Spectral Theory and the Trace Formula (Expanded Text)

BY DANIEL BUMP

We give an account of a portion of the spectral theory $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$, particularly the Selberg trace formula, emphasizing ideas from representation theory. The last section is of a different nature, intended to show a simple application of the trace formula to a lifting problem.

I would like to thank Yonatan Gutman and the referee for extensive and very helpful comments on an earlier draft. Preparation of this report was supported in part by NSF grant DMS-9970841.

1 The spectral problem

The group $G = \mathrm{SL}_2(\mathbb{R})$ acts on $\mathfrak{H} = \{x + i\,y | y > 0\}$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} : z \mapsto \frac{a\,z + b}{c\,z + d}. \tag{1}$$

The stabilizer of i is $K = \mathrm{SO}(2)$ so $\mathfrak{H} \cong G/K$. The *noneuclidean Laplacian*

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

is a G-invariant differential operator. Let Γ be a discrete cocompact subgroup of $G = \mathrm{SL}_2(\mathbb{R})$. Then $X = \Gamma\backslash\mathfrak{H}$ is a compact Riemann surface. The question at hand is to describe the spectrum of Δ on X.

We may sometimes consider the upper half-plane $\mathfrak{H}$ to be embedded in the Riemann sphere $\mathfrak{R} = \mathbb{C} \cup \{\infty\}$. The *boundary* of $\mathfrak{H}$ is then $\mathbb{R} \cup \{\infty\}$. A *geodesic* in $\mathcal{H}$ for the noneuclidean hyperbolic geometry on $\mathfrak{H}$ is a circle perpendicular to the boundary at the two points of intersection. As a special case, a vertical line is a geodesic.

Proposition 1. *The group Γ has a fundamental domain $\mathcal{F}$ on $\mathfrak{H}$ whose boundary consists of pairs of geodesic arcs α_i and $\gamma_i(\alpha_i)$, with $\gamma_i \in \Gamma$. When the boundary is traversed counterclockwise the congruent arcs α_i and $\gamma_i(\alpha_i)$ are traversed in opposite directions.*

Proof. (Sketch) Choose a point $P \in \mathfrak{H}$ which is not fixed by any element of Γ except $\pm\, I$. Let $\mathcal{F}$ be the set of points which are nearer P than to $\gamma(P)$ in the noneuclidean metric for any $\gamma \in \Gamma$. Let $N = \{\gamma_1, \cdots, \gamma_n\}$ be the set elements of Γ such that $\gamma_i(\mathcal{F})$ is adjacent to $\mathcal{F}$.

At first we assume no $\gamma_i^2 = 1$. Evidently each $\gamma_i^{-1} \in N$, so N has an even number $2h$ of elements. We arrange it so that the γ_i so that $\gamma_{i+h} = \gamma_i^{-1}$ $(1 \leqslant i \leqslant h)$. Let α_i be the intersection of the geodesic consisting of the set of points which are equidistant from P and $\gamma_i^{-1}(P)$ with the closure of $\mathcal{F}$. Then $\gamma_i(\alpha_i) = \alpha_{i+h}$ and $\alpha_1, \cdots, \alpha_h$ satisfy our requirements.

If some $\gamma_i^2 = 1$ then the arc α_i is self-congruent by γ_i and its midpoint is fixed by γ_i. So we split such arcs in two at their midpoints and we are done. □

The Laplacian Δ acts on $C^\infty(\Gamma\backslash\mathfrak{H})$.

Proposition 2. *Δ is symmetric and positive with respect to the invariant metric $y^{-2} d\,x \wedge d\,y$.*

Proof. Symmetry means that

$$\langle \Delta f, g\rangle - \langle f, \Delta g\rangle = 0, \qquad f, g \in C^\infty(\Gamma\backslash\mathfrak{H}).$$

Taking a fundamental domain $\mathcal{F}$ as in Proposition 1, this equals

$$\int_{\mathcal{F}} \left(f\left(\frac{\partial^2 \bar{g}}{\partial x^2} + \frac{\partial^2 \bar{g}}{\partial y^2}\right) - \bar{g}\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right)\right) d\,x \wedge d\,y = \int_{\mathcal{F}} d\omega$$

where

$$\omega = f\frac{\partial \bar{g}}{\partial x} d\,y - f\frac{\partial \bar{g}}{\partial y} d\,x - \bar{g}\frac{\partial f}{\partial x} d\,y + \bar{g}\frac{\partial f}{\partial y} d\,x.$$

By Stoke's theorem, this is equal to the integral of ω around the boundary of $\mathcal{F}$. By Proposition 1, the contributions of the boundary arcs cancel in pairs.

Positivity means that $\langle \Delta f, f\rangle \geqslant 0$ with equality only if f is constant. We compute

$$\left[\left|\frac{\partial f}{\partial x}\right|^2 + \left|\frac{\partial f}{\partial y}\right|^2 + \left(\frac{\partial f^2}{\partial x^2} + \frac{\partial f^2}{\partial y^2}\right)\bar{f}\right] d\,x \wedge d\,y = d\left(\bar{f}\left(\frac{\partial f}{\partial x} d\,y - \frac{\partial f}{\partial y} d\,x\right)\right).$$

By Stoke's theorem we thus have

$$\langle \Delta f, f\rangle \geqslant -\int_{\partial\mathcal{F}} \bar{f}\left(\frac{\partial f}{\partial x} d\,y - \frac{\partial f}{\partial y} d\,x\right), \tag{3}$$

with equality only if $\partial f/\partial x = \partial f/\partial y = 0$ identically, that is, if f is constant. Using the Cauchy-Riemann equations and the chain rule if one may check that if $x + i\,y$ and $u + i\,v$ are related by a holomorphic mapping, such as a linear fractional transformation then $(\partial f/\partial x) d\,y - (\partial f/\partial y) d\,x = (\partial f/\partial u) d\,v - (\partial f/\partial v) d\,u$. It follows that the contributions of congruent boundary arcs cancel and (3) is zero. □

We will see later that Δ extends to a self-adjoint unbounded operator on $\mathfrak{L} = L^2(\Gamma\backslash\mathfrak{H}, y^{-2} d\,x \wedge d\,y)$. This means that there is a dense subspace D of $\mathfrak{L}$ containing $C^\infty(\Gamma\backslash\mathfrak{H})$ such that Δ extends to D, and that if $v \in \mathfrak{L}$ is such that $u \to \langle v, \Delta u\rangle$ is continuous on D, then $v \in D$ and $\langle \Delta v, u\rangle = \langle v, \Delta u\rangle$.

Since Δ is a self-adjoint operator it has a nice spectral theory, which we want to develop. We will accomplish this by introducing integral operators which commute with Δ. We will show that these operators are of trace class, and we will prove the Selberg trace formula for them.

Finally we will consider the more difficult case where Γ has a cusp. In this case there are both discrete and continuous spectra, and the theory of Eisenstein series is an essential feature.

2 Rings of integral operators

The group G acts on functions $f: G \to \mathbb{C}$ by the *right regular action* $(\rho(g) f)(x) = f(x g)$. Since we will soon be discussing differential operators let us at first restrict ourselves to $f \in C^\infty(G)$; however f can be any locally integrable function in this notation.

Because $\mathfrak{H}$ is canonically identified with the homogeneous space G/K, any function f on $\mathfrak{H}$ may be regarded as a function on G which is right invariant by K. The relevance of the regular representation of G to functions on $\mathfrak{H}$ may not be immediately clear, because the property of right K invariance is not preserved under right translation. However, there *is* a relevance which we now explain.

The Lie algebra of G consists of the vector space $\mathfrak{g}$ of 2×2 real matrices of trace zero. It acts on smooth functions f on G by:

$$X f(g) = \frac{d}{d t} f(g\, e^{t X})\Big|_{t=0}, \qquad X \in \mathfrak{g}.$$

This is a Lie algebra representation. This means that $[X, Y] = X\,Y - Y\,X$ ($X\,Y =$ matrix multiplication) has the same effect as $X \circ Y - Y \circ X$ ($X \circ Y =$ composition of operators). This representation of $\mathfrak{g}$ is the differential form of the regular representation.

The universal enveloping algebra $U(\mathfrak{g})$ is the associative $\mathbb{R}$-algebra generated by $\mathfrak{g}$ modulo the relations

$$[X, Y] - (X \cdot Y - Y \cdot X) = 0$$

($X \cdot Y =$ multiplication in $U(\mathfrak{g})$). Any Lie algebra representation extends uniquely to a representation of $U(\mathfrak{g})$. In particular the regular representation of $\mathfrak{g}$ extends to a representation of $U(\mathfrak{g})$. Thus $U(\mathfrak{g})$ is realized as a ring of differential operators on $C^\infty(G)$.

If D is an element of the *center* of $U(\mathfrak{g})$ then it commutes with the regular representation. This is intuitively reasonable and proved in Bump [5], Proposition 2.2.4. In particular, if f is fixed under $\rho(K)$, then so is $D f$. Therefore D acts on $C^\infty(\mathfrak{H})$.

A particular element of the center of $U(\mathfrak{g})$ is the *Laplace-Beltrami* or *Casimir* element

$$-4\Delta = H \cdot H + 2R \cdot L + 2L \cdot R, \tag{4}$$

$$R = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad L = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

In fact, $\mathbb{C}[\Delta]$ is the center of $U(\mathfrak{g})$. We have used the same letter Δ that we previously used for the noneuclidean Laplacian, for when this element of $U(\mathfrak{g})$ is interpreted as a differential operator on $\mathfrak{H}$ they are the same. See Bump [5], Proposition 2.2.5.

Let $\mathcal{H}$ be the convolution ring of smooth, compactly supported functions on G. Let $\mathcal{H}^\circ$ be the subring of K-bi-invariant functions. These are rings without unit. We call $\mathcal{H}^\circ$ the *spherical Hecke algebra* but caution the reader that there are other natural and closely related rings which have also been called this. The ring $\mathcal{H}$ is noncommutative, but:

Theorem 3. *The ring $\mathcal{H}^\circ$ is commutative.*

Proof. Matrix transposition preserves K so it induces an involution ι on $\mathcal{H}^\circ$ such that $\iota(\phi * \psi) = \iota(\psi) \circ \iota(\phi)$, where $(\iota f)(g) = f(g^t)$. Every double coset in $K\backslash G/K$ has a diagonal representative. So ι is the identity map. □

This theorem of Gelfand has a representation theoretic meaning. If (π, V) is a representation of G on a Banach space we will denote by V^K the vector subspace of K-fixed vectors. The algebra $\mathcal{H}$ acts on V by

$$\pi(\phi)v = \int_G \phi(g)\pi(g)v\,d\,g, \qquad \phi \in \mathcal{H}, v \in V.$$

Here $\int_G d\,g$ is the Haar integral. If $\phi \in \mathcal{H}^\circ$ then $\pi(\phi)v \in V^K$. In particular V^K is a module for $\mathcal{H}^\circ$.

Since K is compact, any irreducible representation ρ of K is finite-dimensional. The *ρ-isotypic part* $V(\rho)$ of V is the direct sum of all K-invariant subspaces isomorphic to ρ as K-modules. A representation (π, V) of G is called *admissible* if $V(\rho)$ is finite-dimensional for every ρ. In particular, if ρ is the trivial representation $V(\rho) = V^K$ so V^K must be finite-dimensional.

Theorem 4. *If (π, V) is an irreducible admissible representation of G then V^K is at most one-dimensional.*

Proof. Since $\mathcal{H}^\circ$ is commutative, its finite-dimensional irreducible modules are one-dimensional. Thus it is sufficient to show that V^K (if nonzero) is an irreducible module for $\mathcal{H}^\circ$. Suppose $L \subset V^K$ is a closed nonzero $\mathcal{H}^\circ$-invariant subspace. If $v \in V^K$ we will show that $v \in L$.

Let $\epsilon > 0$ be given. Since V is irreducible and L is nonzero, the closure of $\pi(\mathcal{H})L$ is V and so there exists $\phi \in \mathcal{H}$, $w \in L$ such that $\pi(\phi)w = v_1$, where $|v_1 - v| < \epsilon$. Let $v_2 = \int_K \pi(k)v_1 d\,k$. Then v_2 is K-fixed and since v is K-fixed,

$$|v_2 - v| = \left| \int_K \pi(k)(v_1 - v)\,d\,k \right| \leqslant \int_K |\pi(k)(v_1 - v)|\,d\,k = \int_K |v_1 - v|\,d\,k < \epsilon.$$

(We normalize the Haar measure so K has volume 1.) Since w is K-fixed,

$$\pi(k)\pi(\phi)\pi(k')w = \pi(k)v_1$$

for all $k, k' \in K$. Integrating over k and k', we thus get $\pi(\phi_0)\,w = v_2$ where

$$\phi_0(g) = \int_{K\times K} \phi(k\,g\,k')\,d\,k\,d\,k'.$$

Now $\phi_0 \in \mathcal{H}^\circ$ and since L is $\mathcal{H}^\circ$-invariant this implies that $v_2 \in L$. We can therefore approximate v arbitrarily closely by elements of L, and since $L \subset V^K$ is finite-dimensional, hence closed, this implies that $v \in L$. □

More generally, if (π, V) is a representation of G and $k \in \mathbb{Z}$ we will denote by $V(k)$ the subspace of $v \in K$ satisfying

$$\pi(\kappa_\theta)\,v = e^{ik\theta}v, \qquad \kappa_\theta = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix} \in K.$$

(This is the ρ-isotypic subspace where ρ is the character $e^{ik\theta}$ of K.) If (π, V) is irreducible, then since $-I \in K$ is central, it acts by a scalar $(-1)^\epsilon$ where $\epsilon = 0$ or 1. Evidently the parity of k must be the same as ϵ for $V(k)$ to be nonzero. We will call π *even* if $\epsilon = 0$, and *odd* if $\epsilon = 1$.

Proposition 5. *If (π, V) is an irreducible admissible representation of G then $V(k)$ is at most one-dimensional.*

Proof. (Sketch) This may be proved along the same lines as Theorem 4. Because $K = \mathrm{SO}(2)$ is commutative, it may be seen that the convolution ring of smooth, compactly supported functions which satisfy $f(\kappa_\theta g \kappa_{\theta'}) = e^{ik(\theta+\theta')} f(g)$ is commutative. See Bump [5], Proposition 2.2.8. □

We note that while Theorem 4 generalizes directly to arbitrary reductive Lie groups, Proposition 5 does not. Thus if (π, V) is an irreducible admissible representation of a reductive Lie group, the multiplicity of the trivial representation of its maximal compact subgroup is at most one; the other irreducible representations of the maximal compact each occur with finite multiplicity (this is admissibility), but not necessarily multiplicity one. Actually irreducible representations are automatically admissible, though this fact is not needed in the theory of automorphic forms, where admissibility of automorphic representations can be proved directly.

If (π, V) is an irreducible representation of G, we will denote by V_{fin} the direct sum of the $V(k)$. This is the space of *K-finite* vectors. It is not invariant under the action of G, but it is invariant under the actions of the Lie algebra $\mathfrak{g}$ of G and of K. It is therefore a *$(\mathfrak{g}, K)$-module*. See Bump [5], p. 200 for a discussion of this concept.

When $\pi = \rho$ is the right regular representation, we have

$$\rho(\phi) f(x) = \int_G \phi(g) f(x g) \, d g.$$

This formula make sense for any locally integrable function f on G. The operator $\rho(\phi)$ is convolution with the function $g \to \phi(g^{-1})$. These integral operators are important for us because they commute with Δ, yet they are easier to study than Δ. We note that $L^2(\Gamma \backslash G)$ is invariant under $\rho(\phi)$, and if $\phi \in \mathcal{H}^\circ$ then $\rho(\phi)$ preserves the property of right invariance by K, so it can be regarded as an integral operator on $L^2(\Gamma \backslash \mathfrak{H})$.

Let $\phi \in \mathcal{H}$, and let

$$K_\phi(x, y) = \sum_{\gamma \in \Gamma} \phi(x^{-1} \gamma y). \tag{5}$$

At first we regard (x, y) as an element of $G \times G$. If x and y are restricted to a compact set C, then $\phi(x^{-1} g\, y)$ vanishes for g off a compact set C'. Therefore only finitely many γ contribute. It follows that (5) is convergent and defines a smooth function of x and y.

A change of variables shows that $K_\phi(x, y)$ is invariant if either x or y is changed on the left by an element of γ, so we may regard this kernel as defined on either $G \times G$ or on $\Gamma \backslash G \times \Gamma \backslash G$, and it is a continuous function. If $\phi \in \mathcal{H}^\circ$ then K_ϕ may even be regarded as a function on $X \times X$. (Recall that $X = \Gamma \backslash \mathfrak{H}$.)

Proposition 6. *We have*

$$(\rho(\phi) f)(x) = \int_{\Gamma\backslash G} K_\phi(x, y) f(y)\, d y. \tag{6}$$

Proof. The left side equals

$$\int_G \phi(y) f(x y)\, d y = \int_G \phi(x^{-1} y) f(y)\, d y = \sum_{\gamma \in \Gamma} \int_{\Gamma\backslash G} \phi(x^{-1} \gamma y) f(y)\, d y,$$

where we have used $f(\gamma y) = f(y)$. Interchanging sum and integral gives (6). □

If H is a Hilbert space, an operator $T: H \to H$ is *compact* if T maps bounded sets into compact sets.

Theorem 7. *$\rho(\phi)$ is a compact operator on $L^2(\Gamma\backslash G)$.*

See Bump [5], Section II.3, particularly Theorem 2.3.2 and Proposition 2.3.1 for fuller details.

Proof. The kernel K_ϕ is continuous on the compact space $(\Gamma\backslash G) \times (\Gamma\backslash G)$, so it is certainly in $L^2((\Gamma\backslash G) \times (\Gamma\backslash G))$. The well-known theorem of Hilbert and Schmidt asserts that if Z is any locally compact Borel measure space such that $L^2(Z)$ is a separable Hilbert space then integral operator

$$(T f)(x) = \int_Z K(x, y) f(y)\, d y$$

with the kernel $K \in L^2(Z \times Z)$ is compact. □

If

$$\phi(g^{-1}) = \overline{\phi(g)} \tag{7}$$

then $K_\phi(x, y) = \overline{K_\phi(y, x)}$, so $\rho(\phi)$ is self-adjoint.

Theorem 8. *Let T be a compact self-adjoint operator on a separable Hilbert space H. Then H has an orthonormal basis ϕ_i ($i = 1, 2, 3, \cdots$) of eigenvectors of T, so that $T\phi_i = \mu_i \phi_i$. The eigenvalues $\mu_i \to 0$ as $i \to \infty$.*

Proof. This is the Spectral Theorem for compact operators. See Bump [5], Theorem 2.3.1. □

Thus if (7) is true then $\rho(\phi)$ is a self-adjoint compact operator whose nonzero eigenvalues $\mu_i \to 0$. The Hilbert-Schmidt property implies more: $\sum |\mu_i|^2 < \infty$. Later we will see that more is true: $\sum |\mu_i| < \infty$. This means that $\rho(\phi)$ is *trace class*. This fact is important because of the Selberg trace formula.

Theorem 9. *$L^2(X)$ has a basis consisting of eigenvectors of Δ.*

Proof. The operators $\rho(\phi)$ with ϕ satisfying (7) are a commuting family of self-adjoint compact operators so they can be simultaneously diagonalized. By the spectral theorem the nonzero eigenspaces are finite-dimensional; there is no nonzero vector on which the operators $\rho(\phi)$ are all zero, since ϕ can be chosen to be positive, of mass one and concentrated near the identity in which case $\rho(\phi)f$ approximates f. Therefore the simultaneous eigenspaces of $\mathcal{H}^\circ$ are finite-dimensional.

Let V be such an eigenspace. Since Δ commutes with the $\rho(\phi)$, it preserves V, and since it is symmetric it induces a self-adjoint transformation on V. Choose an orthonormal basis for each such V consisting of eigenvectors of Δ and put these together for all V. □

Closely related to Theorem 9 is a representation-theoretic statement about $L^2(\Gamma\backslash G)$. The regular representation ρ is a unitary representation on this space.

Lemma 10. *Let H be a closed nonzero G-invariant subspace of $L^2(\Gamma\backslash G)$. Then H contains an irreducible subspace.*

Proof. (Langlands) Since H is G-invariant, each $\rho(\phi)$ induces an endomorphism of H. We show first that the restriction of $\rho(\phi)$ to H is nonzero for suitable ϕ satisfying (7). If $0 \neq \xi \in H$ then for g near the identity $\rho(g)\xi$ is near ξ. Thus if we take ϕ satisfying (7) such that $\phi > 0$, $\int_G \phi(g)\,d\,g = 1$ and such that the support of ϕ is nonzero then $\rho(\phi)\xi$ is near ξ so $\rho(\phi)$ is nonzero on H, and $\rho(\phi)$ is self-adjoint.

Let $L \subset H$ be the eigenspace of a nonzero eigenvalue of $\rho(\phi)$. It is finite-dimensional by Theorem 8. Let L_0 be a nonzero subspace minimal with respect to property of being the intersection of L with a nonzero closed invariant subspace of H. Let V be the smallest closed invariant subspace of H such that $L \cap V = L_0$. We show V is irreducible. If not, let V_1 be a proper, nonzero closed invariant subspace and let V_2 be its orthogonal complement, so $V = V_1 \oplus V_2$. Let $0 \neq f \in L_0$. Write $f = f_1 + f_2$ with $f_i \in V_i$. Since $0 = \rho(\phi)f - \lambda f = (\rho(\phi)f_1 - \lambda f_1) + (\rho(\phi)f_2 - \lambda f_2)$ and $\rho(\phi)f_i - \lambda f_i \in V_i$ we have $\rho(\phi)f_i - \lambda f_i = 0$. Thus $f_i \in L \cap V_i$. By the minimality of L_0, $L_0 = L \cap V_i$ for some i, say $L_0 = L \cap V_1$. Now the minimality of V is contradicted. □

Theorem 11. *$L^2(\Gamma\backslash G)$ decomposes as a direct sum of closed, irreducible subspaces. Each affords an irreducible admissible representation of G.*

Proof. By Zorn's Lemma, let S be a maximal set of orthogonal closed irreducible subspaces. Let $H = \bigoplus_{V \in S} V$. If H is proper, applying Lemma 10 to its orthogonal complement contradicts the maximality of S. We leave admissibility to the reader. □

Each of these closed irreducible subspaces has at most one K-fixed vector by Theorem 4.

Theorem 11 may be thought of as a more satisfactory extension of Theorem 9. Indeed, if ϕ is an eigenfunction of Δ occurring in $L^2(\Gamma\backslash\mathfrak{H})$ then its right translates span an irreducible subspace of $L^2(\Gamma\backslash G)$. Conversely, Δ acts by a scalar on each irreducible subspace. If that subspace happens to have a K-fixed vector in it, that vector will be one of the basis elements in Theorem 9.

There will, however, be some irreducible subspaces of $L^2(\Gamma\backslash G)$ which have no K-fixed vectors. These can be constructed from *holomorphic modular forms* as follows. Let $f\colon\Gamma\backslash\mathfrak{H}\to\mathbb{C}$ be a holomorphic function satisfying

$$f\left(\frac{a\,z+b}{c\,z+d}\right)=(c\,z+d)^k\,f(z),\qquad\left(\begin{array}{cc}a&b\\c&d\end{array}\right)\in\Gamma.$$

If k is sufficiently large, such f will always exist, as may be shown from the Riemann-Roch theorem. Regarded as a function on G, the function f is right invariant by K, but it is not left invariant by Γ, so it is not a function on $\Gamma\backslash G$. However we may modify it, sacrificing the right invariance by K to obtain true left invariance by Γ. Define

$$F\left(\begin{array}{cc}a&b\\c&d\end{array}\right)=(c\,i+d)^{-k}\,f\left(\frac{a\,i+b}{c\,i+d}\right).$$

Then $F(\gamma g)=F(g)$ for $\gamma\in\Gamma$, while $F(g\,\kappa_\theta)=e^{ik\theta}F(g)$. The irreducible representation spanned by F has not K-fixed vector. This is the weight k *holomorphic discrete series* representation.

The representation-theoretic approach has another generalization, its extension to automorphic forms on adele groups. Assume that $\Gamma=\mathrm{SL}_2(\mathbb{Z})$. Let $\mathbb{A}$ be the adele ring of $\mathbb{Q}$. The inclusion of $\mathrm{SL}_2(\mathbb{R})\to\mathrm{GL}_2(\mathbb{A})$ at the infinite place induces a homeomorphism

$$\Gamma\backslash G\to\mathrm{GL}_2(\mathbb{Q})Z_{\mathbb{A}}\backslash\mathrm{GL}_2(\mathbb{A})/\textstyle\prod_p K_p,$$

where $Z_{\mathbb{A}}$ is the center of $\mathrm{GL}_2(\mathbb{A})$ and $K_p=\mathrm{GL}_2(\mathbb{Z}_p)$ is a maximal compact subgroup of $\mathrm{GL}_2(\mathbb{Q}_p)$. Thus functions on $\Gamma\backslash G$ may be reinterpreted as functions on $\mathrm{GL}_2(\mathbb{Q})Z_{\mathbb{A}}\backslash\mathrm{GL}_2(\mathbb{A})$, and in particular we may embed

$$L^2(\Gamma\backslash G)\to L^2(\mathrm{GL}_2(\mathbb{Q})Z_{\mathbb{A}}\backslash\mathrm{GL}_2(\mathbb{A})).$$

Now the study of $L^2(\mathrm{GL}_2(\mathbb{Q})Z_{\mathbb{A}}\backslash\mathrm{GL}_2(\mathbb{A}))$ may be carried out along exactly the same lines as we've applied above. The class of integral operators is larger now, however. In addition to the ring $\mathcal{H}$, we have its p-adic analogs, which are rings of *Hecke operators*. To see this, fix a prime p. Let $\mathcal{H}_p$ be the convolution ring of smooth (i.e. locally constant) compactly supported functions on $\mathrm{GL}_2(\mathbb{Q}_p)$. This ring is not commutative, but the subring $\mathcal{H}_p^\circ$ of functions which are K_p bi-invariant is commutative. (Compare Theorem 3.) For example, the characteristic function of $K_p\left(\begin{array}{cc}p&0\\0&1\end{array}\right)K_p$ is an element of this ring, and this is the adelization of classical Hecke operator T_p which picks off the p-th Fourier coefficient of a Hecke eigenform. We will return to this point of view in the final section when we study such operators using the trace formula.

