Chapter 11
$\mathrm{SL}_2(\mathbb{R})$

The group $\mathrm{SL}_2(\mathbb{R})$ is the simplest case of a so called reductive Lie group. Harmonic analysis on these groups turns out to be more complex then the previous cases of abelian, compact, or nilpotent groups. On the other hand, the applications are more rewarding. For example, via the theory of automorphic forms, in particular the Langlands program, harmonic analysis on reductive groups has become vital for number theory. In this chapter we prove an explicit Plancherel Theorem for functions in the Hecke algebra of the group $G = \mathrm{SL}_2(\mathbb{R})$. We apply the trace formula to a uniform lattice and as an application derive the analytic continuation of the Selberg zeta function.

11.1 The Upper Half Plane

Let $G = \mathrm{SL}_2(\mathbb{R})$ denote the *special linear group of degree 2*, i.e.

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

The locally compact group $\mathrm{SL}_2(\mathbb{R})$ acts on the *upper half plane*

$$\mathbb{H} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$$

by *linear fractionals*, i.e., for $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ and for $z \in \mathbb{H}$ one defines

$$gz = \frac{az + b}{cz + d}.$$

To see that this is well-defined one has to show that $cz + d \neq 0$. If $c = 0$ then $d \neq 0$ and so the claim follows. If $c \neq 0$ then $\mathrm{Im}(cz + d) = c\mathrm{Im}(z) \neq 0$. Next one has to show that gz lies in $\mathbb{H}$ if z does and that $(gh)z = g(hz)$ for $g, h \in \mathrm{SL}_2(\mathbb{R})$. The latter is an easy computation, for the former we will now derive an explicit formula for the imaginary part of gz. Multiplying numerator and denominator by $c\bar{z} + d$ one gets $gz = \frac{ac|z|^2 + bd + 2bc\mathrm{Re}(z) + z}{|cz+d|^2}$, so in particular,

A. Deitmar, S. Echterhoff, *Principles of Harmonic Analysis,* Universitext,

DOI 10.1007/978-3-319-05792-7_11,

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

which is strictly positive if $\mathrm{Im}(z)$ is. Note that the action of the central element $-1 \in \mathrm{SL}_2(\mathbb{R})$ is trivial.

If $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in G$ stabilizes the point $i \in \mathbb{H}$, then $\frac{ai+b}{ci+d} = i$, or $ai + b = -c + di$, which implies $a = d$ and $b = -c$. So the stabilizer of the point $i \in \mathbb{H}$ is the *rotation group*:

$$K = \mathrm{SO}(2) = \left\{ \begin{pmatrix} a & -b \\ b & a \end{pmatrix} : a, b \in \mathbb{R},\ a^2 + b^2 = 1 \right\},$$

which also can be described as the group of all matrices of the form

$$\begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \quad \text{for} \quad t \in \mathbb{R}.$$

The operation of G on $\mathbb{H}$ is transitive, as for $z = x + iy \in \mathbb{H}$ one has

$$z = \begin{pmatrix} \sqrt{y} & \frac{x}{\sqrt{y}} \\ 0 & \frac{1}{\sqrt{y}} \end{pmatrix} i.$$

It follows that via the map $G/K \to \mathbb{H}$, given by $gK \mapsto gi$, the upper half plane $\mathbb{H}$ can be identified with the quotient G/K.

Theorem 11.1.1 (Iwasawa Decomposition). *Let A be the group of all diagonal matrices in G with positive entries. Let N be the group of all matrices of the form $\left(\begin{smallmatrix} 1 & s \\ 0 & 1 \end{smallmatrix}\right)$ for $s \in \mathbb{R}$. Then one has $G = ANK$. More precisely, the map*

$$\psi : A \times N \times K \to G,$$
$$(a, n, k) \mapsto ank$$

is a homeomorphism.

Proof Let $g \in G$, and let $gi = x + yi$. Then, with

$$a = \begin{pmatrix} \sqrt{y} & \\ & 1/\sqrt{y} \end{pmatrix} \quad \text{and} \quad n = \begin{pmatrix} 1 & x/y \\ & 1 \end{pmatrix},$$

one has $gi = ani$ and so $g^{-1}an$ lies in K, which means that there exists $k \in K$ with $g = ank$. Using the explicit formula for gz above in the case $z = i$, one constructs the inverse map to ψ as follows. Let $\phi : G \to A \times N \times K$ be given by $\phi(g) = (\underline{a}(g), \underline{n}(g), \underline{k}(g))$, where

$$\underline{a}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{c^2+d^2}} & \\ & \sqrt{c^2+d^2} \end{pmatrix},$$

$$\underline{n}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & ac+bd \\ & 1 \end{pmatrix},$$

$$\underline{k}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{\sqrt{c^2 + d^2}} \begin{pmatrix} d & -c \\ c & d \end{pmatrix}.$$

A straightforward computation shows that $\phi\psi = \mathrm{Id}$ and $\psi\phi = \mathrm{Id}$. □

For $g \in \mathrm{SL}_2(\mathbb{R})$ we shall use throughout this chapter the notation $\underline{a}(g)$, $\underline{n}(g)$, and $\underline{k}(g)$ as explained above. Moreover, for $x, t, \theta \in \mathbb{R}$, we shall write

$$a_t \stackrel{\text{def}}{=} \begin{pmatrix} e^t & \\ & e^{-t} \end{pmatrix} \in A$$

$$n_x \stackrel{\text{def}}{=} \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} \in N$$

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

A function $f : G \to \mathbb{C}$ is called *smooth* if the map $\mathbb{R}^3 \to \mathbb{C}$ given by

$$(t, x, \theta) \mapsto f(a_t n_x k_\theta)$$

is infinitely differentiable. We denote the space of smooth functions by $C^\infty(G)$. The space of smooth functions of compact support is denoted by $C_c^\infty(G)$.

Theorem 11.1.2 *The group $G = \mathrm{SL}_2(\mathbb{R})$ is unimodular.*

Proof Let $\phi : G \to \mathbb{R}_+^\times$ be a continuous group homomorphism. We show that $\phi \equiv 1$. First note that $\phi(K) = 1$ as K is compact. As ϕ restricted to A is a continuous group homomorphism, there exists $x \in \mathbb{R}$ such that $\phi(a_t) = e^{tx}$ for every $t \in \mathbb{R}$. Let $w = \left(\begin{smallmatrix} & -1 \\ 1 & \end{smallmatrix} \right)$, then $w a_t w^{-1} = a_{-t}$, and therefore $e^{tx} = \phi(a_t) = \phi(w a_t w^{-1}) = e^{-tx}$ for every $t \in \mathbb{R}$, which implies $x = 0$ and so $\phi(A) = 1$. Similarly, $\phi(n_x) = e^{rx}$ for some $r \in \mathbb{R}$. As we have $a_t n_x a_t^{-1} = n_{e^{2t}x}$ it follows $e^{rs} = e^{re^{2t}s}$ for every $t \in \mathbb{R}$, which implies $r = 0$, so $\phi(N) = 1$ and by the Iwasawa decomposition, we conclude $\phi \equiv 1$. □

We write $\underline{t}(g)$ for the unique $t \in \mathbb{R}$ with $\underline{a}(g) = a_t$, i.e., one has $\underline{a}(g) = a_{\underline{t}(g)}$.

Theorem 11.1.3 *For any given Haar measures on three of the four groups G,A,N,K, there is a unique Haar measure on the fourth such that for $f \in L^1(G)$ the decomposition formula*

$$\int_G f(x)\, dx = \int_A \int_N \int_K f(ank)\, dk\, dn\, da$$

holds. For $\phi \in L^1(K)$ and $x \in G$ one has

$$\int_K \phi(k)\, dk = \int_K \phi(\underline{k}(kx))\, e^{2\underline{t}(kx)}\, dk.$$

From now on we normalize Haar measures as follows. On K we normalize the volume to be one. On A we choose the measure $2dt$, where $t = \underline{t}(a)$, and on N we choose $\int_{\mathbb{R}} f(n_s)ds$. The factor 2 is put there to match the usual invariant measure $\frac{dx\,dy}{y^2}$ on the upper half plane.

Proof Let $B = AN$, the subgroup of G consisting of all upper triangular matrices with positive diagonal entries. Then an easy computation shows that $db = dadn$ is a Haar measure on B and that B is not unimodular. Indeed, one has $\Delta_B(a_x n) = e^{-2x}$, which follows from the equation $a_t n_x a_s n_y = a_{t+s} n_{y+e^{-2s}x}$. Let $\underline{b} : G \to B$ be the projection $\underline{b}(g) = \underline{a}(g)\underline{n}(g)$. The map $B \to G/K \cong \mathbb{H}$ mapping b to bK is a B-equivariant homeomorphism. Any G-invariant measure on G/K gives a Haar measure on B and as both these types of measures are unique up to scaling one gets that every B-invariant measure on G/K is already G-invariant. So the formula $\int_G f(x)\,dx = \int_{G/K}\int_K f(xk)\,dk\,dx$ leads to $\int_G f(x)\,dx = \int_B \int_K f(bk)\,dk\,db$. Since $db = da\,dn$, the integral formula follows.

For the second assertion let $\phi \in L^1(K)$. Let $\eta \in L^1(B)$ and set $g(bk) = \eta(b)\phi(k)$. Then g lies in $L^1(G)$. As G is unimodular, for $y \in G$ one has

$$\begin{aligned}
\int_B \int_K \eta(b)\phi(k)\,dk\,db &= \int_G g(y)\,dy = \int_G g(yx)\,dy \\
&= \int_G \eta(\underline{b}(yx))\phi(\underline{k}(yx))\,dy \\
&= \int_B \int_K \eta(\underline{b}(bkx))\phi(\underline{k}(kx))\,dk\,db \\
&= \int_B \int_K \eta(b\underline{b}(kx))\phi(\underline{k}(kx))\,dk\,db \\
&= \int_B \int_K \eta(b)\Delta_B(\underline{b}(kx))^{-1}\phi(\underline{k}(kx))\,dk\,db \\
&= \int_B \int_K \eta(b)e^{2\underline{t}(kx)}\phi(\underline{k}(kx))\,dk\,db,
\end{aligned}$$

where we used the facts that $\underline{k}(bg) = \underline{k}(g)$ for all $b \in B$, $g \in G$ and $\underline{t}(\underline{b}(kx)) = \underline{t}(kx)$ for all $b \in B, k \in K$, and $x \in G$. Varying η, we get the claim of the theorem. □

Hyperbolic Geometry

Let $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ be in $G = \mathrm{SL}_2(\mathbb{R})$. For $z \in \mathbb{H}$ one gets

$$\frac{d}{dz} gz = \frac{d}{dz}\frac{az+b}{cz+d} = \frac{a(cz+d) - c(az+b)}{(cz+d)^2} = \frac{1}{(cz+d)^2}.$$

Since on the other hand, $\mathrm{Im}(gz) = \frac{\mathrm{Im}(z)}{|cz+d|^2}$, we get $\left|\frac{d}{dz}gz\right| = \frac{\mathrm{Im}(gz)}{\mathrm{Im}(z)}$, or

$$\frac{\left|\frac{d}{dz}gz\right|}{\mathrm{Im}(gz)} = \frac{\left|\frac{d}{dz}z\right|}{\mathrm{Im}(z)}.$$

That is to say, the Riemannian metric $\frac{dx^2+dy^2}{y^2}$ is invariant under the group action of G on $\mathbb{H}$. For a continuously differentiable path $p : [0,1] \to \mathbb{H}$ we get the induced *hyperbolic length* defined by

$$L(p) = \int_0^1 \frac{|p'(t)|}{\mathrm{Im}(p(t))}\, dt.$$

Then it follows that $L(p) = L(g \circ p)$ for every $g \in G$, i.e., the length is G-invariant. The *hyperbolic distance* of two points $z, w \in \mathbb{H}$ is defined by

$$\rho(z,w) = \inf_p L(p),$$

where the infimum is extended over all paths p with $p(0) = z$ and $p(1) = w$.

