Lecture 3: From modular forms on $\mathbb{H}$ to automorphic forms on $\mathrm{SL}_2(\mathbb{R})$

Contents

We have finished our discussion of Hecke characters of number fields. One may think of these as the *automorphic* characters of the group $\mathrm{GL}_1(\mathbb{A}_K) = \mathbb{A}_K^\times$, where by "automorphic" we mean invariant under the group of rational points $\mathrm{GL}_1(K) = K^\times$. We saw that the theory was rich enough to encompass complex Galois characters and compatible families of ℓ-adic Galois characters, with room to spare.

For the next few lectures, we shall motivate and introduce the theory of automorphic representations of $\mathrm{GL}_2(\mathbb{A}_K)$. This will lie much deeper than the abelian case of GL_1. We shall do this through the traditional progression of ideas:

(1) recalling the classical theory of automorphic forms on $\Gamma\backslash\mathbb{H}$, where Γ is a congruence subgroup of $\mathrm{SL}_2(\mathbb{Z})$ (a lattice of arithmetic significance in $\mathrm{SL}_2(\mathbb{R})$);
(2) lifting automorphic forms on $\Gamma\backslash\mathbb{H}$ to automorphic forms on $\Gamma\backslash\mathrm{SL}_2(\mathbb{R})$, and adelizing them to the quotient $\mathrm{SL}_2(\mathbb{Q})\backslash\mathrm{SL}_2(\mathbb{A}_\mathbb{Q})$;
(3) understanding some basic representation theory of $\mathrm{SL}_2(\mathbb{R})$;
(4) understanding some basic (complex) representation theory of $\mathrm{SL}_2(\mathbb{Q}_p)$, including the statement of the local Langlands correspondence for $\mathrm{GL}_2(\mathbb{Q}_p)$.

We shall spend much more time on the representation theory of p-adic groups, such as $\mathrm{SL}_2(\mathbb{Q}_p)$, than that of real groups, such as $\mathrm{SL}_2(\mathbb{R})$, in order to emphasize arithmetic aspects.

1. Lattices in $\mathrm{SL}_2(\mathbb{R})$

We begin by reviewing the theory of lattices in $\mathrm{SL}_2(\mathbb{R})$.

Let Γ be a *lattice* in $\mathrm{SL}_2(\mathbb{R})$, i.e., a discrete subgroup admitting a fundamental domain in $\mathbb{H}$ with finite area relative to the invariant hyperbolic measure $dxdy/y^2$. Note that a lattice in this setting also goes by the name of a Fuchsian group of the first kind. A lattice is said to be *uniform* if the quotient $\Gamma\backslash\mathbb{H}$ is non-compact, and *non-uniform* otherwise.

Let us look at some examples.

1.1. **Congruence subgroups of** $\mathrm{SL}_2(\mathbb{Z})$**.** The modular group $\mathrm{SL}_2(\mathbb{Z})$ and its well-known congruence subgroups

$$\Gamma(N) = \ker(\mathrm{SL}_2(\mathbb{Z}) \xrightarrow{\text{red}} \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}))$$
$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : c \equiv 0 \bmod N \right\}$$
$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : c \equiv 0,\ d \equiv 1 \bmod N \right\}$$

are all examples of lattices in $\mathrm{SL}_2(\mathbb{R})$. (Note that the reduction map mod N is surjective, so that $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}) \simeq \mathrm{SL}_2(\mathbb{Z})/\Gamma(N)$.) In general, a subgroup of $\mathrm{SL}_2(\mathbb{Z})$ is said to be *of congruence type* if it contains the principal congruence subgroup $\Gamma(N)$ for some $N \geq 1$.

1.2. **Arithmetic subgroups of** $\mathrm{SL}_2(\mathbb{Z})$**.** Any finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$ is an example of a larger class of lattices, called *arithmetic lattices.*

Lemma 1.1. *There exist arithmetic non-congruence subgroups of* $\mathrm{SL}_2(\mathbb{Z})$*.*

Proof. Recall that $\mathrm{PSL}_2(\mathbb{Z})$ is freely generated by elements of order 2 and 3, the images under $\mathrm{SL}_2(\mathbb{Z}) \to \mathrm{PSL}_2(\mathbb{Z})$ of the elements typically denoted $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $ST = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 1 \end{smallmatrix}\right)$. Thus $\mathrm{PSL}_2(\mathbb{Z})$ admits a surjection onto any group generated by elements of order 2 and 3. An example of such a group is A_n, for $n \geq 9$ (see Dey and Wiegold, *Generators for alternating and symmetric groups*, 1971). Call Γ the kernel of such a surjection $\mathrm{SL}_2(\mathbb{Z}) \to A_n$, so that Γ is a normal arithmetic subgroup of $\mathrm{SL}_2(\mathbb{Z})$ satisfying $A_n \simeq \mathrm{SL}_2(\mathbb{Z})/\Gamma$. We shall show that Γ is non-congruence.

If Γ contained some $\Gamma(N)$ then A_n would be a quotient of $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$. But this is not the case for $n \geq 6$. Indeed, if A_n were a quotient of $\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$, where $N = \prod_i p^{r_i}$, it would be a simple quotient of some $\mathrm{SL}_2(\mathbb{Z}/p^{r_i}\mathbb{Z})$. But one can write down the composition series of the last group, and find that any non-abelian simple quotient of it is isomorphic to $\mathrm{PSL}_2(\mathbb{Z}/p\mathbb{Z})$. But $|\mathrm{PSL}_2(\mathbb{Z}/p\mathbb{Z})| = (p^2-1)p/2$ is equal to $|A_n| = n!/2$ only for $p = n = 5$. $\square$

Any finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$ is a non-uniform lattice, as its fundamental domain is just a finite number of translates of the non-compact fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$. Generalizing the previous paragraph, we now give the following

Definition 1. A non-uniform lattice is said to be *arithmetic* if a conjugate of it inside $\mathrm{SL}_2(\mathbb{R})$ is commensurable with $\mathrm{SL}_2(\mathbb{Z})$.

Recall that two lattices Γ_1 and Γ_2 of $\mathrm{SL}_2(\mathbb{R})$ are called *commensurable* if their intersection $\Gamma_1 \cap \Gamma_2$ is of finite index in both Γ_1 and Γ_2. The definition above, which allows for a conjugation, is called *commensurabile in the wide sense.*

1.3. **Uniform arithmetic lattices.** One can create (a wide class of) uniform arithmetic lattices from through division quaternion algebras over $\mathbb{Q}$. We shall briefly review some of the theory.

Recall that a quaternion algebra over a field F is a central simple algebra A over F of dimension 4. More concretely, when F is of characteristic not equal to 2 (which, for simplicity, we shall always assume), A is isomorphic to the associative F-algebra $\left(\frac{a,b}{F}\right)$ having basis $1, i, j, k$ verifying the multiplication laws $i^2 = a, j^2 = b$, $ij = k = -ij$ for elements $a, b \in F^{\times}$. For example, $\left(\frac{1,1}{F}\right) = M_2(F)$, where

$$1 = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix}, \quad i = \begin{pmatrix} 1 & \\ & -1 \end{pmatrix}, \quad j = \begin{pmatrix} & 1 \\ 1 & \end{pmatrix}, \quad k = \begin{pmatrix} & 1 \\ -1 & \end{pmatrix}.$$

Of course the prototypical example of a quaternion *division algebra* is given by Hamilton's quaternions $\mathbf{H} = \left(\frac{-1,-1}{\mathbb{R}}\right)$. A basic fact is that if A is not a division algebra then A is isomorphic to $M_2(F)$.

The quaternion algebra is $A = \left(\frac{a,b}{F}\right)$ is naturally endowed with an involution sending $x = \alpha + \beta i + \gamma j + \delta k$ to its conjugate $\bar{x} = \alpha - \beta i - \gamma j - \delta k$. Then the map $\mathrm{Nm} : A \to F$, $x \mapsto x\bar{x}$ is called the *reduced norm*, and $\mathrm{Tr} : A \to F$, $x \mapsto x + \bar{x}$ is called the *reduced trace*. (Here we are identifying F with the center of A.) For example, if $A = M_2(F)$, the involution is

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

and the reduced norm and trace are the determinant and trace.