3 Green's functions and the spectral resolvent

References for this section are Hejhal [19], Section 6 and Bump [5], Chapter 2 Section 3. We are interested in functions f on $\mathfrak{H}$ with the following

Property S *The function f is left invariant by $K = \mathrm{SO}(2)$, possibly singular at K's fixed point at i, and is an eigenfunction of the Laplacian.*

We will see that for each eigenvalue λ of Δ there are two linearly independent such functions, one of which (the *Green's function*) has nice behavior at the boundary, the other of which (the *spherical function*) is continuous at i. Each has its uses. Since $\mathcal{H} \cong G/K$, these may be regarded as functions on $K\backslash G/K$ (possibly undefined on K) which are eigenfunctions of the Laplace Beltrami operator.

We can map the upper half plane into the unit disk $\mathfrak{D}$ by the Cayley transform $z \mapsto w = (z - i)/(z + i)$. Let $r = |w|$. Then since f as in Property S is left invariant by K, $f(z)$ depends only on r. Denote $f(z) = W(r)$. Thus a function f with Property S is determined by the function W on $(0,1)$ such that

$$W(r) = f\begin{pmatrix} y^{1/2} & \\ & y^{-1/2} \end{pmatrix},$$

where $r = (y - 1)/(y + 1) \in (0, 1)$. The eigenvalue property amounts to the differential equation

$$W''(r) + \frac{1}{r}W'(r) + \frac{4\lambda}{(1 - r^2)^2}W(r) = 0. \tag{8}$$

This differential equation has regular singular points at $(0,1)$ and there are two solutions of interest. One is nicely behaved at 0, the other at 1. (We assume familiarity with regular singular points of second order linear differential equations, particularly the indicial equation, for which see Whittaker and Watson [45], Section 10.3.) In this section we will be concerned with the solution which has nice behavior on the boundary, that is, at $r = 1$; the other one will occupy the next section.

Let λ be a negative real number. At $r = 1$ the roots of the indicial equation of (8) for the singularity at $r = 1$ are $\frac{1}{2}(1 \pm \sqrt{1 - 4\lambda})$. With $\lambda < 0$ exactly one root $\alpha = \frac{1}{2}(1 + \sqrt{1 - 4\lambda})$ is positive, so there is a unique (up to multiple) solution g_λ to (8) which vanishes near the boundary. We can use it to study the resolvent of the Laplacian.

Lemma 12. *If λ is a negative real number, then $g_\lambda(r)$ has a logarithmic singularity at $r = 0$. Near $r = 1$ we have $g_\lambda(r) \sim c(1 - r)^\alpha$ where $\alpha = \frac{1}{2}(1 + \sqrt{1 - 4\lambda}) > 1$, and c is a nonzero constant.*

Proof. The roots of the indicial equation at $r = 0$ are 0 with multiplicity 2, so one solution has a logarithmic singularity, another is analytic. If g_λ does not have the logarithmic singularity, then $g_\lambda(r)$ is real and analytic on $[-1, 1]$ hence has a maximum or minimum. At such a point $g_\lambda'(r) = 0$ and since $\lambda < 0$, equation (8) implies that g_λ and g_λ'' have the same sign, impossible at the maximum or minimum because $g_\lambda(-1) = g_\lambda(1) = 0$. This proves the existence of the logarithmic singularity at $r = 0$. The behavior at $r = 1$ is clear from the definition of g_λ. $\square$

Let

$$g_\lambda(z, \zeta) = g_\lambda\left(\left|\frac{z - \zeta}{z - \bar{\zeta}}\right|\right).$$

$z, \zeta \in \mathfrak{H}$. This is a *Green's function.*

Theorem 13. *Let λ be a negative real number. Then*

$$\left[-y^2\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)-\lambda\right]g_\lambda(z,\zeta)=0;$$

$$g_\lambda(z,\zeta) \text{ is singular on the diagonal } z=\zeta;$$

$$g_\lambda(z,\zeta)\to 0 \text{ as } y\to 0;$$

$$g_\lambda(z,\zeta)=g_\lambda(\zeta,z);$$

$$g_\lambda(h(z),h(\zeta))=g_\lambda(z,\zeta), \qquad h\in \mathrm{SL}_2(\mathbb{R}). \tag{9}$$

Proof. The first property boils down to (8), the second property comes from Lemma 12, the third follows from the boundary behavior of g_λ, and the last two properties follow since $|(z-\zeta)/(z-\bar{\zeta})|$ is unchanged if z and ζ are interchanged, or if an element of $\mathrm{SL}_2(\mathbb{R})$ is applied to both. See Bump [5], Proposition 2.3.4 on p. 181 for fuller details. □

Since g_λ has a logarithmic singularity at 0 it can be normalized so $g_\lambda(r)-\frac{1}{2\pi}\log(r)$ is bounded as $r\to 0$. It follows that $g_\lambda'(r)-\frac{1}{2\pi r}$ is analytic near $r=0$.

Theorem 14. *Let λ be a negative real number. If $f\in C_c^\infty(\mathfrak{H})$ then (writing $\zeta=\xi+i\eta$)*

$$\int_{\mathfrak{H}} g_\lambda(z,\zeta)\left[-\eta^2\left(\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\eta^2}\right)-\lambda\right]f(\zeta)\frac{d\xi\wedge d\eta}{\eta^2}=f(z) \qquad (\zeta=\xi+i\eta).$$

Proof. See Bump [5], Proposition 2.3.4 on p. 181 for fuller details. We review the proof quickly. Let $w=(z-\zeta)/(z-\bar{\zeta})=u+i\,v\in\mathfrak{D}$. Let $F:\mathfrak{D}\to\mathbb{C}$ be defined by $F(w)=f(z)$. In the w coordinates we must prove

$$\int_{\mathfrak{D}} g_\lambda(|w|)\left[-\left(\frac{\partial^2}{\partial u^2}+\frac{\partial^2}{\partial v^2}\right)-\frac{4\lambda}{(1-|w|^2)^2}\right]F(w)\,d\,u\wedge d\,v=F(0).$$

Let B_r be the disk of radius r, and let $R<1$ be large enough that the support of F is contained in B_R. The left side equals

$$\lim_{\epsilon\to 0}\int_{B_R-B_\epsilon} g_\lambda(|w|)\left[-\left(\frac{\partial^2}{\partial u^2}+\frac{\partial^2}{\partial v^2}\right)-\frac{4\lambda}{(1-|w|^2)^2}\right]F(w)\,d\,u\wedge d\,v.$$

Using Stoke's theorem as in Proposition 2, this is

$$\lim_{\epsilon\to 0}\int_{B_R-B_\epsilon} F(w)\left[-\left(\frac{\partial^2}{\partial u^2}+\frac{\partial^2}{\partial v^2}\right)-\frac{4\lambda}{(1-|w|^2)^2}\right]g_\lambda(|w|)\,d\,u\wedge d\,v \;+$$
$$\lim_{\epsilon\to 0}\int_{C_\epsilon} F(w)\left(\frac{\partial g_\lambda(|w|)}{\partial u}d\,v-\frac{\partial g_\lambda(|w|)}{\partial v}d\,u\right) \;-$$
$$\lim_{\epsilon\to 0}\int_{C_\epsilon} g_\lambda(|w|)\left(\frac{\partial F(w)}{\partial u}d\,v-\frac{\partial F(w)}{\partial v}d\,u\right),$$

where C_ϵ is the path circling the origin counterclockwise around the circle with radius ϵ. (There would also be terms integrating around C_R, but these are zero because they lie outside the support of F.)

The first term vanishes by Theorem 13. The last term vanishes because the length of the arc shrinks faster than g_λ blows up (logarithmically).

To evaluate the middle term let $w = r\,e^{i\theta}$. By the chain rule,

$$\frac{\partial g_\lambda(|w|)}{\partial u} d\,v - \frac{\partial g_\lambda(|w|)}{\partial v} d\,u = r\,g_\lambda'(r)\,d\theta.$$

We obtain

$$\lim_{\epsilon \to 0} \int_0^{2\pi} F(\epsilon e^{i\theta})\,d\theta\,\epsilon\,g_\lambda'(\epsilon) = F(0),$$

since $g_\lambda'(\epsilon) \sim 1/(2\pi\epsilon)$ as $\epsilon \to 0$. □

Proposition 15. *Let λ be a negative real number. The series*

$$G_\lambda(z,\zeta) = \sum_{\gamma \in \{\pm 1\}\backslash\Gamma} g_\lambda(z,\gamma(\zeta)) = \sum_{\gamma \in \{\pm 1\}\backslash\Gamma} g_\lambda(\gamma(z),\zeta).$$

is absolutely convergent provided z and ζ are not Γ-equivalent.

Proof. Let B_r be a ball of radius r with center at the origin. If $r < 1$ then $B_r \subset \mathfrak{D}$. With ζ fixed we use the Cayley transform $z \longmapsto \mathcal{C}_\zeta(z) = (z-\zeta)/(z-\bar{\zeta})$ to transfer g_λ to the unit disk. The volume of B_r in the hyperbolic metric on $\mathfrak{D}$ is $4\pi r^2/(1-r^2)$, so the number of $\gamma \in \Gamma$ with $(\gamma z - \zeta)/(\gamma z - \bar{\zeta}) \in B_r$ is asymptotically $c r^2/(1-r^2)$, where $c = 4\pi/V$, with V the volume of the fundamental domain of $\mathcal{C}_\zeta \Gamma \mathcal{C}_\zeta^{-1}$.

Near $r=1$, it follows from Lemma 12 that $g_\lambda(r) \sim (1-r)^\alpha$ where $\alpha > 1$. Thus the convergence of the series amounts to the convergence of

$$\int_0^1 g_\lambda(r) \frac{d}{dr} \frac{r^2}{(1-r^2)} dr \quad \text{or} \quad \int_0^1 (1-r)^\alpha \frac{d}{dr} \frac{r^2}{(1-r^2)} dr < \infty. \qquad \square$$

$G_\lambda(z,\zeta)$ is the *automorphic Green's function*. We will see that it is an integral kernel for the resolvent of the Laplacian.

Theorem 16. *G_λ is defined and real analytic for all values of (z,ζ) except where $\zeta = \gamma(z)$ for some $\gamma \in \Gamma$. We have*

$$\left[-y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) - \lambda\right] G_\lambda(z,\zeta) = 0;$$

$$G_\lambda(h(z),h(\zeta)) = G_\lambda(z,\zeta), \qquad h \in G,$$

$$G_\lambda(z,\zeta) = G_\lambda(\gamma(z),\zeta) = G_\lambda(z,\gamma(\zeta)), \qquad \gamma \in \Gamma,$$

$$G_\lambda(z,\zeta) = \overline{G_\lambda(\zeta,z)},$$

and $G_\lambda(z,\zeta) \sim \frac{e}{2\pi}\log|z-\zeta|$ *near* $z=\zeta$*, where* e *is the order of the isotropy subgroup of* ζ *in* Γ*. For* $f \in C^\infty(\Gamma\backslash\mathfrak{H})$

$$\int_{\Gamma\backslash\mathfrak{H}} G_\lambda(z,\zeta)\left[-\eta^2\left(\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\eta^2}\right)-\lambda\right]f(\zeta)\,\frac{d\xi\wedge d\eta}{\eta^2} = f(z) \qquad (\zeta = \xi + i\eta). \tag{10}$$

See Bump [5] Proposition 2.3.5 and Hejhal [19], Proposition 6.5 on p. 33.

Proof. Most of these properties follow from the corresponding properties of g_λ. We prove (10). We will need a function $F \in C_c^\infty(\mathfrak{H})$ such that $f(z) = \sum_{\gamma\in\Gamma} F(\gamma z)$. To construct F, let u be a function on $\mathfrak{H}$ which is smooth, nonnegative, and has compact support containing a fundamental domain of Γ. Then for all z, the function $\sum_{\gamma\in\Gamma} u(\gamma z)$ is positive, and for z restricted to a compact set this sum is finite. It is thus a smooth, positive valued function and we can divide by it. Now

$$F(z) = \frac{u(z)\,f(z)}{\sum_{\gamma\in\Gamma} u(\gamma z)}$$

has the required property.

Substituting this and the definition of G_λ and using (9) gives

$$\sum_{\gamma,\delta\in\Gamma} \int_{\Gamma\backslash\mathfrak{H}} g_\lambda(\delta(z),\gamma(\zeta))\left[-\eta^2\left(\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\eta^2}\right)-\lambda\right]F(\gamma(\zeta))\,\frac{d\xi\wedge d\eta}{\eta^2}.$$

One of the summations may be collapsed with the integration to give

$$\sum_{\gamma\in\Gamma} \int_{\mathfrak{H}} g_\lambda(\delta(z),\zeta)\left[-\eta^2\left(\frac{\partial^2}{\partial\xi^2}+\frac{\partial^2}{\partial\eta^2}\right)-\lambda\right]F(\zeta)\,\frac{d\xi\wedge d\eta}{\eta^2} = \sum_{\delta} F(\delta(z)) = f(z).$$

This completes the proof. □

Let $\lambda < 0$. As is easily checked, the logarithmic singularity along the diagonal is not sufficient to cause divergence of the integral

$$\int_{\Gamma\backslash\mathfrak{H}}\int_{\Gamma\backslash\mathfrak{H}} |G_\lambda(z,\zeta)|^2 \frac{dx\wedge dy}{y^2}\frac{d\xi\wedge d\eta}{\eta^2} < \infty.$$

Thus the corresponding integral operator, which we shall denote $R(\lambda,\ \Delta)$, is Hilbert-Schmidt.

Theorem 17. *(i) The eigenvalues* λ_i *of* Δ *on* $L^2(\Gamma\backslash\mathfrak{H})$ *tend to* ∞*, and satisfy* $\sum \lambda_i^{-2} < \infty$*. (We exclude the eigenvalue* $\lambda_0 = 0$ *corresponding to the constant function from this summation.)*
(ii) The Laplacian Δ *has an extension to a self-adjoint operator on the Hilbert space* $L^2(\Gamma\backslash\mathfrak{H})$*.*
(iii) If $\lambda < 0$ *then the compact operator* $R(\lambda,\Delta)$ *is a bounded inverse to* $\Delta - \lambda I$*.*

We express (iii) by saying that $R(\lambda,\Delta) = (\Delta-\lambda I)^{-1}$ is the *resolvent* of Δ.

Proof. Let $\lambda < 0$ be an arbitrary negative real number. By Theorem 9, let ϕ_i be a basis of $H = L^2(\Gamma\backslash\mathfrak{H})$ consisting of eigenvectors of Δ, with corresponding eigenvalues λ_i. It follows from (10) that ϕ_i is also an eigenfunction of $R(\lambda, \Delta)$ with eigenvalue $(\lambda_i - \lambda)^{-1}$. Since $R(\lambda, \Delta)$ is Hilbert-Schmidt, $\sum_i (\lambda_i - \lambda)^{-2} < \infty$, whence $\sum_i \lambda_i^{-2} < \infty$.

We prove (ii). Let $\mathfrak{D}_\Delta$ be the linear subspace of $L^2(\Gamma\backslash\mathfrak{H})$ consisting of elements of the form $\sum a_i\phi_i$ such that $\sum \lambda_i^2|a_i|^2 < \infty$; on this space, define

$$\Delta\Big(\sum a_i\phi_i\Big) = \sum \lambda_i a_i \phi_i.$$

Since the λ_i tend to infinity, and in particular are bounded away from zero, it is not hard to check that this operator is closed and in fact self-adjoint. This proves (ii).

We have already checked that if $R(\lambda, \Delta)\phi_i = (\lambda_i - \lambda)^{-1}\phi_i$. Since the ϕ_i are an orthonormal basis of H, this implies (iii). □

We arrange the eigenvalues of Δ in ascending order:

$$\lambda_0 = 0 < \lambda_1 \leqslant \lambda_2 \leqslant \lambda_3 \leqslant \cdots.$$

The eigenvalue λ_0 corresponds to the constant function and has multiplicity exactly one. Eigenvalues in the range $(0, \frac{1}{4})$ are called *exceptional eigenvalues*. They are qualitatively different. For example, the spherical functions corresponding to exceptional eigenvalues are nontempered – they grow faster than spherical functions corresponding to $\lambda \geqslant \frac{1}{4}$. (See Section 4). Exceptional eigenvalues correspond to zeros of the Selberg zeta function on the real line between 0 and 1. By contrast zeros corresponding to $\lambda \geqslant \frac{1}{4}$ satisfy the Riemann hypothesis. (See Section 7.) Randol [36] proved that for some X, exceptional eigenvalues do occur. On the other hand, Selberg [40] conjectured that exceptional eigenvalues do not occur in the cuspidal spectrum of congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$, and it would follow from this that they do not occur in the spectrum of compact quotients $\Gamma\backslash\mathfrak{H}$ associated to quaternion division algebras.

4 Spherical functions

As we pointed out, Property S reduces to a second order differential equation which has two independent solutions. One solution, having nice behavior at the boundary, is the Green's function. Another, having nice behavior at i, is the so-called spherical function. The substance of Lemma 12 is that these two solutions are not the same. In this section we will study the spherical function.

Definition 18. *Let λ be a complex number. We call a function σ on $\mathrm{SL}_2(\mathbb{R})$ a λ-spherical function if it is smooth, K-bi-invariant, and $\Delta\sigma = \lambda\sigma$.*

Here Δ is the Laplace-Beltrami operator. Before we show that such a function exists and is unique up to constant multiple, let us explain briefly how such a function fits into the representation theory.

Suppose that (π, V) is an irreducible representation of G. Let $(\hat{\pi}, \hat{V})$ be its contragredient. Thus there exists an invariant bilinear pairing $\langle , \rangle: V \times \hat{V} \rightarrow \mathbb{C}$.

By Theorem 4, if V^K and $\hat{V}^K$ are nonzero, they are one-dimensional, and we will assume this. (Actually if one is nonzero the other is too.) Let $v^\circ \in V^K$ and $\hat{v}^\circ \in \hat{V}^K$ be nonzero K-fixed vectors. The function

$$\sigma(g) = \langle \pi(g) v^\circ, \hat{v}^\circ \rangle$$

is evidently K-bi-invariant. Moreover, Regarding Δ as an element of the center of the universal enveloping algebra $U(\mathfrak{g})$, it acts by a scalar λ on V. and σ inherits this property. So it is a spherical function.

Without reference to this construction, we now show that spherical functions exist and are unique.

Theorem 19. *(i) Let $\lambda \in \mathbb{C}$. Then there is a unique smooth K-bi-invariant function ω_λ on $\mathrm{SL}_2(\mathbb{R})$ such that $\Delta\omega_\lambda = \lambda\omega_\lambda$ and $\omega_\lambda(1) = 1$. (ii) If $f: G \rightarrow \mathbb{C}$ is any smooth function such that $\Delta f = \lambda f$, then*

$$\int_{K \times K} f(k g k') \, d k \, d k' = f(1) \omega_\lambda(g). \tag{11}$$

(iii) If f is right K-invariant and $\Delta f = \lambda f$, then

$$\int_K f(h k g) \, d k = f(h) \omega_\lambda(g). \tag{12}$$

Proof. To satisfy $\Delta\omega = \lambda\omega$ we need

$$W(r) = \omega\begin{pmatrix} y^{1/2} & \\ & y^{-1/2} \end{pmatrix}, \qquad r = \frac{y-1}{y+1}$$

to satisfy (8). As we have seen, this differential equation has a regular singular point at the origin, and one solution is bounded there, whereas the other has a logarithmic singularity. Hence ω_λ, if it exists, is unique.

To show that such a function exists, let f be any continuous function on G which is an eigenfunction of Δ. The left hand side of (11) is a K-bi-invariant function which is an eigenfunction of Δ. If $f(1) = 1$, this will satisfy (i), proving existence. For example we can take $f = f_s$ where $\lambda = s(1-s)$ and

$$f_s\left(\begin{pmatrix} y^{1/2} & * \\ & y^{-1/2} \end{pmatrix} k\right) = y^s, \qquad y > 0, \; k \in K. \tag{13}$$

Now that the existence and uniqueness of ω are established, we note that the left sides of both (11) and (12) are smooth K-bi-invariant eigenfunctions of Δ, hence constant multiples of ω_λ. In both cases, the constant may be evaluated by taking $g = 1$. This proves both (ii) and (iii). □

Theorem 20. *Suppose that f is a smooth function on G which is right invariant by K and such that $\Delta f = \lambda f$. Then for $\phi \in \mathcal{H}^\circ$ we have*

$$\rho(\phi) f = \chi_\lambda(\phi) f \tag{14}$$

where

$$\chi_\lambda(\phi) = \int_G \phi(g) \omega_\lambda(g) \, d g. \tag{15}$$

Proof. By Theorem 19, we have

$$\int_K f(h k g) \, d k = f(h) \omega_\lambda(g). \tag{16}$$

We note that $\rho(\phi) f$ is an average of right translates of f, and right translation commutes with left translation. Hence we may apply $\rho(\phi)$ to both sides of (16) to obtain

$$\int_K (\rho(\phi) f)(h k g) \, d k = f(h) (\rho(\phi) \omega_\lambda)(g).$$

We take $g = 1$ in this identity. Since $\rho(\phi) f$ is right K-invariant, the integrand on the left side becomes constant when $g = 1$ and so the left side becomes just $(\rho(\phi) f)(h)$. On the other hand $(\rho(\phi)\omega_\lambda)(1)$ equals the integral (15), so $\rho(\phi) f(h) = \chi_\lambda(\phi) f(h)$. □

Theorem 21. *If ϕ_1 and $\phi_2 \in \mathcal{H}^\circ$, then*

$$\chi_\lambda(\phi_1 * \phi_2) = \chi_\lambda(\phi_1) \, \chi_\lambda(\phi_2).$$

Proof. This follows from Theorem 20 on applying $\rho(\phi)$ to any eigenfunction f, for example f_s as in (13). □

Thus the function $\chi_\lambda \colon \mathcal{H}^\circ \to \mathbb{C}$ is a *character* of $\mathcal{H}^\circ$. We return to the point of view introduced at the beginning of the section to explain the meaning of $\chi(\phi)$ in terms of representations. First, we recall the construction and parametrization of a class of irreducible representations of $\mathrm{SL}_2(\mathbb{R})$, the *principal series*.

Let s be a complex number. Let P_s^+ be the set of functions $f \colon G \to \mathbb{C}$ such that

$$f\left(\begin{pmatrix} y^{1/2} & x y^{-1/2} \\ & y^{-1/2} \end{pmatrix} g \right) = (-1)^{\epsilon \operatorname{sgn}(y)} y^s f(g), \tag{17}$$

where $\epsilon = 0$, and such that the restriction of f to K is square integrable. Similarly let P_s^- be the space of functions satisfying (17) with $\epsilon = 1$. G acts on these spaces by right translation. Let $(\pi_s^\pm, P_s^\pm)$ be these representations. The representation $\pi_s^\pm$ is irreducible except in the case where $2s$ is an integer and $2s$ is even for π_s^+, odd for π_s^-. We call π_s^+ and π_s^- the *odd* and *even principal series* of representations, respectively.

Suppose that $\mathrm{re}(s) > \frac{1}{2}$. An *intertwining integral* $M(s) \colon P_s^\pm \to P_{1-s}^\pm$ is given by

$$(M(s) f)(g) = \int_{-\infty}^{\infty} f\left(\begin{pmatrix} & -1 \\ 1 & \end{pmatrix} \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} g \right) d x. \tag{18}$$

It may be checked that the integral is absolutely convergent. Now $M(s)$ extends by analytic continuation to an intertwining map for all s such that $2s$ is not an integer congruent to ϵ mod 2. Actually we only claim that it extends to an infinitesimal equivalence, that, is, an isomorphism of the underlying $(\mathfrak{g}, K)$ module, but it is a true isomorphism in the important special case $\mathrm{re}(s) = \frac{1}{2}$. Thus $\pi^{\pm}_{1/2+it} \cong \pi^{\pm}_{1/2-it}$.

Let $k \equiv \epsilon$ modulo 2. Let

$$f_{s,k}\left(\begin{pmatrix} y^{1/2} & x \\ & y^{-1/2} \end{pmatrix}\kappa_\theta\right) = \mathrm{sgn}(y)^\epsilon y^s e^{ik\theta}. \tag{19}$$

This is a K-finite vector in $P_s^{\pm}$ where $\pm$ is $(-1)^\epsilon$. Let $\tilde{f}_{1-s,k} = M(s) f_{s,k}$. One may compute

$$\tilde{f}_{k,1-s} = (-i)^k \sqrt{\pi}\, \frac{\Gamma(s)\Gamma\left(s - \frac{1}{2}\right)}{\Gamma\left(s + \frac{k}{2}\right)\Gamma\left(s - \frac{k}{2}\right)} f_{k,1-s}. \tag{20}$$

See Bump [5], Proposition 2.6.3.

The representation $\pi^{\pm}_{1/2+it}$ is unitary if t is real. These representations are called the *unitary principal series* representations. Also π_s^+ is unitary if s is a real number between 0 and 1. These are the *complementary series* representations. They correspond to exceptional eigenvalues of the Laplacian in the automorphic spectrum. There are a few other irreducible representations, namely the discrete series (related to holomorphic modular forms) and the trivial representations. But only the even principal series π_s^+ have K-fixed vectors. All of these facts are proved in Bump [5], Chapter 2.