Lemma 11.1.4 *For any two point $z, w \in \mathbb{H}$ there exists $g \in G$ such that $gz = i$ and $gw = yi$ for some $y \geq 1$.*

Proof As we have seen in the beginning of this chapter, the group action of G on $\mathbb{H}$ is transitive, hence there exists $h \in G$ with $hz = i$. We next apply an element $k \in K$ so that $g = kh$ does the job. For this we have to show that for any given $z \in \mathbb{H}$ there exists $k \in K$ such that $kz = yi$ for some $y \geq 1$. The map $\theta \mapsto k_\theta z$ is continuous, for $\theta = 0$ we have $k_\theta z = z$ and for $\theta = \pi/2$ we have $k_\theta = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ so that $k_{\pi/2}z = -1/z$. Hence the real parts of z and $k_{\pi/2}z$ have opposite sign, by the intermediate value theorem there exists $k \in K$ such that $\mathrm{Re}(kz) = 0$. If now $kz = yi$ with $y < 1$, then we replace k with $k_{\pi/2}k$ to finish the proof □

Lemma 11.1.5 *The hyperbolic distance is a metric on $\mathbb{H}$. It is G-invariant, i.e., $\rho(gz, gw) = \rho(z,w)$ holds for all $z, w \in \mathbb{H}$, $g \in G$. For $z, w \in \mathbb{H}$ one has*

$$\rho(z,w) = \log \frac{|z - \overline{w}| + |z - w|}{|z - \overline{w}| - |z - w|},$$

and

$$2\cosh \rho(z,w) = 2 + \frac{|z-w|^2}{\mathrm{Im}(z)\mathrm{Im}(w)}.$$

Proof The G-invariance follows from the invariance of the length. The axioms of a metric are immediate from the definition. For the explicit formulae, we start with the special case $z = i$ and $w = yi$ for $y \geq 1$. For any path p with $p(0) = i$ and $p(1) = yi$ we get

$$L(p) = \int_0^1 \sqrt{\mathrm{Re}(p'(t))^2 + \mathrm{Im}(p'(t))^2}\,\frac{dt}{\mathrm{Im}(p(t))}.$$

This is minimized by the path $p(t) = ity$, since for any path $p = \mathrm{Re}(p) + i\,\mathrm{Im}(p)$ the path $i\,\mathrm{Im}(p)$ will also connect the points i and yi. So one gets $\rho(i, yi) = \log y$, which also equals the right hand side of the first assertion in this case. Next the equivalence of the first and second formula are easy, as is the G-invariance of the right hand side of the second formula, which then concludes the proof. □

11.2 The Hecke Algebra

Let A^+ be the subset of A consisting of all diagonal matrices with entries e^t, e^{-t}, where $t > 0$. Let $\overline{A^+} = A^+ \cup \{1\}$ be its closure in G.

Theorem 11.2.1 (Cartan Decomposition). *The group G can be written as $G = K\overline{A^+}K$, i.e. every $x \in G$ is of the form $x = k_1 a k_2$ with $a \in \overline{A^+}$, $k_1, k_2 \in K$. The element a is uniquely determined by x. If $a \neq 1$, which means that $x \notin K$, then also k_1 and k_2 are uniquely determined up to sign, i.e., if $k_1 a k_2 = k_1' a k_2'$, then either $(k_1, k_2) = (k_1', k_2')$ or $(k_1, k_2) = (-k_1', -k_2')$.*

For $f \in L^1(G)$ we have the integral formula

$$\int_G f(x)\,dx = 2\pi \int_K \int_0^\infty \int_K f(ka_t l)\left(e^{2t} - e^{-2t}\right)\,dk\,dt\,dl.$$

Proof For $x \in G$ the matrix xx^t is symmetric and positive definite. As it has determinant one, it follows from linear algebra, that there exists $k \in K$ and $t \geq 0$, such that kxx^tk^t is the diagonal matrix with entries e^{2t}, e^{-2t}. For two elements $x, x_1 \in G$ the condition $xx^t = x_1x_1^t$ is equivalent to $1 = x^{-1}(x_1x_1^t)(x^t)^{-1} = (x^{-1}x_1)(x^{-1}x_1)^t$. The last is equivalent to $x^{-1}x_1 \in K$. So there is $k' \in K$ with $x = k^{-1}a_tk'$.

This shows existence of the decomposition. For the uniqueness note that e^{2t} is the larger of the two eigenvalues of xx^t and thus determined by x. For the uniqueness of k_1, k_2 assume that $a \in A^+$ and $k_1ak_2 = l_1al_2$ with $k_1, k_2, l_1, l_2 \in K$. Then one has $ak_2l_2^{-1} = k_1^{-1}l_1a$. But the equation $ak = k'a$ with $k, k' \in K$ implies $k = k' = \pm 1$ as we show now. Let $a = a_t, k = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}, k' = \begin{pmatrix} c & -d \\ d & c \end{pmatrix}$. Then

$$\begin{aligned}\begin{pmatrix} e^ta & -e^tb \\ e^{-t}b & e^{-t}a \end{pmatrix} &= \begin{pmatrix} e^t & \\ & e^{-t} \end{pmatrix}\begin{pmatrix} a & -b \\ b & a \end{pmatrix} = ak = k'a \\ &= \begin{pmatrix} c & -d \\ d & c \end{pmatrix}\begin{pmatrix} e^t & \\ & e^{-t} \end{pmatrix} = \begin{pmatrix} e^tc & -e^{-t}d \\ e^td & e^{-t}c \end{pmatrix}\end{aligned}$$

Consider the norm of the first column of this matrix to get

$$e^{2t} = e^{2t}(c^2 + d^2) = e^{2t}a^2 + e^{-2t}b^2 = e^{2t}a^2 + e^{-2t}(1 - a^2),$$

or $e^{2t}(1-a^2) = e^{-2t}(1-a^2)$, which implies $a = \pm 1$ and therefore $b = 0$. But then also $d = 0$ and the claim follows. So this means $k_2 l_2^{-1} = k_1^{-1} l_1 = \pm 1$ and therefore the claimed uniqueness up to sign.

Let $M = \{\pm 1\} \subseteq K$. In order to verify the integral formula, consider the map $\phi : K/M \times A^+ \to AN \smallsetminus \{1\}$ defined by

$$\phi(kM, a) = \underline{a}(ka)\underline{n}(ka).$$

Lemma 11.2.2 *The map ϕ is a C^1 diffeomorphisms. In the coordinates $\mathbb{R}/\pi\mathbb{Z} \times \mathbb{R}_{>0} \ni (\theta, s) \mapsto (k_\theta M, a_s)$ on $K/M \times A$ and $(t, x) \mapsto a_t n_x$ on AN one has for the differential matrix*

$$|\det(D\phi)(\theta, s)| = |e^{2s} - e^{-2s}|.$$

Proof A computation shows that

$$\phi(k_\theta, a_s) = \underline{a}(k_\theta a_s)\underline{n}(k_\theta, a_s) = a_t n_x,$$

where

$$\begin{aligned} t &= -\frac{1}{2}\log\left(e^{2s}\sin^2\theta + e^{-2s}\cos^2\theta\right) \\ x &= \left(e^{2s} - e^{-2s}\right)\sin\theta\cos\theta \end{aligned}$$

According to the Cartan decomposition, the map $K/M \times A^+ \to (G \smallsetminus K)/K$ is bijective. By the Iwasawa decomposition, the map $AN \to G/K$ is bijective as well, hence ϕ is bijective.

The map ϕ is continuously differentiable. Once we have shown the claimed formula for the differential, it follows that the differential matrix is invertible and so the inverse function is continuously differentiable as well. We have

$$\det(D\phi)(\theta, s) = \det\begin{pmatrix} \frac{\partial t}{\partial\theta} & \frac{\partial t}{\partial s} \\ \frac{\partial x}{\partial\theta} & \frac{\partial x}{\partial s} \end{pmatrix} = \frac{\partial t}{\partial\theta}\frac{\partial x}{\partial s} - \frac{\partial t}{\partial s}\frac{\partial x}{\partial\theta}.$$

From this one gets the lemma by a computation □

The transformation formula for the variables $(x, t) = \phi(\theta, s)$ shows

$$\begin{aligned} \int_G f(g)\,dg &= 2\int_{\mathbb{R}}\int_{\mathbb{R}}\int_K f(a_t n_y l)\,dl\,dy\,dt \\ &= 2\int_0^{\pi}\int_0^{\infty}\int_K f(k_\theta a_s l)\left(e^{2s} - e^{-2s}\right)\,dl\,ds\,d\theta \\ &= \int_0^{2\pi}\int_0^{\infty}\int_K f(k_\theta a_s l)\left(e^{2s} - e^{-2s}\right)\,dl\,ds\,d\theta \\ &= 2\pi\int_K\int_0^{\infty}\int_K f(k a_s l)\left(e^{2s} - e^{-2s}\right)\,dl\,ds\,dk, \end{aligned}$$

for every $f \in L^1(G)$, where the transition from the integral over $[0, \pi]$ to the integral over $[0, 2\pi]$ in the middle equation is justified by the fact that $k_{\theta+\pi} a_s = k_\theta a_s m$ with $m = \pm 1 \in M$ for all $\theta \in \mathbb{R}$ and $s > 0$. This finishes the proof of the theorem. □

Corollary 11.2.3 *The map*

$$K \backslash G / K \to [2, \infty), \qquad x \mapsto \operatorname{tr}(x^t x)$$

is a bijection.

Proof The map is a bijection when restricted to $\overline{A^+}$, so the corollary follows from the theorem. □

Definition A function f on G is said to be *K-bi-invariant* if it factors through $K \backslash G / K$. We define the *Hecke algebra* $\mathcal{H}$ of G to be the set of K-bi-invariant functions f on G, which are in $L^1(G)$. So we can characterize $\mathcal{H}$ as the space of all $f \in L^1(G)$ with $L_k f = f = R_k f$ for every $k \in K$, where $L_k f(x) = f(k^{-1}x)$ and $R_k f(x) = f(xk)$. We know that for $f, g \in L^1(G)$,

$$L_k(f * g) = (L_k f) * g \quad \text{and} \quad R_k(f * g) = f * (R_k g).$$

We conclude that $\mathcal{H}$ is a convolution subalgebra of $L^1(G)$. Further, $\mathcal{H}$ is stable under the involution $f^*(x) = \overline{f(x^{-1})}$, so $\mathcal{H}$ is a $*$-subalgebra of $L^1(G)$.

Theorem 11.2.4

(a) *The Hecke algebra $\mathcal{H}$ is commutative.*

(b) *For every irreducible unitary representation π of G the space of K-invariants,*

$$V_\pi^K = \{v \in V_\pi : \pi(k)v = v \ \forall k \in K\}$$

is zero or one dimensional.

