Hecke's precursor to endoscopy

Bill Casselman
University of British Columbia
cass@math.ubc.ca

Ngô Bảo Châu recently won a Fields Medal for his work on the Fundamental Lemma, conjectured by Langlands many years ago. But the relationship between the result proved by him and Langlands' original conjecture might not be so clear to someone who has not followed the development closely. The original idea involved what Langlands calls **endoscopy**—to explain harmonic analysis on reductive groups over local fields as well as on arithmetic quotients in terms of what Langlands calls stable analysis on the original group and certain other, smaller, groups associated to it. These smaller groups are called *endoscopic*. The subject is in rapid flux, and the present state of exposition is necessarily, in view of our present understanding, generally very technical. However, I believe that the basic idea of endoscopy can be explained by looking at what was in fact the very first place in which it showed up, in work of Erich Hecke.

Endoscopy decomposes conjugate-invariant distributions on reductive groups over local fields as well as on arithmetic quotients into what Langlands calls *stable* components. Already for $\mathrm{SL}_2(\mathbb{Q})$ and associated local groups this is a non-trivial business, explored in detail in [Labesse-Langlands:1979]. One of the main results of that paper is that subspaces of automorphic forms on the adelic quotient $\mathrm{SL}_2(\mathbb{Q})\backslash\mathrm{SL}_2(\mathbb{A})$ arising from quadratic extensions of $\mathbb{Q}$ *via* theta functions can be characterized in terms of local invariants and quadratic reciprocity.

As the paper [Labesse-Langlands:1979] already briefly mentions, certain simple cases of endoscopy were known already to Hecke, in about 1930. In this paper I'll explain Hecke's results in modern terms, although still following his argument closely. This argument will avoids the technicalities of representation theory of local groups and also the Selberg trace formula. Instead it deals only with harmonic analysis on certain finite groups, and applies the fixed point formula of Atiyah-Bott. All the serious analytic difficulties vanish (or are at least hidden in [Atiyah-Bott:1964]), but at the same time many of the important phenomena involved in endoscopy, and its purpose, already appear to an extent that I hope to be instructive.

At the end, I'll say something about the relationship between the argument here and that in [Langlands-Labesse:1972].

I wish to thank audiences at the Tata Institute in Mumbai and the universities of Minnesota and McGill, as well as Mark Goresky, for helping me work things out.

Contents

1. Modular forms

The group $\mathrm{PSL}_2(\mathbb{Z}/N) = \mathrm{SL}_2(\mathbb{Z}/N)/\{\pm I\}$ acts on the holomorphic cusp forms of weight 2 on $\mathcal{H}$ with respect to the principal congruence group $\Gamma(N)$ of level N:

$$\pi(\gamma)f = f \,\big|\, [\gamma^{-1}]_2$$

where (using Shimura's notation)

$$\big(f \,\big|\, [g]_k\big)(z) = f\big(g(z)\big)(cz+d)^{-k} \text{ if } g = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

The group $\Gamma(N)$ itself acts trivially by definition, and so does $\pm I$, so the action passes to the quotient $\mathrm{SL}_2(\mathbb{Z})/\Gamma(N)\cdot\{\pm I\}$, which is canonically isomorphic to $\mathrm{PSL}_2(\mathbb{Z}/N)$. A natural question apparently first investigated by Hecke is: *what is the irreducible decomposition of this representation?* At the time he asked this, the classification of the irreducible representations of $\mathrm{PSL}_2(\mathbb{Z}/N)$ had been worked out by Frobenius in the case $N = p$, a prime, and this was the case Hecke looked at.

Incidentally, in reading Hecke's paper it should be kept in mind that he works with the *right* action $f \mapsto f\,\big|[\gamma]_k$ instead of the one I do.

Let $\overline{X}_p$ be the compactification of X_p by cusps. The forms of weight two with respect to $\Gamma(p)$ may be identified with the space of holomorphic differential forms on $\overline{X}_p$. Let π be the representation of $\mathrm{PSL}_2(\mathbb{Z}/p)$ on this space. It was known to Hecke, as it is to us, that the representation $\pi \oplus \overline{\pi}$ is that of $\mathrm{PSL}_2(\mathbb{Z}/p)$ on the rational cohomology of $\overline{X}_p$. Therefore the sum of π and its complex conjugate has to be rational. But the representation π itself is not necessarily even real, and so it makes sense to ask, *what can one say about the difference between π and $\overline{\pi}$?* This is the question that most intrigued Hecke.