The K-fixed vector in P_s^+ is precisely the function f_s in (13). Now if $\phi \in \mathcal{H}^\circ$ then since ϕ is K-bi-invariant, the operator $\pi^+(\phi)$ maps all of P_s^+ onto the unique K-fixed vector f_s. Thus the trace of the rank one operator $\pi_s^+(\phi)$ is just its eigenvalue on this vector f_s, so

$$\mathrm{tr}\, \pi_s^+(\phi) = \chi_\lambda(\phi). \tag{21}$$

5 The Plancherel formula

Let $\phi \in \mathcal{H}^\circ$. Define

$$g(u) = e^{u/2} \int_{-\infty}^{\infty} \phi\left(\begin{pmatrix} e^{u/2} & \\ & e^{-u/2} \end{pmatrix}\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix}\right) dx, \tag{22}$$

and let

$$h(t) = \int g(u) e^{iut} dt \tag{23}$$

be its Fourier transform.

Theorem 22. *The functions g and h are even, and g is compactly supported. If $\lambda = \frac{1}{4} + t^2$, then*

$$\chi_\lambda(\phi) = h(t). \tag{24}$$

Proof. Let f_s be as in (13) with $s = \frac{1}{2} + i\,t$ so that $\lambda = \frac{1}{4} + t^2 = s(1-s)$. By Theorem 19,

$$\int_K f_s(k\,g)\,d\,k = \omega_\lambda(g),$$

and

$$\chi_\lambda(\phi) = \int_G \int_K \phi(g)\,f_s(g\,k)\,d\,k\,d\,g.$$

Interchanging the order of integration and making the variable change $g \to g\,k^{-1}$, since ϕ is right k invariant, we obtain

$$\chi_\lambda(\phi) = \int_G \phi(g)\,f_s(g)\,d\,g.$$

Now we use the coordinates

$$g = \begin{pmatrix} e^{u/2} & \\ & e^{-u/2} \end{pmatrix}\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix}\kappa_\theta, \qquad d\,g = \frac{1}{2\pi}d\,u\,d\,x\,d\theta. \tag{25}$$

Noting that in these coordinates $f_s(g) = e^{u/2}e^{i\,u\,t}$, we obtain

$$\chi_\lambda(\phi) = \int_\infty^\infty g(u)\,e^{i\,t\,u}\,d\,u,$$

proving (24). We note that ω_λ, and therefore the character $\chi_\lambda(\phi)$ is unchanged if $s \to 1-s$, that is, if $t \to -t$. Hence (24) implies that h is an even function. By Fourier inversion, so is g. □

Theorem 23. *(i) We have*

$$\phi(1) = \frac{1}{2\pi}\int_0^\infty t\,h(t)\tanh(\pi t)\,d\,t.$$

(ii) If ϕ_1, ϕ_2 are K-invariant and compactly supported then

$$\int_G \phi_1(g)\,\overline{\phi_2(g)}\,d\,g = \frac{1}{2\pi}\int_0^\infty t\,h_{\phi_1}(t)\,\overline{h_{\phi_2}(t)}\tanh(\pi t)\,d\,t.$$

Proofs may be found in Knapp [25], Chapter 11, Gelfand, Graev and Piatetski-Shapiro [14], Chapter 2 Section 6 and Varadarajan [41] Theorem 39 on p. 205. We will give a proof (after some Lemmas) which uses no Lie theory. A portion of the argument parallels Proposition 4.1 on p. 15 of Hejhal [18], and we have made our notation consistent with his, and with Selberg [38], (3.1).

Theorem 23 (i) is the Fourier inversion formula on the noncommutative group $\mathrm{SL}_2(\mathbb{R})$. It is sometimes called the *Plancherel formula* because it implies (ii), which is the true Plancherel formula. The measure $\frac{1}{2\pi}t\tanh(\pi t)\,d\,t$ is called the *Plancherel measure* on the even unitary principal series. (Since we have only considered $\phi \in \mathcal{H}^\circ$ we do not need the other irreducible unitary representations.) The Plancherel measure is closely related to the intertwining integrals (18). Indeed, $M(\frac{1}{2} - i\,t) \circ M(\frac{1}{2} + i\,t)$ is an G-equivariant endomorphism of the irreducible space $P_{1/2+it}^+$, so by Schur's Lemma it is a scalar. One may check using (20) that the reciprocal of this scalar is $\frac{1}{\pi}t\tanh(\pi t)$. See Knapp and Stein [24].

Lemma 24. *Let $\eta \geqslant y > 1$. Then*

$$\phi\left(\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix}\begin{pmatrix} y^{1/2} & \\ & y^{-1/2} \end{pmatrix}\right) = \phi\begin{pmatrix} \eta^{1/2} & \\ & \eta^{-1/2} \end{pmatrix},$$

where

$$x = \pm \sqrt{y(\eta + \eta^{-1} - y - y^{-1})}.$$

Proof. Identifying $K\backslash G/K$ with $K\backslash\mathcal{H}$, we want to find x such that the images ηi and $z = x + i\,y$ of the two matrices in $\mathcal{H}$ are in the same SO(2) orbit. Mapping $\mathcal{H}$ to the unit disk by the Cayley transform $z \to (z - i)/(z + i)$, their images must therefore be equidistant from the origin. That is,

$$\frac{\eta - 1}{\eta + 1} = \left|\frac{z - i}{z + i}\right| = \sqrt{\frac{a - 2y}{a + 2y}}, \qquad a = x^2 + y^2 + 1.$$

Applying the map $t \to (1 + t^2)/(1 - t^2)$ to both sides of this equation, $\eta + \eta^{-1} = a/y$. This equation can now be solved for x. □

Define a function Φ on $\mathbb{R}^+$ by

$$\Phi(e^u + e^{-u} - 2) = \phi\begin{pmatrix} e^{u/2} & \\ & e^{-u/2} \end{pmatrix}. \tag{26}$$

Lemma 25. *If $U = e^u + e^{-u} - 2$, then $g(u) = Q(U)$ where*

$$Q(U) = \int_U^\infty \frac{\Phi(V)\,d\,V}{\sqrt{V - U}}. \tag{27}$$

Proof. Let $V = e^v + e^{-v} - 2$. With $y = e^u$, $\eta = e^v$ and x as in Lemma 24, think of x as function of $V \geqslant U$. Then $d\,x = \frac{1}{2} e^u (V - U)^{-1/2} d\,V$. We integrate V from U to infinity and double the result to account for both positive and negative x. □

Lemma 26. *We have*

$$\Phi(U) = -\frac{1}{\pi}\int_U^\infty \frac{Q'(V)\,d\,V}{\sqrt{V - U}}. \tag{28}$$

Proof. We'd like to differentiate under the integral sign in (27) but since the left endpoint depends on U we must be careful. Integrate (27) by parts to obtain

$$Q(U) = -2\int_U^\infty \Phi'(V)\sqrt{V - U}\,d\,V.$$

The integrand now vanishes at the left endpoint, so we may differentiate under the integral sign, then integrate by parts again to obtain

$$Q'(U) = \int_U^\infty \frac{\Phi'(V)}{\sqrt{V - U}} d\,V = -2\int_U^\infty \Phi''(V)\sqrt{V - U}\,d\,V.$$

We substitute this into the right side of (28), then switch the order of integration:

$$-\frac{1}{\pi}\int_U^\infty \frac{Q'(V)\,dV}{\sqrt{V-U}} = \frac{2}{\pi}\int_U^\infty \int_V^\infty \sqrt{\frac{W-V}{V-U}}\,\Phi''(W)\,dW\,dV = \frac{2}{\pi}\int_U^\infty \int_U^W \sqrt{\frac{W-V}{V-U}}\,dV\,\Phi''(W)\,dW. \tag{29}$$

To evaluate the inner integral, let $V = U + (W-U)\,v$. We have

$$\int_U^W \sqrt{\frac{W-V}{V-U}}\,dV = (W-U)\int_0^1 \sqrt{\frac{1-v}{v}}\,dv = \frac{\pi}{2}(W-U)$$

since $\int_0^1 v^{-1/2}(1-v)^{1/2}\,dv = B(\frac{1}{2},\frac{3}{2}) = \frac{\pi}{2}$. Thus (29) equals

$$\int_U^\infty (U-W)\,\Phi''(W)\,dW = -\int_U^\infty \Phi'(W)\,dW = \Phi(U),$$

where we have integrated twice by parts. □

Proof of Theorem 23. Take $U=0$ and write Lemma 26 in the form

$$\phi(1) = \Phi(0) = -\frac{1}{\pi}\int_0^\infty \frac{g'(u)\,du}{e^{u/2}-e^{-u/2}}. \tag{30}$$

Since g is even and g' is odd, we have the Fourier inversion formula

$$g(u) = \frac{1}{\pi}\int_0^\infty h(t)\,e^{-itu}\,dt, \qquad g'(u) = \frac{1}{i\pi}\int_0^\infty t\,h(t)\,e^{-itu}\,dt,$$

and we may change the limits in (30) to $(-\infty,\infty)$, dividing by 2, then interchange the order of integration to obtain

$$\phi(1) = \frac{1}{2\pi}\int_0^\infty t\,h(t)\,\frac{i}{\pi}\int_{-\infty}^\infty \frac{e^{-iut}\,du}{e^{u/2}-e^{-u/2}}\,dt.$$

The inner integral passes through the pole at $u=0$ of the integrand and is interpreted as the principal value. The Plancherel formula now follows from

$$\int_{-\infty}^\infty \frac{e^{-iut}\,du}{e^{u/2}-e^{-u/2}} = -i\pi\tanh(\pi t), \tag{31}$$

which we may prove as follows. Since $t>0$, the numerator e^{-iut} is small for u in the lower half plane, and we may move the path of integration downwards. The left side of (31) $-2\pi i$ times the sum of the residues at $u = -2\pi i\,n$ of the integrand in the lower half plane. The residue at $u=0$ is only counted half since the path of integration passes through this point. We get

$$-2\pi i\left(\frac{1}{2} + \sum_{k=1}^\infty (-1)^k e^{-2\pi k t}\right) = -\pi i\tanh(\pi t)$$

proving (i).

To prove (ii) define $\phi_2'(g) = \overline{\phi_2(g^{-1})}$. Since ϕ_2 is K-bi-invariant, and since the K-double cosets are stable under $g \to g^{-1}$ this equals $\overline{\phi_2(g)}$. Now apply (i) to $\phi = \phi_1 * \phi_2'$, then $\phi(0) = \langle \phi_1, \phi_2 \rangle_2$. On the other hand $h_\phi(t) = h_{\phi_1}(t)\,\overline{h_{\phi_2}(t)}$ by Theorem 21 so (ii) follows. □

Proposition 27. *Let g be an even, compactly supported function on $\mathbb{R}$. There exists $\phi \in \mathcal{H}^\circ$ such that (22) is true.*

Proof. Let Q: $\mathbb{R}^+ \to \mathbb{C}$ be defined by $Q(e^u + e^{-u} - 2) = g(u)$. We claim that Q has derivatives of all orders that are continuous on $\mathbb{R}^+$. The only issue is continuity at 0. Write $U = e^u + e^{-u} - 2$. We find that

$$u = 2\log\left(\frac{1}{2}(\sqrt{U} + \sqrt{U+4})\right) = \sqrt{U} - \frac{U^{3/2}}{24} + \frac{3U^{5/2}}{640} - \dots .$$

Since only odd powers of $U^{1/2}$ appear, when we substitute u into the smooth, even function g we obtain a function of v that has continuous derivatives of all orders at $v = 0$. We define Φ: $\mathbb{R}_+ \to \mathbb{C}$ by (28) or, integrating by parts and substituting $W = V - U$,

$$\Phi(U) = \frac{2}{\pi}\int_0^\infty Q''(W+U)\sqrt{W}\,dW.$$

We may differentiate under the integral sign arbitrarily many times, so Φ has continuous derivatives of all orders, even at $U = 0$. Now $u \to \Phi(e^u + e^{-u} - 2)$ is a smooth, even, compactly supported function. We will show that this implies that there is a unique smooth, compactly supported K-bi-invariant ϕ satisfying (26). Every double coset of SO(2) has a representative of the form $\begin{pmatrix} e^{u/2} & \\ & e^{-u/2} \end{pmatrix}$. The only two such representatives are for values u and $-u$, and since Φ is even there is a uniquely determined function such that (22) is true. The issue is to show that this function is smooth. Since it is constant on cosets of $G/\mathrm{SO}(2) \cong \mathfrak{H}$, we can transfer it to the upper half plane by the map $G \longrightarrow \mathfrak{H}$ given by $g \longmapsto g(i)$ and it is sufficient to show that the corresponding function on $\mathfrak{H}$ is smooth. Furthermore, we may then transfer the function to the unit disk by the Cayley transform. If $re^{i\theta}$ are polar coordinates on the disk, then the resulting function depends only on r, and its value is

$$\Phi(e^u + e^{-u} - 2) = \Phi\left(\frac{4r^2}{1-r^2}\right), \qquad u = \log\left(\frac{1-r}{1+r}\right).$$

Since Φ is a compactly supported function on $\mathbb{R}_+$ with continous derivatives of all orders, $re^{i\theta} \longmapsto \Phi\left(\frac{4r^2}{1-r^2}\right)$ defines a smooth compactly supported function on the disk. Thus proves that $\phi \in \mathcal{H}^\circ$. □

6 The Selberg trace formula

We have already shown that the integral operators $\rho(\phi)$ are Hilbert-Schmidt, hence compact. More is true: they are *trace class*. A compact operator is *trace class* if it can be factored as the composite of two Hilbert-Schmidt operators. If it is self-adjoint, and has eigenvalues λ_i, it is easy to see that this is equivalent to $\sum |\lambda_i| < \infty$. Lang's [27] contains much useful material about trace class operators.

Let f_i be an orthonormal basis of $L^2(\Gamma\backslash\mathfrak{H})$ consisting of eigenfunctions of $\rho(\phi)$ which are also eigenfunctions of Δ. Assuming that $\phi \in \mathcal{H}^\circ$ satisfies (7), let μ_i be the eigenvalues of $\rho(\phi)$ with respect to this basis. Making a Fourier expansion we have

$$K_\phi(z,w) = \sum \mu_i f_i(z)\, \overline{f_i(w)}. \tag{32}$$

Initially this expansion is only valid in $L^2(\Gamma\backslash\mathfrak{H} \times \Gamma\backslash\mathfrak{H})$, but we will now show that the right-hand side represents the kernel in the sense of uniform convergence.

Lemma 28. (Dini's Theorem) *Let f_i be a sequence of functions on the compact set X such that $\sum_{i=1}^\infty |f_i(x)|$ converges pointwise to a continuous function $F(x)$. Then the series $\sum_{i=1} f_i(x)$ converges absolutely and uniformly.*

Proof. For each $x \in X$ there exists N_x such that $F(x) - \sum_{i=1}^n |f_i(x)| < \varepsilon$ of $n \geqslant N_x$. Since the sequence of functions

$$g_n(n) = F(x) - \sum_{i=1}^n |f_i(x)|$$

are continuous and decrease monotonely to zero, there exists a neighborhood U_x of x such that $g_n(y) < \varepsilon$ for all $y \in U_x$ and $n \geqslant N_x$. Since X is compact, it is covered by a finite number of the sets U_{x_i}, $i = 1, \cdots, r$. Now take $N = \max(N_{x_i})$. We see that if $n \geqslant N$ then $g_n(x) < \varepsilon$ on X. This shows that the convergence of the series $\sum_{i=1}^\infty |f_i(x)|$ is uniform, so the series $\sum_i f_i(x)$ converges absolutely and uniformly. ☐

Theorem 29. (Hilbert and Schmidt) *Let $f \in C^\infty(\Gamma\backslash\mathfrak{H})$, and let f_i be an orthonormal basis of $L^2(\Gamma\backslash\mathfrak{H})$ consisting of eigenfunctions of Δ. Then the expansion*

$$f(z) = \sum_{i=1}^\infty c_i f_i(z), \qquad c_i = \langle f, f_i \rangle \tag{33}$$

is absolutely and uniformly convergent. More precisely, let λ be any negative number, and let $\varepsilon > 0$, $C > 0$ be positive constants. Then there exists a constant $N_{\varepsilon,C}$, independent of f such that

$$\left| f(z) - \sum_{i=1}^\infty c_i f_i(z) \right| < \varepsilon \quad \text{if } \langle h, h \rangle \leqslant C,\ n > N_{\varepsilon,C}, \qquad h = (\Delta - \lambda) f. \tag{34}$$

Proof. Let λ be any negative number. We have

$$G_\lambda(z,\zeta)=\sum (\lambda_i-\lambda)^{-1} f_i(z)\,\overline{f_i(w)},$$

where $\Delta f_i=\lambda_i f_i$. Indeed, if we expand $G_\lambda(z,\zeta)=\sum \nu_i f_i(z)\overline{f_i(w)}$, then the coefficient ν_i may be evaluated by taking $f=f_i$ in (10); we find that $\nu_i=(\lambda_i-\lambda)^{-1}$. Let

$$\gamma(z)=\int_{\Gamma\backslash\mathfrak{H}} |G_\lambda(z,\zeta)|^2 \frac{d\xi\wedge d\eta}{\eta^2}, \qquad \zeta=\xi+i\eta\in\mathfrak{H},$$

which is easily seen to be a continuous function on $\Gamma\backslash\mathfrak{H}$. We have

$$\sum_i |\lambda_i-\lambda|^{-2}|f_i(z)|^2=\gamma(z). \tag{35}$$

This follows from the orthogonormality of the f_i. Both sides are continuous and so by Dini's theorem (Lemma 28) the convergence in (35) is uniform in z.

By (10) we have

$$f(z)=\int_{\Gamma\backslash\mathfrak{H}} G_\lambda(z,\zeta)h(\zeta)\,\frac{d\xi\wedge d\eta}{\eta^2}, \tag{36}$$

where $h(\zeta)=(\Delta-\lambda)f$. If $h=\sum a_i f_i$, then since h is continuous, it is square integrable, and

$$\sum_i |a_i|^2=\langle h,h\rangle<\infty. \tag{37}$$

By (36) we have

$$c_i=a_i(\lambda_i-\lambda)^{-1}.$$

Now by the Cauchy-Schwartz inequality and (35), (37) we have

$$\sum_i |c_i f_i(z)|=\sum_i |a_i(\lambda_i-\lambda)^{-1} f_i(z)|\leqslant\sqrt{\sum_i |\lambda_i-\lambda|^{-2}|f_i(z)|^2}\cdot\sqrt{\sum_i |a_i|^2}<\infty. \tag{38}$$

Since (35) is uniformly convergent, convergence of this series is also uniformly convergent. Since $\sum c_i f_i=f$ in $L^2(\Gamma\backslash\mathfrak{H})$, we have $\sum c_i f_i(z)=f(z)$ almost everywhere. The uniform convergence of $\sum |c_i f_i(z)|$ implies the uniform convergence of $\sum c_i f_i(z)$. As the uniform limit of continuous functions, $\sum c_i f_i$ is continuous and so is f, and so $\sum c_i f_i(z)=f(z)$ for all z. Finally, if h is allowed to vary with $\langle h,h\rangle<C$, we may replace $\sqrt{\sum_i |a_i|^2}$ by C in (38), and it is clear that $N_{\varepsilon,C}$ can be chosen to make (34) true. □

Theorem 30. *The expansion on the right-hand side of (32) converges absolutely and uniformly to* K_ϕ.

Proof. If w is fixed, then applying Theorem 29 to $K_\phi(z,w)$ shows that the convergence is absolute and uniform in z. Moreover, taking λ to be negative and applying $\Delta-\lambda$ to $K_\phi(z,w)$ (in the z parameter) gives a continuous function on the compact set $\Gamma\backslash\mathfrak{H}\times\Gamma\backslash\mathfrak{H}$ that is bounded. We regard this as family of functions of z, indexed by w, that are bounded in the L^∞ norm and *a fortiori* in the L^2 norm, so (34) shows that the convergence of (32) to $K_\phi(z,w)$ is actually uniform in both variables. □

Theorem 31. *If $\phi \in \mathcal{H}^\circ$ then $\rho(\phi)$ is trace class.*

Proof. A linear combination of trace class operators is trace class. Hence it is sufficient to prove this with ϕ replaced by $\frac{1}{2}\big(\phi(g) + \overline{\phi(g^{-1})}\big)$ and by $\frac{1}{2i}\big(\phi(g) - \overline{\phi(g^{-1})}\big)$. We may thus assume that ϕ satisfies (7) and so $\rho(\phi)$ is self-adjoint. Let μ_i be its (nonzero) eigenvalues. Let λ_i be the corresponding eigenvalues of Δ. Thus $\sum \lambda_i^{-2} < \infty$.

Applying Δ to $K_\phi(z, w)$ in the first variable gives a new kernel $\Delta_z K_\phi$. We will show that

$$(\Delta_z K_\phi)(z, w) = \sum \mu_i \lambda_i f_i(x)\, \overline{f_i(y)}. \tag{39}$$

Formally this follows from (32) by termwise differentiation. However this must be justified – the uniform convergence of Theorem 30 would of course justify term-by-term integration but not differentiation.

Note that $\Delta_z K_\phi(z, w)$ is continuous hence has an expansion

$$(\Delta_z K_\phi)(z, w) = \sum \nu_i f_i(x)\, \overline{f_i(y)}.$$

Let $\lambda < 0$. Consider

$$\int_{\Gamma\backslash\mathfrak{H}} G_\lambda(z, \zeta)\, [(\Delta_\zeta K_\phi)(\zeta, w) - \lambda\, K_\phi(\zeta, w)] \frac{d\xi \wedge d\eta}{\eta^2}.$$

On the one hand by (10) this is just (32). On the other hand, the term in square brackets is $\sum_i\, (\nu_i - \lambda\mu_i) f_i(\zeta)\, \overline{f_i(w)}$, and since the resolvent is compact, may apply it term by term. Using $R(\lambda, \Delta) f_i = (\lambda_i - \lambda)^{-1} f_i$ we get $(\nu_i + \lambda\mu_i)(\lambda_i - \lambda)^{-1} = \mu_i$, so $\nu_i = \lambda_i \mu_i$ proving (39).

Since this function $\Delta_z K(z, w)$ is continuous, it is Hilbert-Schmidt, and so we obtain the bound

$$\sum |\mu_i \lambda_i|^2 < \infty. \tag{40}$$

Now $\sum |\mu_i| < \infty$ follows from $\sum |\lambda_i|^{-2} < \infty$ and (40) by Cauchy-Schwarz. □

The trace $\mathrm{tr}\, T$ of a self-adjoint trace class operator T is by definition the sum of its eigenvalues.

Theorem 32. *If $\phi \in \mathcal{H}^\circ$ satisfies (7), and if μ_i are the eigenvalues of $\rho(\phi)$, the trace*

$$\mathrm{tr}\, \rho(\phi) = \int_{\Gamma\backslash\mathfrak{H}} K_\phi(z, z) \frac{dx \wedge dy}{y^2}. \tag{41}$$

Proof. This follows from orthonormality on integrating (32). □

The Selberg trace formula is a more explicit formula for its trace. Let $\{\gamma\}$ denote a set of representatives for the conjugacy classes of Γ. Let $Z_\Gamma(\gamma)$ denote the centralizer in Γ of γ.

Theorem 33. *We have*

$$\mathrm{tr}\,\rho(\phi) = \sum_{\{\gamma\}} \int_{Z_\Gamma(\gamma)\backslash G} \phi(g^{-1}\gamma g)\,d\,g. \tag{42}$$

Proof. We rewrite the right side of (41) as

$$\sum_{\gamma\in\Gamma} \int_G \phi(g^{-1}\gamma g)\,d\,g = \sum_{\{\gamma\}} \sum_{\delta\in Z_\Gamma\backslash\Gamma} \int_G \phi\big(g^{-1}\delta^{-1}\gamma\delta\,g\big)\,d\,g.$$

Combining the integral and the summation gives (42). □

This is a primitive form of the trace formula. To make this more explicit, we must study more explicitly the *orbital integrals* on the right side. An element $1 \neq \gamma \in \Gamma$ is *hyperbolic* if its eigenvalues are real, *elliptic* if complex of absolute value 1. Γ is *hyperbolic* if each $1 \neq \gamma \in \Gamma$ is hyperbolic. For example, let X be a compact Riemann surface of genus $\geqslant 2$. Its universal cover is $\mathfrak{H}$ and $\Gamma = \pi_1(X)$ acts with quotient X. These examples are precisely the hyperbolic groups.

We assume now that Γ is a hyperbolic group. If $1 \neq \gamma \in \Gamma$ define $N = N(\gamma)$ by asking that γ be conjugate to

$$\begin{pmatrix} N^{1/2} & \\ & N^{-1/2} \end{pmatrix}. \tag{43}$$

for some N. Let $N_0 = N_0(\gamma)$ be such that

$$\begin{pmatrix} N_0^{1/2} & \\ & N_0^{-1/2} \end{pmatrix}$$

is conjugate to a generator of $Z_\Gamma(\gamma)$. We may obviously assume that N and N_0 are > 1. We note that $Z_G(\gamma)$ is conjugate to the diagonal subgroup. Its image in X is a closed geodesic. So the numbers $N(\gamma)$ are thus the lengths of closed geodesics, and the numbers $N_0(\gamma)$ are the lengths of prime geodesics.