(c) *For every irreducible representation π of G and for every $f \in \mathcal{H}$ we have $\pi(f) = P_K \pi(f) P_K$, where $P_K : V_\pi \to V_\pi^K$ denotes the orthogonal projection.*

Proof For $x \in G$ the Cartan decomposition implies that $KxK = Kx^{-1}K$, as this is the case for $x \in A$, since conjugating $x \in A$ with $\begin{pmatrix} & -1 \\ 1 & \end{pmatrix} \in K$ gives x^{-1}. This implies that for every $f \in \mathcal{H}$ one has $f(x^{-1}) = f(x)$. For general $f \in L^1(G)$ let $f^\vee(x) = f(x^{-1})$, then $(f * g)^\vee = g^\vee * f^\vee$ for all $f, g \in L^1(G)$. For $f, g \in \mathcal{H}$, one has $f^\vee = f$ and likewise for g and $f * g$, so that

$$f * g = (f * g)^\vee = g^\vee * f^\vee = g * f.$$

So $\mathcal{H}$ is commutative, which proves (a).

For (b) assume $V_\pi^K \neq 0$. The Hecke algebra acts on V_π^K. We show that V_π^K is irreducible under $\mathcal{H}$, so let $U \subset V_\pi^K$ a closed, $\mathcal{H}$-stable subspace. We show that $U = 0$ or $U = V_\pi^K$. For this assume $U \neq 0$, then, as π is irreducible, one has

$\overline{\pi(L^1(G))U} = V_\pi$. Let $P_K : V_\pi \to V_\pi^K$ be the orthogonal projection. Then $P_K v = \int_K \pi(k)v\,dk$ for $v \in V_\pi$ (see Proposition 7.3.3), as we normalize the Haar measure on K to have volume one. For $f \in L^1(G)$ let

$$\tilde{f}(x) = \int_K \int_K f(kxl)\,dk\,dl \ \in \ \mathcal{H}.$$

It follows that $P_K \pi(f) P_K = \pi(\tilde{f})$. Let $u \in U$ and $f \in L^1(G)$. Then

$$P_K \pi(f) u = P_K \pi(f) P_K u = \pi(\bar{f}) u \ \in \ U.$$

So we conclude that $U = P_K V_\pi = V_\pi^K$ and thus V_π^K is irreducible. Finally, to see that every irreducible *-representation $\eta : \mathcal{H} \to \mathcal{B}(V_\eta)$ on a Hilbert space V_η is one-dimensional, observe that for each $f \in \mathcal{H}$ the operator $\eta(f)$ commutes with the self-adjoint irreducible set $\eta(\mathcal{H}) \subset \mathcal{B}(V_\eta)$, since $\mathcal{H}$ is commutative. Thus $\eta(\mathcal{H}) \subset \mathbb{C}\mathrm{Id}$ by Schur's Lemma (Theorem 5.1.6). As η is irreducible, it must be one dimensional.

For (c) observe that $\tilde{f} = f$ for every $f \in \mathcal{H}$. Thus it follows from the above computations that $\pi(f) = P_K \pi(f) P_K$ for every $f \in \mathcal{H}$. □

Let $\widehat{G}_K$ be the set of all $\pi \in \widehat{G}$ such that the space V_π^K of K-invariants is non-zero. We will now give a list of the $\pi \in \widehat{G}_K$. For $\lambda \in \mathbb{C}$ let V_λ be the Hilbert space of all functions $\phi : G \to \mathbb{C}$ with firstly, $\phi(m a_t n x) = e^{t(2\lambda+1)}\phi(x)$ for $m = \pm 1 \in G$, $a_t \in A$, $n \in N$ and $x \in G$. By the Iwasawa decomposition, such ϕ is uniquely determined by its restriction to K. We secondly insist that $\phi|_K$ be in $L^2(K)$. We equip V_λ with the inner product of $L^2(K)$. The group G acts on this space by $\pi_\lambda(y)\phi(x) = \phi(xy)$. Note that the restriction to the subgroup K of the representation π_λ is the induced representation $\mathrm{Ind}_M^K(1)$ as in Sect. 7.4. The Frobenius reciprocity (Theorem 7.4.1) implies that

$$\pi_\lambda|_K \ \cong \ \bigoplus_{l \in \mathbb{Z}} \varepsilon_{2l},$$

where for $l \in \mathbb{Z}$ the character ε_l on K is defined by

$$\varepsilon_l \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = e^{il\theta}.$$

Proposition 11.2.5 If $\lambda \in i\mathbb{R}$, *then the representation* π_λ *is unitary.*

Proof The map $\phi \mapsto \phi|_K$ yields an isomorphism of Hilbert spaces, $V_\lambda \cong L^2(\bar{K})$, where $\bar{K} = K/\pm 1$. The representation π_λ can, on $L^2(\bar{K})$, be written as $\pi_\lambda(y)\phi(k) = e^{\underline{t}(ky)(2\lambda+1)}\phi(\underline{k}(ky))$. To see this, recall that $t = \underline{t}(ky)$ is the unique real number such that $a_t = \underline{a}(ky)$ in the Iwasawa decomposition $ky = \underline{a}(ky)\underline{n}(ky)\underline{k}(ky)$, and therefore

$$\pi_\lambda(y)\phi(k) = \phi(ky) = \phi\left(a_{\underline{t}(ky)}\underline{n}(ky)\underline{k}(ky)\right) = e^{\underline{t}(ky)(2\lambda+1)}\phi(\underline{k}(ky)).$$

It follows that $|\pi_\lambda(y)\phi(k)|^2 = e^{\underline{t}(ky)(4\mathrm{Re}(\lambda)+2)}|\phi(\underline{k}(ky))|^2$. By the second assertion of Theorem 11.1.3, one sees that π_λ is indeed unitary if $\lambda \in i\mathbb{R}$. □

Definition For a general representation (π, V_π) of G we let $V_{\pi,K}$ denote the space of all K*-finite vectors*, i.e., the space of all vectors $v \in V_\pi$ such that $\pi(K)v$ spans a finite dimensional space in V_π. The vector space $V_{\pi,K}$ is in general not stable under G, but is always stable under K. Since V_π has a decomposition $V_\pi = \widehat{\bigoplus}_{i\in I} U_i$, where U_i is an irreducible (hence finite-dimensional) K-representation, it follows that $V_{\pi,K}$ is dense in V_π.

The representations π_λ for $\lambda \in i\mathbb{R}$ are called the *unitary principal series* representations. One can show that π_λ is irreducible and unitarily equivalent to $\pi_{-\lambda}$ if $\lambda \in i\mathbb{R}$. These are the only equivalences that occur. One can show that for $0 < \lambda < 1/2$ there is an inner product on the space $V_{\lambda,K}$ such that the completion of $V_{\lambda,K}$ with respect to this inner product is the space of a unitary representation of G. By abuse of notation, this representation is again denoted (π_λ, V_λ). These are called the *complementary series* representations. The set $\widehat{G}_K$ consists of

- the trivial representation,
- the unitary principal series representations π_{ir}, where $r \geq 0$, and
- the complementary series π_λ for $0 < \lambda < 1/2$.

No two members of this list are equivalent. The proofs of these facts can be found in [Kna01], Chapter II.

Note that the one dimensional space V_λ^K is spanned by the element p_λ with

$$p_\lambda(mank) = e^{\underline{t}(a)(2\lambda+1)}.$$

By Corollary 11.2.3 there exists for every $f \in \mathcal{H}$ a unique function ϕ_f on $[0,\infty)$ such that

$$f(x) = \phi_f\left(\mathrm{tr}\,(x^t x) - 2\right).$$

Consider the special case $x \in AN$, say $x = a_t n_s$, then $\mathrm{tr}\,(x^t x) = (s^2+1)e^{2t} + e^{-2t}$. For $f \in \mathcal{H}$ there exists a function h_f such that

$$\pi_{ir}(f)p_{ir} = h_f(r)p_{ir}.$$

The function h_f is called the *eigenvalue function* of f. Here ir can vary in $\mathrm{i}\mathbb{R} \cup (0, 1/2)$. Since $p_{ir}(1) = 1$, we can compute $h_f(r)$ as follows

$$h_f(r) = \pi_{ir}(f)p_{ir}(1) = \int_G f(x)p_{ir}(x)\,dx$$

Lemma 11.2.6 *The map $f \mapsto h_f$ is injective on $\mathcal{H}$. We have $h_f(r) = \mathrm{tr}\,\pi_{ir}(f)$, and for $f, g \in \mathcal{H}$ the formula*

$$h_{f*g} = h_f h_g$$

holds. The function h_f can be computed via the following integral transformations. First set

$$q_f(x) = A(\phi_f)(x) \stackrel{\text{def}}{=} \int_{\mathbb{R}} \phi_f\left(x+s^2\right)\,ds, \qquad x \geq 0.$$

The map $\phi \mapsto A(\phi)$ *is called the* Abel transform. *Next define*

$$g_f(u) \stackrel{\text{def}}{=} q_f\left(e^u + e^{-u} - 2\right), \qquad u \in \mathbb{R}.$$

Then one has

$$h_f = \int_{\mathbb{R}} g_f(u) e^{iru}\, du.$$

Proof For the injectivity take an $f \in \mathcal{H}$ with $h_f = 0$. Then $\pi(f) = 0$ for every $\pi \in \widehat{G}$. By the Plancherel Theorem the representation $(R, L^2(G))$ is a direct integral of irreducible representations and so it follows that $R(f) = 0$. In particular it follows that $g * f = 0$ for every $g \in C_c(G)$. Letting g run through a Dirac net, it follows $f = 0$.

The equation $h_f(r) = \operatorname{tr} \pi_{ir}(f)$ is a consequences of Theorem 11.2.4 and $h_{f*g} = h_f h_g$ follows from $\pi(f * g) = \pi(f)\pi(g)$ for all $f, g \in L^1(G)$.

Using Iwasawa coordinates and the K-invariance of f, we compute

$$\begin{aligned} h_f(r) &= \int_{AN} f(an) e^{t(a)(2ir+1)}\, da\, dn \\ &= 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi_f\left(e^{2t} + e^{-2t} + s^2 - 2\right) e^{2tir}\, ds\, dt, \end{aligned}$$

where we used the transformation $s \mapsto e^{-t}s$. As $q_f(x) = A\phi_f(x)$ and g is even, we have

$$h_f(r) = 2\int_{-\infty}^{\infty} q_f\left(e^{2t} + e^{-2t} - 2\right) e^{2tir}\, dt = \int_{\mathbb{R}} g_f(u) e^{iru}\, du. \qquad \square$$

Definition Let $\mathcal{S}_{[0,\infty)}$ be the space of all infinitely differentiable functions ϕ on $[0, \infty)$ such that the function $x^n \phi^{(m)}(x)$ is bounded for all $m, n \geq 0$.

Lemma 11.2.7 *The Abel transform is invertible in the following sense: Let* ϕ *be continuously differentiable on* $[0, \infty)$ *such that*

$$|\phi\left(x + s^2\right)|, |s\phi'\left(x + s^2\right)| \;\leq\; g(s)$$

for some $g \in L^1([0, \infty))$, *then* $q = A(\phi)$ *is continuously differentiable and*

$$\phi = \frac{-1}{\pi} A(q').$$

Moreover, the Abel transform maps $\mathcal{S}_{[0,\infty)}$ *to itself and defines a bijection* $A : \mathcal{S}_{[0,\infty)} \to \mathcal{S}_{[0,\infty)}$.