Now let $E = F[i]$. If E is not a field then E contains zero divisors so A is isomorphic to $M_2(F)$. Otherwise, we may view A as a 2-dimensional right vector space over E: $A = E \oplus jE$. One can then define the left regular representation $A \to \mathrm{End}_E(A)$ in which an element $x \in A$ is sent to the endomorphism given by left-multiplication by x. Using the E-basis $\{1, j\}$ of E, we identify $\mathrm{End}_E(A)$ with $M_2(E)$, in which case the above map sends

$$i \mapsto \begin{pmatrix} i & \\ & -i \end{pmatrix}, \qquad j \mapsto \begin{pmatrix} & b \\ 1 & \end{pmatrix},$$

by the defining multiplicative relations. Note that the resulting map

$$\begin{aligned} \Phi : \left(\frac{a,b}{F}\right) &\longrightarrow M_2(E) \\ \alpha + i\beta + j\gamma + k\delta &\longmapsto \begin{pmatrix} \alpha + \beta i & b(\gamma + \delta i) \\ \gamma - \delta i & \alpha - \beta i \end{pmatrix} \end{aligned}$$

is an injective algebra homomorphism commuting with the respective norm and trace maps.

Now take F to be $\mathbb{Q}$. Recall that a $\mathbb{Z}$-lattice in a finite dimensional $\mathbb{Q}$-vector space V is a finitely generated $\mathbb{Z}$ submodule L of V such that $L \otimes_{\mathbb{Z}} \mathbb{Q} = V$. We shall apply this definition to $V = A$, a quaternion algebra over $\mathbb{Q}$. A $\mathbb{Z}$-order $\mathcal{O}$ in A is a $\mathbb{Z}$-lattice of A which is a subring of A. For example, if $A = M_2(\mathbb{Q})$ then $\mathcal{O} = M_2(\mathbb{Z})$ is a $\mathbb{Z}$-order. More generally, if $A = \left(\frac{a,b}{\mathbb{Q}}\right)$, with a, b taken to be in $\mathbb{Z} \setminus \{0\}$, then $\mathcal{O} = \mathbb{Z} \oplus \mathbb{Z}i \oplus \mathbb{Z}j \oplus \mathbb{Z}k$ is a $\mathbb{Z}$-order.

With this brief refresher behind us, we now come to the construction of a wide class of uniform arithmetic lattices. Let A be an indefinite quaternion algebra over $\mathbb{Q}$. "Indefinite" here means that $A \otimes_{\mathbb{Q}} \mathbb{R} = M_2(\mathbb{R})$ (say, by requiring that $a > 0$). Thus the field E described above is real quadratic (since $i^2 = a > 0$) so that (choosing an embedding $E \hookrightarrow \mathbb{R}$) the image of Φ lies in $M_2(\mathbb{R})$. Let $\mathcal{O}$ be a $\mathbb{Z}$-order in A and write $\mathcal{O}^1$ for the group of elements having reduced norm 1.

Lemma 1.2. *With notations as above, $\Gamma = \Phi(\mathcal{O}^1)$ is a lattice in $\mathrm{SL}_2(\mathbb{R})$. It is cocompact if, and only if, A is a division algebra.*

Note that we recover $\Gamma_0(N)$ by taking $\Phi(\mathcal{O}^1)$, where $\mathcal{O}$ is the set of integral matrices in $A = M_2(\mathbb{Q})$ whose lower left corner is divisible by N.

Remark 1. More general constructions of uniform arithmetic lattices are possible, by using a totally real field F, and a quaternion algebra which is definite at all places but one. We shall not discuss this; see Definition 8.2.5 of Maclachlan–Reid, *The arithmetic of hyperbolic* 3*-manifolds.* We also recommend the article, *Introduction to arithmetic Fucshsian groups,* by C. Maclachlan, in the book Topics on Riemann Surfaces and Fuchsian Groups.

1.4. **Arithmetic vs non-arithmetic lattices.** One way to think about arithmetic lattices is as those lattices which admit a rich source of Hecke operators. These arise through the commensurator of Γ.

Recall that the *commensurator* of Γ is

$$\mathrm{Com}(\Gamma) = \{\alpha \in \mathrm{GL}_2^+(\mathbb{R}) : \Gamma \cap \alpha^{-1}\Gamma\alpha \text{ is of finite index in } \Gamma \text{ and } \alpha^{-1}\Gamma\alpha\}.$$

For every $\alpha \in \mathrm{Com}(\Gamma)$ we may decompose the double coset $\Gamma\alpha\Gamma$ into a finite union of left-cosets

$$\Gamma\alpha\Gamma = \bigcup_{j=1}^{n} \Gamma\alpha\delta_j, \tag{1.1}$$

where the $\delta_j \in \Gamma$ form a system of representatives of $(\Gamma \cap \alpha^{-1}\Gamma\alpha)\backslash\Gamma$. In the next section, we shall review how such double coset decompositions give rise to Hecke operators.

We have the following famous theorem of Margulis.

Theorem 1.1 (Margulis)**.** *The lattice Γ is arithmetic if, and only if,* $\mathrm{Com}(\Gamma)/\Gamma$ *is infinite.*

For example, if $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ then $\mathrm{GL}_2^+(\mathbb{Q})$ is the commensurator $\mathrm{Com}(\mathrm{SL}_2(\mathbb{Z}))$. In particular, taking $\alpha = \left(\begin{smallmatrix} p & \\ & 1 \end{smallmatrix}\right)$, we have the decomposition

$$\mathrm{SL}_2(\mathbb{Z}) \begin{pmatrix} p & \\ & 1 \end{pmatrix} \mathrm{SL}_2(\mathbb{Z}) = \mathrm{SL}_2(\mathbb{Z}) \begin{pmatrix} p & \\ & 1 \end{pmatrix} \bigcup \bigcup_{j=0}^{p-1} \mathrm{SL}_2(\mathbb{Z}) \begin{pmatrix} 1 & j \\ & p \end{pmatrix}. \tag{1.2}$$

This gives rise to the familiar Hecke operator T_p.

Finally we observe that most lattices in $\mathrm{SL}_2(\mathbb{R})$ are not arithmetic: the moduli space of genus $g \geq 2$ cocompact lattices is of (complex) dimension $3g-3 > 0$. We shall always restrict to Γ arithmetic in this course, and later, when we move to the adelic language, we shall in fact only consider lattices of congruence type.

2. Classical modular forms on $\mathbb{H}$

Recall the action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{H}$, given by

$$g.z = \frac{az+b}{cz+d}, \qquad \text{if } g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ and } z \in \mathbb{H}. \tag{2.1}$$

For $k \in \mathbb{N}$, we recall the slash operator

$$(f|_k g)(z) = j(g,z)^{-k} f(g.z), \tag{2.2}$$

where $j(g;z) = cz+d$ is the *automorphy factor*.

Henceforth in this lecture, we shall take Γ to be an arithmetic lattice, which, when it is non-uniform, we shall take to be contained in $\mathrm{SL}_2(\mathbb{Z})$. To define modular forms in the latter case, shall need to discuss the geometry at infinity of $\Gamma\backslash\mathbb{H}$. We may extend the action of $\mathrm{SL}_2(\mathbb{R})$ on $\mathbb{H}$ by fractional linear transformations to an action on the boundary $\mathbb{P}^1(\mathbb{R}) = \mathbb{R} \cup \{\infty\}$. As $\mathrm{SL}_2(\mathbb{Z})$ acts with one orbit on $\mathbb{P}^1(\mathbb{Q})$, the set of Γ-orbits on $\mathbb{P}^1(\mathbb{Q})$ for the finite index subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ is finite. The *cusps* of Γ are precisely these orbits.

Definition 2. A function $f : \mathbb{H} \to \mathbb{C}$ is a *modular form* of weight k with respect to Γ if

- f is holomorphic on $\mathbb{H}$;
- $f|_k\gamma = f$ for all $\gamma \in \Gamma$;
- f is holomorphic at the cusps (not needed if Γ is cocompact).