Theorem 34. *With g the function in (22),*

$$\int_{Z_\Gamma(\gamma)\backslash G} \phi(g^{-1}\gamma g)\,d\,g = \frac{\log N_0}{N^{1/2} - N^{-1/2}}\, g(\log N). \tag{44}$$

Proof. We may assume that γ equals (43). Using Iwasawa coordinates (25) the integral is

$$\int_0^{\log N_0} d\,u \int_{-\infty}^{\infty} \phi\left(\begin{pmatrix} 1 & -x \\ & 1 \end{pmatrix}\begin{pmatrix} N^{1/2} & \\ & N^{-1/2} \end{pmatrix}\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix}\right) d\,x \;=$$

$$\log(N_0)|\int_{-\infty}^{\infty} \phi\left(\begin{pmatrix} N^{1/2} & \\ & N^{-1/2} \end{pmatrix}\begin{pmatrix} 1 & (1-N^{-1})x \\ & 1 \end{pmatrix}\right) d\,x$$

and a change of variables proves (44). □

Theorem 35. *Let g be a smooth, even, compactly supported function and let h be its Fourier transform, defined by (23). If $\frac{1}{4}+t_i^2$ are the eigenvalues of Δ on $\Gamma\backslash\mathfrak{H}$, and if $\log(N)$ runs through the lengths of closed geodesics of Γ, where for each N we let N_0 be the length of the corresponding prime geodesic, we have*

$$\sum h(t_i) = \frac{vol(\Gamma\backslash\mathfrak{H})}{4\pi}\int_{-\infty}^{\infty} t\,h(t)\tanh(\pi t)\,dt + \sum_N \frac{\log N_0}{N^{1/2}-N^{-1/2}}g(\log N). \tag{45}$$

This is the *Selberg trace formula*. See Selberg [38] and [39].

Proof. We choose ϕ as in Proposition 27. By Theorem 20 and Theorem 22, the $h(t_i)$ are the eigenvalues of $\rho(\phi)$ on the eigenfunctions of the Laplacian, so the left side of (45) is the trace of $\rho(\phi)$. By Theorem 23 and Theorem 34, the right side of (45) is the sum of the orbital integrals. Thus the identity follows from Theorem 33. □

7 The Selberg zeta function

In order to get useful applications the class of functions g and h in the trace formula must be expanded.

Theorem 36. *The trace formula Theorem 35 remains true provided h is an even function analytic in the strip $\mathrm{im}(z) \leqslant \frac{1}{2}+\delta$, such that $h(r) = O(1+|r|)^{-2-\delta}$ in this strip. This assumption implies that the Fourier transform $g(u) = O(e^{-(\frac{1}{2}+\delta)|u|})$.*

Proof. See Hejhal [18], Chapter 1 Section 7 for the proof by an approximation argument that Theorem 35 implies this stronger statement. □

Theorem 37. (Weyl's law) . *The number of j with $t_j \leqslant x$ is asymptotically*

$$\frac{1}{4\pi}\mathrm{vol}(\Gamma\backslash\mathfrak{H})\,x^2.$$

This must be mentioned as the first significant application of the trace formula. Noting that $\lambda_j = \frac{1}{4}+t_j^2$, this means that the number of $\lambda_j \leqslant x$ is asymptotically $\frac{1}{4\pi}\mathrm{vol}(\Gamma\backslash\mathfrak{H})\,x$.

Proof. (Sketch) Briefly, taking $h(t) = e^{-t^2T}$, the hyperbolic contributions in Theorem 35 are of smaller magnitude and the first term in (45) predominates. One obtains $\sum e^{-\lambda_j T} \cong \frac{1}{4\pi T}\mathrm{vol}(\Gamma\backslash\mathfrak{H})$, and Weyl's law follows from a Tauberian theorem. See Hejhal [18], Chapter 2 Section 2. □

The trace formula may be regarded as a duality between the length spectrum of $\Gamma\backslash\mathfrak{H}$ (that is, the set of lengths N of closed geodesics) and the numbers t_i such that $\frac{1}{4}+t_i^2$ are the eigenvalues of the Laplacian. A similar duality, which we next discuss, pertains between the set of prime numbers and the zeros of the Riemann zeta function.

Theorem 38. *Let g be compactly supported, smooth and even, and let*

$$h(t) = \int_{-\infty}^{\infty} e^{itu} g(u)\, du.$$

Let $\frac{1}{2} + i t_i$ denote the zeros of ζ in the cricital strip. Then

$$h\left(-\frac{i}{2}\right) - \sum h(t_i) + h\left(\frac{i}{2}\right) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} h(t) \frac{\Gamma'\left(\frac{1+2it}{4}\right)}{\Gamma\left(\frac{1+2it}{4}\right)} dt$$
$$+ g(0)\log(\pi) + 2\sum \frac{\log(p)}{\sqrt{p^n}} g(\log(p^n)). \qquad (46)$$

Since h is assumed even $h(-i/2) = h(i/2)$. However there are good reasons for writing the formula this way. (See Remark 4 below.) Special cases ("explicit formulae") were found by Riemann, von Mangoldt, Hadamard, de la Vallee Poussin, and Ingham. These are discussed in Ingham [21]. Weil [43] formalized the duality.

Proof. To prove this, let $\xi(s) = \Gamma_{\mathbb{R}}(s)\zeta(s)$, where $\Gamma_{\mathbb{R}}(s) = \pi^{-s/2}\Gamma(\frac{s}{2})$. The function $h(t)$ is entire and we write $h(t) = H(\frac{1}{2} + it)$, $H(s) = h(-i(s-1/2))$. Consider, for $\delta > 0$

$$\frac{1}{2\pi i}\int_{1+\delta-i\infty}^{1+\delta+i\infty} \frac{\xi'(s)}{\xi(s)} H(s)\, ds.$$

Moving the path of integration to $\mathrm{re}(s) = -\delta$ and using the functional equation $\xi(s) = \xi(1-s)$, we obtain the negative of this integral plus the sum of the residues; so

$$\frac{1}{\pi i}\int_{1+\delta-i\infty}^{1+\delta+i\infty} \frac{\xi'(s)}{\xi(s)} H(s)\, ds = \sum h(t_i) - h\left(\frac{i}{2}\right) - h\left(-\frac{i}{2}\right).$$

We have $\xi'/\xi = \Gamma_{\mathbb{R}}'/\Gamma_{\mathbb{R}} + \zeta'/\zeta$, and the integral over $\Gamma_{\mathbb{R}}$ can be moved left to $\mathrm{re}(s) = \frac{1}{2}$. We have

$$\frac{1}{\pi i}\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} \frac{\Gamma_{\mathbb{R}}'(s)}{\Gamma_{\mathbb{R}}(s)} h\left(-i(s-\frac{1}{2})\right) ds = -g(0)\log(\pi) + \frac{1}{2\pi}\int_{-\infty}^{\infty} h(t) \frac{\Gamma'\left(\frac{1+2it}{4}\right)}{\Gamma\left(\frac{1+2it}{4}\right)} dt.$$

On the other hand, we have $-\zeta'(s)/\zeta(s) = \sum \Lambda(n) n^{-s}$ where $\Lambda(p^k) = \log(p)$, p prime, while $\Lambda(n) = 0$ if n is not a prime power. We have

$$\frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} H(s) n^{-s} ds = \frac{g(\log(n))}{\sqrt{n}},$$

and assembling the pieces we get (46). □

Remark 1. Moving the line of integration requires knowing that a path from $2 + i\,T$ to $-1 + i\,T$ can be found for arbitrarily large T where ζ'/ζ is not too large. This can be accomplished by choosing the path to lie about half way between a pair of zeros that are not too close together. See Ingham [21], Theorem 26 on p. 71, where he proves that one can always find t near T such that ζ/ζ' is $O(\log^2(T))$ on the line from $2 + i\,t$ to $-1 + i\,t$.

Remark 2. In practice the condition that g be compactly supported is too strong. It is sufficient that g and h be as in Theorem 36.

Remark 3. If H is not too big in the left half plane we have

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} h(t)\frac{\Gamma'\left(\frac{1+2i\,t}{4}\right)}{\Gamma\left(\frac{1+2i\,t}{4}\right)}d\,t = \frac{1}{2\pi i}\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty} H(s)\frac{\Gamma'\left(\frac{s}{2}\right)}{\Gamma\left(\frac{s}{2}\right)}d\,s = 2\sum_{n=0}^{\infty} H(-2n),$$

where the sum is over the residues at the poles of $\Gamma(s/2)$.

Remark 4. Let X be a nonsingular complete curve of genus g over the finite field $k = \mathbb{F}_{\mathrm{q}}$. If P is a prime divisor of degree $d(P)$ we denote $N(P) = q^{d(P)}$. The zeta function of X is

$$\prod (1 - N(P)^{-s})^{-1} = \frac{\prod_{j=1}^{2g} (1 - \alpha_j q^{-s})^{-1}}{(1 - q^{-s})(1 - q^{1-s})},$$

where α_i are the eigenvalues of the Frobenius map in $H^1(X, \mathbb{Q}_l)$, for any prime $l \neq p$. The Riemann hypothesis is that $|\alpha_j| = \sqrt{q}$. We write $\alpha_j = q^{\frac{1}{2}+i\theta_j}$. Let $g(n)\colon \mathbb{Z} \to \mathbb{C}$ be a sequence which is even and nonzero for only finitely many n, and let

$$h(t) = \sum_n g(n)\, e^{itn\log(q)}.$$

Thus h is entire, even, and periodic with period $2\pi/\log(q)$. Then

$$h\left(-\frac{i}{2}\right) - \sum h(\theta_j) + h\left(\frac{i}{2}\right) = -(2g-2)\,g(0) + \sum_P \sum_{m=1}^{\infty} \frac{d(P)\,g(m\,d(P))}{\sqrt{q^{m\,d(P)}}}.$$

This function field analog of Theorem 38 may be proved by considering

$$\frac{1}{2\pi i}\int_2^{2+2\pi it/\log(q)} \left(\frac{Z'(s)}{Z(s)} + (g-1)\log(q)\right)H(s)\,d\,s,$$

where $h(t) = H(\frac{1}{2} + i\ t)$. It may be regarded as an application of the Lefschetz-Grothendieck fixed point formula applied to a correspondence on X, and the three contributions $h(-i/2)$, $\sum h(\theta_j)$ and $h(i/2)$ are the traces of the correspondence applied to $H^0(X)$, $H^1(X)$ and $H^2(X)$. See Patterson [35], Chapter 5 and the first part of Connes [9].

The analogy between the Selberg trace formula and the explicit formulae has been the source of much speculation. Whether this analogy is misleading or not remains to be seen. One major difference between the explicit formulae and the Selberg trace formula is that in the trace formula, $\sum h(t_i)$ appears with a positive sign, while in (46), it appears with a negative sign. As Connes [9] points out, this difference may be very significant.

One tangible fruit of the analogy is Selberg's discovery of a zeta function which bears a relationship to the trace formula similar to that of the Riemann zeta function to the explicit formulae. With notation as in the previous section, Selberg considered

$$Z(s) = \prod_{\{N_0\}} \prod_{k=0}^{\infty} (1 - N_0^{-s-k}),$$

where $\log(N_0)$ runs through the lengths of prime geodesics.

Theorem 39. *$Z(s)$ has analytic continuation to all s, with zeros at the negative integers and at $\frac{1}{2} \pm i\,t_k$, where $\frac{1}{4} + t_k^2$ are the eigenvalues of the Laplacian on $\Gamma\backslash\mathfrak{H}$.*

Proof and discussion. To motivate the introduction of the Selberg zeta function, we would like to take

$$g(u) = e^{-|u|(s-1/2)}$$

in the trace formula. Unfortunately we cannot use this function, but if we could, the geometric side of the trace formula would be the logarithmic derivative of $Z(s)$:

Lemma 40.

$$\sum_{\{N\}} \frac{\log(N_0)\, N^{-(s-1/2)}}{N^{1/2} - N^{-1/2}} = \frac{Z'(s)}{Z(s)}.$$

Proof. Writing $N = N_0^m$

$$\sum_{\{N_0\}} \sum_{m=1}^{\infty} \frac{\log(N_0)\, N_0^{-m(s-1/2)}}{N_0^{m/2} - N_0^{-m/2}} = \sum_{\{N_0\}} \sum_{m=1}^{\infty} \frac{\log(N_0)\, N_0^{-ms}}{1 - N_0^{-m}}.$$

Substituting $(1 - N_0^{-m})^{-1} = \sum_{k=0}^{\infty} N_0^{-ks}$ and interchanging the order of summation, this equals

$$\sum_{\{N_0\}} \frac{\log(N_0)\, N_0^{-(s+k)}}{1 - N_0^{-(s+k)}} = \frac{Z'(s)}{Z(s)}.$$

This completes the proof. □

Although we cannot use $g(u) = e^{-|u|(s-1/2)}$, we may use

$$g(u) = \frac{e^{-|u|(s-1/2)}}{2s-1} - \frac{e^{-|u|(\sigma-1/2)}}{2\sigma-1},$$

where s and σ are distinct. This device of subtraction eliminates the discontinuity in g' at $u=0$. We will hold σ fixed and vary s. Initially, they are both assumed to have large real part. The Fourier transform $h(t)=\int_{-\infty}^{\infty} g(u)e^{itu}du$ is

$$h(t)=\frac{1}{(s-\frac{1}{2})^2+t^2}-\frac{1}{(\sigma-\frac{1}{2})^2+t^2}.$$

Next we prove, assuming that s and σ have real part $\geqslant\frac{1}{2}$, that

$$\int_{-\infty}^{\infty} h(t)\,t\tanh(\pi t)\,dt=2\sum_{k=0}^{\infty}\left[\frac{1}{s+k}-\frac{1}{\sigma+k}\right]. \tag{47}$$

To prove this, use the partial fraction decomposition

$$\frac{t}{(s-\frac{1}{2})^2+t^2}=\frac{1}{2i}\left[\frac{1}{(s-\frac{1}{2})-it}-\frac{1}{(s-\frac{1}{2})+it}\right].$$

The left side of (47) equals

$$\frac{1}{2i}\int_{-\infty}^{\infty}\left[\frac{1}{(s-\frac{1}{2})-it}-\frac{1}{(\sigma-\frac{1}{2})-it}\right]\tanh(\pi t)\,dt$$
$$-\frac{1}{2i}\int_{-\infty}^{\infty}\left[\frac{1}{(s-\frac{1}{2})+it}-\frac{1}{(\sigma-\frac{1}{2})+it}\right]\tanh(\pi t)\,dt.$$

Since tanh is odd, the two contributions are equal, and the first integral may be evaluated by moving the path of integration up into the upper half-plane, where the only poles are at the poles $i(k+\frac{1}{2})$ of $\tanh(\pi t)$, $k=0,1,2,\cdots$. The residue of $\tanh(\pi t)$ at these points is π^{-1}, and we obtain (47).

We now use the Selberg Trace Formula. We note that the volume of the fundamental domain is $4\pi(g-1)$. So for these functions g and h we get:

$$\frac{1}{2s-1}\frac{Z'(s)}{Z(s)}-\frac{1}{2\sigma-1}\frac{Z'(\sigma)}{Z(\sigma)}=$$
$$-(2g-2)\sum_{k=0}^{\infty}\left[\frac{1}{s+k}-\frac{1}{\sigma+k}\right]+\sum\left[\frac{1}{(s-\frac{1}{2})^2+t_i^2}-\frac{1}{(\sigma-\frac{1}{2})^2+t_i^2}\right].$$

Fixing σ and varying s, the right side has meromorphic continuation to all s. Multiplying by $2s-1$ see that the poles of Z'/Z are simple and have integer residues. Hence

$$Z(s)=\exp\int_{\infty}^{s}\frac{Z'(s)}{Z(s)}ds$$

has analytic continuation to all s.

8 Cusps

There are discontinuous subgroups Γ of $G=\mathrm{SL}_2(\mathbb{R})$ such that $\Gamma\backslash\mathfrak{H}$ is noncompact but has finite volume. A well-known example is $\Gamma=\mathrm{SL}_2(\mathbb{Z})$.

We recall that $\mathrm{SL}_2(\mathbb{C})$ acts on the Riemann sphere $\mathfrak{R} = \mathbb{P}^1(\mathbb{C}) = \mathbb{C} \cup \{\infty\}$ by linear fractional equations, extending (1) to complex matrices. Now $\mathfrak{R}$ contains both the upper half plane $\mathfrak{H}$ and the unit disk $\mathfrak{D} = \{z \in \mathbb{C} \,|\, |z| < 1\}$. We can map $\mathfrak{H}$ to the unit disk $\mathfrak{D}$ by the Cayley transform

$$\mathcal{C} = \frac{1}{\sqrt{2i}} \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix} : z \longmapsto \frac{z-i}{z+i}.$$

We have

$$\mathcal{C}\,\mathrm{SL}_2(\mathbb{R})\,\mathcal{C}^{-1} = \mathrm{SU}(1,1) = \left\{ \begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix} \Big| |a|^2 + |b|^2 = 1 \right\}.$$

Thus if Γ is a discontinuous group of $\mathrm{SL}_2(\mathbb{R})$ acting on $\mathfrak{H}$ such that $\Gamma\backslash\mathfrak{H}$ has finite volumen, then $\Gamma_{\mathfrak{D}} = \mathcal{C}\Gamma\mathcal{C}^{-1}$ a discontinuous group of $\mathrm{SU}(1,1)$ acting on $\mathfrak{D}$ such that $\Gamma_{\mathfrak{D}}\backslash\mathfrak{D}$ has finite volume. The advantage of working with $\Gamma_{\mathfrak{D}}$ in the following discussion is that $\mathfrak{D}$ is bounded, so we can draw better pictures showing the behavior of the boundary.

If $\Gamma\backslash\mathfrak{H}$ or equivalently $\Gamma_{\mathfrak{D}}\backslash\mathfrak{D}$ is noncompact, we can still find a fundamental domain $\mathcal{F}$ for such a group whose boundary arcs are pairs of congruent geodesics, as in Proposition 1, by the same method. As in Proposition 1, if *one traverses the boundary in a counterclockwise direction, the congruent boundary arcs are always traversed in opposite directions.*

We will call such a domain *polygonal*. With the removal of the assumption that $\Gamma\backslash\mathfrak{H}$ is compact, there is an important difference. Now the fundamental domain can go down to the boundary in one or more places. Let us call the point of the boundary of $\mathfrak{H}$ or $\mathfrak{D}$ where the fundamental domain $\mathcal{F}$ of Γ or $\Gamma_{\mathfrak{D}}$ touches it a *cusp* of $\mathcal{F}$.

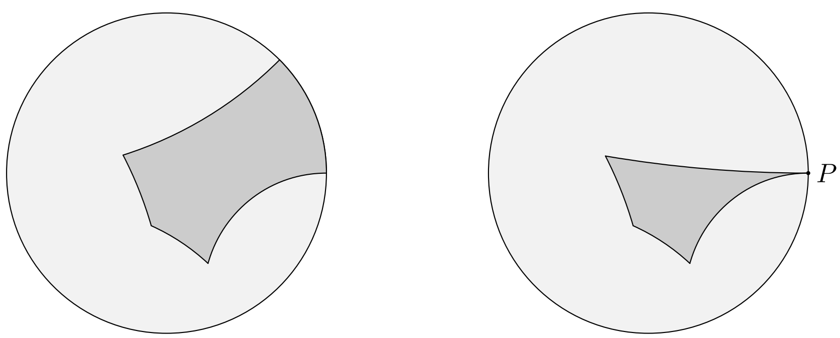

Figure 1. Noncompact fundamental domains in $\mathfrak{D}$. Left: infinite volume. Right: finite volume with one cusp (at the point P.)

Figure 1 shows two polygonal domains inside of $\mathfrak{D}$ bounded by noneuclidean polygons, whose boundaries are geodesic arcs as in Proposition 1. It is easy to see that the domain on the left has infinite volume, so only the type of boundary behavior in the second figure is permitted for groups where $\Gamma\backslash\mathfrak{H}$ (or, equivalently $\Gamma_{\mathfrak{D}}\backslash\mathfrak{D}$) has finite volume. If $\mathcal{F}$ is a fundamental domain for Γ, then of course $\mathcal{F}_{\mathfrak{D}} = \mathcal{C}\mathcal{F}$ is a fundamental domain for $\Gamma_{\mathfrak{D}}$. We will consider these two groups and fundamental domains as equivalent, but our pictures will usually draw $\mathfrak{D}$ and $\mathcal{F}_{\mathfrak{D}}$.

Proposition 41. *We may choose the fundamental domain $\mathcal{F}$ so that if P and Q are distinct cusps of $\mathcal{F}$, then there is no element $\gamma \in \Gamma$ such that $\gamma(P) = Q$.*

Proof. (sketch) We suppose that there is such an element γ. Then we may modify the fundamental domain by cutting off a piece $\mathcal{N}$ of $\mathcal{F}$ near Q and replacing it with $\gamma^{-1}(\mathcal{N})$. This procedure is illustrated in Figure 2.

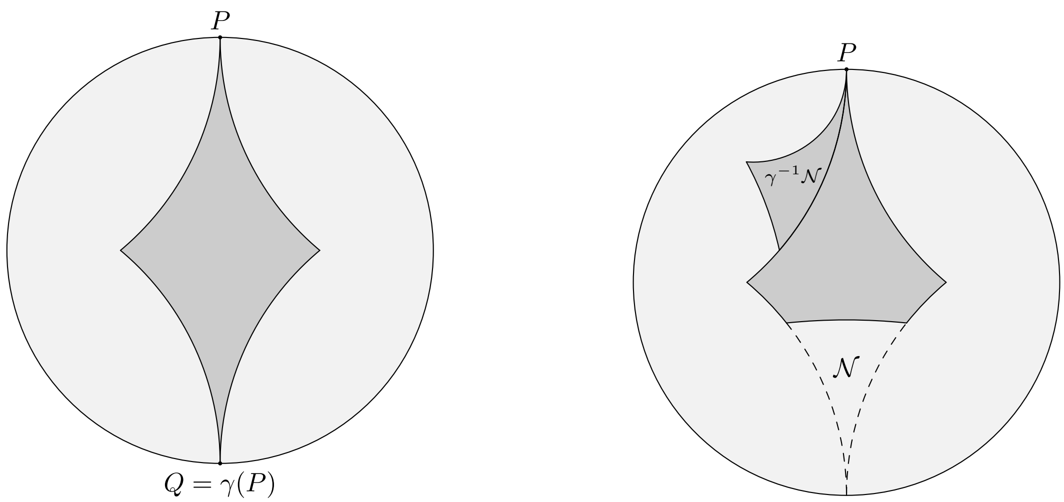

Figure 2. Left: a fundamental domain containing two congruent boundary points. Right: fundamental domain for the same group after moving a boundary piece from the vicinity of Q up to P.

It may be confirmed that this procedure does not affect the fact that the boundary consists of congruent geodesic arcs that are traversed in opposite directions when the boundary is navigated counterclockwise.

The diagram in Figure 2 suggests that we may perform this operation in such a way as to produce a connected fundamental domain. This is true, though we have not proved it yet – one might have to move the piece $\gamma^{-1}\mathcal{N}$ around a bit more. Suppose that we have "missed" and obtained a disconnected fundamental domain as in Figure 3.

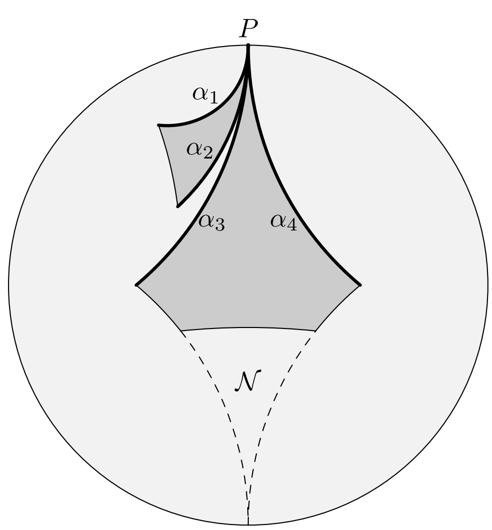

Figure 3.

We have labeled the four arcs through P in order as α_1, α_2, α_3 and α_4. The boundary arcs occur in congruent pairs, so α_1 must be congruent to another arc, which must obviously pass through a cusp. Since we have already arranged that P is not congruent to any other cusp of $\mathcal{F}$, we see that α_1 must be congruent to another arc through P, that is, to α_2, α_3 or α_4. It cannot be α_3 that it is congruent to, since we recall that if the boundary is traversed in a clockwise direction, congruent arcs are traversed oppositely. Thus there is an element of Γ that can move α_1 to α_2 or α_4. In either case, this operation will reduce the "gap" and repeating the process if necessary will eventually produce a connected fundamental domain. □

We will henceforth assume that the fundamental domain is chosen in this way.

We classify an element $g \neq 1$ of $\mathrm{SL}_2(\mathbb{R})$ as *hyperbolic, parabolic* or elliptic depending on how many fixed points they have in $\mathfrak{R}$. If there are two fixed points on the boundary $\mathbb{R} \cup \{\infty\}$, then g is *hyperbolic*; if there is a single fixed point on the boundary, g is *parabolic*; and if there are a pair of complex conjugate fixed points, one each in the interior of $\mathcal{H}$ and $\bar{\mathcal{H}}$ then g is *elliptic*. We have a similar classification of elements of SU(1, 1) in terms of the fixed points in $\mathfrak{D}$ and its boundary; thus $g \in \mathrm{SL}_2(\mathbb{R})$ will have the same classification as $\mathcal{C}g\mathcal{C}^{-1} \in \mathrm{SU}(1,1)$.

Proposition 42. *If $g \in \mathrm{SL}_2(\mathbb{R})$ or $\mathrm{SU}(1,1)$ then g is*

$$\begin{cases} \textit{hyperbolic} & \textit{if } |\mathrm{tr}(g)| > 2; \\ \textit{parabolic} & \textit{if } |\mathrm{tr}(g)| = 2; \\ \textit{elliptic} & \textit{if } |\mathrm{tr}(g)| < 2. \end{cases}$$

Proof. First suppose $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$. If z is a fixed point then $z = \frac{az+b}{cz+d}$ so z is a root of

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

The discriminant of this quadratic is $(d-a)^2 + 4bc = (a+d)^2 - 4$. So there are two real roots if $|a+d| > 2$, one real root of $|a+d| = 2$ and two complex conjugate roots if $|a+d| < 2$. (This argument must be modified slightly if $c = 0$, when ∞ is a fixed point.)