Proof We first show that for any ϕ satisfying the conditions we have $\lim_{x\to\infty}\phi(x) = 0$. To see this, let $h(s) = s\phi'(s^2)$. Then h is integrable on $[0,\infty]$. It follows that

$$\phi(y) - \phi(0) = \int_0^y \frac{h(\sqrt{t})}{\sqrt{t}}\,dt = 2\int_0^{\sqrt{y}} h(u)\,du.$$

Letting $y \to \infty$, we see that $\lim_{x\to\infty}\phi(x)$ exists and since $\phi(x+s^2)$ is integrable, this limit is zero.

Next by the theorem of dominated convergence one sees that q is continuously differentiable and that $q' = A(\phi')$. Using polar coordinates, we compute

$$\begin{aligned} -\frac{1}{\pi}\int_{\mathbb{R}}\int_{\mathbb{R}} \phi'\left(x+s^2+t^2\right)\,ds\,dt &= -2\int_0^\infty r\phi'\left(x+r^2\right)\,dr \\ &= -\phi\left(x+r^2\right)|_0^\infty = \phi(x). \end{aligned}$$

It is easy to see that the Abel transform as well as its inverse map $\mathcal{S}_{[0,\infty)}$ to itself. The lemma follows. □

Lemma 11.2.8 *Let E be the space of all entire functions h such that $x^n h^{(m)}(x+ki)$ is bounded in $x \in \mathbb{R}$ for all $m,n \geq 0$ and every $k \in \mathbb{R}$. Let F be the space of all smooth functions g on $\mathbb{R}$ such that $(e^u + e^{-u})^n g^{(m)}(u)$ is bounded for all $m,n \geq 0$. Then the Fourier transform*

$$\Phi(h)(u) \stackrel{\text{def}}{=} \frac{1}{2\pi}\int_{\mathbb{R}} h(r)e^{-iru}\,dr, \qquad h \in E,$$

defines a linear bijection $\Phi : E \to F$. Its inverse is given by

$$\Phi^{-1}(g)(r) = \int_{\mathbb{R}} g(u)e^{iru}\,du.$$

The map Φ maps the subspace of even functions E^{ev} in E to the space of even functions F^{ev} in F.

Proof By some simple estimates, the space F can also be characterized as the space of all smooth g such that $e^{-ku}g^{(m)}(u)$ is bounded for every $k \in \mathbb{R}$ and every $m \geq 0$.

The space F is a subspace of the Schwartz space $\mathcal{S}(\mathbb{R})$ be definition. By the identity theorem of holomorphic functions, the restriction $h \mapsto h|_{\mathbb{R}}$ is an injection of E into $\mathcal{S}(\mathbb{R})$. As the Fourier transform is a bijection on $\mathcal{S}(\mathbb{R})$, it suffices to show that it maps E to F and vice versa.

For $h \in E$ let $g = \Phi(h)$. With $k \in \mathbb{R}$, and $m \geq 0$ compute

$$e^{-ku}g^{(m)}(u) = e^{-ku}\frac{1}{2\pi i^m}\int_{\mathbb{R}} h(r)r^m e^{-iru}\,dr$$

$$= \frac{1}{2\pi i^m} \int_{\mathbb{R}} h(r) r^m e^{-i(r-ik)u}\, dr$$
$$= \frac{1}{2\pi i^m} \int_{\mathbb{R}} h(r+ik)(r+ik)^m e^{-iru}\, dr.$$

The latter is the Fourier transform of a Schwartz function and hence a bounded function in u. It follows that g lies in F.

For the converse, let $g \in F$. Then the Fourier integral

$$h(r) = \int_{\mathbb{R}} g(u) e^{iru}\, du$$

converges for every $r \in \mathbb{C}$, so h extends to a unique entire function. Further, for $m, n \geq 0$ and $k \in \mathbb{R}$ we have

$$x^n h^{(m)}(x+ik) = x^n i^m \int_{\mathbb{R}} u^m g(u) e^{-ku} e^{ixu}\, du.$$

The latter function is bounded in $x \in \mathbb{R}$. So h lies in E as claimed. The last assertion is clear as the Fourier transform preserves evenness. □

Recall the definition of the function h_f for $f \in \mathcal{H}$ as given preceding Lemma 11.2.6.

Proposition 11.2.9 *Let $\mathcal{HS}$ be the space of all smooth functions f on G of the form $f(x) = \phi(\mathrm{tr}\,(x^t x) - 2)$ for some $\phi \in \mathcal{S}_{[0,\infty)}$. Then $\mathcal{HS}$ is a subalgebra of the Hecke algebra $\mathcal{H}$ and the map $\Psi : f \mapsto h_f$ is a bijection onto the space E^{ev}.*

For a given $h \in E^{\mathrm{ev}}$ the function $f = \Psi^{-1}(h)$ is computed as follows. First one defines the even function

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

Then $q : [0,\infty) \to \mathbb{C}$ is defined to be the unique function with $g(u) = q(e^u + e^{-u} - 2)$. Further one sets $\phi = -\frac{1}{\pi} A(q')$. Then

$$f(x) = \phi\left(\mathrm{tr}\,(x^t x) - 2\right).$$

Proof First note that that the map $q \mapsto g$ with $g(u) = q(e^u + e^{-u} - 2)$ is a bijection between $\mathcal{S}_{[0,\infty)}$ and the space F^{ev}. Finally, Lemma 11.2.7 and 11.2.8 give the claim. □

11.3 An Explicit Plancherel Theorem

The Plancherel Theorem says that there exists a measure μ on $\widehat{G}$ such that for $g \in L^1(G) \cap L^2(G)$ one has

$$\|g\|_2^2 = \int_{\widehat{G}} \|\pi(g)\|_{\mathrm{HS}} \, d\mu(\pi).$$

The techniques developed so far allow us as a side-result, to give an explicit measure on $\widehat{G}_K$, for which this equation holds with $f \in \mathcal{H}_{\mathrm{sym}}$. Any such computation is called an *Explicit Plancherel Theorem.*

Theorem 11.3.1 *For every $g \in \mathcal{H}_{\mathrm{sym}}$ one has*

$$\|g\|_2^2 = \frac{1}{4\pi} \int_{\mathbb{R}} \|\pi_{ir}(g)\|_{\mathrm{HS}}^2 \, r \tanh(\pi r) \, dr.$$

Moreover, for every $f \in \mathcal{H}_{\mathrm{sym}}$ one has

$$f(1) = \frac{1}{4\pi} \int_{\mathbb{R}} \mathrm{tr}\,(\pi_{ir}(f)) \, r \tanh(\pi r) \, dr.$$

Proof We show the second assertion first. Let $h = h_f$, $\phi = \phi_f$ and $g = g_f$ be as in the discussion at the end of the previous section. Recall in particular that

$$A\phi\left(e^u + e^{-u} - 2\right) = g(u) = \frac{1}{2\pi} \int_{\mathbb{R}} h(r) e^{iru} \, dr$$

(since h is even), from which it follows that $g'(u) = \frac{i}{2\pi} \int_{\mathbb{R}} r h(r) e^{iru} \, dr$. Using this and Lemma 11.2.7 we compute

$$f(1) = \phi(0) = -\frac{1}{\pi} \int_{\mathbb{R}} (A\phi)'(x^2) \, dx.$$

As $g_f(u) = A\phi(e^u + e^{-u} - 2) = A\phi((e^{u/2} - e^{-u/2})^2)$, we get $g'(u) = (A\phi)'((e^{u/2} - e^{-u/2})^2)(e^u - e^{-u})$. Putting $x = e^{u/2} - e^{-u/2}$ in the above integral, we get

$$\begin{aligned} f(1) &= -\frac{1}{2\pi} \int_{\mathbb{R}} \frac{g'(u)}{e^{u/2} - e^{-u/2}} \, du \\ &= -\frac{i}{4\pi^2} \int_{\mathbb{R}} \int_{\mathbb{R}} r h(r) \frac{e^{iru}}{e^{u/2} - e^{-u/2}} \, dr \, du. \end{aligned}$$

As h is even, the latter equals

$$-\frac{i}{8\pi^2} \int_{\mathbb{R}} r h(r) \int_{\mathbb{R}} \frac{e^{iru} - e^{-iru}}{e^{u/2} - e^{-u/2}} \, du \, dr.$$

The first step of the following computation is justified by the fact that the integrand is even. We compute

$$
\begin{aligned}
-\frac{i}{8\pi^2}\int_{\mathbb{R}} \frac{e^{iru}-e^{-iru}}{e^{u/2}-e^{-u/2}}\,du &= \frac{1}{4\pi^2 i}\int_0^\infty \frac{e^{iru}-e^{-iru}}{e^{u/2}-e^{-u/2}}\,du \\
&= \frac{1}{4\pi^2 i}\int_0^\infty e^{-u/2}\frac{e^{iru}-e^{-iru}}{1-e^{-u}}\,du \\
&= \frac{1}{4\pi^2 i}\int_0^\infty e^{-u/2}\left(e^{iru}-e^{-iru}\right)\sum_{n=0}^\infty e^{-nu}\,du \\
&= \frac{1}{4\pi^2 i}\sum_{n=0}^\infty \int_0^\infty e^{-u(n+\frac12-ir)}\,du - \int_0^\infty e^{-u(n+\frac12+ir)}\,du \\
&= \frac{1}{4\pi^2 i}\sum_{n=0}^\infty \frac{1}{n+\frac12-ir} - \frac{1}{n+\frac12+ir}.
\end{aligned}
$$

For this latter expression we temporarily write $\psi(r)$. Then

$$
\psi\left(i\left(r+\frac{1}{2}\right)\right) = \frac{1}{4\pi^2 i}\sum_{n=0}^\infty \frac{1}{n+1+r} - \frac{1}{n-r} = \frac{1}{4\pi i}\cot(\pi r).
$$

The last step is the well known Mittag-Leffler expansion of the cotangent function. We conclude

$$
\psi(r) = \frac{1}{4\pi i}\tan(\pi i r) = \frac{1}{4\pi}\tanh(\pi r).
$$

The second assertion of the theorem follows. For the first, put $f = g * g^*$ and apply the theorem to this f. Then, on the one hand, $f(1) = g * g^*(1) = \|g\|_2^2$, and on the other, for $\pi \in \widehat{G}$,

$$
\operatorname{tr}\pi(f) = \operatorname{tr}\pi(g)\pi(g)^* = \|\pi(g)\|_{\mathrm{HS}}^2.
$$

This implies the theorem. $\square$

11.4 The Trace Formula

For $g \in \mathrm{SL}_2(\mathbb{R})$ the two eigenvalues in $\mathbb{C}$ must be inverse to each other as the determinant is one. Since g is a real matrix, its characteristic polynomial is real, and so the eigenvalues are either both real, or complex conjugates of each other. Let $g \neq \pm 1$. There are three cases.

- g is in $\mathrm{SL}_2(\mathbb{C})$ conjugate to a diagonal matrix with entries $\varepsilon, \bar{\varepsilon}$ for some $\varepsilon \in \mathbb{C}$ of absolute value one. In this case, g is called an *elliptic element* of G; or
- g is conjugate to $\pm\begin{pmatrix}1 & 1\\ & 1\end{pmatrix}$. In this case g is called a *parabolic element*; or
- g is conjugate to a diagonal matrix with entries $t, 1/t$ for some $t \in \mathbb{R}$, in which case g is called a *hyperbolic element*.

Let g be elliptic, say g is conjugate to $\begin{pmatrix} a+bi & \\ & a-bi \end{pmatrix}$. As an element of G, the element g is conjugate to some $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ in K. This implies that g has a unique fixed point in $\mathbb{H}$.