We explain the last condition. Let $s \in \mathbb{P}^1(\mathbb{Q})$. There is a $\sigma_s \in \mathrm{SL}_2(\mathbb{Z})$ such that $\sigma_s\infty = s$. We call σ_s an *integral scaling matrix.* Letting Γ_s denote the stabilizer of s in Γ, it follows from arithmeticity that $(\pm\sigma_s^{-1}\Gamma\sigma_s)_\infty$ depends only on the Γ-orbit $\mathfrak{a} = \Gamma.s$ and is of finite index $h_\mathfrak{a}$ in $\mathrm{SL}_2(\mathbb{Z})_\infty = \pm\langle\left(\begin{smallmatrix} 1 & 1 \\ & 1 \end{smallmatrix}\right)\rangle$. This index $h_\mathfrak{a}$, called the *width* of the cusp $\mathfrak{a}$, is characterized by

$$\pm\,(\sigma_s^{-1}\Gamma\sigma_s)_\infty = \pm\left\langle \begin{pmatrix} 1 & h_\mathfrak{a} \\ & 1 \end{pmatrix} \right\rangle. \tag{2.3}$$

Note that an integral scaling matrix of any element in $\mathfrak{a} = \Gamma.s$ differs from σ_s by left-multiplication by Γ (recall our assumption $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$). (See also

Remark 2 below.) By Γ-invariance of $f|_k$, it makes sense to write $f_{\mathfrak{a}} = f|_k\sigma_s$. Then $f_{\mathfrak{a}}|_k\gamma = f_{\mathfrak{a}}$ for all $\gamma \in \sigma_s^{-1}\Gamma\sigma_s$. Moreover, it follows from (2.3) that $(\sigma_s^{-1}\Gamma\sigma_s)_\infty$ is one of either

$$\pm\left\langle\begin{pmatrix}1 & h_{\mathfrak{a}}\\ & 1\end{pmatrix}\right\rangle, \quad \left\langle\begin{pmatrix}1 & h_{\mathfrak{a}}\\ & 1\end{pmatrix}\right\rangle, \quad \text{or} \quad \left\langle-\begin{pmatrix}1 & h_{\mathfrak{a}}\\ & 1\end{pmatrix}\right\rangle.$$

In the first two cases, we have $f_{\mathfrak{a}}(z+h_{\mathfrak{a}}) = f_{\mathfrak{a}}(z)$. In the third case we have $f_{\mathfrak{a}}(z+h_{\mathfrak{a}}) = -f_{\mathfrak{a}}(z)$, so that $f_{\mathfrak{a}}(z+2h_{\mathfrak{a}}) = f_{\mathfrak{a}}(z)$. We set $h'_{\mathfrak{a}} = h_{\mathfrak{a}}$ or $2h'_{\mathfrak{a}}$ according to this division, and call it the *period* of f at $\mathfrak{a}$. Thus,

$$f_{\mathfrak{a}}(z+h'_{\mathfrak{a}}) = f_{\mathfrak{a}}(z)$$

and we may consider the Fourier expansion relative to this invariance:

$$f_{\mathfrak{a}}(z) = \sum_{n\in\mathbb{Z}} \hat{f}_{\mathfrak{a}}(n) q^{n/h'_{\mathfrak{a}}},$$

where $q = e^{2\pi i z}$.

Remark 2. One must be careful with the dependence on the Fourier coefficients $\hat{f}_{\mathfrak{a}}(n)$ on the integral scaling matrix σ_s taking ∞ to s. Indeed, any two integral scaling matrices differ from each other by right-multiplication by an element in $\mathrm{SL}_2(\mathbb{Z})_\infty$, and by left-multiplication by Γ_s. The latter ambiguity is anodyne, as we have already discussed. But right-multiplication of σ_s by $u = \pm\left(\begin{smallmatrix}1 & 1\\ & 1\end{smallmatrix}\right)$ gives $(f|_k\sigma_s u)(z) = (\pm 1)^k (f|_k\sigma_s)(z+1)$. From this it follows easily that the Fourier coefficients of $f|_k\sigma_s u$ are those of $f|_k\sigma_s$, multiplied by an $h'_{\mathfrak{a}}$-th root of unity. Note that when the width of the cusp is 1 this ambiguity disappears.

In view of the above remark, the *vanishing* of the Fourier coefficients of $f_{\mathfrak{a}}$ is independent of the choice of integral scaling matrix. We may therefore make the following definition.

Definition 3. We say that f is *holomorphic* at $\mathfrak{a}$ if $\hat{f}_{\mathfrak{a}}(n) = 0$ for all $n < 0$.

We write $M_k(\Gamma)$ for the space of weight k modular forms with respect to Γ. Note that one can give a more general definition of modular form, by allowing a character of Γ in the relation $f|_k\gamma = f$.

Definition 4. Let $f \in M_k(\Gamma)$. We say that f is *cuspidal* if $\hat{f}_{\mathfrak{a}}(0) = 0$ for every cusp $\mathfrak{a}$ of Γ. We write $S_k(\Gamma)$ for the space of weight k cuspidal forms with respect to Γ.

Lemma 2.1. *We have*

$$S_k(\Gamma) \subset \left\{ f \in M_k(\Gamma) : \int_{\Gamma\backslash\mathbb{H}} |f(z)|^2 y^k \frac{dxdy}{y^2} < \infty \right\}.$$

Proof. One has to check the convergence of the integral at each cusp $\mathfrak{a}$. Using the Fourier expansion at $\mathfrak{a}$, we find that

$$f_{\mathfrak{a}}(z) = \sum_{n\geq 1} \hat{f}_{\mathfrak{a}}(n) e^{2\pi i n(x+iy)/h'_{\mathfrak{a}}} \ll e^{-2\pi y/h'_{\mathfrak{a}}}.$$

The contribution of the cusp at s is therefore

$$\int_{y\gg 1}\int_0^{h'_{\mathfrak{a}}} |f_{\mathfrak{a}}(z)|^2 y^{k-2}\,dx\,dy \ll \int_{y\gg 1} e^{-4\pi y/h'_{\mathfrak{a}}} y^{k-2}\,dy < \infty. \qquad \square$$

Some well-known examples:

(1) for any Γ, we have $M_0(\Gamma) = \mathbb{C}$ and $S_0(\Gamma) = 0$;

(2) the full level weight $2k$ Eisenstein series

$$E_{2k} = \frac{1}{2}\sum_{(m,n)\in\mathbb{Z}^2-0}(mz+n)^{-2k}$$

lies in $M_{2k}(\mathrm{SL}_2(\mathbb{Z}))$;

(3) we have $S_{12}(\mathrm{SL}_2(\mathbb{Z})) = \mathbb{C}\Delta$, where

$$\Delta = (G_4^3 - G_6^2)/1728 = \sum_{n\geq 1}\tau(n)q^n$$

and $G_{2k} = \zeta(2k)^{-1}E_{2k}$.

2.1. **Hecke operators.** Recall the commensurator $\mathrm{Com}(\Gamma)$ from §1.4.

For every $\alpha \in \mathrm{Com}(\Gamma)$ we would like to define a *Hecke operator*

$$T_\alpha : M_k(\Gamma) \to M_k(\Gamma)$$

using the decomposition (1.1). Since the matrices $\alpha\delta_j$ are in $\mathrm{GL}_2^+(\mathbb{R})$, to do this we must extend the transformation rule (2.2) to the group $\mathrm{GL}_2^+(\mathbb{R})$, by inserting an additional factor of $\det(g)^{k-1}$ and extending (2.1) by the same formula (the upper half-plane is preserved since the determinant of g is positive).[1]

Having done this, we may define

$$(T_\alpha f)(z) = \sum_{j=1}^{n}(f|_k\alpha\delta_j)(z),$$

for any $f \in M_k(\Gamma)$. One may prove that T_α preserves the cuspidal subspace.