If $g \in \mathrm{SU}(1,1)$, then g has the same classification as $\mathcal{C}^{-1}g\mathcal{C} \in \mathrm{SL}_2(\mathbb{R})$. Since conjugation does not change the trace, the statement follows. □

Proposition 43. *Suppose that P is a point of the boundary of $\mathfrak{H}$ (resp. $\mathfrak{D}$) that is a cusp of $\mathcal{F}$ (resp. $\mathcal{F}_{\mathfrak{D}}$). Then Γ (resp. $\Gamma_{\mathfrak{D}}$) contains a parabolic element fixing P.*

Proof. We will treat the case of a discontinuous group Γ acting on $\mathfrak{H}$. The case of the disk is clearly equivalent.

Let α and β be the two boundary arcs passing through P. Since the boundary arcs occur in congruent pairs, there is some $\gamma \in \Gamma$ such that $\gamma(\alpha)$ is another boundary arc, and evidently $\gamma(P)$ is a cusp of Γ. Since $\mathcal{F}$ is chosen as in Proposition 41, $\gamma(P) = P$ and so $\gamma(\alpha) = \beta$.

We must show that γ is parabolic. It has one fixed point on the boundary of $\mathfrak{H}$, namely P and so we must show that it has no other. Arguing by contradiction, let Q be a second boundary point such that $\gamma(Q) = Q$. Applying a linear fractional transformation, we may assume that $P = \infty$ and $Q = 0$. Thus $\gamma(z) = a\,z$ for some positive real number a. We may assume that $a < 1$, since if $a > 1$ we may interchange α and β, and hence replace γ by γ^{-1}. Now it is clear that the sequence of fundamental domains $\mathcal{F}$, $\gamma\mathcal{F}$, $\gamma^2\mathcal{F}$, $\cdots$ have accumulation points on the positive imaginary axis. This is a contradiction, since Γ is a discontinuous group. □

With Proposition 43 in mind, we may now give a more satisfactory definition of a cusp. Let $S \subset \mathbb{R} \cup \infty$ (resp. the unit circle) be the set of boundary a points of $\mathfrak{H}$ (resp. $\mathfrak{D}$) such that $\gamma(a) = a$ for some parabolic element. Clearly Γ (resp. $\Gamma_{\mathfrak{D}}$) acts transitively on S. We call an orbit of Γ on S a *cusp*. This is consistent with our previous terminology since, given Proposition 43, it is not hard to see that every cusp of a fundamental domain that satisfies the conclusion of Proposition 41 must contain exactly one representative from each orbit in S. So this notion of a cusp gives the same set of cusps as our previous definition, but is intrinsic in the sense that it does not depend on the choice of a fundamental domain.

9 Fredholm equations

We will deduce the meromorphic continuation of the Eisenstein series from the meromorphic continuation of the resolvent of an operator. In this section we will prove a statement that is sufficient for our purposes.

We begin by recalling the following property of compact operators.

Theorem 44. (Fredholm Alternative) *Let H be a Hilbert space, and let $T: H \longrightarrow H$ be a compact operator. Let $0 \neq \lambda \in \mathbb{C}$. If $T - \lambda I$ is not invertible, then $Tx = \lambda x$ for some nonzero $x \in H$.*

We will give a proof of this well-known fact in a special case, assuming that T is a self-adjoint operator of Hilbert-Schmidt type; this proof will give more information since it will also show that the *resolvent* $(T - \lambda I)^{-1}$ is a meromorphic function of λ. More precisely, $(T - \lambda I)^{-1} - \lambda^{-1} I$ is a Hilbert-Schmidt operator that can be represented by a kernel that is meromorphic as a function of λ.

Thus, let X be a locally compact Hausdorff space with a positive Borel measure, and let $K \in L^2(X \times X)$. We define an operator T on $H = L^2(X)$ by

$$Tf(x) = \int_X K(x, y)\, f(y)\, dx.$$

The operator T is compact, and assuming $K(x, y) = \overline{K(y, x)}$ it is self-adjoint. The spectral theorem for compact operators guarantees that H has an orthonormal basis $\{\phi_i\}$ of eigenvectors of T, and if $T\phi_i = \lambda_i \phi_i$ then

$$K(x, y) = \sum_i \lambda_i \phi_i(x)\, \overline{\phi_i(y)}.$$

The λ_i are real, $\lambda_i \longrightarrow 0$ as $i \longrightarrow \infty$ and

$$\sum \lambda_i^2 = \int_{X \times X} |K(x,y)|^2 \, dx\, dy < \infty. \tag{48}$$

Define another kernel $K_\lambda(x,y)$ by

$$K_\lambda(x,y) = K(x,y) + \sum_i \frac{\lambda_i^2}{\lambda - \lambda_i} \phi_i(x) \overline{\phi_i(y)}.$$

Note that the series is absolutely convergent by (48), so K_λ defines another Hilbert-Schmidt operator. It is a meromorphic function of λ, with poles at the λ_i.

Theorem 45. *In this setting, let λ be a nonzero complex number which is not among the λ_i. Then*

$$R_\lambda f(x) = -\lambda^{-1} f - \lambda^{-2} \int_X K_\lambda(x,y) f(y) \, dx$$

defines a bounded inverse of $T - \lambda I$.

Proof. If f is in the nullspace of T, then f is orthogonal to the ϕ_i and it follows that $R_\lambda f = -\lambda^{-1} f$, so $(T - \lambda I) R_\lambda f = R_\lambda (T - \lambda I) f = f$. To prove the result, it is thus sufficient to check that $R_\lambda \phi_i = (\lambda_i - \lambda)^{-1} \phi_i$. Indeed,

$$\begin{aligned} R_\lambda \phi_i = \left[-\lambda^{-1} - \lambda^{-2} \left(\lambda_i + \frac{\lambda_i^2}{\lambda - \lambda_i} \right) \right] \phi_i &= \\ \lambda^{-2} (\lambda_i - \lambda)^{-1} \left[\lambda(\lambda - \lambda_i) - \lambda_i(\lambda_i - \lambda) + \lambda_i^2 \right] \phi_i &= (\lambda_i - \lambda)^{-1} \phi_i. \end{aligned}$$

The proof is now complete. ◻

A *Fredholm integral equation* is one that may be written

$$\int_X K(x,y) f(y) \, dy - \lambda f(x) = u(x),$$

where f is the "unknown" and u is a given function. Since the left-hand side is $(T - \lambda I) f$, we have proved that if K is a Hilbert-Schmidt kernel satisfying $K(x,y) = \overline{K(y,x)}$, and if λ is not in the spectrum of T, then this equation has a unique solution $R_\lambda u$.

10 Groups with one cusp

If the discontinuous subgroup Γ of $G = \mathrm{SL}_2(\mathbb{R})$ is of cofinite volume but has cusps, the spectral theory is complicated by a continuous spectrum, coming from the Eisenstein series. In this section we will consider the case where Γ has *one* cusp. We may assume that $-I \in \Gamma$, though of course it acts trivially on $\mathfrak{H}$. Without loss of generality, we may assume that this cusp is at ∞ and the stabilizer of ∞ in Γ is

$$\Gamma_\infty = \left\{ \pm \begin{pmatrix} 1 & n \\ & 1 \end{pmatrix} \middle| n \in \mathbb{Z} \right\}.$$

We will also denote $G_\infty = \left\{ \pm \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} \Big| x \in \mathbb{R} \right\}$.

For some positive real number T_0, $\mathcal{F}$ will contain the set

$$\mathcal{F}_T = \{ z = x + iy \in \mathcal{F} \,|\, y > T \}$$

for all $T \geqslant T_0$, and $\mathcal{F} - \mathcal{F}_T$ will be compact. We will also let $\mathcal{G}$ and $\mathcal{G}_T$ be the preimages of $\mathcal{F}$ and $\mathcal{F}_T$ in G under the map $g \longmapsto g(i)$; thus $\mathcal{G}$ is a fundamental domain for $\Gamma \backslash G$. Finally, let

$$\mathcal{H}_T = \{ z = x + iy \in \mathcal{H} \,|\, y > T \}, \qquad G_T = \{ g \in G \,|\, g(i) \in \mathcal{H}_T \}.$$

Then $\mathcal{F}_T$ (resp. $\mathcal{G}_T$) is a fundamental domain for $\Gamma_\infty \backslash \mathcal{H}_T$ (resp. $\Gamma_\infty \backslash G_T$).

In contrast with Theorem 7, the operators $\rho(\phi)$ are bounded but no longer compact. In order to obtain compact operators, we introduce *truncation*. Truncation was systematically studied by Arthur, but for rank one groups it was already used by Selberg, and is used in his Göttingen lectures. If f is a locally integrable function on $\Gamma \backslash G$, and if T is a sufficiently large real number, let f_0 be its *constant term*

$$f_0(g) = \int_{\Gamma_\infty \backslash G_\infty} f(ug)\, du = \int_0^1 f\left(\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} g \right) dx.$$

Let us introduce the notation

$$y(g) = \mathrm{im}(g(i)).$$

Thus $y(g) = f_{1,0}(g)$ in the notation (19). Let

$$\delta_T(g) = \begin{cases} 1 & \text{if } y(g) > T; \\ 0 & \text{otherwise.} \end{cases} \tag{49}$$

Lemma 46. *Suppose that $T \geqslant T_0$ and that $g \in \mathcal{G}$, $\gamma \in \Gamma$. Then*

$$\delta_T(\gamma g) = \begin{cases} 1 & \textit{if } y(g) > T \textit{ and } \gamma \in \Gamma_\infty; \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. If $y(\gamma g) > T$ then for suitable $\gamma_\infty \in \Gamma_\infty$ we have $\gamma_\infty \gamma g \in \mathcal{G}$. Since $\mathcal{G}$ is a fundamental domain and g, $\gamma_\infty \gamma g$ are both in it, this implies that $\gamma_\infty \gamma = 1$, so $\gamma \in \Gamma_\infty$. □

Let

$$\Lambda^T f(g) = f(g) - \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} \delta_T(\gamma g)\, f_0(\gamma g). \tag{50}$$

On the fundamental domain $\mathcal{G}$, the truncation operator is easy to characterize: the constant term is subtracted above T. That is, we have

$$\Lambda^T f(g) = \begin{cases} f(g) & \text{if } g \in \mathcal{G} - \mathcal{G}_T; \\ f(g) - f_0(g) & \text{if } g \in \mathcal{G}_T. \end{cases} \tag{51}$$

Indeed, this follows immediately from Lemma 46.

The truncation operator has the following adjointness property.

Proposition 47. *Let f and h be two locally integrable functions on $\Gamma\backslash G$. Then*

$$\langle \Lambda^T f, h \rangle = \langle f, \Lambda^T h \rangle \tag{52}$$

provided that the integrals defining the inner products $\langle f, h\rangle$ and $\langle f_0, h\rangle$ are absolutely convergent.

Proof. We have

$$\langle \Lambda^T f, h \rangle = \int_{\mathcal{G}} f(g)\overline{h(g)}\,dg - \int_{\mathcal{G}_T} f_0(g)\overline{h(g)}\,dg,$$

while

$$\langle f, \Lambda^T h \rangle = \int_{\mathcal{G}} f(g)\overline{h(g)}\,dg - \int_{\mathcal{G}_T} f(g)\overline{h_0(g)}\,dg.$$

It is thus enough to show that

$$\int_{\mathcal{G}_T} f_0(g)\overline{h(g)}\,dg = \int_{\mathcal{G}_T} f(g)\overline{h_0(g)}\,dg.$$

Indeed, we may write

$$\int_{\mathcal{G}_T} f_0(g)\overline{h(g)}\,dg = \int_{\Gamma_\infty\backslash G_T} f_0(g)\overline{h(g)}\,dg = \int_{\Gamma_\infty\backslash G_T}\int_{\Gamma_\infty\backslash G_\infty} f(ug)\overline{h(g)}\,du\,dg.$$

Interchanging the order of integration and making the variable change $g \longmapsto u^{-1}g$ proves (52). □

Let $\phi \in C_c^\infty(G)$ and let $K_\phi(g,h) = \sum \phi(g^{-1}\gamma h)$ as in (5). Define another kernel $K_\phi^{\Lambda,T}$ by

$$K_\phi^{\Lambda,T}(g,h) = \Lambda_g^T \Lambda_h^T K_\phi(g,h),$$

where the meaning of the notation is as follows. We recall that K_ϕ is automorphic in each variable separately. Thus we may apply the truncation operator to K_ϕ in both g and h.

Theorem 48. *Let $\phi \in C_c^\infty(G)$. Then $\rho^{\Lambda,T}(\phi) = \Lambda^T \circ \rho(\phi) \circ \Lambda^T$ is a compact operator on $L^2(\Gamma\backslash G)$. We have, for $f \in L^2(\Gamma\backslash G)$*

$$\rho^{\Lambda,T}(\phi) f(g) = \int_{\Gamma\backslash G} K_\phi^{\Lambda,T}(g,h)\, f(h)\,dh.$$

The kernel $K_\phi^{\Lambda,T}(g,h)$ is of weakly rapid decay in the following sense. Let us denote

$$g = \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix}\begin{pmatrix} y^{1/2} & \\ & y^{-1/2} \end{pmatrix}\kappa_g, \qquad h = \begin{pmatrix} 1 & u \\ & 1 \end{pmatrix}\begin{pmatrix} v^{1/2} & \\ & v^{-1/2} \end{pmatrix}\kappa_h,$$

where κ_g and $\kappa_h \in \mathrm{SO}(2)$. There exist constants A and B, depending only on ϕ such that if $g, h \in \mathcal{G}$, with then $K_\phi^{\Lambda,T}(g,h) = 0$ unless $A < y(g)y(h)^{-1} < B$. Moreover, if $N > 0$ is any constant, then there is a constant $C_N > 0$ such that

$$|K_\phi^{\Lambda,T}(g,h)| < C_N\, y(g)^{-N}$$

as $y(g) \longrightarrow \infty$ *uniformly in* h.

Proof. By (6) we have

$$\rho^{\Lambda,T}(\phi)f(g) = \int_{\Gamma\backslash G} \Lambda_g^T K_\phi(g,h)\Lambda^T f(h)\,dh.$$

Now by (52) we may move the truncation in h from f to $\Lambda_\phi^T K_\phi$, whence (48).

Our aim is to prove that the kernel $K_\phi^{\Lambda,T}$ is Hilbert-Schmidt, so that $\rho^{\Lambda,T}(\phi)$ is a compact operator.

Let

$$K_\phi^\infty(g,h) = \sum_{\gamma\in\Gamma_\infty} \phi(g^{-1}\gamma h).$$

The first step will be to show that there is a compact subset Ω of $\mathcal{G}\times\mathcal{G}$ such that

$$(g,h)\in(\mathcal{G}\times\mathcal{G})-\Omega \text{ implies } K_\phi(g,h) = K_\phi^\infty(g,h). \tag{53}$$

Since $\operatorname{supp}(\phi)$ is compact, there exists a constant $\varepsilon>0$ such that if $t<\varepsilon$ then

$$\phi\left(\kappa\begin{pmatrix} t^{1/2} & \\ & t^{-1/2}\end{pmatrix}\kappa'\right) = 0$$

for all $\kappa,\kappa'\in\mathrm{SO}(2)$. There exists a constant $c_1>0$ such that if $g\in\mathcal{G}$ then $y=y(g)>c_1$. Moreover, there exists a constant constant U such that if $\gamma\in\Gamma-\Gamma_\infty$, and if $g\in\mathcal{G}$, $h\in\mathcal{G}_U$ then $y(\gamma h)<c_1\varepsilon$. This means that

$$g^{-1}\gamma h = \kappa_g^{-1}\begin{pmatrix} t^{1/2} & \\ & t^{-1/2}\end{pmatrix}\kappa'$$

where $t = y^{-1}y(\gamma h)<\varepsilon$, and so

$$\phi(g^{-1}\gamma h)\neq 0, \qquad g\in\mathcal{G}, h\in\mathcal{G}_U \qquad \text{implies} \qquad \gamma\in\Gamma_\infty.$$

Therefore if $h\in\mathcal{G}_U$ and $g\in\mathcal{G}$ then $K_\phi(g,h) = K_\phi^\infty(g,h)$. Similarly, there exists a U' such that if $g\in\mathcal{G}_{U'}$ and $h\in\mathcal{G}$ then $K_\phi(g,h) = K_\phi^\infty(g,h)$. We may therefore take Ω to be the complement in $\mathcal{G}\times\mathcal{G}$ of $(\mathcal{G}\times\mathcal{G}_U)\cup(\mathcal{G}_{U'}\times\mathcal{G})$, which is compact, and (53) is proved.

Next we show that if $K=K_\phi$ or $K_\phi^{\Lambda,T}$ then there are positive constants $A<B$ such that for $g,h\in\mathcal{G}$ we have $K(g,h)=0$ unless

$$K(g,h)\neq 0 \quad \text{implies} \quad A<\frac{y(g)}{y(h)}<B. \tag{54}$$

Thus if g or h goes to the cusp at infinity within the fundamental domain $\mathcal{G}$, and (g,h) remains in the support of K, then g and h *both* go to the cusp, and at the same rate. It is easy to see that if f is any locally integrable function on $\Gamma\backslash G$ then $y(\operatorname{supp}\Lambda^T f)\subseteq y(\operatorname{supp} f)$; applying this fact to K in both variables shows that if (54) is true for $K=K_\phi$ then it is also true for $K_\phi^{\Lambda,T}$ with the same constants A and B.

Thus to prove (54) we may assume that $K = K_\phi$. Clearly it is enough to check this off the compact set Ω, and so we may actually assume that $K = K_\phi^\infty$. When $\gamma \in \Gamma_\infty$ we have

$$g^{-1}\gamma h = \kappa_g^{-1}\begin{pmatrix} (y^{-1}v)^{1/2} & * \\ & (y^{-1}v)^{-1} \end{pmatrix}\kappa_h^{-1}$$

where the value $*$ is unimportant. It is easy to see that if this is to lie in a given compact subset of G, particularly $\mathrm{supp}(\phi)$, then $y^{-1}v$ is restricted to a compact subset of $\mathbb{R}_+^\times$, which is the content of (54).

Now to show that the kernel $K_\phi^{\Lambda,T}$ is Hilbert-Schmidt, it is sufficient to show that it is of rapid decay as $y(g)$, $y(h) \longrightarrow \infty$, because $K_\phi^{\Lambda,T}$ is obviously bounded on any compact subset of $\mathcal{G} \times \mathcal{G}$. In view of (54), $y(g)$, $y(h)$ must go to ∞ at about the same rate. If $y(g)$ and $y(h)$ are sufficiently large then, in view of (53) we have

$$K_\phi(g,h) = \sum_{\xi \in \mathbb{Z}} F_{g,h}(-x+u+\xi)$$

where

$$F_{g,h}(\xi) = \phi\left(\kappa_g^{-1}\begin{pmatrix} y^{-1/2} & \\ & y^{1/2} \end{pmatrix}\begin{pmatrix} 1 & \xi \\ & 1 \end{pmatrix}\begin{pmatrix} v^{1/2} & \\ & v^{-1/2} \end{pmatrix}\kappa_h\right).$$

Now in $\mathcal{G}_T$, truncation in g subtracts the constant term producing

$$\sum_{\xi \in \mathbb{Z}} F_{g,h}(-x+u+\xi) - \int_0^1 \sum_{\xi \in \mathbb{Z}} F_{g,h}(-x+u+\xi+t)\,dt \;=\; \sum_{\xi \in \mathbb{Z}} F_{g,h}(-x+u+\xi) - \int_{-\infty}^{\infty} F_{g,h}(-x+u+t)\,dt.$$

As a function of h, this function has no constant term since, it is easy to see, integrating u from 0 to 1 produces zero. Thus the second truncation in h does not change the result, so

$$K_\phi^{\Lambda,T}(g,h) = \sum_{\xi \in \mathbb{Z}} F_{g,h}(-x+u+\xi) - \int_{-\infty}^{\infty} F_{g,h}(-x+u+t)\,dt.$$

Now by the Poisson summation formula we have

$$K_\phi^{\Lambda,T}(g,h) = \sum_{n \neq 0} \int_{-\infty}^{\infty} F_{g,h}(-x+u+t)e^{-2\pi i n t}\,dt.$$

The n-th term in this sum is

$$\sqrt{y}\,e^{2\pi i n(u-x)} \int_{-\infty}^{\infty} \phi\left(\kappa_g^{-1}\begin{pmatrix} 1 & t \\ & 1 \end{pmatrix}\begin{pmatrix} (y^{-1}v)^{1/2} & * \\ & (y^{-1}v)^{-1} \end{pmatrix}\kappa_h\right)e^{-2\pi i\sqrt{y}\,n t}\,dt.$$

As the Fourier transform of a smooth, compactly supported function, this is rapidly decreasing as $y \longrightarrow \infty$; that is, it is $O(y^{-N})$ for all N; the decay is uniform in κ_g, κ_h and $y^{-1}v$, all of which are restricted to compact sets, remembering (54).

It is now evident that

$$\int_{\mathcal{G}\times\mathcal{G}} \|K_\phi^{\Lambda,T}(g,h)\|^2\,dg\,dh < \infty,$$

so $\Lambda^T \circ \rho(\phi) \circ \Lambda^T$ is a Hilbert-Schmidt operator. In particular, it is compact. □

Let $L_0^2(\Gamma\backslash G)$ be the subspace of "cusp forms" $f \in L^2(\Gamma\backslash G)$ satisfying

$$\int_0^1 f\left(\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} g\right) d x = 0 \tag{55}$$

almost everywhere. It is easy to see that this space is invariant under the action of G, and is $\rho(\phi)$-stable.

Also let $L_0^2(\Gamma\backslash\mathfrak{H})$ be the right K-invariant elements of $L_0^2(\Gamma\backslash G)$, which may be regarded as functions on $\mathfrak{H}$. It is *not* true that $\rho(\phi)$ is a compact operator. However, its restriction to the space of cusp forms is compact.

Theorem 49. (Gelfand, Graev and Piatetski-Shapiro) *If $\phi \in \mathcal{H}$, the restriction of $\rho(\phi)$ to $L_0^2(\Gamma\backslash G)$ is a compact operator. Indeed, the restriction of $\rho(\phi)$ to $L_0^2(\Gamma\backslash G)$ coincides with $\rho(\phi)^{\Lambda,T}$.*

Proof. This follows from Theorem 48 since the truncation operator Λ^T clearly coincides with the identity operator on the space of cusp forms. □

If f is a function on $\Gamma\backslash G$, then we say that f is of *weakly rapid decay* if for all $N >$ 0 there exists a constant C_N such that $f(g) < C_N y(g)^{-N})$ for all $g \in \mathcal{G}$. If f is smooth, and if for all $D \in U(\mathfrak{g})$ (regarded as a ring of differential operators on G) the function Df is of weakly rapid decay, then we say that f is of *rapid decay*.

Theorem 50. *(i) $L_0^2(\Gamma\backslash\mathfrak{H})$ has a basis consisting of eigenfunctions of Δ.*
(ii) $L_0^2(\Gamma\backslash G)$ decomposes as a direct sum of irreducible invariant subspaces.
(iii) Any K-finite element of an irreducible invariant subspace $L_0^2(\Gamma\backslash\mathfrak{H})$ is smooth and of rapid decay.

Proof. For parts (i) and (ii), the proofs Theorem 9 and Theorem 11 are easily adapted.

For (iii), it is sufficient to show that if $V \subset L_0^2(\Gamma\backslash G)$ is an irreducible subspace and $f \in V(k)$ then f is of rapid decay. We may find $\phi \in \mathcal{H}$ satisfying $\phi(\kappa_\theta g \kappa_\sigma) = e^{ik(\theta+\sigma)}\phi(g)$, and such that

$$y \rightarrow \phi\begin{pmatrix} y^{1/2} & \\ & y^{-1/2} \end{pmatrix}$$

is a positive function of mass 1 concentrated near $y = 1$. Then $\rho(\phi) f$ is near f, therefore nonzero, and it is in $V(k)$, which is one-dimensional by Proposition 5, so it is proportional to f. Since $\rho(\phi) f$ is the convolution of f with a smooth function, it is smooth as a function on G. It follows from the weakly rapid decay of the kernel $K_\phi^{\Lambda,T}(g, h)$ that $\rho(\phi) f$ is of weakly rapid decay, and thus so is f. Moreover by Proposition 2.4.5 of Bump [5], the SO(2)-finite vectors in V are smooth vectors, the space of SO(2)-finite vectors is closed under the action of $U(\mathfrak{g})$. Thus or $D \in U(\mathfrak{g})$ the function Df satisfies the same assumptions as f, and is of weakly rapid decay, proving (iii). □

As in the case of compact quotient, the Laplacian acts by scalars on each irreducible one-dimensional subspace. The cuspidal spectrum behaves much as the entire spectrum in the compact case. On the other hand the orthogonal complement of $L_0^2(\Gamma\backslash G)$ contains a continuous spectrum. The eigenfunctions of the Laplacian relevant to the spectral theorem – Eisenstein series – are themselves not square integrable.

To understand how this can be, and to get some intuition as to the nature of the continuous spectrum, consider the following example. The group $\mathbb{R}$ acts on itself by translation, and the Laplacian $-d^2/dx^2$ is an invariant differential operator. It has eigenfunctions $f_a(x) = e^{2\pi i a x}$ with eigenvalues a^2. Any L^2 function has a Fourier expansion

$$\phi(x) = \int_{-\infty}^{\infty} \hat{\phi}(a)\, f_a(x)\, d\, a,$$

but f_a is itself not L^2. If $T \subset \mathbb{R}$ is measurable, the Fourier transforms of L^2 functions supported on T form an invariant subspace. There are no minimal invariant subspaces, so $L^2(\mathbb{R})$ doesn't decompose as a direct sum of irreducible representations.