Proposition 11.4.1 *A uniform lattice* $\Gamma \subset G$ *contains no parabolic elements.*

Proof Consider the map $\eta : \mathbb{H} \to [0, \infty)$ given by

$$\eta(z) = \inf\{\rho(z, \gamma z) : \gamma \in \Gamma,\ \gamma \neq \pm 1,\ \gamma \text{ not elliptic}\}.$$

It is easy to see that the map η is continuous. Further it is Γ-invariant and therefore it constitutes a continuous function $\Gamma\backslash\mathbb{H} \to (0, \infty)$. Since $\Gamma\backslash\mathbb{H} \cong \Gamma\backslash G/K$ is compact, the function η attains its minimum, hence there exists $\theta > 0$ such that $\eta(z) \geq \theta$ for all $z \in \mathbb{H}$. Now *assume* that Γ contains a parabolic element, say $p = g \begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix} g^{-1} \in \Gamma$ for some $g \in G$. Then for $y > 1$ we have

$$\rho(g(yi), pg(yi)) = \rho(g(yi), g(yi + 1)) = \rho(yi, yi + 1)$$

and the latter tends to zero as $y \to \infty$, which follows from

$$\rho(yi, yi + 1) \leq \int_0^1 |p'(t)| \frac{dt}{\mathrm{Im}(p(t))} = \frac{1}{y},$$

where $p(t) = yi + t$. We therefore have a *contradiction!* Hence Γ does not contain any parabolic element. □

For a hyperbolic element g, with eigenvalues $\lambda, 1/\lambda$ for $|\lambda| > 1$, define the *length* of g as $l(g) = 2 \log |\lambda|$.

Let $\Gamma \subset G$ be a uniform lattice. For convenience we will assume that Γ contains no elliptic elements. Then Γ consists, besides ± 1, of hyperbolic elements only. We call such a group a *hyperbolic lattice*. In [Bea95], there are given many examples of uniform lattices in G without elliptic elements. For instance, every Riemannian manifold of genus $g \geq 2$ is a quotient of the upper half plane by a hyperbolic lattice in G.

So let Γ be a hyperbolic lattice in G. Let $r_0 = \frac{i}{2}$, and let $(r_j)_{j\geq 1}$ be a sequence in $\mathbb{C}$ such that $ir_j \in i\mathbb{R} \cup (0, \frac{1}{2})$ with the property that π_{ir_j} is isomorphic to a subrepresentation of $(R, L^2(\Gamma\backslash G))$ and the value $r = r_j$ is repeated in the sequence as often as $N_\Gamma(\pi_{ir})$ times, i.e, as often as π_{ir_j} appears in the decomposition of R. Let $f \in \mathcal{H}$ such that the operator $R(f)$ is of trace class, and define $\phi = \phi_f, g = g_f$ and $h = h_f$ as in the previous two sections (See Lemma 11.2.6). Recall that $f(x) = \phi(\mathrm{tr}\,(x^t x) - 2)$, $g(u) = A\phi(e^u + e^{-u} - 2)$, where A denotes the Abel transform, and $h(r) = \int_{\mathbb{R}} g(u) e^{iru}\, du$. Recall from Lemma 11.2.6 that $h(r) = \mathrm{tr}\, \pi_{ir}(f)$ for every $ir \in i\mathbb{R} \cup (0, \frac{1}{2})$. Moreover, it follows from Theorem 11.2.4 that $\mathrm{tr}\, \pi(f) = 0$ for all $\pi \in \widehat{G} \setminus \widehat{G}_K$. We therefore get $\mathrm{tr}\, R(f) = \sum_{j=0}^{\infty} h(r_j)$. Suppose that the trace formula of Theorem 9.3.2 is valid for the function f. Then

$$\sum_{j=0}^{\infty} h(r_j) = \sum_{[\gamma]} \mathrm{vol}(\Gamma_\gamma \backslash G_\gamma) \mathcal{O}_\gamma(f).$$

Recall the hyperbolic tangent function $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

An element $\gamma \in \Gamma \smallsetminus \{1\}$ is called *primitve*, if it is not a power in Γ, i.e., if the equation $\gamma = \sigma^n$ with $n \in \mathbb{N}$ and $\sigma \in \Gamma$ implies $n = 1$.

Lemma 11.4.2 *If* Γ *is a torsion free uniform lattice, every element* γ of $\Gamma \smallsetminus \{1\}$ *is a positive power of a uniquely determined primitive element* γ_0. *This element generates the centralizer* Γ_γ of γ *in* Γ. *We call it the primitive element underlying* γ.

Proof Let $\gamma \in \Gamma \smallsetminus \{1\}$. By assumption, γ is hyperbolic. Replacing Γ with a conjugate group we may assume that γ is the diagonal matrix with entries e^t, e^{-t} for some $t > 0$, as the other case of $\gamma = -\mathrm{diag}(e^t, e^{-t})$ gives the same result. Then the centralizer G_γ of γ in G equals $\pm A$, the group of all diagonal matrices in G and $\Gamma_\gamma = \Gamma \cap \pm A$. As $-1 \notin \Gamma$, since Γ is torsion-free, it follows that there is $\gamma_0 \in \Gamma$ such that the centralizer Γ_γ in Γ is equal to $\langle \gamma_0 \rangle$. Replacing γ_0 by γ_0^{-1} if necessary, we can assume that $\gamma = \gamma_0^n$ for some $n \in \mathbb{N}$. It follows that γ_0 is primitive. □

Theorem 11.4.3 *Assume that* Γ *is a torsion free uniform lattice in* $\mathrm{SL}(2, \mathbb{R})$. *Let* $\varepsilon > 0$, *and let h be a holomorphic function on the strip* $\{|\mathrm{Im}(z)| < \frac{1}{2} + \varepsilon\}$. *Suppose that h is even, i.e.,* $h(-z) = h(z)$ *for every z, and that* $h(z) = O(|z|^{-2-\varepsilon})$ *as* $|z|$ *tends to infinity. Let* $g(u) = \frac{1}{2\pi} \int_{\mathbb{R}} h(r) e^{-iru}\, dr$. *Then one has*

$$\sum_{j=0}^{\infty} h(r_j) = \frac{\mathrm{vol}(\Gamma \backslash G)}{4\pi} \int_{\mathbb{R}} r h(r) \tanh(\pi r)\, dr + \sum_{[\gamma] \neq 1} \frac{l(\gamma_0)}{e^{l(\gamma)/2} - e^{-l(\gamma)/2}} g(l(\gamma)),$$

where for $\gamma \neq 1$, γ_0 *is the primitive element underlying* γ.

Proof We start with functions $f \in \mathcal{H}_{\mathrm{sym}}$, for which the trace formula holds and then we extend the range of the trace formula up to the level of the theorem. So let $f \in \mathcal{H}_{\mathrm{sym}}$ such that the trace formula is valid for f. For instance, $f \in C_c^\infty(G)^2 = C_c^\infty(G) * C_c^\infty(G)$ will suffice. At first we consider the class $[\gamma]$ with $\gamma = 1$. Then

$$\mathrm{vol}(\Gamma_\gamma \backslash G_\gamma) \mathcal{O}_\gamma(f) = \mathrm{vol}(\Gamma \backslash G)\, f(1).$$

Theorem 11.3.1 tells us that

$$f(1) = \frac{1}{4\pi} \int_{\mathbb{R}} \mathrm{tr}\,(\pi_{ir}(f))\, r \tanh(\pi r)\, dr = \frac{1}{4\pi} \int_{\mathbb{R}} h(r)\, r \tanh(\pi r)\, dr.$$

Next let γ be an element of Γ with $\gamma \neq 1$ and recall that this implies $G_\gamma = \pm A$. If $\gamma_0 = \mathrm{diag}(e^t, e^{-t})$ and if we identify A with $\mathbb{R}$ via the exponential map, the group

$\Gamma_\gamma = \langle \gamma_0 \rangle$ corresponds to the subgroup $t\mathbb{Z}$. It follows that $\mathrm{vol}(\Gamma_\gamma \backslash G_\gamma) = 2t = l(\gamma_0)$, where the factor 2 is due to the normalization of Haar measure on A.

By the Iwasawa decomposition, the set $G_\gamma \backslash G$ can be identified with $NK/\pm 1$. As f is K-bi-invariant and $f(x) = \phi(\mathrm{tr}\,(x^t x) - 2)$ for every $x \in G$, the orbital integral $\mathcal{O}_\gamma(f)$ equals

$$\int_{\mathbb{R}} f(n_s^{-1} \gamma n_s)\, ds = \int_{\mathbb{R}} \phi\left(e^{2t} + e^{-2t} + s^2(e^t - e^{-t})^2 - 2\right) ds,$$

so that

$$\begin{aligned}\mathcal{O}_\gamma(f) &= \frac{1}{e^t - e^{-t}} A\phi\left(e^{2t} + e^{-2t} - 2\right)\\ &= \frac{1}{e^t - e^{-t}} g(2t) = \frac{1}{e^{l(\gamma)/2} - e^{-l(\gamma)/2}} g(l(\gamma)).\end{aligned}$$

By the general trace formula as stated before the theorem, we see that the theorem holds if $f \in \mathcal{H}_{\mathrm{sym}}$ is admissible for the trace formula.

We now derive the trace formula for the special case of the *heat kernel*. Let

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

Note that $h_t \in E$ and so Proposition 11.2.9 applies. One gets $g(u) = \frac{e^{-t/4}}{\sqrt{4\pi t}} e^{-\frac{u^2}{4t}}$. Let $f_t = \Psi^{-1}(h_t)$ with $\Psi : \mathcal{H}_{\mathrm{sym}} \to E^{\mathrm{ev}}$ as in Proposition 11.2.9. Recall from Sect. 9.2 the definition of the space $C_{\mathrm{unif}}(G)$ of uniformly integrable functions on G.

Proposition 11.4.4 *The function f_t lies in $C_{\mathrm{unif}}(G)^2$, so the trace formula is valid for f.*

Proof Note that $h_t = h_{t/2}^2$, which means $f_t = f_{t/2} * f_{t/2}$ and so, in order to show that the trace formula is valid for f_t, it suffices to show that $f_t \in C_{\mathrm{unif}}(G)$ for every $t > 0$, as it then follows that $f_t \in C_{\mathrm{unif}}(G)^2$. Let $r > 0$ and define

$$U(r) \stackrel{\mathrm{def}}{=} K\{a_s : 0 \le s < r\}K \ \subset G.$$

Then $U(r)$ is an open neighborhood of the unit. Note that $U(r) = \{x \in G : \mathrm{tr}\,(x^t x) < e^{2r} + e^{-2r}\}$ and the boundary satisfies

$$\partial U(r) = \{x \in G : \mathrm{tr}\,(x^t x) = e^{2r} + e^{-2r}\} = K a_r K.$$

Lemma 11.4.5 *For $0 < r < \frac{s}{2}$ we have*

$$U(r) a_s U(r) \subset U(s + 2r) \smallsetminus U(s - 2r).$$

Proof As $U(r)$ is invariant under K-multiplication from both sides, it suffices to show everything modulo K-multiplication on both sides. Suppose that for $0 \le y < r$

and $k \in K$ we can show that $a_y k a_x$ and $a_x k a_y$ both lie in $U(x+r) \smallsetminus U(x-r)$. Then, modulo K-multiplication one has $a_y k a_s = a_t$ for $s - r < t < s + r$. Iterating the argument with t taking the part of s, one gets for $0 \le y' < r$,

$$a_y k a_s k' a_{y'} = a_t k'' a_{y'} \in U(t+r) \smallsetminus U(t-r) \subset U(s+2r) \smallsetminus U(s-2r).$$