Example 1. If $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ then the commensurator $\mathrm{Com}(\mathrm{SL}_2(\mathbb{Z}))$ is $\mathrm{GL}_2^+(\mathbb{Q})$. For $\alpha \in \mathrm{GL}_2^+(\mathbb{Q})$ one may show (see Diamond–Shurman, Section 5.5) that the adjoint of T_α, with respect to the natural inner product

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

[1] Some authors use a different convention for the determinant dependence. For example, Shimura, in his famous book, Introduction to the arithmetic theory of automorphic functions, uses $\det(g)^{k/2}$. In this way the center of $\mathrm{GL}_2^+(\mathbb{R})$ acts trivially under the $|_k$ operator. In that case, the Hecke operators are normalized differently as well, with an additional factor of $\det(g)^{k/2-1}$. The end result is the same.

is $T_{\alpha'}$, where $\alpha' = \det(\alpha)\alpha^{-1}$. In particular, for $\alpha = \left(\begin{smallmatrix} p & \\ & 1 \end{smallmatrix}\right)$, we have $\alpha' = \alpha$. We write the corresponding self-adjoint Hecke operator as T_p. Using the double coset decomposition (1.2) we find

$$(T_p f)(z) = (f|_k \left(\begin{smallmatrix} p & \\ & 1 \end{smallmatrix}\right))(z) + \sum_{j=0}^{p-1} (f|_k \left(\begin{smallmatrix} 1 & j \\ & p \end{smallmatrix}\right))(z).$$

2.2. **Ramanujan–Petersson.** Recall the following fundamental result of the last century: the proof of the Ramanujan–Petersson conjecture. We assume familiarity with the theory of newforms for congruence subgroups of $\mathrm{SL}_2(\mathbb{Z})$.

Let f be a primitive cusp form relative to $\Gamma_0(N)$ and of weight $k \geq 1$. The Fourier expansion at the cusp ∞ (of width 1) takes the form

$$f(z) = \sum_{n \geq 1} a_f(n) q^n. \tag{2.4}$$

For $p \nmid N$ let $\alpha_p, \beta_p \in \mathbb{C}$ be complex numbers satisfying

$$(1 - a_f(p)p^{1-s-k/2} + p^{1-2s})^{-1} = (1 - \alpha_p p^{-s})^{-1}(1 - \beta_p p^{-s})^{-1}.$$

Thus

$$\alpha_p + \beta_p = a_f(p)p^{1-k/2} \quad \text{and} \quad \alpha_p \beta_p = p.$$

We write $\alpha_p = p^{s_p}$ for $s_p \in \mathbb{C}$. Then $\beta_p = p^{1-s_p}$.

Observe that since $a_f(p) \in \mathbb{R}$,

$$\text{either} \;\; s_p \in \mathbb{R} \;\; \text{or} \;\; s_p \in \frac{1}{2} + i\mathbb{R}. \tag{2.5}$$

Note that in the latter case, we obtain $|a_f(p)| \leq 2p^{(k-1)/2}$. The Hecke bound $|a_f(p)| \leq Cp^{k/2}$ roughly states that $\mathrm{Re}(s_p) \in [0, 1]$.

Theorem 2.1 (Deligne ($k \geq 2$), Deligne–Serre ($k = 1$))**.** *For all $p \nmid N$ we have $s_p \in \frac{1}{2} + i\mathbb{R}$.*

Deligne proved this result (for $k \geq 2$) by first proving a conjecture of Serre, namely that associated with f is a compatible system of Galois representations $\rho_{f,\lambda}$ on a 2-dimensional $K_{f,\lambda}$ vector space, where K_f is the number field generated by the Fourier coefficients of f and $K_{f,\lambda}$ the completion at a prime ideal λ lying over a prime ℓ, such that $\mathrm{tr}\, \rho_{f,\lambda}(\mathrm{Frob}_p) = a_f(p)$ for all primes $p \nmid N\ell$. In this way holomorphic modular forms of weight $k \geq 2$ play an analogous role in this theory to the algebraic Hecke characters we saw in the last lecture.

2.3. **Hecke L-function.** Let f be a primitive weight k cusp form relative to $\Gamma_0(N)$. We briefly dicuss the Hecke L-function one may associate with f. For simplicity, we shall take $N = 1$.

Using the q-expansion of f at ∞ described in (2.4) we define the Dirichlet series

$$L(s, f) = \sum_{n \geq 1} a_f(n) n^{-s},$$

which converges absolutely on the right half-plane $\operatorname{Re} s > (k+1)/2$, in view of the Deligne bound. We put $\Lambda(s,f) = \Gamma_{\mathbb{C}}(s)L(s,f)$, where $\Gamma_{\mathbb{C}}(s) = (2\pi)^{-s}\Gamma(s)$.

Theorem 2.1. *With notations as above, the completed L-function $\Lambda(s,f)$ extends to an entire function of $s \in \mathbb{C}$, and satisfies the functional equation $-i^k\Lambda(k-s,f) = \Lambda(s,f)$.*

Proof. Consider the global zeta integral

$$Z(s,f) = \int_0^\infty f(iy)y^s \frac{dy}{y}. \tag{2.6}$$

We write this as $Z(s,f) = Z_0(s,f) + Z_\infty(s,s)$ according to the decomposition of the integration range $\int_0^1 + \int_1^\infty$. The integral $Z_\infty(s,f)$ converges absolutely, in view of the rapid decay of the cusp form f at the cusp ∞. For $Z_0(s,f)$, we use the modular property $f|_k\gamma = f$ for $\gamma = w$, where w is the order 2 element $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$. This gives the relation

$$(iy)^{-k}f(i/y) = (iy)^{-k}f(-1/iy) = (f|_k w)(iy) = f(iy). \tag{2.7}$$

From this it follows that $f(iy)$ decays rapidly as $y \to 0$, establishing the absolute convergence of $Z_0(s,f)$. In this way, $Z(s,f)$ defines an analytic function of $s \in \mathbb{C}$. Note furthermore that by changing variables $y \mapsto 1/y$ in the defining expression (2.6) for $Z(s,f)$, and inserting (2.7), we obtain the functional equation $Z(s,f) = -i^k Z(k-s,f)$.

We now insert the Fourier expansion (2.4) and exchange the order of summation and integration (which is justified by the absolute convergence of both the integral and sum), obtaining

$$Z(s,f) = \int_0^\infty \sum_{n\geq 1} a_f(n)e^{-2\pi n y}y^s\frac{dy}{y} = \sum_{n\geq 1} a_f(n)\int_0^\infty e^{-2\pi n y}y^s\frac{dy}{y}.$$

Changing variables yields

$$Z(s,f) = (2\pi)^{-s}\sum_{n\geq 1} a_f(n)n^{-s}\int_0^\infty e^{-y}y^s\frac{dy}{y} = (2\pi)^{-s}\Gamma(s)L(s,f) = \Lambda(s,f).$$

This proves the Hecke integral representation of $\Lambda(s,f)$. From it we deduce the analytic continuation and functional equation of $\Lambda(s,f)$. $\square$

3. Classical Maass forms on $\mathbb{H}$

Just as there are many more Hecke characters than the algebraic ones, there are more classical modular forms than the holomorphic ones. To describe them, we begin by recalling the example from the first lecture, namely the real analytic Eisenstein series

$$E(s,z) = \frac{1}{2}\sum_{(m,n)\in\mathbb{Z}^2-0}\frac{y^s}{|cz+d|^{2s}}, \qquad \operatorname{Re}(s) > 1.$$

Amongst other properties, we saw that E_s satisfies

$$(\Delta + s(1-s))E_s = 0, \quad \text{where} \quad \Delta = y^2(\partial_x^2 + \partial_y^2).$$

This annihilation by the operator $\Delta + s(1-s)$ is a substitute for the holomorphy of modular forms. Recall the elliptic regularity theorem which states that any function annihilated by an elliptic operator with analytic coefficients (such as $\Delta + s(1-s)$) is analytic.

Definition 5. Let Γ be a lattice in $\mathrm{SL}_2(\mathbb{R})$. Let $s \in \mathbb{C}$. The space $M_s(\Gamma)$ of *Maass forms* relative to Γ and s is the space of functions $f : \mathbb{H} \to \mathbb{C}$ such that

- $(\Delta + s(1-s))f = 0$,
- $f(\gamma.z) = f(z)$ for all $\gamma \in \Gamma$,
- f is of moderate growth at all cusps.

Moderate growth here means that there is $A \in \mathbb{R}$ such that for any cusp $\mathfrak{a}$ of Γ, then $f(\sigma_{\mathfrak{a}}^{-1}z) \ll y^A$ as $y \to +\infty$, where $z = x + iy$.