Proposition 51. *The series*

$$E(z,s) = \sum_{\gamma \in \Gamma_\infty\backslash\Gamma} \frac{y^s}{|c z + d|^{2s}}, \qquad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

converges absolutely if $\mathrm{re}(s) > 1$ *when* $z \in \mathfrak{H}$.

Proof. Let σ be the real part of s. What we need to show is that if $\sigma > 1$ then

$$\sum_{\Gamma_\infty\backslash\Gamma} \frac{y^\sigma}{|c z + d|^{2\sigma}} < \infty. \tag{56}$$

We define a measure μ_σ on the upper half plane by $\mu_\sigma = y^{\sigma-2} d\,x \wedge d\,y$. Let B be a small neighborhood around $z \in \mathfrak{H}$. The Jacobian of the map $z \to \gamma(z)$, where $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{R})$ is $|c\,z + d|^{-4}$. Hence the μ_σ-volume of $\gamma(B)$ is (approximately) $|c\,z + d|^{-2\sigma}\mathrm{vol}(B)$.

We may choose the representatives $\gamma \in \Gamma_\infty\backslash\Gamma$ so that the images $\gamma(B)$ all lie within the rectangle $0 \leqslant x \leqslant 1$, $0 < y \leqslant C$ for some constant C. (Actually this is not *quite* true. If one $\gamma(B)$ happens to lie on the left or right edge of this region, cut it into two pieces along this edge and move one piece by ± 1 back into the region.) This rectangle has finite volume, so $\sum_{\Gamma_\infty\backslash\Gamma} |c\,z + d|^{-2\sigma} < \infty$, which implies (56). □

If $g \in G$, we will also denote $E(g,s) = E(z,s)$ where $z = g(i)$. Thus

$$E(g,s) = \sum_{\gamma \in \Gamma_\infty\backslash\Gamma} f_{s,0}(\gamma g). \tag{57}$$

We will eventually prove that this has meromorphic continuation to all $s \in \mathbb{C}$. More generally, if f_s is any $\mathrm{SO}(2)$-finite element of P_s^+ we can define

$$E(g, f_s) = \sum_{\gamma \in \Gamma_\infty\backslash\Gamma} f_s(\gamma g).$$

As s varies, we may organize the functions f_s into a family by requiring that the restriction of f_s to SO(2) is independent of s. Since f_s is to be SO(2)-finite, this means that f_s is a finite linear combination of the functions $f_{s,k}$ with k an even integer, that were defined by (19). When $\mathrm{re}(s) > 1$, the series is analytic function of s, and in this generality $E(g, s, f_s)$ is an analytic function of s. More generally still, we may introduce a unitary character $\chi: \Gamma \longrightarrow \mathbb{C}^\times$ such that $\chi(-I) = (-1)^\varepsilon$ with $\varepsilon = 0$ or 1; then let $f_s \in P_s^\pm$ be a family of SO(2)-finite vector, where we use P_s^+ if $\varepsilon = 0$ and P_s^- if $\varepsilon = 1$. Then we may consider

$$E(g, f_s, \chi) = \chi(\gamma) \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} f_s(\gamma g),$$

and in the same way one has meromorphic continuation as a function of s.

The analytic continuation of the Eisenstein series is closely connected with the analytic continuation of its *constant term*

$$E_0(g, f_s, \chi) = \int_0^1 E\left(\begin{pmatrix} 1 & t \\ & 1 \end{pmatrix} g, f_s, \chi \right) dt. \tag{58}$$

We recall that an intertwining operator $M(s): P_s^\pm \rightarrow P_{1-s}^\pm$ defined by (18).

Proposition 52. *Assume* $\mathrm{re}(s) > 1$. *There exists an analytic function* $c(s)$ *independent of the choice of* $f_s \in P_s^\pm$ *which is bounded on vertical strips (to the right of* 1*) such that*

$$E_0(g, f_s, \chi) = f_s(g) + c(s) M(s) f_s. \tag{59}$$

Proof. Substitute the definition of $E(g, f_s, \chi)$ into (58). The coset Γ_∞ in $\Gamma_\infty \backslash \Gamma$ contributes $f_{s,k}$. The remaining terms contribute

$$\int_0^1 \sum_{\substack{\gamma \in \Gamma_\infty \backslash \Gamma / \Gamma_\infty \\ \gamma \notin \Gamma_\infty}} \sum_{\delta \in \Gamma_\infty} \overline{\chi(\gamma\delta)} f_s\left(\gamma\delta \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} g \right) dx$$
$$= \sum_{\substack{\gamma \in \Gamma_\infty \backslash \Gamma / \Gamma_\infty \\ \gamma \notin \Gamma_\infty}} \overline{\chi(\gamma)} \int_{-\infty}^\infty f_s\left(\gamma \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} g \right) dx.$$

If $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \notin \Gamma_\infty$ then $c = c(\gamma) \neq 0$ and

$$\gamma = \begin{pmatrix} c^{-1} & a \\ & c \end{pmatrix} \begin{pmatrix} & -1 \\ 1 & \end{pmatrix} \begin{pmatrix} 1 & d/c \\ & 1 \end{pmatrix},$$

so the variable change $x \rightarrow x - d/c$ shows that

$$\int_{-\infty}^\infty f_s\left(\gamma \begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} g \right) dx = |c|^{-2s} M(s) f_{s,k}.$$

Thus (59) is satisfied where

$$c(s) = \sum_{\substack{\gamma \in \Gamma_\infty \backslash \Gamma / \Gamma_\infty \\ \gamma \notin \Gamma_\infty}} \chi(\gamma) |c(\gamma)|^{-2s}. \tag{60}$$

Since we are within the region of absolute convergence of the Eisenstein series this Dirichlet series is convergent if $\mathrm{re}(s) > 1$, and (60) shows that it is an analytic function bounded in vertical strips. □

Example. Let us consider

$$E(g, s) = \sum_{\Gamma_\infty \backslash \Gamma} f_{s,0}(\gamma g)$$

where $\Gamma = \mathrm{SL}(2, \mathbb{Z})$. Then in (60), the number of γ with $c(\gamma) = c$ is $\phi(c)$, where (in this example only) ϕ denotes the Euler totient function. Thus

$$c(s) = \sum_{c=1}^{\infty} \phi(c)\, c^{-2s} = \frac{\zeta(2s-1)}{\zeta(2s)}.$$

The analytic continuation of $c(s)$ can be proved directly, and so can the analytic continuation of the Eisenstein series. For example, let $E^*(z, s) = \pi^{-s}\Gamma(s)\zeta(2s)E(g, s)$ where $z = g(i) = x + i\,y \in \mathcal{H}$ and $t > 0$. Let

$$\Theta(t) = \sum_{(m,n) \in \mathbb{Z}^2} e^{-\pi |mz+n|^2 t/y}.$$

It follows from Euler's integral for the Gamma function

$$E^*(z, s) = \frac{1}{2} \int_0^\infty (\Theta(t) - 1)\, t^s \frac{dt}{t}.$$

The Poisson summation formula implies that $\Theta(t) = t^{-1}\Theta(t^{-1})$. From this one gets

$$E^*(z, s) = \frac{1}{2} \int_0^\infty \Theta(t)\,(t^s + t^{1-s}) \frac{dt}{t} - \frac{1}{2s} - \frac{1}{2-2s}.$$

This expression gives the analytic continuation and functional equation. Such methods do not work in general, which is why Selberg's proof of the analytic continuation of the Eisenstein series for arbitrary Γ in 1953 was a breakthrough.

The literature on the analytic continuation of Eisenstein series is quite extensive. Proofs may be found in Borel [4], Cohen and Sarnak [7], Colin de Verdier [8], Efrat [10], Elstrodt [11], Fadeev [12], Harish-Chandra [17], Hejhal [19] (Chapter 6 and Appendix F, with discussion of the literature on p. 225), Kubota [26], Jacquet [23], Langlands [30], Lax and Phillips [31], Moeglin and Waldspurger [32], Müller [33], Osborne and Warner [34], Venkov [42] and Wong [46]. The most general treatments are Langlands' historically important work [30] and the careful modern treatise of Moeglin and Waldspurger [32].

This body of literature all owes something to Selberg, who found three proofs, in 1953, 1957 and 1967. Generally speaking, one shows the analytic continuation of the Eisenstein series and its constant term simultaneously. The basic principle is that the resolvent of an operator has analytic continuation to the complement of its spectrum. Applied as for example in Kubota [26], Venkov [42] or Appendix IV of Langlands [30] to the resolvent of Δ, this gives the analytic continuation of the Eisenstein series to the region $\mathrm{re}(s) > \frac{1}{2}$, $s \notin (\frac{1}{2}, 1]$. In this approach, similar to Selberg's earlier proofs, obtaining the meromorphic continuation to the entire plane then presents some difficulties.

Selberg and Bernstein realized independently that these difficulties could be avoided by combining the resolvent principle with another one, the insight that if a system of inhomogeneous linear equations having analytic coefficients has a *unique* solution, then the solution has meromorphic continuation to wherever the coefficients in the linear equations do. In this method of proof, the linear equations are integral equations of Fredholm type Bernstein's work in the 1980's came later than Selberg's 1967 proof, which was not published by Selberg but shown to Hejhal and to Cohen and Sarnak, and which influenced Efrat [10] and Wong [46]. Selberg's published comments on the idea of using Fredholm equations are in his introduction (written in 1988) to his 1955 Göttingen lectures [39].

Despite Selberg's priority in the use of Fredholm equations to prove the analytic continuation of Eisenstein series, Bernstein's rediscovery was been extremely important in clarifying the issues to the world at large. Moreover, Bernstein also simplified the analytic continuation of Eisenstein series in several variables, and in addition to the analytic continuation of Eisenstein series he gave other applications of the idea, such as to the analytic continuation of the intertwining integrals $M(s)$ and their p-adic analogs ([1], [2]), and his method has become an standard technique in the representation theory of p-adic groups. Thus his work has been extremely influential.

In this section we will prove the meromorphicity of $E(z, s)$, or equivalently the function $E(g, s)$ defined by (57). Thus $f_s = f_{s,0}$ is the SO(2)-fixed vector in P_s^+. The proof we give is based on Jacquet [23]. Instead of using the resolvent of Δ, we will use the analytic continuation of the resolvent of the compact operators $\rho^{\Lambda,T}(\phi)$ with $\phi \in C_c^\infty(G)$, which is a Fredholm operator.

In order that we may treat $E(g, s)$ as a vector element of some function space, we will write $E_s(g) = E(g, s)$ interchangeably. It does not live in $L^2(\Gamma\backslash G)$, but it does live in the space of locally square-integrable functions $L^2_{\text{loc}}(\Gamma\backslash G)$. This is a space whose topology is defined by the set of semi-norms given by the L^2 norm on compact sets. It is a complete, locally convex space with this topology, in other words a Frechet space.

In this section we will denote

$$\hat{\phi}(s) = \chi_{s(1-s)}(\phi),$$

where χ_λ is the character of $\mathcal{H}^\circ$ defined by (15).

Proposition 53. *If* $\text{re}(s) > 1$ *we have*

$$\rho(\phi) E_s = \hat{\phi}(s) E_s. \tag{61}$$

Proof. By (14) we have

$$\rho(\phi) f_{s,0} = \chi_{s(1-s)}(\phi) f_{s,0} = \hat{\phi}(s) f_{s,0}(g)$$

Since $\rho(\phi)$ is an average of *right* translates of $f_{s,0}$, this operator commutes with left translation by γ. Therefore

$$(\rho(\phi) f_{s,0})(\gamma g) = \hat{\phi}(s) f_{s,0}(\gamma g)$$

for all $\gamma \in \Gamma$. Summing over γ we obtain (61). □

Lemma 54. *Let* $\phi \in \mathcal{H}^\circ$. *Then* $\rho(\phi)$ *is self-adjoint if* ϕ *is real-valued.*

Proof. If $g \in \mathrm{SL}_2(\mathbb{R})$, then g and g^{-1} lie in the same SO(2) double coset. Indeed, we can write $g = \kappa d \kappa'$, where $\kappa, \kappa' \in \mathrm{SO}(2)$ and d is diagonal, and

$$d = w d^{-1} w^{-1}, \qquad w = \begin{pmatrix} & -1 \\ 1 & \end{pmatrix} \in \mathrm{SO}(2).$$

Thus if ϕ is real-valued then $\phi(g^{-1}) = \phi(g)$, which implies that (7) is satisfied, so ϕ is self-adjoint. □

Lemma 55. *Let Ω be a compact region of $\mathbb{C}$. Then there exists $\phi \in \mathcal{H}^\circ$ which is real-valued and such that $\hat{\phi}$ is nonvanishing on Ω. Moreover, if $s = \frac{1}{2} + it \in \Omega$ then unless t is real or pure imaginary (so s is real) we may choose ϕ so that $\hat{\phi}(s)$ is not real.*

Proof. From the Taylor expansion $\sin(x)/x = 1 - \frac{1}{3!}x^2 + \frac{1}{5!}x^4 - \ldots$ we see that if $\varepsilon > 0$ is sufficiently small, then $\sin(\varepsilon x)/(\varepsilon x)$ is nonvanishing on Ω. Moreover, if $s = \frac{1}{2} + it \in \Omega$ is given such that t is not real or purely imaginary, then we can also choose ε so that $\sin(\varepsilon x)/(\varepsilon x)$ is not real at $x = s$.

Now we let g_i be a sequence of smooth, even, compactly supported functions that converge uniformly on $\mathbb{R}$ to the $1/2\varepsilon$ times the characteristic function of the interval $[-\varepsilon, \varepsilon]$. By Proposition 27 there is a corresponding sequence of functions $\phi_i \in \mathcal{H}^\circ$ related to the g_i by (22). By Theorem 22 we have

$$\hat{\phi}_i\left(\frac{1}{2} + it\right) = \int_{-\infty}^{\infty} g_i(u) e^{iut} du \longrightarrow \frac{1}{2\varepsilon} \int_{-\varepsilon}^{\varepsilon} e^{iut} du = \frac{1}{\varepsilon t} \sin(\varepsilon t)$$

uniformly on Ω. The statement is now clear. □

Let us denote

$$\mu(s) = c(s) \sqrt{\pi} \frac{\Gamma\left(s - \frac{1}{2}\right)}{\Gamma(s)},$$

so that by (20) with $k = 0$ and (59) the constant term

$$E_0(g, s) = f_{s,0}(g) + \mu(s) f_{1-s,0}(g). \tag{62}$$

Let E_s^T be the function

$$E_s^T(g) = \sum_{\Gamma_\infty \backslash \Gamma} \delta_T(\gamma g) f_{s,0}(\gamma g)$$

where δ_T is defined as in (49).

Lemma 56. *If $T > T_0$ and $g \in \mathcal{G}$ we have*

$$E_s^T(g) = \begin{cases} f_s(g) & \text{if } g \in \mathcal{G}_T; \\ 0 & \text{otherwise.} \end{cases} \tag{63}$$

The function $E_s^T(g)$ is entire as a function of s and is square-integrable if $\mathrm{re}(s) < \frac{1}{2}$. *We have*

$$\Lambda^T E_s = E_s - E_s^T - \mu(s) E_{1-s}^T. \tag{64}$$

Proof. The characterization (63) follows from Lemma 46, the fact that E_s^T is entire is clear, and the square-integrability when $\mathrm{re}(s) < \frac{1}{2}$ is easily checked. Now (64) follows from (51) and (62). □

Proposition 57. *If $\theta \in L_0^2(\Gamma\backslash\mathfrak{H})$, and if $\mathrm{re}(s) > 1$ we have*

$$\int_{\Gamma\backslash\mathfrak{H}} \theta(g) E(g, s) dg = 0.$$

(The integral is absolutely convergent.)

Thus we say that the Eisenstein series is "orthogonal" to the cusp forms, though this is slightly wrong since E_s is not in $L^2(\Gamma\backslash G)$.

Proof. The convergence of the integral follows from the rapid decay of f in Theorem 50. We unfold the integral, using the automorphicity of θ:

$$\begin{aligned} \int_{\Gamma\backslash\mathfrak{H}} \theta(g) \sum_{\Gamma_\infty\backslash\Gamma} f_{s,0}(\gamma g) dg &= \sum_{\Gamma_\infty\backslash\Gamma} \int_{\Gamma\backslash\mathfrak{H}} \theta(\gamma g) f_{s,0}(\gamma g) dg \\ &= \int_{\Gamma\infty\backslash\mathfrak{H}} \theta(g) f_{s,0}(g) dg \\ &= \int_{G_\infty\backslash\mathfrak{H}} \int_{\Gamma_\infty\backslash G_\infty} \theta(ug) f_{s,0}(ug) du dg. \end{aligned}$$

since $f_{s,0}(ug) = f_{s,0}(g)$ this vanishes by the cuspidality of θ. □

Theorem 58. (Selberg) *The Eisenstein series $E(g, s)$ has meromorphic continuation to all s and satisfies the functional equation*

$$E(g, s) = \mu(s) E(g, 1 - s). \tag{65}$$

For all s such that $E(g, s)$ does not have a pole, it is a smooth function of g, and $\Lambda^T E_s$ is square-integrable. The values of s where $E(g, s)$ has a pole are the same as the values of s for which $\mu(s)$ has a pole. There are only a finite number of poles with $\mathrm{re}(s) \geqslant \frac{1}{2}$, and these all lie on the real line. If $E(g, s)$ has a pole with $\mathrm{re}(s) > \frac{1}{2}$, then the residue is square-integrable. If θ is an element of $L_0^2(\Gamma\backslash\mathfrak{H})$, and if s is not a pole of E_s, then

$$\int_{\Gamma\backslash G} E(g, s) \theta(g) dg = 0. \tag{66}$$

(The integral is absolutely convergent.) The function $\mu(s)$ satisfies

$$\mu(s) \mu(1 - s) = 1. \tag{67}$$

If $\mathrm{re}(s) = \frac{1}{2}$, then $|\mu(s)| = 1$.

Proof. Let Ω be a connected, relatively compact open subset of $\mathbb{C}$ that intersects the domain $\{\mathrm{re}(s) > 1\}$ of absolute convergence of $E(g, s)$. By Lemmas 54 and 55 we may choose $\phi \in \mathcal{H}^\circ$ such that $\rho(\phi)$ is self-adjoint and $\hat{\phi}$ is nonvanishing on Ω. We will prove that we have meromorphic continuation of both E_s and μ to Ω.

We will show that $\Lambda^T E_s$ and is meromorphic as a function $\Omega \longrightarrow L^2(\Gamma\backslash\mathfrak{H})$. We will also see that $\mu(s)$ is meromorphic and does not have a pole at s unless $\Lambda^T E_s$ also has a pole there. If this is established it will follow that E_s is a smooth function of g where it does not have a pole in s, since it will satisfy (61); thus $E_s = \Lambda^T E_s + E_s^T + \mu(s) E_{1-s}^T$ is the convolution of a locally integrable function with a compactly supported smooth function, hence smooth.

Let us define

$$F_s^T = \Lambda^T \rho(\phi) E_s^T.$$

Lemma 59. *The function $F_s^T(g)$ is bounded and compactly supported modulo Γ as a function of g. It is entire as a function of s.*

Proof. Let $g \in \mathcal{G}_T$. If $y(g)$ is sufficiently large, then $y(gh) > T$ for all h in the support of ϕ. For such g we have

$$\rho(\phi) E_s^T\left(\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} g\right) = \int \phi(h) E_s^T\left(\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} g h\right) dh = \int \phi(h) f_s\left(\begin{pmatrix} 1 & x \\ & 1 \end{pmatrix} g h\right) dh.$$

This is independent of x, so $\rho(\phi) E_s^T(g)$ agrees with its own constant term when $y(g)$ is sufficiently large. Thus $F_s^T(g) = 0$ for all such g. This proves that $F_s^T(g)$ is compactly supported. On the other hand, it is clearly dominated by $\rho(\phi) f_{|s|}$, which is continuous, and so it is bounded. □

Working in $L^2(\Gamma\backslash\mathfrak{H})$, let

$$e_s^T = -(\Lambda^T \rho(\phi) \Lambda^T - \hat{\phi}(s))^{-1} F_s^T. \tag{68}$$

This is the unique solution to the Fredholm equation

$$(\Lambda^T \rho(\phi) \Lambda^T - \hat{\phi}(s)) e_s^T = F_s^T.$$

By Theorem 45, e_s^T is a meromorphic function of s, analytic as long as $\hat{\phi}(s)$ is not an eigenvalue of the compact operator $\Lambda^T \rho(\phi) \Lambda^T$. We will show that if $\mathrm{re}(s) > 1$ we have

$$\Lambda^T E_s = e_s^T + \mu(s) e_{1-s}^T. \tag{69}$$

Since $\Lambda^T \rho(\phi) \Lambda^T - \hat{\phi}(s)$ is invertible when $\hat{\phi}(s) \notin \mathrm{Spec}(\Lambda^T \rho(\phi) \Lambda^T)$, it is sufficient to show that applying this operator to both sides has the same result. In other words, (69) follows immediately from

$$(\Lambda^T \rho(\phi) \Lambda^T - \hat{\phi}(s)) \Lambda^T E_s = -F_s^T + \mu(s) F_{1-s}^T,$$

which in turn follows from (64) and (61), and the definition of F_s^T.

Since e_s^T is a meromorphic function $\Omega \longrightarrow L^2(\Gamma\backslash\mathfrak{H})$, by (69) the function $\Lambda^T E_s$ is meromorphic wherever $\mu(s)$ is meromorphic. Now let us show that $\mu(s)$ is meromorphic on Ω. Define

$$\tilde{E}_s = E_s^T + e_s^T.$$

Observe that $\tilde{E}_s$ is a meromorphic analytic function of $s \in \Omega$ taking values in $L^2_{\text{loc}}(\Gamma\backslash\mathfrak{H})$, the Frechet space of locally square-integrable functions, since E_s^T is an entire function taking values in $L^2_{\text{loc}}(\Gamma\backslash G)$, and e_s^T is meromorphic on Ω, taking values in $L^2(\Gamma\backslash\mathfrak{H})$. Moreover, $\tilde{E}_s$ is square-integrable if $\text{re}(s) < \frac{1}{2}$ (since then E_s^T is L^2). By (64) and (69) we have

$$E_s = \tilde{E}_s^T + \mu(s)\tilde{E}_{1-s}^T \tag{70}$$

Apply $\rho(\phi)$ to (70). By (61) we have

$$\mu(s)(\rho(\phi) - \hat{\phi}(s))\tilde{E}_{1-s}^T = -(\rho(\phi) - \hat{\phi}(s))\tilde{E}_s^T. \tag{71}$$

Since $(\rho(\phi) - \hat{\phi}(s))\tilde{E}_s^T$ is meromorphic with values in $L^2_{\text{loc}}(\Gamma\backslash\mathfrak{H})$, this proportionality gives the meromorphicity of $\mu(s)$ provided $(\rho(\phi) - \hat{\phi}(s))\tilde{E}_{1-s}$ is not identically zero. To see this, we may take $\text{re}(s) > \frac{1}{2}$ such that $\hat{\phi}(s)$ is not real. Then $\tilde{E}_{1-s}$ is a vector in the Hilbert space $L^2(\Gamma\backslash G)$, and since $\rho(\phi)$ induces a self-adjoint operator, its eigenvalues are real, and $\hat{\phi}(s)$ is not, so $\tilde{E}_s$ cannot be a eigenvector.

By (71) and the fact that $\hat{\phi}(s) = \hat{\phi}(1-s)$ we have (67), which together with (70) implies the functional equation (65).

Let $\theta \in L^2_0(\Gamma\backslash\mathfrak{H})$ be a cusp form. Then (66) is proved in Proposition 57, when $\text{re}(s) > 1$. but both sides are meromorphic functions of s. Indeed, the integral on the left side is convergent where E_s does not have a pole, since $E_s = \Lambda^T E_s + E_s^T + \mu(s)E_{1-s}^T$, where $\Lambda^T E_s$ is square integrable while $E_s^T(g)$ and $\mu(s)E_{1-s}^T(g)$ are of polynomial growth in $y = y(g)$ as $y \longrightarrow \infty$, while θ is of rapid decay. Thus (66) remains true by analytic continuation for all s where E_s is not polar.

Next we show that if E_s has a pole at $s = s_0$ then so does $\mu(s)$. If not, observe that the constant term (62) has no pole, so the residue $\theta = \text{res}_{s=s_0} E_s$ is in $L^2_0(\Gamma\backslash\mathfrak{H})$. Consider the integral

$$\int_{\Gamma\backslash\mathfrak{H}} E_s(z)\overline{\theta(z)}\frac{dx \wedge dy}{y^2}.$$

By (66), this is zero, but taking the residue at $s = s_0$ produces the L^2 norm of $\theta(z)$, which is not zero. This contradiction proves that the poles of E_s and the poles of $\mu(s)$ are the same.

We will show that if $\text{re}(s_0) > \frac{1}{2}$ and E_s has a pole at $s = s_0$, then s_0 is real. If not, then by Lemma 55 we can choose ϕ so that $\hat{\phi}(s_0)$ is not real. Then since $\Lambda^T\rho(\phi)\Lambda^T$ is self-adjoint, $\hat{\phi}(s_0)$ cannot be an eigenvector, and so by Theorem 45 and (68), e_s^T has no pole at s_0. On the other hand, E_s^T is entire. Thus $\tilde{E}_s^T = E_s^T + e_s^T$ has no pole at $s = s_0$. Similarly, since $\hat{\phi}(1-s) = \hat{\phi}(s)$, e_{1-s}^T has no pole at $s = s_0$. This means that $\tilde{E}_{1-s}^T = e_{1-s}^T + E_{1-s}^T$ has no pole at $s = s_0$. Since $\text{re}(1-s_0) < \frac{1}{2}$, $E_{1-s_0}^T$ is square-integrable, and so $\tilde{E}_{1-s_0} \in L^2(\Gamma\backslash\mathfrak{H})$. Now $\rho(\phi)$, though not compact, is a self-adjoint bounded operator, so its eigenvalues are real, and $\hat{\phi}(\rho)$ is not. Therefore $(\rho(\phi) - \hat{\phi}(s_0))\tilde{E}_{1-s_0}^T$ is nonzero. Since we are assuming E_s has a pole at $s = s_0$, $\mu(s)$ has a pole there. We see that the left-hand side of (71) has a pole but the right-hand side does not, which is a contradiction.