So it suffices to show that for $k \in K$ one has $a_y k a_s, a_s k a_y \in U(s+r) \smallsetminus U(s-r)$ for $0 \le y < r$ and arbitrary s. For $x \in G$ let $T(x) = \operatorname{tr}(x^t x)$. Note that

$$T \begin{pmatrix} a & b \\ c & d \end{pmatrix} = a^2 + b^2 + c^2 + d^2.$$

We have to show that

$$e^{2(s-r)} + e^{2(r-s)} < T(a_y k a_s) < e^{2(s+r)} + e^{-2(s+r)}.$$

Now any $k \in K$ can be written as $k = \left(\begin{smallmatrix} a & -b \\ b & a \end{smallmatrix}\right)$ for some $a, b \in \mathbb{R}$ with $a^2 + b^2 = 1$. Then

$$\begin{aligned} T(a_y k a_s) &= T \begin{pmatrix} e^{y+s} a & -e^{y-s} b \\ e^{s-y} b & e^{-(y+s)} a \end{pmatrix} \\ &= e^{2(y+s)} + e^{-2(y+s)} + b^2 (e^{2(y-s)} + e^{2(s-y)} - e^{2(y+s)} - e^{-2(y+s)}). \end{aligned}$$

Here we have used $a^2 = 1 - b^2$. Now $b \in [-1, 1]$ and the above is a quadratic polynomial in b, which takes its extremal values at the zero of its derivative, i.e., at $b = 0$ or at $b = \pm 1$. In both cases we get the claim. □

The proof of the proposition now proceeds as follows: One notes that the function ϕ_t with $f_t(x) = \phi_t\left(\operatorname{tr} x^t x - 2\right)$ is monotonically decreasing. This follows from $\phi_t = -\frac{1}{\pi} A(q_t')$. Hence Lemma 11.4.5 implies that $(f_t)_{U(r)}(a_s) \le f_t(a_{s-2r})$ for $0 \le r < s/2$. (Recall the notation $f_U(x) = \sup |f(UxU)|$.) Therefore it suffices to show that for any $r \ge 0$,

$$\int_{\{x \in G : T(x) \ge 2r\}} \phi_t(\operatorname{tr}(x^t x) - 2 - 2r)\, dx < \infty.$$

For this we use the integration formula of the Cartan decomposition in Theorem 11.2.1, which shows that the integral equals

$$2\pi \int_{e^{2x} + e^{-2x} - 2 > 2r} \phi_t\left(e^{2x} + e^{-2x} - 2 - 2r\right)\left(e^{2x} - e^{-2x}\right) dx.$$

Substituting $u = e^{2x} + e^{-2x}$ this becomes

$$\pi \int_{u > 2r+2} \phi_t(u - 2 - 2r)\, du = \pi \int_0^\infty \phi_t(x)\, dx.$$

As $h_t \in E^{\mathrm{ev}}$, the function ϕ_t lies in $\mathcal{S}_{[0,\infty)}$ and so this integral is indeed finite. □

The trace formula for the function f_t says

$$\sum_{j=0}^{\infty} e^{-\left(\frac{1}{4}+r_j^2\right)t} = \frac{\mathrm{vol}(\Gamma\backslash G)}{4\pi} \int_{\mathbb{R}} r e^{-\left(\frac{1}{4}+r^2\right)t} \tanh(\pi r)\, dr + \sum_{[\gamma]\neq 1} \frac{l(\gamma_0)}{e^{l(\gamma)/2} - e^{-l(\gamma)/2}} \frac{1}{2\pi} \int_{\mathbb{R}} e^{-(r^2+1/4)t} e^{irl(\gamma)}\, dr,$$

where we used the equation $g_t(u) = \frac{1}{2\pi}\int_{\mathbb{R}} h_t(r)e^{iru}\, dr$, which follows from inverse Fourier transform and the fact that h_t is even. Let $\mu(t)$ denote either side of this equation. For a complex number s with $\mathrm{Re}(s^2) < -\frac{1}{4}$ let

$$\alpha(s) \stackrel{\mathrm{def}}{=} \int_1^{\infty} \mu(t) e^{t\left(s^2+\frac{1}{4}\right)}\, dt.$$

By realizing μ via the left hand side of the trace formula gives

$$\alpha(s) = \sum_{j=0}^{\infty} \frac{e^{s^2-r_j^2}}{r_j^2 - s^2}$$

and using the right hand side of the trace formula gives

$$\alpha(s) = \frac{\mathrm{vol}(\Gamma\backslash G)}{4\pi} \int_{\mathbb{R}} r \frac{e^{s^2-r^2}}{r^2-s^2} \tanh(\pi r)\, dr + \sum_{[\gamma]\neq[1]} \frac{l(\gamma_0)}{e^{l(\gamma)/2} - e^{-l(\gamma)/2}} \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{s^2-r^2}}{r^2-s^2} e^{irl(\gamma)}\, dr.$$

Now take h as in the assumptions of the theorem, but with the stronger growth condition $h(z) = O(\exp(-a|z|^4))$ for some $a > 0$ and $|\mathrm{Im}(z)| < \frac{1}{2} + \varepsilon$. For $T > 0$, let R_T denote the positively oriented rectangle with vertices $\pm T \pm i\frac{\varepsilon+1}{2}$. By the Residue Theorem we can compute

$$\frac{1}{2\pi i} \int_{R_T} \frac{e^{s^2-r^2}}{r^2-s^2} s h(s)\, ds = \frac{1}{2}(h(r) + h(-r)) = h(r)$$

whenever r lies in the interior of the rectangle, and 0 else. For $T \to \infty$ this converges to $h(r)$ for every $r \in \mathbb{R} \cup i(0, \frac{1}{2})$. Thus, using the realization $\alpha(s) = \sum_{j=0}^{\infty} \frac{e^{s^2-r_j^2}}{r_j^2-s^2}$ it follows that that $\frac{1}{2\pi i}\int_{R_T} \alpha(s) s h(s)\, ds$ converges to the right hand side of Theorem 11.4.3 if $T \to \infty$. On the other hand, using the realization of $\alpha(s)$ given by the left hand side of the trace formula and interchanging the order of integration, which is justified by the growth condition on h, shows that $\frac{1}{2\pi i}\int_{R_T} \alpha(s) s h(s)\, ds$ equals

$$\frac{\mathrm{vol}(\Gamma\backslash G)}{4\pi} \int_{-T}^{T} r h(r) \tanh(\pi r)\, dr + \sum_{[\gamma]\neq[1]} \frac{l(\gamma_0)}{e^{l(\gamma)/2} - e^{-l(\gamma)/2}} \frac{1}{2\pi} \int_{-T}^{T} h(r) e^{irl(\gamma)}\, dr.$$

This converges to the left hand side of Theorem 11.4.3 if $T \to \infty$ since $g(l(\gamma)) = \frac{1}{2\pi}\int_{\mathbb{R}} h(r)e^{irl(\gamma)}\,dr$.

This proves the theorem for h satisfying the stronger growth condition. For arbitrary h, let $a > 0$ and set $h_a(z) = h(z)\exp(-az^4)$. Then the function h_a satisfies the stronger growth condition for $|\mathrm{Im}(z)| < \frac{1}{2} + \varepsilon$ and the limit $a \to 0$, using Lebesgue's convergence theorem for the integrals, gives the claim. □

11.5 Weyl's Asymptotic Law

In the proof of the trace formula, we have used the "heat kernel" $h_t(r) = e^{-(\frac{1}{4}+r^2)t}$. The reason for this being called so is the following. The *Laplace operator* for hyperbolic geometry on $\mathbb{H}$,

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

is invariant under G, i.e., $\Delta L_g = L_g\Delta$ for every $g \in G$. Therefore Δ defines a differential operator on the quotient $\Gamma\backslash\mathbb{H}$, which we denote by the same letter. It can be shown that its eigenvalues are $\lambda_j = (\frac{1}{4} + r_j^2)$ for $j \geq 0$. Since this requires additional arguments from Lie theory and is not essential for our purposes, we will not give the proof, but only mention the fact as an explanation for the terminology. The interested reader may consult Helgason's book [Hel01].

The hyperbolic *heat operator* is $e^{-t\Delta}$ for $t > 0$. This is an integral operator whose kernel $k_t(z, w)$ describes the amount of heat flowing in time t from point z to point w. Therefore

$$\sum_{j=0}^{\infty} e^{-\left(\frac{1}{4}+r_j^2\right)t} = \sum_{j=0}^{\infty} e^{-t\lambda_j} = \operatorname{tr} e^{-t\Delta}$$

is the *heat trace* on $\Gamma\backslash\mathbb{H}$.

Proposition 11.5.1 *As $t \to 0$, one has*

$$t\sum_{j=0}^{\infty} e^{-\left(\frac{1}{4}+r_j^2\right)t} \to \frac{\mathrm{vol}(\Gamma\backslash\mathbb{H})}{4\pi}.$$

Proof As $g(u) = \frac{e^{-t/4}}{\sqrt{4\pi t}}e^{-\frac{u^2}{4t}}$, the trace formula for the heat kernel gives

$$t\sum_{j=0}^{\infty} e^{-t(\frac{1}{4}+r_j^2)} = t\,\frac{\mathrm{vol}(\Gamma\backslash\mathbb{H})}{4\pi}\int_{-\infty}^{\infty} re^{-(\frac{1}{4}+r^2)t}\tanh(\pi r)\,dr$$

$$+\,t\sum_{[\gamma]\neq 1}\frac{l(\gamma_0)}{e^{l(\gamma)/2} - e^{-l(\gamma)/2}}\,\frac{e^{-\frac{t}{4}-\frac{l(\gamma)^2}{4t}}}{\sqrt{4\pi t}}.$$

Substituting r with $r/\sqrt{t}$ shows that the first summand equals

$$\frac{\mathrm{vol}(\Gamma\backslash\mathbb{H})}{4\pi}e^{-t/4}\int_{\mathbb{R}} re^{-r^2}\tanh\left(\pi\frac{r}{\sqrt{t}}\right)dr.$$

The integral equals $\int_0^\infty 2re^{-r^2}\tanh\left(\pi\frac{r}{\sqrt{t}}\right)dr$. As $t \to 0$, the tanh-term tends to 1 monotonically from below; therefore the integral tends to

$$\int_0^\infty 2re^{-r^2}\,dr = -e^{-r^2}\Big|_0^\infty = 1.$$

It remains to show that

$$\sqrt{t}\sum_{[\gamma]\neq 1}\frac{l(\gamma_0)}{e^{l(\gamma)/2}-e^{-l(\gamma)/2}}\frac{e^{-\frac{t}{4}-\frac{l(\gamma)^2}{4t}}}{\sqrt{4\pi}}$$

tends to zero as $t \to 0$. This is clear as the sum is finite for every $0 < t < 1$ and each summand tends to zero monotonically as soon as $t < l/2$, where l is the minimal length $l(\gamma)$ for $\gamma \in \Gamma\smallsetminus\{1\}$. □

We use this proposition to derive Weyl's asymptotic formula.

Theorem 11.5.2 *For $T > 0$, let $N(T)$ be the number of eigenvalues $\lambda_j = \frac{1}{4} + r_j^2$ of Δ that are $\leq T$. Then, as $T \to \infty$, one has*

$$N(T) \sim \frac{\mathrm{vol}(\Gamma\backslash\mathbb{H})}{4\pi}T,$$

where the asymptotic equivalence $\sim$ means that the quotient of the two sides tends to 1, as $T \to \infty$.