Remark 3. Note that Δ extends to a self-adjoint negative semi-definite operator on the Hilbert space $L^2(\Gamma\backslash\mathbb{H})$. Its spectrum is therefore real and non-positive. In other words, $s(1-s) \geq 0$. We deduce that

$$s \in \frac{1}{2} + i\mathbb{R} \cup [0, 1]$$

Notice that $s(1-s) \geq \frac{1}{4}$ precisely when $s \in \frac{1}{2} + i\mathbb{R}$.

We have an analogous theory of q-expansion, but it is more complicated, since holomorphicity is gone. Let $\mathcal{W}_s$ denote the space of functions W on $\mathbb{H}$ such that

$$(\Delta + s(1-s))W = 0, \quad W(z + x) = e(x)W(z), \quad \exists A \in \mathbb{R} : W(iy) \ll y^A.$$

This is the space of Whittaker functions of parameter s.

Lemma 3.1. *For all $s \in \mathbb{C}$ we have* $\dim \mathcal{W}_s = 1$.

Proof. If $W \in \mathcal{W}_s$, then by separation of variables we have $W(z) = F_s(y)e(x)$, with F_s satisfying the degree 2 ODE

$$\left\{ y^2 \frac{d^2}{dy^2} + y\frac{d}{dy} - \left(y^2 + \left(s - \frac{1}{2} \right)^2 \right) \right\} F_s = 0.$$

The space of solutions is of dimension 2, generated by two fundamental solutions, one of which decays exponentially and the other grows exponentially. (The equation is a perturbation of $(\frac{d^2}{dy^2} - 1)F_s = 0$, whose solutions are generated by e^y and e^{-y}.) $\square$

In fact, $\mathcal{W}_s = \mathbb{C}\mathbb{W}_s$, where

$$\mathbb{W}_s(z) = y^{1/2}K_{s-1/2}(2\pi y)e^{2\pi i x}$$

and $K_\nu(y)$ is the K-Bessel funciton, and we can write any $f \in M_s(\Gamma)$ in its Fourier–Whittaker expansion

$$f(\sigma_{\mathfrak{a}} z) = \sum_{n\in\mathbb{Z}} \hat{f}_{\mathfrak{a}}(n)\mathbb{W}_s(nz).$$

We say that f is a *Maass cusp form* if it vanishes at every cusp; let $S_s(\Gamma)$ be the space of Maass cusp forms. Similarly to Lemma 2.1, from the Fourier–Whittaker expansion (and the exponential decay of $\mathbb{W}_s(iy)$ for y large) it follows that $f \in S_s(\Gamma)$ implies that $f \in L^2(\Gamma\backslash\mathbb{H})$.

3.1. **Ramanujan–Petersson–Selberg.** We now come to the analogue for Maass forms of the Ramanujan–Petersson conjecture. Let $f \in S_s(\Gamma_0(N))$ be primitive, with Fourier expansion at infinity given by

$$f(z) = \sum_{n\in\mathbb{Z}} a_f(n)|n|^{-1/2}\mathbb{W}_s(nz).$$

For $p \nmid N$ let $\alpha_p, \beta_p \in \mathbb{C}$ be complex numbers satisfying

$$(1 - a_f(p)p^{1/2-s} + p^{-2s})^{-1} = (1-\alpha_p p^{-s})^{-1}(1-\beta_p p^{-s})^{-1}.$$

Thus $\alpha_p + \beta_p = a_f(p)p^{1/2}$ and $\alpha_p\beta_p = 1$. We once again write $\alpha_p = p^{s_p}$ for $s_p \in \mathbb{C}$. With this normalization, we deduce the same constraints as in (2.5). Note that in the latter case, we obtain $|a_f(p)| \leq 2$. The Hecke bound states that $\mathrm{Re}(s_p) \in [0,1]$.

Conjecture 1. *Then*

1. *(Ramanujan-Petersson) for $p \nmid N$ we have $s_p \in \frac{1}{2} + i\mathbb{R}$.*
2. *(Selberg) $s \in \frac{1}{2} + i\mathbb{R}$.*

Unlike the case for modular forms this conjecture is still open and is considered one of the most important in the subject.

Remark 4. We only have bounds which improve upon the Hecke bound $a_n \ll n^{1/2}$. Note the following implication

$$\sum_{n\leq N} |a_n|^{2k} \ll N \quad \forall k \in \mathbb{N}^* \implies a_n \ll_\epsilon n^\epsilon.$$

Presently, we only know the above bound for $k = 1, 2, 3, 4$, which yields $a_n \ll n^{1/8+\epsilon}$. In a later lecture, we shall show the bound $a_n \ll n^{1/4+\epsilon}$, using a relative trace formula and bounds on Kloosterman sums.

Remark 5. The Selberg conjecture can be expressed as saying that $\lambda_f(\infty) \geq 1/4$, where $\lambda_f(\infty) = s(1-s)$ is the eigenvalue of $-\Delta$. Equivalently,

$$\lambda_1(\Gamma_1(N)\backslash\mathbb{H}) \geq 1/4,$$

where λ_1 is the smallest non-zero eigenvalue of the Laplacian on $L^2(\Gamma_1(N)\backslash\mathbb{H})$. Recalling the Rayleigh quotient for a compact Riemannian manifold

$$\lambda_1(Y) = \inf_{\int_Y f = 0} \frac{\int_Y |\nabla f|^2}{\int_Y |f|^2},$$

we can see that $\lambda_1(Y)$ measures the connectivity of Y.

4. From functions on $\mathbb{H}$ to functions on G

Set $G = \mathrm{SL}_2(\mathbb{R})$. We want to pass to a more representation theoretic approach by passing from functions on $\mathbb{H}$ to functions on G. Note that since G acts transitively on $\mathbb{H}$, and the stabilizer of i is $K = \mathrm{SO}(2)$, we may identify $\mathbb{H}$ with G/K.

The main difference to observe is that $\Gamma\backslash\mathbb{H}$ does not admit an action by G, but $\Gamma\backslash G$ does, by right translation

$$G \times \Gamma\backslash G \to \Gamma\backslash G, \qquad (g, \Gamma x) \mapsto \Gamma x g.$$

In fact $\Gamma\backslash G$ is a homogeneous space for G. The right G-action on the space $\Gamma\backslash G$ gives rise to a left action of G on functions $\mathfrak{F}(\Gamma\backslash G)$, in which (g, f) is sent to the function $R_g f : \Gamma x \mapsto f(\Gamma x g)$. Thus, if $f \in \mathfrak{F}(\Gamma\backslash G)$ we obtain a vector space

$$V_f = \langle f \rangle = \{R_g f : g \in G\},$$

which receives a linear action of G. This is then a representation of G, which is generally of infinite dimension. (It is not very useful to us, unless we put an appropriate topology on this space, which we will do in a later lecture.)

An additional benefit of passing to the group will be the fact that the Hecke integral for the L-function in (2.6) will be written in the form

$$\int_0^\infty \phi_f\left(\begin{pmatrix} y & \\ & 1 \end{pmatrix}\right) y^s \frac{dy}{y}.$$

The above integral can then be interpreted as a period integral, in the sense of being the integral of ϕ_f along the orbit under a subgroup of $\mathrm{PGL}_2(\mathbb{R})$, in this case the diagonal torus A.

4.1. Passage from $\mathfrak{F}(\mathbb{H})$ to $\mathfrak{F}(G)$.

For $k \in \mathbb{N}$, we recall the slash operator (2.2), which, thanks to the cocycle relation

$$j(gh, z) = j(g, h.z)j(h, z) \tag{4.1}$$

verifies the right-action rule $f|_k(gh) = (f|_k g)|_k h$.

Definition 6. For $k \in \mathbb{N}$, we define the *lifting map*

$$\Phi_k : \mathfrak{F}(\mathbb{H}) \to \mathfrak{F}(G), \qquad f \mapsto \phi_f(g) = f(g.i)j(g,i)^{-k} = (f|_k g)(i). \tag{4.2}$$

Let us first establish some basic properties of Φ_k.

