.

[Ve] A. B. Venkov, *Spectral theory of automorphic functions and its applications*, Translated from the Russian by N. B. Lebedinskaya. Mathematics and its Applications (Soviet Series), 51. Kluwer Academic Publishers Group, Dordrecht, 1990.