The case $p = 2$ is not interesting. Nor is the case $p \equiv 1 \bmod 4$, since it follows from Frobenius' character tables that in this case π is always isomorphic to its conjugate. What is interesting is the remaining case in which $p \equiv 3 \bmod 4$. Understanding what happens for these primes leads to an instructive prototype of later results of Labesse and Langlands. Since $\mathrm{PSL}_2(\mathbb{Z}/p)$ is a finite group, phenomena are more transparent than they are for $\mathrm{SL}_2(F)$, and it is remarkable that many of the features explored by Labesse and Langlands occur already here. Among other things, much like the later results, Hecke's observation has ties to quadratic reciprocity.

The case $p = 3$ is exceptional, and also uninteresting. So from now on I assume that $p > 3$, $p \equiv 3 \bmod 4$. One immediate consequence of this assumption is that -1 is not a square in $\mathbb{Z}/p$. I'll first discuss the representations of $\mathrm{SL}_2(\mathbb{Z}/p)$, then explain how this relates to the representation on holomorphic forms on $\overline{X}_p$.

2. Representations of PSL(2,Z/p)

Let $F = \mathbb{Z}/p$, $E = F(\sqrt{-1})$, Let $G = \mathrm{PSL}_2(F)$. I'll recall here (following Hecke) the classification of conjugacy classes and irreducible representations of G.

Representatives of conjugacy classes are

$$\begin{aligned}
&\pm\begin{bmatrix}1&0\\0&1\end{bmatrix}\\
&\pm\begin{bmatrix}t&0\\0&1/t\end{bmatrix}\sim\pm\begin{bmatrix}1/t&0\\0&t\end{bmatrix}\quad(t\neq 1)\\
&\pm\begin{bmatrix}0&-1\\1&0\end{bmatrix}\\
&\pm\begin{bmatrix}a&-b\\b&a\end{bmatrix}\sim\pm\begin{bmatrix}a&b\\-b&a\end{bmatrix}\quad(a^2+b^2=1,a\neq 0)\\
&\pm\begin{bmatrix}1&1\\0&1\end{bmatrix}\\
&\pm\begin{bmatrix}1&-1\\0&1\end{bmatrix}.
\end{aligned}$$

The classification of semi-simple conjugacy classes will be dealt with in detail later on, but I point out here that the map

$$a+b\sqrt{-1}\longmapsto\begin{bmatrix}a&-b\\b&a\end{bmatrix}$$

is an embedding of the group $N^1_{E/F}$ into $\mathrm{SL}_2(F)$, inducing a map from $N^1_{E/F}/\{\pm1\}$ into G. The group $N^1_{E/F}/\{\pm1\}$ is cyclic of order $(p+1)/2$. Since $p\equiv 3 \bmod 4$, it has a unique non-trivial element ε_0 of order two, as well as a unique non-trivial character ρ_0 of order two. As for the unipotent classes, the important fact is that

$$\begin{bmatrix}t&0\\0&1/t\end{bmatrix}\begin{bmatrix}1&x\\0&1\end{bmatrix}\begin{bmatrix}t&0\\0&1/t\end{bmatrix}^{-1}=\begin{bmatrix}1&t^2x\\0&1\end{bmatrix},$$

so that since -1 is not a square in F the two unipotent classes listed above are distinct. Among all classes, they have the unique property that although distinct in $\mathrm{PSL}_2(F)$ they fuse in $\mathrm{PSL}_2(E)$. This is the simplest example of the distinction between ordinary conjugacy and what Langlands calls **stable conjugacy** in reductive groups—two elements are said to be stably conjugate if they are conjugate over an algebraic closure.

The matrix

$$\iota=\begin{bmatrix}-1&0\\0&1\end{bmatrix}$$

in $\mathrm{GL}_2(F)$ acts by conjugation on G. All except the two unipotent classes in G are fixed by this conjugation, but the unipotent classes are swapped by it. This feature is valid for all the groups $\mathrm{SL}_n\subset\mathrm{GL}_n$, and makes the analysis of stable conjugacy in SL_n relatively simple.

As for representations, the group G has

- one representation of dimension 1 (the trivial representation)
- one representation of dimension p (called the **Steinberg** representation)
- representations $\pi(\chi)$ of dimension $p+1$ parametrized by characters χ of $F^\times$ (the principal series)
- representations $\pi(\rho)$ of dimension $p-1$ parametrized by characters $\rho\neq\rho_0$ of $N^1_{E/F}$ (cuspidal representations)
- two representations of dimension $(p-1)/2$, which I'll call $\pi_\pm$, corresponding to the unique character ρ_0 of order two of $N^1_{E/F}$

The trivial and Steinberg representations together decompose the representation of G on the space $\mathbb{C}\big(\mathbb{P}^1(F)\big)$. It is only the last two we are really interested in. In general, there is a representation $\pi(\rho)$ of G associated

to every ρ of $N^1_{E/F}/\{\pm 1\}$, and there exists a G-ismorphism T of $\pi(\rho)$ with $\pi(1/\rho)$. For $\rho = \rho_0$ this is a non-trivial G-automorphism of $\pi(\rho_0)$ of order two, and the representations $\pi_\pm$ are its eigenspaces.