Proof We need a lemma. Recall the definition of the Γ-function from Sect. 11.2.6.

Lemma 11.5.3 *Let μ be a Borel measure on $[0,\infty)$ such that*

$$\lim_{t\to 0} t\int_{[0,\infty)} e^{-t\lambda}\,d\mu(\lambda) = C$$

for some $C > 0$. Then the following hold.

(a) *If f is a continuous function on $[0, 1]$, then*

$$\lim_{t\to 0} t\int_{[0,\infty)} f(e^{-t\lambda})e^{-t\lambda}\,d\mu(\lambda) = C\int_0^\infty f(e^{-x})e^{-x}\,dx.$$

(b) *One has*

$$\lim_{t\to 0} t\int_{[0,\frac{1}{t}]} d\mu(\lambda) = C.$$

Proof (a) By The Stone-Weierstraß Theorem A.10.1, the set of polynomials is dense in $C([0,1])$. We first show that it suffices to prove the lemma for polynomials in the role of f. So let $f_n \to f$ be a convergent sequence in $C([0,1])$ and assume the lemma holds for each f_n. We have to show that it holds for f as well. Let $\varepsilon > 0$. Then there exists n_0 such that $\|f_n - f\|_{[0,1]} < \varepsilon$ for every $n \geq n_0$. For such n one gets

$$\left| t \int_{[0,\infty)} \left(f_n(e^{-t\lambda}) - f(e^{-t\lambda})\right) e^{-t\lambda}\, d\mu(\lambda) \right| < \varepsilon t \int_{[0,\infty)} e^{-t\lambda}\, d\mu(\lambda),$$

and the latter tends to εC as $t \to 0$.

On the other hand,

$$\left| C \int_0^\infty \left(f_n(e^{-x}) - f(e^{-x})\right) e^{-x}\, dx \right| < \varepsilon C.$$

So it suffices to prove the lemma for a polynomial and indeed for $f(x) = x^n$, in which case it comes down to

$$\begin{aligned} \lim_{t\to 0} t \int_{[0,\infty)} e^{-t(n+1)\lambda}\, d\mu &= (n+1)^{-1} \lim_{t\to 0} t \int_{[0,\infty)} e^{-t\lambda}\, d\mu(\lambda) \\ &= \frac{C}{(n+1)} \\ &= C \int_0^\infty e^{-(n+1)t}\, dt. \end{aligned}$$

Now for (b). Consider any continuous function $f \geq 0$ on the interval such that $f(x) = \frac{1}{x}$ for $x \geq e^{-1}$. Then

$$t \int_{[0,\frac{1}{t}]} f\left(e^{-t\lambda}\right) e^{-t\lambda}\, d\mu(\lambda) = t \int_{[0,\frac{1}{t}]} d\mu(\lambda),$$

so that for the limit superior we have the bound

$$\begin{aligned} \limsup_{t\to 0} t \int_{[0,\frac{1}{t}]} d\mu(\lambda) &\leq \lim_{t\to 0} t \int_{[0,\infty)} f\big(e^{-t\lambda}\big)\, e^{-t\lambda}\, d\mu(\lambda) \\ &= C \int_0^\infty f\big(e^{-x}\big)\, e^{-x}\, dx \\ &= C + C \int_1^\infty f\big(e^{-x}\big)\, e^{-x}\, dx. \end{aligned}$$

As the last integral can be chosen arbitrarily small, by using the Monotone Convergence Theorem we get that the limit superior in question is $\leq C$. Similarly, by choosing $f(x)$ to vanish for $x \leq e^{-1}$ and satisfy $0 \leq f(x) \leq 1/x$ one gets

$$\liminf_{t\to 0} t \int_{[0,\frac{1}{t}]} d\mu(\lambda) \geq C.$$

□

To get the theorem, we apply part (b) of the last lemma to the measure $\mu = \sum_{j=0}^{\infty} \delta_{\lambda_j}$. Indeed, substituting $T = \frac{1}{t}$, the left hand side of the above equation becomes $\lim_{T\to\infty} \frac{1}{T} N(T)$, while, by the proposition, we have $C = \lim_{t\to 0} t \sum_{j=0}^{\infty} e^{-t\lambda_j} = \frac{\mathrm{vol}(\Gamma\backslash\mathbb{H})}{4\pi}$. □

11.6 The Selberg Zeta Function

As in the previous sections, let Γ be a torsion free hyperbolic uniform lattice in $\mathrm{SL}(2,\mathbb{R})$. The compact surface $\Gamma\backslash\mathbb{H}$ is homeomorphic to a 2-sphere with a finite number of handles attached. The number of handles g is ≥ 2. It is called the *genus* of the surface $\Gamma\backslash\mathbb{H}$ (See [Bea95]).

The Selberg zeta function for Γ is defined for $s \in \mathbb{C}$ with $\mathrm{Re}(s) > 1$ as

$$Z(s) = \prod_{\gamma} \prod_{k\geq 0} \left(1 - e^{-(s+k)l(\gamma)}\right),$$

where the first product runs over all primitive hyperbolic conjugacy classes in Γ.

Theorem 11.6.1 *The product Z(s) converges for* $\mathrm{Re}(s) > 1$ *and the function Z(s) extends to an entire function with the following zeros. For $k \in \mathbb{N}$ the number $s = -k$ is a zero of multiplicity* $2(g-1)(2k+1)$*, where g is the genus of* $\Gamma\backslash\mathbb{H}$*. For every* $j \geq 0$ *the numbers*

$$\frac{1}{2} + ir_j, \quad \textit{and} \quad \frac{1}{2} - ir_j$$

are zeros of $Z(s)$ *of multiplicity equal to the multiplicity* $N_\Gamma(\pi_{ir_j})$*. These are all zeros.*

Proof Let $a, b \in \mathbb{C}$ with real part $> \frac{1}{2}$. Then the function

$$h(r) = \frac{1}{a^2 + r^2} - \frac{1}{b^2 + r^2}$$

satisfies the conditions of the trace formula of Theorem 11.4.3. One computes (Exercise 11.5),

$$g(u) = \frac{1}{2\pi} \int_{\mathbb{R}} h(r) e^{-iru}\, dr = \frac{e^{-a|u|}}{2a} - \frac{e^{-b|u|}}{2b}.$$

We compute, formally at first,

$$\begin{aligned}\frac{Z'}{Z}(s) &= \partial_s \left(\log\left(\prod_{\gamma_0}\prod_{k\geq 0}\left(1 - e^{-(s+k)l(\gamma_0)}\right)\right)\right)\\ &= \partial_s \left(-\sum_{\gamma_0}\sum_{k\geq 0}\sum_{n=1}^{\infty} \frac{e^{-n(s+k)l(\gamma_0)}}{n}\right)\end{aligned}$$

$$= \sum_{\gamma_0} \sum_{k \geq 0} \sum_{n=1}^{\infty} e^{-n(s+k)l(\gamma_0)} l(\gamma_0).$$

If γ_0 runs over all primitive classes, then $\gamma = \gamma_0^n$ will run over all classes $\neq 1$. Using $l(\gamma_0^n) = nl(\gamma_0)$ we get

$$\begin{aligned} \frac{Z'}{Z}(s) &= \sum_{\gamma} \sum_{k \geq 0} e^{-(s+k)l(\gamma)} l(\gamma_0) \\ &= \sum_{\gamma} e^{-sl(\gamma)} \frac{l(\gamma_0)}{1 - e^{-l(\gamma)}} \\ &= \sum_{\gamma} e^{-(s-\frac{1}{2})l(\gamma)} \frac{l(\gamma_0)}{e^{l(\gamma)/2} - e^{-l(\gamma)/2}}. \end{aligned}$$

Up to this point we have ignored questions of convergence. To deal with these, note that the geometric side of the trace formula for our function h equals $\frac{\mathrm{vol}(\Gamma \backslash G)}{4\pi} \int_{\mathbb{R}} rh(r) \tanh(\pi r)\, dr$ plus

$$\frac{1}{2} \sum_{[\gamma] \neq 1} \left(\frac{e^{-al(\gamma)}}{a} - \frac{e^{-bl(\gamma)}}{b} \right) \frac{l(\gamma_0)}{e^{l(\gamma)/2} - e^{-l(\gamma)/2}}.$$

By the trace formula, the latter sum converges absolutely for all complex numbers a, b with $\mathrm{Re}(a), \mathrm{Re}(b) > \frac{1}{2}$. In the special case $b = 2a > 1$ all summands are positive and the estimate

$$\frac{e^{-al(\gamma)}}{a} - \frac{e^{-2al(\gamma)}}{2a} > \frac{e^{-al(\gamma)}}{a} - \frac{e^{-al(\gamma)}}{2a} = \frac{1}{2} \frac{e^{-al(\gamma)}}{a}$$

shows that the series

$$\sum_{[\gamma] \neq 1} e^{-al(\gamma)} \frac{l(\gamma_0)}{e^{l(\gamma)/2} - e^{-l(\gamma)/2}} = \frac{Z'}{Z}(a + \frac{1}{2})$$

converges locally uniformly absolutely for $\mathrm{Re}(a) > \frac{1}{2}$. To be precise, for every $a_0 > \frac{1}{2}$ consider the open set $U = \{\mathrm{Re}(a) > a_0\}$. For $a \in U$ and every $\gamma \in \Gamma \smallsetminus \{1\}$ one has $|e^{-al(\gamma)}| = e^{-\mathrm{Re}(a)l(\gamma)} < e^{-a_0 l(\gamma)}$. This shows locally uniform absolute convergence of the logarithmic derivative

$$\frac{Z'}{Z}(s) = \sum_{\gamma_0} \sum_{k \geq 0} \sum_{n=1}^{\infty} e^{-n(s+k)l(\gamma_0)} l(\gamma_0)$$

for $\mathrm{Re}(s) > 1$. By direct comparison we conclude the absolute locally uniform convergence of the series $-\sum_{\gamma_0} \sum_{k \geq 0} \sum_{n=1}^{\infty} \frac{e^{-n(s+k)l(\gamma_0)}}{n}$, which is the logarithm of Z. This implies the locally uniform convergence of the product $Z(s)$ in the region $\{\mathrm{Re}(s) > 1\}$.

The geometric side of the trace formula for h equals

$$\frac{\mathrm{vol}(\Gamma\backslash G)}{4\pi}\int_{\mathbb{R}} rh(r)\tanh(\pi r)\,dr + \frac{1}{2a}\frac{Z'}{Z}\left(a+\frac{1}{2}\right) - \frac{1}{2b}\frac{Z'}{Z}\left(b+\frac{1}{2}\right).$$

The spectral side is

$$\sum_{j=0}^{\infty}\frac{1}{2a}\left(\frac{1}{a+ir_j}+\frac{1}{a-ir_j}\right)-\frac{1}{2b}\left(\frac{1}{b+ir_j}+\frac{1}{b-ir_j}\right).$$

The trace formula implies that this series converges for complex numbers a, b with $\mathrm{Re}(a), \mathrm{Re}(b) > \frac{1}{2}$. Being a Mittag-Leffler series, it converges for all $a, b \in \mathbb{C}$, which are not one of the poles $\pm ir_j$, and it represents a meromorphic function in, say $a \in \mathbb{C}$ with simple poles at the $\pm ir_j$ of residue $1/2a$ times the multiplicity of $\pm ir_j$.