Lemma 4.1. *The following holds.*

(1) *The map Φ_k is G-equivariant relative to the $|_k$ action of G on $\mathfrak{F}(\mathbb{H})$ and left-translation of G on $\mathfrak{F}(G)$:*

$$\Phi_k(f|_k g) = L_g(\Phi_k f).$$

(2) *Let* $\mathfrak{F}_k(G) = \{\phi \in \mathfrak{F}(G) : \phi(gr_\theta) = e^{ik\theta}\phi(g)\}$, *where*

$$r_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \in \mathrm{SO}(2).$$

Then Φ_k *is a bijection of* $\mathfrak{F}(\mathbb{H})$ *onto* $\mathfrak{F}_k(G)$.

Proof. From the cocycle relation (4.1) we have

$$\begin{aligned}\Phi_k(f|_k g)(h) = (f|_k g)(h.i)j(h,i)^{-k} &= j(g,h.i)^{-k} f(gh.i)j(h,i)^{-k} \\ &= f(gh.i)j(gh,i)^{-k} \\ &= L_g(\Phi_k f)(h).\end{aligned}$$

For the second property, recall the Iwasawa decomposition of G, which states that for every $g \in G$ we have a unique decomposition

$$g = \begin{pmatrix} y^{1/2} & xy^{-1/2} \\ & y^{-1/2} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Notice that $g.i = x + iy$. One then calculates

$$j(g,i) = ci + d = y^{-1/2}(\cos\theta - i\sin\theta) = y^{-1/2}e^{-i\theta}. \tag{4.3}$$

Thus

$$\Phi_k(f)(g) = y^{k/2}e^{ik\theta} f(x+iy). \tag{4.4}$$

It follows that $\mathrm{im}(\Phi_k) \subset \mathfrak{F}_k(G)$.

The map $\Phi_k : \mathfrak{F}(\mathbb{H}) \to \mathfrak{F}_k(G)$ is in fact bijective, with inverse sending ϕ to

$$f_\phi(z) = y^{-k/2}\phi\left(\begin{pmatrix} y^{1/2} & xy^{-1/2} \\ & y^{-1/2} \end{pmatrix}\right).$$

Indeed, we calculate

$$\begin{aligned}\phi_{f_\phi}(g) = y^{k/2}e^{ik\theta} f_\phi(x+iy) = e^{ik\theta}\phi&\left(\begin{pmatrix} y^{1/2} & xy^{-1/2} \\ & y^{-1/2} \end{pmatrix}\right) \\ = \phi&\left(\begin{pmatrix} y^{1/2} & xy^{-1/2} \\ & y^{-1/2} \end{pmatrix} r_\theta\right) = \phi(g). \qquad \square\end{aligned}$$

In this way, the map Φ_k intertwines the representation $\mathfrak{F}(\mathbb{H})$ of G (using the $|_k$ operation) and the left regular representation of G on $\mathfrak{F}_k(G)$.

4.2. **Lifting holomorphic functions.** We shall now examine how the lifting map interacts with analytic structures on $\mathfrak{F}(\mathbb{H})$, such as holomorphy. Let

$$\partial_z = \frac{1}{2}\left(\partial_x - i\partial_y\right) \quad \text{and} \quad \partial_{\bar{z}} = \frac{1}{2}\left(\partial_x + i\partial_y\right).$$

Let $\mathrm{Hol}(\mathbb{H})$ denote the space of holomorphic functions on $\mathbb{H}$:

$$\mathrm{Hol}(\mathbb{H}) = \{f \in \mathfrak{F}(\mathbb{H}) : \partial_{\bar{z}} f = 0\}.$$

There is no notion of holomorphy on G (which is odd dimensional, in any case) but we shall provide a substitute, using something that resembles the $\partial_{\bar{z}}$ operator, but which also sees the θ variable. Namely, we put

$$\begin{aligned}\mathbf{L} &= e^{-2i\theta}\left(-iy\partial_x + y\partial_y + \frac{i}{2}\partial_\theta\right)\\ &= e^{-2i\theta}\left(-2iy\partial_{\bar{z}} + \frac{i}{2}\partial_\theta\right).\end{aligned}$$

Lemma 4.2. *If $f \in \mathrm{Hol}(\mathbb{H})$ and $\phi = \Phi_k(f)$ then $\mathbf{L}\phi = 0$.*

Proof. Since, from (4.4), we have $\phi(g) = y^{k/2}e^{ik\theta}f(g.i)$, it will suffice to show that

$$\mathbf{L}(y^{k/2}e^{ik\theta}) = 0 \quad \text{and} \quad \mathbf{L}f(g.i) = 0.$$

By direct calculation, we have

$$\mathbf{L}(y^{k/2}e^{ik\theta}) = e^{-2i\theta}(\frac{k}{2}y^{k/2}e^{ik\theta} + \frac{i}{2}(ik)y^{k/2}e^{ik\theta}) = 0.$$

Since $f(g.i)$ is constant in the θ variable, we have

$$\mathbf{L}f(g.i) = -2iye^{-2i\theta}\partial_{\bar{z}}f(z) = 0,$$

by the holomorphy of f. □

A function on $\mathbb{H}$ which is killed by the operator $\partial_{\bar{z}}$ has a regularity property: it is holomorphic. Do functions in

$$\{\phi \in \mathfrak{F}_k(G) : \mathbf{L}\phi = 0\}$$

also have a regularity property?

Lemma 4.3. *If $f \in \mathrm{Hol}(\mathbb{H})$ then $\phi = \Phi_k(f)$ is real analytic.*

Recall that an analytic function on G is one which admits a power series expansion locally about every point. We denote by $C^\omega(G)$ the space of analytic functions on G, and $C^\omega_k(G) = C^\omega \cap \mathfrak{F}_k(G)$. We deduce from the lemma the following

Corollary 1. *With notations as above,*

$$\mathrm{Im}(\Phi_k) = \{\phi \in C^\omega_k(G) : \mathbf{L}_k\phi = 0\}.$$

4.3. **Higher weight Laplacians.** To prove the lemma one must pass to higher order left-invariant differential operators.

Let $\mathbf{\Delta}$ be the second-order differential operator on $C^2(G)$ defined by

$$\mathbf{\Delta} = y^2(\partial_x^2 + \partial_y^2) - y\partial_x\partial_\theta.$$

Then $\mathbf{\Delta}$ is a *hyperbolic* differential operator. Indeed, writing

$$\mathbf{R} = e^{2i\theta}\left(2iy\partial_z - \frac{i}{2}\partial_\theta\right) \qquad \text{and} \qquad \mathbf{H} = -i\partial_\theta,$$

then we calculate

$$\mathbf{L}\circ\mathbf{R} = 8y^2\partial_{\bar z}\partial_z - 2y\partial_\theta(\partial_{\bar z}+\partial_z) + \frac{1}{2}\partial_\theta^2 + 2i\partial_\theta = -2\mathbf{\Delta} - \frac{1}{2}(\mathbf{H}\circ\mathbf{H}) - 2\mathbf{H}$$
$$\mathbf{R}\circ\mathbf{L} = 8y^2\partial_{\bar z}\partial_z - 2y\partial_\theta(\partial_{\bar z}+\partial_z) + \frac{1}{2}\partial_\theta^2 - 2i\partial_\theta = -2\mathbf{\Delta} - \frac{1}{2}(\mathbf{H}\circ\mathbf{H}) + 2\mathbf{H},$$

so that

$$2(\mathbf{L}\circ\mathbf{R}) + 2(\mathbf{R}\circ\mathbf{L}) + \mathbf{H}\circ\mathbf{H} = 4\mathbf{\Delta}. \tag{4.5}$$

Thus $\mathbf{\Delta} = \sigma_{\mathbf{\Delta}}(\mathbf{L},\mathbf{R},\mathbf{H})$, where

$$\sigma_{\mathbf{\Delta}}(X,Y,Z) = \frac{1}{4}(2XY + 2YX + Z^2) \in \mathbb{C}[X,Y,Z]$$

is the *principal symbol* of $\mathbf{\Delta}$. Viewed as a quadratic form over $\mathbb{R}$, it is of signature $(1,2)$.