Next observe that $\mu(\bar{s}) = \overline{\mu(s)}$, so $|\mu(s)| = |\mu(\bar{s})|$. If $\mathrm{re}(s) = \frac{1}{2}$, this means $|\mu(s)| = |\mu(1-s)|$ so by (67) we have $|\mu(s)| = 1$ on the line $\mathrm{re}(s) = \frac{1}{2}$. This implies that E_s has no poles on the line $\mathrm{re}(s) = \frac{1}{2}$.

Finally, if $\mathrm{re}(s_0) > \frac{1}{2}$ and E_s has a pole at s_0, then (64) shows that the residue is

$$\mathrm{res}_{s=s_0}\Lambda^T E_s + M E^T_{1-s_0},$$

where $M = \mathrm{res}_{s=s_0}\mu(s)$. Both terms are square integrable. □

11 The Maass-Selberg relation

Theorem 60. (Selberg) *We have*

$$\int_{\Gamma\backslash\mathfrak{H}} \Lambda^T E(z,s)\Lambda^T E(z,s')\frac{dx\wedge dy}{y^2} = \frac{T^{s+s'-1} - \mu(s)\mu(s')T^{1-s-s'}}{s+s'-1} + \frac{\mu(s)T^{s'-s} - \mu(s')T^{s-s'}}{s'-s}. \tag{72}$$

Harish-Chandra and Langlands described this as the *Maass-Selberg relation* but it is more correctly attributed to Selberg. However, the name has stuck, and we will refer to it as the Maass-Selberg relation. Selberg gave two proofs. One was in the original version of the Göttingen notes, the other is in a comment added when these were published in his collected works. We will reproduce the latter proof.

Proof. Assuming $\mathrm{re}(s') > \mathrm{re}(s) + 1$ and $\mathrm{re}(s') > 1$, we may proceed as follows. Let $\delta_T(y) = 1$ if $y > T$, 0 otherwise.

$$\int_{\Gamma\backslash\mathfrak{H}} \Lambda^T E(z,s)\Lambda^T E(z,s')\frac{dx\wedge dy}{y^2} = \int_{\Gamma\backslash\mathfrak{H}} E(z,s)\Lambda^T E(z,s')\frac{dx\wedge dy}{y^2} =$$

$$\int_{\mathcal{F}} E(z,s)(E(z,s') - \delta_T(y)y^{s'})\frac{dx\wedge dy}{y^2} - \mu(s')\int_{\mathcal{F}_T} E(z,s)y^{1-s'}\frac{dx\wedge dy}{y^2}.$$

The two terms may be handled separately as follows. The first term equals

$$\int_{\mathcal{F}} E(z,s)(E(z,s') - y^{s'})\frac{dx\wedge dy}{y^2} + \int_{\mathcal{F}_T} E(z,s)y^{s'}\frac{dx\wedge dy}{y^2} =$$

$$\int_{\mathcal{F}} E(z,s)\sum_{\gamma\in\Gamma_\infty\backslash(\Gamma-\Gamma_\infty)} y(\gamma z)^{s'}\frac{dx\wedge dy}{y^2} + \int_{\mathcal{F}_T} E(z,s)y^{s'}\frac{dx\wedge dy}{y^2} =$$

$$\int_{\mathcal{F}_\infty-\mathcal{F}} E(z,s)y^{s'}\frac{dx\wedge dy}{y^2} + \int_{\mathcal{F}_T} E(z,s)y^{s'}\frac{dx\wedge dy}{y^2},$$

where $\mathcal{F}_\infty = \{x + iy | |x| < \frac{1}{2}\}$ is a fundamental domain for Γ_∞. Since the disjoint union of $\mathcal{F}_\infty - \mathcal{F}$ and $\mathcal{F}_T$ is $\{x + iy | |x| < \frac{1}{2}, y < T\}$, this gives

$$\int_0^T (y^s + \mu(s) y^{1-s}) y^{s'} \frac{dy}{y^2} = \frac{T^{s+s'-1}}{s+s'-1} - \frac{\mu(s) T^{s'-s}}{s'-s}.$$

The second term gives

$$\mu(s') \int_T^\infty (y^s + \mu(s) y^{1-s}) y^{1-s'} \frac{dy}{y^2} = -\mu(s') \frac{T^{s-s'}}{s-s'} + \mu(s)\mu(s') \frac{T^{1-s-s'}}{1-s-s'}.$$

The identity is now proved for $\mathrm{re}(s')$ sufficiently large, but since both sides are meromorphic, it now follows for all s' where they are not polar. □

As a special case, we may take $s = \sigma - it$, $s' = \sigma + it$ with σ, t real, to obtain

$$\int_{\Gamma\backslash\mathfrak{H}} \left|\Lambda^T E(z, \sigma + it)\right|^2 \frac{dx \wedge dy}{y^2} = \frac{T^{2\sigma-1} - |\mu(\sigma+it)|^2 T^{1-2\sigma}}{2\sigma - 1} + \frac{\mu(\sigma - it) T^{2it} - \mu(\sigma + it) T^{-2it}}{2it}. \tag{73}$$

Corollary 61. *On the region $\frac{1}{2} < \sigma < \frac{3}{2}$, $t > 1$ the factor $\mu(\sigma + it)$ is bounded.*

Proof. Fix any $T > T_0$, with T_0 as in Section 10. Since (73) is positive, writing $|\mu(\sigma + it)| = M$, we have

$$M^2 < bM + c, \qquad b = \frac{2\sigma - 1}{t} T^{2\sigma-1},$$

where

$$b = \frac{2\sigma - 1}{t} T^{2\sigma - 1}, \qquad c = T^{4\sigma - 2}.$$

Thus M is bounded by the positive root of the quadratic equation $M^2 - bM - c = 0$, that is, by $\frac{1}{2}(b + \sqrt{b^2 + 4c})$. The statement follows since b and c are both bounded in the region in question. □

Corollary 62. *We have*

$$\int_{\Gamma\backslash\mathfrak{H}} \left|\Lambda^T E(z, \frac{1}{2} + it)\right|^2 \frac{dx \wedge dy}{y^2} = 2\log(T) - \frac{\mu'}{\mu}\left(\frac{1}{2} + it\right) + \frac{\mu\left(\frac{1}{2} - it\right) T^{2it} - \mu\left(\frac{1}{2} + it\right) T^{-2it}}{2it}.$$

Proof. This follows from (73) on letting $\sigma \longrightarrow \frac{1}{2}$ and using L'Hôpital's rule, remembering that $\mu(\frac{1}{2} + it)\mu(\frac{1}{2} - it) = 1$. □

Corollary 63. *Let $\sigma \in (\frac{1}{2}, 1]$ be a place where $E(z, s)$ and $\mu(s)$ has a pole. If $\eta \in L^2(\Gamma\backslash\mathfrak{H})$ is the residue of $E(z, s)$ at σ then the residue of μ is $\langle \eta, \eta \rangle$. In particular $\rho > 0$.*

Proof. The Eisenstein series $E(z, s)$ and $\mu(s)$ are both real valued when $s \in \mathbb{R}$, so ρ is real. Let be the residue of $E(z, s)$. Then multiplying (72) by $(s - \sigma)(s' - \sigma)$ and taking the limit as $s, s' \longrightarrow \sigma$ we obtain

$$\frac{T^{s+s'-1} - \mu(s)\mu(s')T^{1-s-s'}}{s+s'-1} + \frac{\mu(s)T^{s'-s} - \mu(s')T^{s-s'}}{s'-s}$$

$$\langle \Lambda^T\eta, \Lambda^T\eta\rangle = \frac{-\rho^2 T^{1-2\sigma}}{2\sigma - 1} + \rho$$

This proves that $\rho > 0$, and taking the limit as $T \longrightarrow \infty$ gives $\langle \eta,\eta\rangle = \rho$. □

12 Spectral expansion

The spectral expansion was obtained by Roelcke [37], modulo the analytic continuation of the Eisenstein series. Numerous accounts are in the literature, of which Godement [15], [16] is a good and influential one. In this section we discuss the spectral expansion for $L^2(\Gamma\backslash\mathfrak{H})$ where as in Section 6 the group has a single cusp. We will assume the analytic continuation of the Eisenstein series (Theorem 61).

Let α be a K-finite element of $C_c^\infty(G_\infty\backslash G)$. The *incomplete theta series*

$$\theta_\alpha(g) = \sum_{\Gamma_\infty\backslash\Gamma} \alpha(\gamma g) \tag{74}$$

is something like an Eisenstein series but it is not a Δ-eigenfunction. (Godement's term "incomplete theta series" seems something of a misnomer.) If $f \in C^\infty(\Gamma\backslash G)$ let

$$f_0(g) = \int_{\Gamma_\infty\backslash G_\infty} f(u\,g)\,d\,u$$

be its constant term, which is in $C^\infty(G_\infty\backslash G)$. As a notational point, if f is itself indexed, i.e. $f = f_i$ we will write f_i^0 instead of $f_{i,0}$ for the constant term.

Proposition 64. *The incomplete theta series and constant term maps are adjoints; that is, if $f \in C^\infty(\Gamma\backslash G)$ and $\alpha \in C_c^\infty(G_\infty\backslash G)$ we have*

$$\int_{\Gamma\backslash G} \theta_\alpha(g)\,\overline{f(g)}\,d\,g = \int_{G_\infty\backslash G} \alpha(g)\,\overline{f_0(g)}\,d\,g. \tag{75}$$

Proof. The left side is

$$\int_{\Gamma\backslash G} \sum_{\gamma\in\Gamma_\infty\backslash\Gamma} \alpha(\gamma g)\,\overline{f(g)}\,d\,g = \int_{\Gamma_\infty\backslash G} \alpha(g)\,\overline{f(g)}\,d\,g =$$

$$\int_{G_\infty\backslash G} \int_{\Gamma_\infty\backslash G_\infty} \alpha(u\,g)\,\overline{f(u\,g)}\,d\,u\,d\,g = \int_{G_\infty\backslash G} \alpha(g) \int_{\Gamma_\infty\backslash G_\infty} \overline{f(u\,g)}\,d\,u\,d\,g$$

which equals the right side. □

The incomplete theta series are compactly supported modulo Γ, hence are square-integrable. For this reason they are easier to work with than the Eisenstein series themselves.

Proposition 65. *$L_0^2(\Gamma\backslash G)$ is the orthogonal complement in $L^2(\Gamma\backslash G)$ of the closed subspace spanned by the incomplete theta series.*

Proof. Immediate from Proposition 64, since the cuspidal spectrum is characterized by vanishing of its constant terms. □

For the remainder of this section, we make the simplifying assumption that the character χ to be trivial and we only consider functions on G which are right invariant by K, that is, which may be regarded as functions on $\mathfrak{H}$. We will therefore denote $E(g, f_{s,0}, 1)$ as just $E(g, s)$, or as $E(z, s)$ where $z = g(i)$. We will also denote $f_s = f_{s,0}$. We will write the constant term in the form

$$E_0(g, s) = f_s(g) + \mu(s)\, f_{1-s}(g). \tag{76}$$

(See Proposition 52.)

Using Theorem 50, let ξ_i be a basis of $L_0^2(\Gamma\backslash G)$. Also, note that by Theorem 61 there may be a finite number of poles of the Eisenstein series $E(z, s)$ at locations $\sigma_j \in (\frac{1}{2}, 1]$ $(j = 1, \cdots, N)$. These are also poles of $\mu(s)$. Let ρ_j be the residue of $\mu(s)$ at $s = \sigma_j$. By Corollary 63, $\rho > 0$ and if we define

$$\eta_j = \frac{1}{\sqrt{\rho_j}} \operatorname{res}_{s=\sigma_j} E(g, s), \tag{77}$$

then the η_j have norm 1 in $L^2(\Gamma\backslash\mathfrak{H})$.

Proposition 66. *The constant term of η_j is $\sqrt{\rho}\, f_{1-\sigma_j}$. The functions η_j are real valued, square integrable and orthogonal to the cusp forms.*

Proof. Part (i) is immediate from (76) since the first term on the right has no pole but the second one does. The η_j are real valued since the σ_j are real. The fact that the residue is square integrable is already noted in Theorem 58, and the fact that the constant term of η_j is $\sqrt{\rho}\, f_{1-\sigma_j}$ follows from (76). Taking the residue in Proposition 57 shows that the η_j are orthogonal to the cusp forms. □

Proposition 67. *We have*

$$\langle \theta_\alpha, \eta_j \rangle = \sqrt{\rho_j}\, \check{\alpha}(\sigma_j).$$

Proof. This follows from Proposition 64, Proposition 66 and from the definition of $\check{\alpha}$. □

Proposition 68. *Let $\alpha \in C_c^\infty(G_\infty\backslash G/K)$. Define*

$$\check{\alpha}(s) = \int_{G_\infty\backslash G} \alpha(g)\, f_{1-s}(g)\, dg. \tag{78}$$

This is an entire function of s and $t \to \breve{\alpha}(\sigma + i\,t)$ is of Schwartz class for all real σ. For real α we have

$$\alpha(g) = \frac{1}{2\pi i}\int_{\sigma - i\infty}^{\sigma + i\infty} \breve{\alpha}(s)\, f_s(g)\, d\,s. \tag{79}$$

(The contour integral is over the vertical line with real part σ.) Moreover if $\sigma > 1$ we have

$$\theta_\alpha(g) = \frac{1}{2\pi i}\int_{\sigma - i\infty}^{\sigma + i\infty} \breve{\alpha}(s)\, E(g, s)\, d\,s \tag{80}$$

Proof. We have

$$\breve{\alpha}(s) = \int_0^\infty \alpha\begin{pmatrix} y^{1/2} & \\ & y^{-1/2} \end{pmatrix} y^{-s}\frac{d\,y}{y}.$$

Since α is compactly supported and smooth, $t \to \breve{\alpha}(\sigma + i\,t)$ is the Fourier transform of a Schwartz function, hence is Schwartz itself. By the Mellin inversion formula

$$\alpha\begin{pmatrix} y^{1/2} & \\ & y^{-1/2} \end{pmatrix} = \frac{1}{2\pi i}\int_{\sigma - i\infty}^{\sigma + i\infty} \breve{\alpha}(s)\, y^s\, d\,s. \tag{81}$$

This verifies (79) on the diagonal, and since both sides are left invariant by G_∞ and right invariant by K, the general case follows. If $\sigma > 1$ we may replace g by γg then sum over $\gamma \in \Gamma_\infty \backslash \Gamma$ to obtain (80). □

Proposition 69. *Let $\alpha, \beta \in C_c^\infty(G_\infty \backslash G / K)$. Then*

$$\int_{\Gamma\backslash G} \theta_\alpha(g)\overline{\theta_\beta(g)}\, dg =$$

$$\sum_{j=1}^N \langle \theta_\alpha(g), \eta_j\rangle \overline{\langle \theta_\beta(g), \eta_j\rangle} + \frac{1}{4\pi}\int_{-\infty}^{\infty} \left\langle \theta_\alpha, E\left(\,\cdot\,, \frac{1}{2} + i\,t\right)\right\rangle \overline{\left\langle \theta_\beta, E\left(\,\cdot\,, \frac{1}{2} + i\,t\right)\right\rangle}\, dt. \tag{82}$$

Proof. Take $\sigma > 1$. Using Proposition 64 and (80), the inner product on the left-hand side equals

$$\int_{G_\infty\backslash G} \theta_\alpha^0(g)\, \overline{\beta(g)}\, d\,g = \int_{G_\infty\backslash G} \frac{1}{2\pi i}\int_{\sigma - i\infty}^{\sigma + i\infty} \breve{\alpha}(s)\, E_0(g, s)\, d\,s\, \overline{\beta(g)}\, d\,g.$$

Since $E_0(g, s) = f_s(g) + \mu(s) f_{1-s}(g)$ with $\mu(s)$ bounded by Corollary 61, there is no difficulty inerchanging the order of integration to obtain

$$\frac{1}{2\pi i}\int_{\sigma - i\infty}^{\sigma + i\infty} \breve{\alpha}(s)\int_{G_\infty\backslash G} (f_s(g) + \mu(s) f_{1-s}(g))\, \overline{\beta(g)}\, d\,g\, d\,s.$$

Moreover, using Proposition 61 and the fact that β is compactly supported while $\breve{\alpha}$ is of rapid decay on vertical lines, it is legitimate to move the path of integration left to $\sigma = \frac{1}{2}$. At each σ_j there is a residue

$$\breve{\alpha}(\sigma_j)\int_{G_\infty\backslash G} \rho_j f_{1-\sigma_j}(g)\, \overline{\beta(g)}\, d\,g = \rho_j\, \breve{\alpha}(\sigma_j)\, \overline{\breve{\beta}(\sigma_j)} = \langle \theta_\alpha, \eta_j\rangle\, \overline{\langle \theta_\beta, \eta_j\rangle},$$

where we have used (78) and Proposition 67. These terms will thus appear, together with

$$\frac{1}{2\pi i}\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty}\check{\alpha}(s)\int_{G_\infty\backslash G}(f_s(g)+\mu(s)f_{1-s}(g))\overline{\beta(g)}\,dg\,ds =$$
$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\check{\alpha}\Big(\frac{1}{2}+it\Big)\Big[\overline{\check{\beta}\Big(\frac{1}{2}+it\Big)}+\mu\Big(\frac{1}{2}+it\Big)\overline{\check{\beta}\Big(\frac{1}{2}-it\Big)}\Big]dt$$

Comparing this with (82), what we need to show is that

$$\frac{1}{4\pi}\int_{-\infty}^{\infty}\Big\langle\theta_\alpha,E(\,\cdot\,,\frac{1}{2}+it)\Big\rangle\overline{\Big\langle\theta_\beta,E(\,\cdot\,,\frac{1}{2}+it)\Big\rangle}dt =$$
$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\check{\alpha}\Big(\frac{1}{2}+it\Big)\Big[\overline{\check{\beta}\Big(\frac{1}{2}+it\Big)}+\mu\Big(\frac{1}{2}+it\Big)\overline{\check{\beta}\Big(\frac{1}{2}-it\Big)}\Big]dt. \tag{83}$$

Proposition 64 implies that

$$\Big\langle\theta_\beta,E(\,\cdot\,,\frac{1}{2}+it)\Big\rangle =$$
$$\int_{G_\infty\backslash G}\beta(g)\Big(f_{\frac{1}{2}-it}(g)+\mu\Big(\frac{1}{2}-it\Big)f_{\frac{1}{2}+it}(g)\Big)dg =$$
$$\check{\beta}\Big(\frac{1}{2}+it\Big)+\mu\Big(\frac{1}{2}-it\Big)\check{\beta}\Big(\frac{1}{2}-it\Big).$$

Multiplying the complex conjugate of this identity by

$$\Big\langle\theta_\alpha,E(\,\cdot\,,\frac{1}{2}+it)\Big\rangle=\int_{\Gamma\backslash G}\theta_\alpha(g)\,E\Big(g,\frac{1}{2}-it\Big)dg,$$

the left-hand side of (83) equals

$$\frac{1}{4\pi}\int_{-\infty}^{\infty}\int_{\Gamma\backslash G}\theta_\alpha(g)\Big[\overline{\check{\beta}\Big(\frac{1}{2}+it\Big)}+\mu\Big(\frac{1}{2}+it\Big)\overline{\check{\beta}\Big(\frac{1}{2}-it\Big)}\Big]E\Big(g,\frac{1}{2}-it\Big)dg\,dt =$$
$$\frac{1}{4\pi}\int_{-\infty}^{\infty}\int_{\Gamma\backslash G}\theta_\alpha(g)\Big[\overline{\check{\beta}\Big(\frac{1}{2}+it\Big)}E\Big(g,\frac{1}{2}-it\Big)+\overline{\check{\beta}\Big(\frac{1}{2}-it\Big)}E\Big(g,\frac{1}{2}+it\Big)\Big]dg\,dt,$$

where we have used the functional equation of the Eisenstein series. Observe that there are two terms, but these are equal since the change of variables $t\longmapsto -t$ interchanges them, so the the left-hand side of (83) equals

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{\Gamma\backslash G}\theta_\alpha(g)\,\overline{\check{\beta}\Big(\frac{1}{2}+it\Big)}E\Big(g,\frac{1}{2}-it\Big)dg\,dt$$

We now use Proposition 64 again to obtain

$$\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{G_\infty\backslash G}\alpha(g)\overline{\breve{\beta}\left(\frac{1}{2}+it\right)}\left[f_{\frac{1}{2}-it}(g)+\mu\left(\frac{1}{2}-it\right)f_{\frac{1}{2}+it}(g)\right]dg\,dt \;=\; \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{G_\infty\backslash G}\overline{\breve{\beta}\left(\frac{1}{2}+it\right)}\left[\breve{\alpha}\left(\frac{1}{2}+it\right)+\mu\left(\frac{1}{2}-it\right)\breve{\alpha}\left(\frac{1}{2}-it\right)\right]dg\,dt.$$

Making the variable change $t\longmapsto -t$ in the second term (but not the first) gives (83). □

Let $L^2_{\text{res}}(\Gamma\backslash\mathfrak{H})$ be the finite-dimensional linear span of the residues η_j of $E(z,s)$; we call it the *residual* part of $L^2(\Gamma\backslash\mathfrak{H})$. The set of eigenvalues ρ_j of Δ that occur in $L^2_{\text{res}}(\Gamma\backslash\mathfrak{H})$ are referred to as the *residual spectrum*. Similarly, the eigenvalues λ_j of Δ that occur in $L^2_0(\Gamma\backslash\mathfrak{H})$ comprise the *cuspidal spectrum*; let ξ_j be an orthonormal basis of cusp forms such that $\Delta\xi_j=\lambda_j\xi_j$. Let $L^2_{\text{disc}}(\Gamma\backslash\mathfrak{H}) = L^2_0(\Gamma\backslash\mathfrak{H}) \oplus L^2_{\text{res}}(\Gamma\backslash\mathfrak{H})$. Thus $L^2_{\text{disc}}(\Gamma\backslash\mathfrak{H})$ decomposes into a direct sum of one-dimensional eigenspaces for Δ; the union of the ρ_j and λ_j are called the *discrete spectrum*. Let $L^2_{\text{cont}}(\Gamma\backslash\mathfrak{H})$ be the orthogonal complement of $L^2_{\text{disc}}(\Gamma\backslash\mathfrak{H})$.

Theorem 70. *Let $f\in L^2(\Gamma\backslash\mathfrak{H})$. Then there exists a function $\hat{f}\in L^2(\mathbb{R})$ that satisfies $\hat{f}(t)=\mu\left(\frac{1}{2}-it\right)\hat{f}(-t)$ such that*

$$\langle f_1,f_2\rangle=\sum_j\langle f_1,\eta_j\rangle\overline{\langle f_2,\eta_j\rangle}+\sum_j\langle f_1,\xi_j\rangle\overline{\langle f_2,\xi_j\rangle}+\frac{1}{4\pi}\int_{-\infty}^{\infty}\hat{f}_1(t)\,\overline{\hat{f}_2(t)}\,dt$$

If f is a cusp form then $\hat{f}(t)=0$, while if f is incomplete theta series we have

$$\hat{f}(t)=\left\langle f,E\left(\cdot\,,\frac{1}{2}+it\right)\right\rangle.$$

The map $f\longmapsto\hat{f}$ is continuous.

Proof. Let H be the Hilbert space

$$\left(\bigoplus_{\rho_j\in\text{residual spectrum}}\mathbb{C}\right)\oplus\left(\bigoplus_{\lambda_j\in\text{cuspidal spectrum}}\mathbb{C}\right)\oplus L^2(\mathbb{R}),$$

and let V be the linear span of $L^2_0(\Gamma\backslash\mathfrak{H})$ and the vector space of incomplete theta series. The function $\hat{f}$ is defined on this dense subspace as in the statement of the Proposition. Then if $f\in V$ and $L(f)\in H$ is

$$\left(\langle f,\eta_1\rangle,\langle f,\eta_2\rangle,\cdots,\langle f,\xi_1\rangle,\langle f,\xi_2\rangle,\cdots,\hat{f}\right),$$

then by Proposition 69, the map $L\colon V\longrightarrow H$ is an isometry. Since V is dense in $L^2(\Gamma\backslash\mathfrak{H})$, we can extend L to all of $L^2(\Gamma\backslash\mathfrak{H})$ in the following way. If $f\in L^2(\Gamma\backslash\mathfrak{H})$ find a sequence $f_i\in V$ converging to f; using the isometry property of L, Lf_i is a Cauchy sequence, so it has a limit in the Hilbert space H, and the Theorem is proved. □

To proceed further, we need a bit more information out of the Maass-Selberg relations. If f is a smooth, compactly supported function, we want to assign a meaning to integrals such as

$$\int_{-\infty}^{\infty} \frac{f(u)}{u} du.$$

We take this to be the principal value

$$\lim_{\varepsilon \longrightarrow 0} \left[\int_{-\infty}^{-\varepsilon} \frac{f(u)}{u} du + \int_{\varepsilon}^{\infty} \frac{f(u)}{u} du \right].$$

This equals

$$\int_{-\infty}^{\infty} \frac{f(u) - f(-u)}{u} du.$$

We will always employ this meaning in the sequel, particularly in the following Lemma.