We want to evaluate the integral

$$\int_{\mathbb{R}} rh(r)\tanh(\pi r)\,dr.$$

The Mittag-Leffler series of tanh equals

$$\tanh(\pi z) = \frac{1}{\pi}\sum_{n=0}^{\infty}\frac{1}{z+i(n+\frac{1}{2})}+\frac{1}{z-i(n+\frac{1}{2})},$$

where the sum converges absolutely locally uniformly outside the set of poles $i(\frac{1}{2}+\mathbb{Z})$. For $n \in \mathbb{N}$ the path γ_n consisting of the interval $[-n, n]$ and the half-circle in $\{\mathrm{Im} z > 0\}$ around zero of radius n will not pass through a pole. Note that the function $\tanh(\pi r)$ is periodic, i.e, $\tanh(\pi(r+2i)) = \tanh(\pi r)$. Further, it is globally bounded on any set of the form $\{z \in \mathbb{C} : |z - i(k+1/2)| \geq \varepsilon\ \forall k \in \mathbb{Z}\}$ for any $\varepsilon > 0$. As $rh(r)$ is decreasing to the power r^{-3}, it follows that the integral $\int_{\gamma_n} rh(r)\tanh(\pi r)\,dr$ converges to the integral in question. By the residue theorem we conclude

$$\int_{\mathbb{R}} rh(r)\tanh(\pi r)\,dr = 2\pi i\sum_{z:\mathrm{Im} z>0}\mathrm{res}_{r=z}(rh(r)\tanh(\pi r)).$$

We have

$$rh(r) = \frac{1}{2}\left(\frac{1}{r+ia}+\frac{1}{r-ia}\right)-\frac{1}{2}\left(\frac{1}{r+ib}+\frac{1}{r-ib}\right).$$

We will assume $a \neq b$, both in $\mathbb{C}\smallsetminus(\frac{1}{2}+\mathbb{Z})$. Then the poles of $rh(r)$ and of $\tanh(\pi r)$ are disjoint and we conclude that the integral equals

$$\pi i\tanh(\pi i a) - \pi i\tanh(\pi i b) + 2i\sum_{n=0}^{\infty}\frac{1}{2}\left(\frac{1}{i(n+\frac{1}{2})+ia}+\frac{1}{i(n+\frac{1}{2})-ia}\right)-(\ldots),$$

where the dots indicate the same term for b instead of a. Plugging in the Mittag-Leffler series of tanh, one shows that the integral equals

$$2\sum_{n=0}^{\infty}\left(\frac{1}{a+\frac{1}{2}+n}-\frac{1}{b+\frac{1}{2}+n}\right).$$

From hyperbolic geometry (see [Bea95] Theorem 10.4.3) we take

Lemma 11.6.2 *The positive number* $\frac{\mathrm{vol}(\Gamma\backslash G)}{4\pi}$ *is an integer. More precisely, it is equal to* $g-1$, *where* $g \geq 2$ *is the genus of the compact Riemann surface* $\Gamma\backslash\mathbb{H}$.

After a change of variables $a \mapsto a+\frac{1}{2}$ and the same for b, comparing the two sides of the trace formula tells us that

$$\frac{Z'}{Z}\left(a+\frac{1}{2}\right)=\frac{a}{b}\frac{Z'}{Z}\left(b+\frac{1}{2}\right)+4a(1-g)\sum_{n=0}^{\infty}\left(\frac{1}{a+\frac{1}{2}+n}-\frac{1}{b+\frac{1}{2}+n}\right)$$
$$+\sum_{j=0}^{\infty}\frac{1}{a+ir_j}+\frac{1}{a-ir_j}-\frac{a}{b}\frac{1}{b+ir_j}+\frac{a}{b}\frac{1}{b-ir_j}.$$

Fixing an appropriate b, this extends to a meromorphic function on $\mathbb{C}$ with simple poles at $a=-n$ and $a=\frac{1}{2}\pm ir_j$. It follows that Z extends to an entire function on $\mathbb{C}$ and by a theorem from Complex Analysis (see [Rud87], Theorem 10.43) it follows that the poles of $\frac{Z'}{Z}$ are precisely the zeros of Z with multiplicity the respective residues. These are $2(2n+1)(g-1)$ for $a=-n$ and 1 in all other cases. □

We define the *Ruelle zeta function* of Γ as the infinite product

$$R(s)=\prod_{[\gamma]}(1-e^{-sl})^{\gamma}.$$

Corollary 11.6.3 *The product defining the Ruelle zeta function converges for* $\mathrm{Re}(s)>1$ *and the so defined Ruelle zeta function extends to a meromorphic function on* $\mathbb{C}$. *Its poles and zeros all lie in the union of* $\mathbb{R}$ *with the two vertical lines* $\mathrm{Re}(s)=\frac{1}{2}$ *and* $\mathrm{Re}(s)=-\frac{1}{2}$. *One has*

$$R(s)=\frac{Z(s)}{Z(s+1)}.$$

Proof The correlation between the Ruelle and the Selberg zeta function is immediate from the Euler product. The rest of the Corollary follows from this and Theorem 11.6.1. □

Note that, as $r_j \in i\left[-\frac{1}{2},\frac{1}{2}\right]\cup\mathbb{R}$, the Selberg zeta function satisfies a weak form of the Riemann hypothesis, as its zeros in the critical strip $\{0<\mathrm{Re}(s)<1\}$ are all in the set $\{\mathrm{Re}(s)=\frac{1}{2}\}$ with the possible exception of finitely many zeros in the interval $[0,1]$.

Note further, that one has a simple zero at $s = 1$ and no other poles or zeros in $\{\mathrm{Re}(s) \geq 1\}$. This information, together with the product expansion, suffices to use standard machinery from analytic number theory as in [Cha68] to derive the following theorem.

Theorem 11.6.4 (Prime Geodesic Theorem). *For $x > 0$ let $\pi(x)$ be the number of hyperbolic conjugacy classes $[\gamma]$ in Γ with $l(\gamma) \leq x$. Then, as $x \to \infty$,*

$$\pi(x) \sim \frac{e^{2x}}{2x}.$$

11.7 Exercises and Notes

Exercise 11.1 Show that $\int_{\mathbb{R}} \frac{e^{-u/2}\sin(ru)}{1+e^{-u}}\, du = \pi \tanh(\pi r)$.

(Hint: Write $\sin(ru) = \frac{1}{2i}(e^{iru} - e^{-iru})$ and thus decompose the integral into the sum of two integrals, each of which can be computed by the residue theorem.)

Exercise 11.2 Show that $g \in G = SL_2(\mathbb{R})$ is

$$\text{hyperbolic} \Leftrightarrow |\mathrm{tr}\,(g)| > 2,$$
$$\text{parabolic} \Leftrightarrow |\mathrm{tr}\,(g)| = 2,$$
$$\text{elliptic} \Leftrightarrow |\mathrm{tr}\,(g)| < 2.$$

Exercise 11.3 Show that a circle or a line in $\mathbb{C}$ is described by the equation $Az\bar{z} + Bz + \overline{Bz} + C = 0$, where $A, C \in \mathbb{R}$. Show that the linear fractional $z \mapsto \frac{az+b}{cz+d}$ for $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in GL_2(\mathbb{C})$ maps circles and lines to circles and lines.

Exercise 11.4 Let $A \in M_n(\mathbb{C})$. Show that

$$\det(\exp(A)) = \exp(\mathrm{tr}\,(A)).$$

Exercise 11.5 Let $a \in \mathbb{C}$ with $\mathrm{Re}(a) > 0$. Show that $\int_{\mathbb{R}} e^{-a|u|} e^{iru}\, dr = \frac{2a}{a^2+r^2}$.

Exercise 11.6 Let G be a locally compact group and K a compact subgroup. The *Hecke algebra* $\mathcal{H} = L^1(K\backslash G/K)$ is defined to be the space of all L^1-functions on G which are invariant under right and left translations from K. Show that $\mathcal{H}$ is an algebra under convolution. Show that for every $(\pi, V_\pi) \in \widehat{G}$ the space V_π^K of K-invariants is either zero, or an irreducible $\mathcal{H}$-module in the sense that it does not contain a closed $\mathcal{H}$-stable subspace.

Exercise 11.7 Continue the notation of the last exercise. The pair (G, K) is called a *Gelfand pair* if $\mathcal{H}$ is commutative. Show that if (G, K) is a Gelfand pair, then for every $(\pi, V_\pi) \in \widehat{G}$ the space V_π^K is at most one dimensional.

Exercise 11.8 Keep the notation of Exercise 11.6. Suppose that there is a continuous map $G \to G$, $x \mapsto x^c$ such that $(xy)^c = y^c x^c$ and $(x^c)^c = x$ as well as $x^c \in KxK$ for every $x \in G$. Show that G is unimodular and that (G, K) is a Gelfand pair.

(Hint: Let μ be the Haar measure on G and set $\mu^c(A) = \mu(A^c)$. Show that μ^c is a right Haar measure and that $\int_G f(x)\,d\mu(x) = \int_G f(x)\,d\mu^c(x)$ holds for every $f \in \mathcal{H}$. Consider the equation $\int_G f(xy)\,d\mu(x) = \Delta(y^{-1}) \int_G f(x)\,d\mu(x)$ for $f \in \mathcal{H}$ and make the integrand on the right hand side K-bi-invariant.)

Exercise 11.9 Let $f \in \mathcal{H}$ with $\operatorname{tr}(\pi_{ir}(f)\pi_{ir}(x)) = 0$ for every $x \in G$ and every $r \in \mathbb{R}$. Show that $f \equiv 0$.

Exercise 11.10 Let Δ denote the hyperbolic Laplace operator. Show that the function $z \mapsto \operatorname{Im}(z)^s$ for $s \in \mathbb{C}$ is an eigenfunction of Δ of eigenvalue $s(1-s)$.

Exercise 11.11 Read and understand the proof of the prime number theorem in [Cha68]. Apply the same methods to give a proof of Theorem 11.6.4.

Notes

The Selberg zeta function has been introduced in Selberg's original paper on the trace formula [Sel56]. It has fascinated mathematicians from the beginning as its relation to the trace formula is similar to the relation of the Riemann zeta function to the Poisson summation formula and, as we have seen, a weak form of the Riemann hypothesis can be proved for the Selberg zeta function. However, Selberg's zeta continues to live in a world separate from Riemann's, and although many tried, no one has found a bridge between these worlds yet.

The name *Prime Geodesic Theorem* for Theorem 11.6.4 is derived from the following geometric facts. On the upper half plane $\mathbb{H}$ there is a Riemannian metric given by $\frac{dx^2+dy^2}{y^2}$, which is left stable by the action of the group G, in other words, G acts by isometries. If $\Gamma \subset G$ is a torsion-free discrete cocompact subgroup, the quotient $\Gamma\backslash\mathbb{H}$ will inherit the metric and thus become a Riemannian manifold, the projection $\mathbb{H} \to \Gamma\backslash\mathbb{H}$ is a covering. A closed geodesic c in $\Gamma\backslash\mathbb{H}$ is covered by geodesics of infinite lengths in $\mathbb{H}$ and any such geodesic is being closed by an element $\gamma \in \Gamma$, which is uniquely determined up to conjugacy. The map $c \mapsto \gamma$ sets up a bijection between closed geodesics and primitive conjugacy classes in Γ. The number $l(\gamma)$ is just the length of the geodesic c. So indeed, Theorem 11.6.4 gives an asymptotic of lengths of closed geodesics. This theorem has been generalized several times, the most general version being a theorem of Margulis [KH95], which gives a similar asymptotic for compact manifolds of strictly negative curvature.