Now let

$$\mathbf{L}_k = -2iy\partial_{\bar z} - \frac{k}{2} \qquad \text{and} \qquad \mathbf{R}_k = 2iy\partial_z + \frac{k}{2}$$

be the restrictions of $e^{2i\theta}\mathbf{L}$ and $e^{-2i\theta}\mathbf{R}$ to $C^1_k(G) = C^1(G)\cap\mathfrak{F}_k(G)$.

Lemma 4.4. *The operators* $\mathbf{L}_k$ *and* $\mathbf{R}_k$ *send* $C^1_k(G)$ *to* $C_{k-2}(G)$ *and* $C_{k+2}(G)$*, respectively.*

Proof. See Bump, Lemma 2.1.1, page 130. □

For this reason, $\mathbf{L}_k$ and $\mathbf{R}_k$ are called the weight k **lowering** and **raising** operators.

The restriction of $\mathbf{\Delta}$ to $C^2_k(G) = C^2(G)\cap\mathfrak{F}_k(G)$ is given by the *weight* k *Laplacian*

$$\mathbf{\Delta}_k = y^2(\partial_x^2 + \partial_y^2) - iky\partial_x. \tag{4.6}$$

Now we can clearly rewrite this last expression as

$$\mathbf{\Delta}_k = \mathbf{R}_{k-2}\mathbf{L}_k - \frac{k}{2}\left(1-\frac{k}{2}\right) = \mathbf{L}_{k+2}\mathbf{R}_k + \frac{k}{2}\left(1+\frac{k}{2}\right). \tag{4.7}$$

From the preceding lemma, we deduce the following

Corollary 2. *The operator* $\mathbf{\Delta}_k$ *takes weight* k *functions to weight* k *functions.*

When restricted to right $K = \mathrm{SO}(2)$-invariant functions on G, the formula (4.6) for $\mathbf{\Delta}_0$ on G returns that of the hyperbolic Laplacian Δ on $\mathbb{H} = G/K$.

The weight k Laplacian (unlike $\mathbf{\Delta}$ itself) is an *elliptic* differential operator, as its highest degree homogenous component is $y^2(\partial_x^2+\partial_y^2)$, whose principal symbol is non-zero away from the origin.

Now we return to the proof of Lemma 4.3. Since $\mathbf{L}_k\phi = 0$ it follows that

$$\left(\mathbf{\Delta}_k + \frac{k}{2}\left(1-\frac{k}{2}\right)\right)\phi = 0.$$

As an eigenfunction for an elliptic differential operator (with analytic coefficients) ϕ is an analytic function on G, by the elliptic regularity theorem.

5. Lifting modular forms

From the left G-equivariance of Φ_k and Corollary 1 we deduce that Φ_k defines a continuous bijective map from the space

$$\{f \in \mathrm{Hol}(\mathbb{H}) : f|_k\gamma = f_k \quad \forall \gamma \in \Gamma\}$$

onto

$$\{\phi \in C_k^\omega(\Gamma\backslash G) : \mathbf{L}\phi = 0\}.$$

Let us now determine the image of $M_k(\Gamma)$, which involves the condition of holomorphy at the cusp, as well as that of the cuspidal subspace $S_k(\Gamma)$. For simplicity, we shall concentrate only on the image of $S_k(\Gamma)$.

5.1. **Square integrability.** The group $G = \mathrm{SL}_2(\mathbb{R})$ is unimodular: a left-invariant Haar measure is right-invariant. Up to scaling, the Haar measure on G is given in Iwasawa coordinates $G = NAK$ as

$$dg = \frac{dxdy}{y^2}\frac{d\theta}{2\pi}.$$

Note that the push-forward of the Haar measure on G to $G/K = \mathbb{H}$ is the invariant hyperbolic measure $dxdy/y^2$.

We write $L^2(\Gamma\backslash G)$ for the Hilbert space of L^2-integrable functions on $\Gamma\backslash G$ with respect to the above measure.

Lemma 5.1. *The image of $S_k(\Gamma)$ under Φ_k lies in $L^2(\Gamma\backslash G)$.*

Proof. Recall the definition (4.2) and the calculation (4.3). For $g \in G$ we have

$$|\phi_f(g)| = \left| f\left(\frac{ai+b}{ci+d}\right)\frac{1}{(ci+d)^k}\right| = |f(z)|(\mathrm{Im}\, z)^{k/2} \qquad (z = g.i),$$

which is right-K-invariant. Thus it suffices to show establish the lemma for f relative to the measure $y^{k-2}dxdy$ on $\mathbb{H}$. But this is Lemma 2.1. $\square$

5.2. **Unipotent integration.** We return to the description of cusps of arithmetic lattices of $\mathrm{SL}_2(\mathbb{R})$.

Recall that $\mathrm{SL}_2(\mathbb{R})$ acts on $\mathbb{P}^1(\mathbb{R}) = \mathbb{R}\cup\{\infty\}$ by fractional linear transformations. As this action is transitive, and the stabilizer of ∞ is the standard Borel subgroup of upper triangular matrices B, we may identify $\mathbb{P}^1(\mathbb{R})$ with $\mathrm{SL}_2(\mathbb{R})/B$. Since all Borel subgroups are conjugate, and the normalizer of B is itself, we may identify the quotient $\mathrm{SL}_2(\mathbb{R})/B$ with the variety of Borel subgroups of $\mathrm{SL}_2(\mathbb{R})$.

In fact, we may make this identification over $\mathbb{Q}$. Thus $\mathbb{P}^1(\mathbb{Q}) = \mathrm{SL}_2(\mathbb{Q})/B$, the variety of $\mathbb{Q}$-Borel sugroups. We can then view the cusps of Γ as the Γ-inequivalent $\mathbb{Q}$-Borel subgroups of SL_2, with the latter being viewed as an algebraic group over $\mathbb{Q}$. We write this correspondence as $\mathfrak{a} \leftrightarrow P_{\mathfrak{a}}$.

Let us reinterpret the cuspidality condition, replacing the vanishing of the constant term at all cusps by the vanishing of certain integrals along unipotent subgroups. Let P be a $\mathbb{Q}$-Borel subgroup of G. Write N_P for its unipotent radical (the largest normal subgroup of P consisting of unipotent elements). We define the *constant term* ϕ_P of ϕ along P to be the function

$$\phi_P(g) = \int_{\Gamma \cap N_P(\mathbb{R}) \backslash N_P(\mathbb{R})} \phi(ng) dn.$$

Proposition 5.1. *Let $\mathfrak{a}$ be a cusp of the arithmetic lattice Γ, of period $h'_{\mathfrak{a}}$. Let $f \in M_k(\Gamma)$. Then*

$$(\phi_f)_{P_{\mathfrak{a}}}(g) = y^{k/2} e^{ki\theta} h'_{\mathfrak{a}} \hat{f}_{\mathfrak{a}}(0),$$

where $z = x + iy$ is given by $z = (\sigma_{\mathfrak{a}}^{-1} g).i$ and $\sigma_{\mathfrak{a}} \in \mathrm{SL}_2(\mathbb{Z})$ is an integral scaling matrix taking ∞ to $\mathfrak{a}$, as in §2.