Lemma 71. *Let ψ be a smooth, compactly supported function. Then*

$$\lim_{T \longrightarrow \infty} \int_{-\infty}^{\infty} \frac{\psi(u) T^{iu}}{iu} du = \pi \psi(0).$$

Proof. We may separate ψ into its odd and even parts, and treat these two cases separately. If ψ is odd, then $\psi_1(u) = \psi(u)/iu$ is continuous at 0, and

$$\int_{-\infty}^{\infty} \psi_1(u) T^{iu} du \longrightarrow 0 = \pi \psi(0)$$

by the Riemann-Lebesgue Lemma. Thus we may assume that ψ is even. In this case the integral is

$$\int_{-\infty}^{\infty} \psi(u) \left(\frac{T^{iu} - T^{-iu}}{2iu} \right) du = \int_{-\infty}^{\infty} \psi(u) \frac{\sin(\log(T) u)}{u} du \longrightarrow \pi \psi(0).$$

The last step needs a bit of justification. We will make use of the identity

$$\int_{-\infty}^{\infty} e^{-u^2} \frac{\sin(xu)}{u} du = \sqrt{\pi} \int_{-x/2}^{x/2} e^{-t^2} dt.$$

Thus

$$\int_{-\infty}^{\infty} \psi(u) \frac{\sin(xu)}{u} du = \int_{-\infty}^{\infty} \frac{\psi(u) - \psi(0) e^{-u^2}}{u} \sin(ux) du + \psi(0) \sqrt{\pi} \int_{-x/2}^{x/2} e^{-t^2} dt.$$

The first term tends to zero by the Riemann-Lebegue Lemma, while the second has the limit $\pi \psi(0)$ as $x = \log(T) \longrightarrow \infty$. □

Theorem 72. *Let ϕ be a smooth, compactly supported function on $\mathbb{R}$ satisfying*

$$\phi(t) = \mu\left(\frac{1}{2} - it \right) \phi(-t), \tag{84}$$

and consider

$$\Phi(g)=\frac{1}{4\pi}\int_{-\infty}^{\infty}\phi(t)E\left(g,\frac{1}{2}+it\right)dt.$$

Then $\Phi\in L^2(\Gamma\backslash\mathfrak{H})$, *and*

$$\int_{\Gamma\backslash\mathfrak{H}}|\Phi(g)|^2dg=\frac{1}{4\pi}\int_{-\infty}^{\infty}|\phi(t)|^2dt.$$

Proof. Let

$$\Phi_T(g)=\frac{1}{4\pi}\int_{-\infty}^{\infty}\phi(t)\Lambda^TE\left(g,\frac{1}{2}+it\right)dt.$$

This is clearly square integrable since Λ^T is. We will prove that

$$\int_{\Gamma\backslash\mathfrak{H}}|\Phi_T(g)|^2dg=\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\frac{\psi(u)T^{iu}}{iu}du \tag{85}$$

where we use the principal value and

$$\psi(u)=\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\phi(u+t)\overline{\phi(t)}dt.$$

By the Maass-Selberg relation, the left-hand side of (85) is

$$\frac{1}{(4\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(t)\overline{\phi(t')}\left[\frac{T^{i(t-t')}-\mu\left(\frac{1}{2}+it\right)\mu\left(\frac{1}{2}-it'\right)T^{-i(t-t')}}{i(t-t')}\right]dtdt' \; -$$
$$\frac{1}{(4\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(t)\overline{\phi(t')}\left[\frac{\mu\left(\frac{1}{2}+it\right)T^{-i(t'+t)}-\mu\left(\frac{1}{2}-it'\right)T^{i(t+t')}}{i(t'+t)}\right]dtdt'.$$

Written this way, these integrands are clearly bounded since $\mu\left(\frac{1}{2}+it\right)$ and $\mu\left(\frac{1}{2}-it\right)$ are complex conjugates of absolute value 1 when t is real. The following manipulations destroy the absolute convergence of the integral but are justified as long as we agree to use the principal value. We make use of $\phi(t)\mu\left(\frac{1}{2}+it\right)=\phi(-t)$ and $\overline{\phi(t')}\,\mu\left(\frac{1}{2}-it'\right)=\overline{\phi(-t')}$ to eliminate each occurrence of μ. Then wherever $\phi(-t)$ or $\overline{\phi(-t')}$ occurs, we replace t or t' by its negative to obtain four equal terms. We get

$$\frac{1}{(2\pi)^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\phi(t)\overline{\phi(t')}\frac{T^{i(t-t')}}{i(t-t')}dtdt',$$

and (85) follows.

Now it follows from the Lemma that

$$\int_{\Gamma\backslash\mathfrak{H}}|\Phi(g)|^2dg=\lim_{T\longrightarrow\infty}\int_{\Gamma\backslash\mathfrak{H}}|\Phi_T(g)|^2dg=\frac{1}{4\pi}\int_{-\infty}^{\infty}|\phi(t)|^2dt.$$

□

Corollary 73. *The image of* $L^2(\Gamma\backslash\mathfrak{H})$ *under the map* $f\longmapsto\hat{f}$ *in (70) is the space* $L^2(\mathbb{R};\mu)$ *of square integrable functions on* $\mathbb{R}$ *that satisfy* $\hat{f}(t)=\mu\left(\frac{1}{2}-it\right)\hat{f}(-t)$.

Proof. The issue is surjectivity. It follows from Theorem 72 that the image of this map contains a dense subspace, since we have constructed the inverse map on the subspace of smooth compactly supported functions in $L^2(\mathbb{R})$ satisfying $\hat{f}(t) = \mu\left(\frac{1}{2} - it\right)\hat{f}(-t)$. □

We now have constructed an isomorphism

$$L^2(\Gamma\backslash\mathfrak{H}) \cong \bigoplus \left(\bigoplus_{\rho_j \in \text{residual spectrum}} \mathbb{C}\right) \oplus \left(\bigoplus_{\lambda_j \in \text{cuspidal spectrum}} \mathbb{C}\right) \oplus L^2(\mathbb{R}; \mu),$$

and so we may write, for any $f \in L^2(\Gamma\backslash\mathfrak{H})$

$$f = \sum_j \langle f, \eta_j \rangle \eta_j + \sum_j \langle f, \xi_j \rangle \xi_j + \frac{1}{4\pi} \int_{-\infty}^{\infty} \hat{f}(t) E\left(\cdot, \frac{1}{2} + it\right) dt.$$

This is the *spectral expansion*.

13 Liftings and the Trace Formula

One of the most interesting applications of the trace formula is to liftings of automorphic forms. The method involves comparison of two different trace formulae, on different groups, leading to the conclusion that automorphic forms on one group can be lifted to automorphic forms on the other.

Jacquet and Langlands [22] gave an early application in a lifting from automorphic forms on a division algebra to GL_2. A variation of this theme noted by Gelbart and Jacquet [13] is probably the simplest example of this type, since in this case neither trace formula involves a continuous spectrum. It will be convenient to switch to an adelic point of view, but we think the reader who has read Section 5 will not have trouble making the transition.

If D_1 and D_2 are central division algebras over a field F, then $D_1 \otimes D_2 \cong \mathrm{Mat}_k(D_3)$ for some D_3 and k, and $D_1, D_2 \to D_3$ is an associative multiplication on the set $B(F)$ of isomorphism classes of central division algebras. Thus $B(F)$ becomes a group, called the *Brauer group*.

If D is a central division algebra over F then the dimension of D is a square d^2, and if E/F is any field extension of degree d which can be embedded in D then $E \otimes D \cong \mathrm{Mat}_d(E)$. Thus a division ring is a Galois twisted form of a matrix ring. The composite map

$$D \to E \otimes D \cong \mathrm{Mat}_d(E) \to E,$$

the last map being either the trace or determinant, takes values in F, and gives us the *reduced trace* or *reduced norm*.

The Brauer group of a local or global field F admits a simple and beautiful description related to the reciprocity laws of class field theory. See Section 1 of Chapter 6 ("Local Class Field Theory" by J.-P. Serre) and Section 9 of Chapter 7 ("Global Class Field Theory" by J. Tate) in Cassels and Fröhlich [6].

Let F be a global field, $\mathbb{A}$ its adele ring, and D a central division algebra of degree p^2 over F, where p is a prime. Let $Z \cong F^\times$ be the center of $D^\times$. Let S be the finite set of places where D_v is a division ring. If $v \notin S$ we identify $D_v = \mathrm{Mat}_p(F_v)$.

Let $\mathcal{H}$ be the set of functions on $D_\mathbb{A}^\times = \prod_v D_v^\times$ which are finite linear combinations of functions of the form $\prod_v \phi_v$ where for each v, ϕ_v: $D_v^\times \to \mathbb{C}$ is smooth and compactly supported modulo Z_v, satisfies $\phi_v(z_v g_v) = \phi_v(g_v)$ when $z_v \in Z_v$, and agrees with the characteristic function of $Z_v \mathrm{Mat}_p(\mathfrak{o}_v)$ for almost all places v of F. The ring $\mathcal{H}$ contains the classical Hecke operators as well as the integral operators introduced in Section 7 above.

$Z_\mathbb{A} D_F^\times \backslash D_\mathbb{A}^\times$ is compact. As with $\mathrm{SL}_2(\mathbb{R})$, $L^2(Z_\mathbb{A} D_F^\times \backslash D_\mathbb{A}^\times)$ admits integral operators $\rho(\phi)$ for $\phi \in \mathcal{H}$:

$$(\rho(\phi) f)(g) = \int_{Z_\mathbb{A} \backslash D_\mathbb{A}^\times} \phi(h) f(g h) d h.$$

Let $\{\gamma\}$ be a set of representatives for the conjugacy classes of $D_F^\times$. We denote by C_γ the centralizer of γ in $D_F^\times$. It is an algebraic group, so $C_\gamma(\mathbb{A}) \subset D_\mathbb{A}^\times$ will denote its points in $\mathbb{A}$.

Theorem 74. ***(Selberg trace formula).***

$$\mathrm{tr}\, \rho(\phi) = \sum_{\{\gamma\}} \mathrm{vol}(C_\gamma \backslash C_\gamma(\mathbb{A})) \int_{C_\gamma(\mathbb{A}) Z_\mathbb{A} \backslash D_\mathbb{A}^\times} \phi(g^{-1} \gamma g) d g. \tag{86}$$

Proof. The proof of Proposition 6 goes through without change, so

$$(\rho(\phi) f)(g) = \int_{Z_\mathbb{A} D_F^\times \backslash D_\mathbb{A}^\times} K_\phi(g, h) f(h) d h,$$

$$K_\phi(g, h) = \sum_{\gamma \in D_F^\times / Z_F^\times} \phi(g^{-1} \gamma h).$$

As with $\mathrm{SL}_2(\mathbb{R})$, the operator $\rho(\phi)$ is thus Hilbert-Schmidt, and with more work may be shown to be trace class. As in Theorem 32,

$$\mathrm{tr}\, \rho(\phi) = \int_{Z_\mathbb{A} D_F^\times \backslash D_\mathbb{A}^\times} K_\phi(g, g) d g$$

Now (86) follows as in Theorem 32. □

The conjugacy classes of $D_F^\times$ are easily described.

Proposition 75. *If $\alpha \in D_F^\times - Z_F$, then $F(\alpha)$ is a field extension of F of degree p. Elements α and β are conjugate in $D_F^\times$ if and only if there is a field isomorphism $F(\alpha) \to F(\beta)$ such that $\alpha \mapsto \beta$. If $F(\alpha)$ is a field extension of degree p, then $F(\alpha)$ may be embedded in D_F if and only if $[F_v(\alpha){:}\, F_v] = p$ for all $v \in S$.*

Proof. The conjugacy of α and β follows from the *Skolem-Noether Theorem* (Herstein [20], p. 99). The last statement follows from (i) $\leftrightarrow$ (ii) in Weil [44], Proposition VIII.5 on p. 253. □

The trace formula can be used to prove functorial liftings in many cases.

Let E be another division algebra of degree p^2, and assume that the set of places where E_v is a division ring agrees with the set S of places where D_v is. If $p = 2$ this implies that D and E are isomorphic, but not in general. (This follows from the computation of the Brauer group in the global class field theory. See Chapters 6 and 7 in [6].) Thus we want $p > 2$.

We will show that two spaces of automorphic forms on D and on E are isomorphic.

Suppose that $\pi = \otimes_v \pi_v$ is an irreducible constituent of $L^2(Z_{\mathbb{A}} D_F^\times \backslash D_{\mathbb{A}}^\times)$. Since $Z_v \backslash D_v^\times$ is compact for $v \in S$, π_v is finite-dimensional at these places. We assume that π_v is trivial when $v \in S$.

If $v \notin S$, then $D_v \cong E_v \cong \mathrm{Mat}_p(F_v)$. We may therefore identify π_v with an irreducible representation π'_v of E_v when $v \notin S$, and if $v \in S$ we take $\pi'_v = 1$. Let $\pi' = \otimes \pi'_v$. It is an irreducible representation of $E_{\mathbb{A}}^\times$.

Theorem 76. *π' occurs in $L^2(Z_{\mathbb{A}} E_F^\times \backslash E_{\mathbb{A}}^\times)$.*

The correspondence $\pi \to \pi'$ is a functorial lift of automorphic forms in the sense of Langlands (Langlands [29], Borel [3]).

Proof. (Sketch.) If $v \in S$, then $Z_v \backslash D_v^\times$ is compact, so the constant function $\phi_v^\circ(g_v) = 1$ is in $C_c^\infty(F_v)$. Let $\mathcal{H}_S$ be the subalgebra of $\mathcal{H}$ spanned by functions $\prod \phi_v$ such that $\phi_v = \phi_v^\circ$ for $v \in S$. It is isomorphic to the corresponding Hecke ring on E. Let $\phi \to \phi'$ denote this isomorphism.

By Proposition 75, noncentral conjugacy classes in $D_F^\times$ and $E_F^\times$ are both in bijection with the set of Galois equivalence classes of elements α of field extensions $[F(\alpha) : F] = p$ such that $[F_v(\alpha) : F_v] = p$ for all $p \in S$. This intrinsic characterization shows that we may identify the conjugacy classes of D_F and E_F, and compare trace formulae to get

$$\mathrm{tr}\, \rho(\phi) = \mathrm{tr}\, \rho(\phi)'. \tag{87}$$

This is almost but not quite as easy as we've made it sound, because one must show that the volumes on the right side of (86) are the same for the two trace formulae.

It follows from (87) that the representations of $\mathcal{H}_S$ on the spaces

$$L^2(Z_{\mathbb{A}} D_F^\times \prod_{v \in S} D_v^\times \backslash D_{\mathbb{A}}^\times) \text{ and } L^2(Z_{\mathbb{A}} E_F^\times \prod_{v \in S} E_v^\times \backslash E_{\mathbb{A}}^\times)$$

are gisomorphic, and the theorem follows. □

Underlying the final step is the fact that two representations of rings are characterized by their traces. For example if R is an algebra over a field of characteristic zero and if M_1, M_2 are finite-dimensional semisimple R-modules, and if for every $\alpha \in R$ the induced endomorphisms of M_1 and M_2 have the same trace, then the modules are isomorphic (Lang [28], Corollary 3.8 on p. 650). This statement is not directly applicable here but it gives the flavor.

For the remainder we take $p = 2$, and review the *Jacquet-Langlands correspondence*. Let D be as before. The Jacquet-Langlands correspondence is a lifting of automorphic representations from $D^\times$ to $\mathrm{GL}_2(F)$.

There is a local correspondence for $v \in S$. $D_v^\times$ is compact modulo its center, so its irreducible representations are finite-dimensional. These lift to irreducible representations of $\mathrm{GL}_2(F_v)$ having the same central character. The lifting was constructed by Jacquet and Langlands by use of the theta correspondence. Indeed, $Z_v\backslash D_v^\times$ is a quotient of the orthogonal group GO(4) while GL_2 is the same as GSp_2, and theta correspondence $\mathrm{GO}(4) \leftrightarrow \mathrm{GSp}_2$ gives the Jacquet-Langlands correspondence. Its image is the square integrable representations (that is, the supercuspidals and the Steinberg representation).

Jacquet and Langlands constructed a global correspondence from automorphic forms on $D^\times$ to automorphic forms on GL_2 first using the converse theorem in Section 14 of their book. To prove functional equations of L-functions on $D^\times$, they use the Godement-Jacquet construction, because the Hecke integral is not available in this context.

Finally, they reconsidered the lifting from the point of view of the trace formula. This allowed them to characterize the image of the lift. They sketched a proof (and later Gelbart and Jacquet completed) of:

Theorem 77. *An automorphic representation* π *of* $\mathrm{GL}_2(\mathbb{A})$ *is the lift of an automorphic representation of* $D_{\mathbb{A}}^\times$ *if and only if* π_v *is square integrable for every* $v \in S$.

The remarkable fact is that their proof of this fact uses so many different techniques which have proved important in the subsequent 30 years: the Hecke and Godement-Jacquet integral constructions of L-functions, the Weil representation and the trace formula.

The trace formula on GL_2 is harder than on the division ring because of the presence of the continuous spectrum. We've avoided this problem by proving Theorem 76 instead of Theorem 77.

Bibliography

[1] J. Bernstein. Letter to Piatetski-Shapiro. unpublished, 1986.

[2] J. Bernstein. Meromorphic continuation of Eisenstein series. unpublished, 1987.

[3] A. Borel. Automorphic L-functions. In *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 2*, Proc. Sympos. Pure Math., XXXIII, pages 27–61. Amer. Math. Soc., Providence, R.I., 1979.

[4] Armand Borel. *Automorphic forms on* $\mathrm{SL}_2(\mathbb{R})$, volume 130 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1997.

[5] Daniel Bump. *Automorphic forms and representations*, volume 55 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1997.

[6] J. Cassels and A. Fröhlich. *Algebraic number theory*. Proceedings of an instructional conference organized by the London Mathematical Society (a NATO Advanced Study Institute) with the support of the Inter national Mathematical Union. Academic Press, London, 1967.

[7] P. Cohen and P. Sarnak. Discontinuous groups and harmonic analysis. unpublished, 1985.

[8] Yves Colin de Verdière. Une nouvelle démonstration du prolongement méromorphe des séries d'Eisenstein. *C. R. Acad. Sci. Paris Sér. I Math.*, 293(7):361–363, 1981.

[9] Alain Connes. Noncommutative geometry and the Riemann zeta function. In *Mathematics: frontiers and perspectives*, pages 35–54. Amer. Math. Soc., Providence, RI, 2000.

[10] Isaac Y. Efrat. The Selberg trace formula for $PSL_2(\mathbb{R})^n$. *Mem. Amer. Math. Soc.*, 65(359), 1987.

[11] Jürgen Elstrodt. Die Resolvente zum Eigenwertproblem der automorphen Formen in der hyperbolischen Ebene. I. *Math. Ann.*, 203:295–300, 1973.

[12] L. D. Faddeev. The eigenfunction expansion of Laplace's operator on the fundamental domain of a discrete group on the Lobačevskiĭ plane. *Trudy Moskov. Mat. Obšč.*, 17:323–350, 1967.

[13] Stephen Gelbart and Hervé Jacquet. Forms of GL(2) from the analytic point of view. In *Automorphic forms, representations and L-functions (Proc. Sympos. Pure Math., Oregon State Univ., Corvallis, Ore., 1977), Part 1*, Proc. Sympos. Pure Math., XXXIII, pages 213–251. Amer. Math. Soc., Providence, R.I., 1979.

[14] I. M. Gel'fand, M. I. Graev, and I. I. Pyatetskii-Shapiro. *Representation theory and automorphic functions*. Translated from the Russian by K. A. Hirsch. W. B. Saunders Co., Philadelphia, Pa., 1969.

[15] R. Godement. The decomposition of $L^2(G/\Gamma)$ for $\Gamma = SL(2, Z)$. In *Algebraic Groups and Discontinuous Subgroups (Proc. Sympos. Pure Math., Boulder, Colo.*, pages 211–224. Amer. Math. Soc., Providence, R.I., 1966.

[16] R. Godement. The spectral decomposition of cusp-forms. In *Algebraic Groups and Discontinuous Subgroups (Proc. Sympos. Pure Math., Boulder, Colo., 1965)*, pages 225–234. Amer. Math. Soc., Providence, R.I., 1966.

[17] Harish-Chandra. *Automorphic forms on semisimple Lie groups*. Notes by J. G. M. Mars. Lecture Notes in Mathematics, No. 62. Springer-Verlag, Berlin, 1968.

[18] Dennis A. Hejhal. *The Selberg trace formula for PSL(2, R). Vol. I*. Springer-Verlag, Berlin, 1976. Lecture Notes in Mathematics, Vol. 548.

[19] Dennis A. Hejhal. *The Selberg trace formula for* PSL(2, ℝ). *Vol. 2*, volume 1001 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1983.

[20] I. N. Herstein. *Noncommutative rings*. The Carus Mathematical Monographs, No. 15. Published by The Mathematical Association of America, 1968.

[21] A. E. Ingham. *The distribution of prime numbers*. Cambridge Tracts in Mathematics and Mathematical Physics, No. 30. Stechert-Hafner, Inc., New York, 1964.

[22] H. Jacquet and R. P. Langlands. *Automorphic forms on GL(2)*. Springer-Verlag, Berlin, 1970. Lecture Notes in Mathematics, Vol. 114.

[23] Hervé Jacquet. Note on the analytic continuation of Eisenstein series: An appendix to "Theoretical aspects of the trace formula for GL(2)" [in *representation theory and automorphic forms (edinburgh, 1996)*, 355–405, Proc. Sympos. Pure Math., 61, Amer. Math. Soc., Providence, RI, 1997] by A. W. Knapp. In *Representation theory and automorphic forms (Edinburgh, 1996)*, volume 61 of *Proc. Sympos. Pure Math.*, pages 407–412. Amer. Math. Soc., Providence, RI, 1997.

[24] A. W. Knapp and E. M. Stein. Intertwining operators for semisimple groups. *Ann. of Math. (2)*, 93, 1971.

[25] Anthony W. Knapp. *Representation theory of semisimple groups*, volume 36 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ, 1986. An overview based on examples.

[26] Tomio Kubota. *Elementary theory of Eisenstein series*. Kodansha Ltd., Tokyo, 1973.

[27] Serge Lang. $SL_2(\mathbb{R})$. Addison-Wesley Publishing Co., Reading, Mass.-London-Amsterdam, 1975.

[28] Serge Lang. *Algebra*, volume 211 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2002.

[29] R. P. Langlands. Problems in the theory of automorphic forms. In *Lectures in modern analysis and applications, III*, pages 18–61. Lecture Notes in Math., Vol. 170. Springer, Berlin, 1970.

[30] Robert P. Langlands. *On the functional equations satisfied by Eisenstein series*. Springer-Verlag, Berlin, 1976. Lecture Notes in Mathematics, Vol. 544.

[31] Peter D. Lax and Ralph S. Phillips. *Scattering theory for automorphic functions*. Princeton Univ. Press, Princeton, N.J., 1976. Annals of Mathematics Studies, No. 87.

[32] C. Mœglin and J.-L. Waldspurger. *Spectral decomposition and Eisenstein series*, volume 113 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, 1995. Une paraphrase de l'Écriture [A paraphrase of Scripture].

[33] W. Müller. On the analytic continuation of rank one Eisenstein series. *Geom. Funct. Anal.*, 6(3):572–586, 1996.

[34] M. Scott Osborne and Garth Warner. *The theory of Eisenstein systems*, volume 99 of *Pure and Applied Mathematics*. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1981.

[35] S. J. Patterson. *An introduction to the theory of the Riemann zeta-function*, volume 14 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1988.

[36] Burton Randol. Small eigenvalues of the Laplace operator on compact Riemann surfaces. *Bull. Amer. Math. Soc.*, 80:996–1000, 1974.

[37] Walter Roelcke. Über die Wellengleichung bei Grenzkreisgruppen erster Art. *S.-B. Heidelberger Akad. Wiss. Math.-Nat. Kl.*, 1953/1955:159–267 (1956), 1953/1955.

[38] A. Selberg. Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. *J. Indian Math. Soc. (N.S.)*, 20:47–87, 1956.

[39] A. Selberg. Harmonic Analysis. This is the annotated surviving portion of the 1955 Göttingen lecture notes. In *Selberg's Collected Works, Vol. I*, pages 626–674. Springer-Verlag, 1989.

[40] Atle Selberg. On the estimation of Fourier coefficients of modular forms. In *Proc. Sympos. Pure Math., Vol. VIII*, pages 1–15. Amer. Math. Soc., Providence, R.I., 1965.

[41] V. S. Varadarajan. *An introduction to harmonic analysis on semisimple Lie groups*, volume 16 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1989.

[42] A. B. Venkov. Spectral theory of automorphic functions, the Selberg zeta function and some problems of analytic number theory and mathematical physics. Translated in Russian Mathematical Surveys **34** (1979), 79–153. *Uspekhi Mat. Nauk*, 34(3(207)):69–135, 248, 1979.

[43] André Weil. Sur les "formules explicites" de la théorie des nombres premiers. *Comm. Sém. Math. Univ. Lund [Medd. Lunds Univ. Mat. Sem.]*, 1952(Tome Supplementaire):252–265, 1952.

[44] André Weil. *Basic number theory*. Die Grundlehren der mathematischen Wissenschaften, Band 144. Springer-Verlag New York, Inc., New York, 1967.

[45] E. T. Whittaker and G. N. Watson. *A course of modern analysis. An introduction to the general theory of infinite processes and of analytic functions: with an account of the principal transcendental functions*. Fourth edition. Reprinted regularly. Cambridge University Press, New York, 1927.

[46] Shek-Tung Wong. The meromorphic continuation and functional equations of cuspidal Eisenstein series for maximal cuspidal groups. *Mem. Amer. Math. Soc.*, 83(423), 1990.