Proof. Write $N_{\mathfrak{a}}$ for the unipotent radical of $P_{\mathfrak{a}}$. Since $\sigma_{\mathfrak{a}}\infty = \mathfrak{a}$ we have $N_{\mathfrak{a}} = \sigma_{\mathfrak{a}} N \sigma_{\mathfrak{a}}^{-1}$, where N is group of upper triangular unipotent matrices (the unipotent radical of the standard Borel B). Thus

$$(\phi_f)_{P_{\mathfrak{a}}}(g) = \int_{\Gamma \cap N_{\mathfrak{a}}(\mathbb{R}) \backslash N_{\mathfrak{a}}(\mathbb{R})} \phi_f(ng) dn = \int_{\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}} \cap N(\mathbb{R}) \backslash N(\mathbb{R})} \phi_f(\sigma_{\mathfrak{a}} n \sigma_{\mathfrak{a}}^{-1} g) dn.$$

Recall from §2 that $f_{\mathfrak{a}} = f|_k \sigma_{\mathfrak{a}}$. Using the definition $\phi_f(g) = (f|_k g)(i)$ and the rule $f|_k(gh) = (f|_k g)|_k h$, we note that

$$\phi_f(\sigma_{\mathfrak{a}} n \sigma_{\mathfrak{a}}^{-1} g) = (f|_k \sigma_{\mathfrak{a}} n \sigma_{\mathfrak{a}}^{-1} g)(i) = ((f_{\mathfrak{a}})|_k n)|_k (\sigma_{\mathfrak{a}}^{-1} g)(i).$$

Furthermore, for any function F on $\mathbb{H}$ we have

$$F|_k(\sigma_{\mathfrak{a}}^{-1} g)(i) = j(\sigma_{\mathfrak{a}}^{-1} g, i)^{-k} F(\sigma_{\mathfrak{a}}^{-1} g.i) = j(\sigma_{\mathfrak{a}}^{-1} g, i)^{-k} F(z) = y^{k/2} e^{ki\theta},$$

where we have used (4.3) and the definition of z in the statement of the proposition. Now $\Gamma_{\mathfrak{a}} \cap N_{\mathfrak{a}}(\mathbb{R}) = \Gamma \cap N_{\mathfrak{a}}(\mathbb{R})$, so that $(\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}})_\infty \cap N(\mathbb{R}) = \sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}} \cap N(\mathbb{R})$. Thus, using the description of $(\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}})_\infty$ from §2, we have

$$\begin{aligned}(\phi_f)_{P_{\mathfrak{a}}}(g) &= y^{k/2} e^{ki\theta} \int_{\sigma_{\mathfrak{a}}^{-1}\Gamma\sigma_{\mathfrak{a}} \cap N(\mathbb{R}) \backslash N(\mathbb{R})} ((f_{\mathfrak{a}})|_k n)(z) dn \\ &= y^{k/2} e^{ki\theta} \int_{h'_{\mathfrak{a}}\mathbb{Z} \backslash \mathbb{R}} (f_{\mathfrak{a}})(z + x) dx \\ &= y^{k/2} e^{ki\theta} \int_0^{h_{\mathfrak{a}}} \sum_{n \geq 0} \hat{f}_{\mathfrak{a}}(n) e^{2\pi i n (z+x)/h'_{\mathfrak{a}}} dx.\end{aligned}$$

Only the zero frequency Fourier coefficient survives, yielding the result. □

5.3. **Characterizing the image of cusp forms.**

Theorem 5.1. *For $k \geq 1$, the image of $S_k(\Gamma)$ under Φ_k is the space of functions $\phi \in L^2(\Gamma \backslash G)$ such that*

(1) $\phi(g r_\theta) = e^{ik\theta} \phi(g)$;
(2) $(\mathbf{\Delta} + \frac{k}{2}(1 - \frac{k}{2}))\phi = 0$;

(3) *for every $\mathbb{Q}$-Borel subgroup P of G, the constant term ϕ_P vanishes almost everywhere on G.*

Proof. We have already shown that the image of $S_k(\Gamma)$ lies in the above space.

Conversely, we know from Lemma 4.1 that Γ-invariant functions satisfying condition (1) admit preimages by Φ_k to functions on $\mathfrak{H}$ satisfying the modularity condition $f|_k\gamma = f$ for all $\gamma \in \Gamma$. Furthermore, in Proposition 5.1 we showed that (3) is equivalent to the classical cuspidality condition defining $S_k(\Gamma)$ inside $M_k(\Gamma)$; see Definition 4.

It remains to examine condition (2) and relate it to holomorphy. For this, recall the expression (4.7). From this we see that if we grant ourselves condition (1), then condition (2) is equivalent to $\mathbf{RL}\phi = 0$. Thus, to show that Φ_k is onto, we must show that if $\phi \in L^2(\Gamma\backslash G)$ satisfies conditions (1) and (3), as well as $\mathbf{RL}\phi = 0$, then $\mathbf{L}\phi = 0$. We will do this by showing that the integral $\int_{\Gamma\backslash G} |\mathbf{L}\phi(g)|^2 dg$ vanishes. Ultimately, this will an application of Stokes' theorem on the non-compact space $\Gamma\backslash G$.

We begin by writing $|\mathbf{L}\phi(g)|^2 dg$ as the derivative of a 2-form ω on $\Gamma\backslash G$. This will be possible precisely because of our assumption that $\mathbf{RL}\phi = 0$. Firstly, we note that

$$|\mathbf{L}\phi|^2 = \mathbf{L}\phi.\overline{\mathbf{L}\phi} = \mathbf{L}\phi.\mathbf{R}\overline{\phi},$$

since $\mathbf{L}$ and $\mathbf{R}$ are complex conjugates of one another. Now the Leibniz rule for $\mathbf{R}$ together with the hypothesis $\mathbf{RL}\phi = 0$ give

$$\mathbf{R}\left((\mathbf{L}\phi).\overline{\phi}\right) = \mathbf{L}\phi.\mathbf{R}\overline{\phi} + \overline{\phi}\,(\mathbf{RL}\phi) = \mathbf{L}\phi.\mathbf{R}\overline{\phi}.$$

Thus we obtain $|\mathbf{L}\phi|^2 = \mathbf{R}\left((\mathbf{L}\phi).\overline{\phi}\right)$. Then an explicit calculation shows[2] that, if we set

$$\omega = -e^{-2i\theta}\left((\mathbf{L}\phi).\overline{\phi}\right)\left(\frac{d\bar{z}d\theta}{y} + i\frac{dxdy}{2y^2}\right),$$

a 2-form on $\Gamma\backslash G$, then $\mathbf{R}\big((\mathbf{L}\phi).\overline{\phi}\big)\, dg = d\omega$. Thus

$$\int_{\Gamma\backslash G} |\mathbf{L}\phi(g)|^2 dg = \int_{\Gamma\backslash G} d\omega$$

We break up $\Gamma\backslash G$ into (horocyclic) cuspidal neighborhoods $\mathcal{N}_\epsilon$ (for $\epsilon > 0$ small) and its complement $\mathcal{C}_\epsilon = \Gamma\backslash G - \mathcal{N}_\epsilon$. Then

$$\int_{\Gamma\backslash G} d\omega = \int_{\mathcal{C}_\epsilon} d\omega + \int_{\mathcal{N}_\epsilon} d\omega = \int_{\partial\mathcal{C}_\epsilon} \omega + \int_{\mathcal{N}_\epsilon} d\omega, \tag{5.1}$$

the last equality by Stokes' theorem. We want to show that both integrals go to zero as $\epsilon \to 0$. For the first (resp. second), this is the same thing as to show that $(\mathbf{L}\phi).\overline{\phi}$ (resp. $\mathbf{R}\big((\mathbf{L}\phi).\overline{\phi}\big)$) decays as g enters each cusp.

---

[2]See `https://virtualmath1.stanford.edu/~conrad/conversesem/Notes/L5.pdf`, pages 23-24

We use an important lemma of Harish-Chandra, which applies in particular to any ϕ satisfying the conditions of the theorem. It says that for any neighborhood U of the identity in G there exists an $\alpha_0 \in C_c^\infty(G)$ such that $\phi = \alpha_0 * \phi$ (convolution product on G). The equality $D\phi = D(\phi * \alpha_0) = \phi * (D\alpha_0)$ then allows us to replace $\mathbf{L}\phi$ by $\phi * (\mathbf{L}\alpha_0)$.

Now integration by parts shows that for any $\alpha \in C^\infty(G)$ and any $n \geq 0$ there exists $c_n(\alpha) > 0$ such that

$$|(\phi * \alpha)(\sigma_{\mathfrak{a}} g)| \leq c_n(\alpha) y^{-n} \|\phi\|_2,$$

where $g.i = x + iy$ and $\sigma_{\mathfrak{a}}(\infty) = \mathfrak{a}$. Applying this to $\alpha = \alpha_0$ (for ϕ) and $\alpha = \mathbf{L}\alpha_0$ (for $\mathbf{L}\phi$) yields the rapid decay of the integral $\int_{\partial \mathcal{C}_\epsilon} \omega$ in (5.1). A similar reasoning shows the same for the integral $\int_{\mathcal{N}_\epsilon} d\omega$. $\square$