The representations $\pi_\pm$ have the unique feature that neither is isomorphic to its complex conjugate. Instead, complex conjugation interchanges them. What is especially important for us is that *their characters differ only on the two unipotent classes of G*. In other words, the unusual representations $\pi_\pm$ are in some way matched with the unusual unipotent conjugacy classes. Details of this matching appear in the following character table:

REPRESENTATION / CONJUGACY CLASS	I	STEINBERG	$\pi(\chi)$	$\pi(\rho)$	$\pi_+(\rho_0)$	$\pi_-(\rho_0)$
$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	1	p	$p+1$	$p-1$	$(p-1)/2$	$(p-1)/2$
$\begin{bmatrix} t & 0 \\ 0 & 1/t \end{bmatrix}$	1	1	$\chi(t)+\chi(1/t)$	0	0	0
$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$	1	-1	0	$-2\rho(\varepsilon_0)$	$-\rho_0(\varepsilon_0)$	$-\rho_0(\varepsilon_0)$
$\begin{bmatrix} a & -b \\ b & a \end{bmatrix}$	1	-1	0	$-\rho(\varepsilon)-\rho(1/\varepsilon)$	$-\rho_0(\varepsilon)$	$-\rho_0(\varepsilon)$
$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$	1	0	1	-1	$\overline{\mathfrak{G}}$	$\mathfrak{G}$
$\begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$	1	0	1	-1	$\mathfrak{G}$	$\overline{\mathfrak{G}}$

Here

$$\mathfrak{G} = \sum_{(x/p)=1} e^{2\pi i x/p}, \quad \overline{\mathfrak{G}} = \sum_{(x/p)=-1} e^{2\pi i x/p}$$

are partial Gauss sums, $\varepsilon = a + b\sqrt{-1}$, and $\varepsilon_0 = \sqrt{-1}$. There is a slight difference between my notation and Hecke's. Because his representations are right actions, his characters are the conjugates of the ones in this table.

The representations $\pi_\pm$ are to be thought of as twins, distinguished only by which of $\mathfrak{G}$ and $\overline{\mathfrak{G}}$ they correspond to. One thing that can be read from this table is that π^ι is isomorphic to $\overline{\pi}$ for all π, and isomorphic to π itself for all but $\pi_\pm$, for which

$$\pi^\iota_\pm \cong \overline{\pi}_\pm \cong \pi_\mp \,.$$

One consequence of these observations is a very close analogue of the Fundamental Lemma. The space of conjugation-invariant 'distributions' D on G such that $D = -D^\iota$ has dimension one. It includes

$$f \longmapsto \operatorname{trace} \pi_+(f) - \operatorname{trace} \pi_-(f)$$

as well as the difference of 'orbital integrals'

$$f \longmapsto \sum_{x \sim \nu} \big(f(x) - f(x^\iota)\big)\,,$$

where

$$\nu = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \bmod p\,.$$

These two are therefore proportional to each other. This is exactly what the Fundamental Lemma is all about, but for p-adic groups.

For our purposes another direct consequence is more relevant.

2.1. Proposition. *Suppose π to be any finite-dimensional representation of G. Let $m_\pm$ be the multiplicity of $\pi_\pm$ in its irreducible decomposition. Then*

$$m_+ - m_- = \frac{\operatorname{trace}\pi(\nu) - \operatorname{trace}\overline{\pi}(\nu)}{\overline{\mathfrak{G}} - \mathfrak{G}}\,.$$

Proof. Say

$$\pi = \sum_k m_k \pi_k$$

is the decomposition of π into irreducibles. Then for g in G

$$\begin{aligned}
\operatorname{trace}\pi(g) &= \sum m_k \operatorname{trace}\pi_k(g)\\
\operatorname{trace}\overline{\pi}(g) &= \sum m_k \operatorname{trace}\overline{\pi}_k(g)\\
\operatorname{trace}\pi(g) - \operatorname{trace}\overline{\pi}(g) &= \sum m_k\big(\operatorname{trace}\pi_k(g) - \operatorname{trace}\overline{\pi}_k(g)\big)\,.
\end{aligned}$$

All terms in this last sum vanish unless g lies in the 'stable conjugacy class' of unipotents, and in that case the only terms that don't vanish are those for the pair $\pi_\pm$. Then

$$\begin{aligned}
\operatorname{trace}\pi(\nu) - \operatorname{trace}\overline{\pi}(\nu) &= m_+\big(\operatorname{trace}\pi_+(\nu) - \operatorname{trace}\overline{\pi}_-(\nu)\big) + m_-\big(\operatorname{trace}\pi_-(\nu) - \operatorname{trace}\overline{\pi}_+(\nu)\big)\\
&= (m_+ - m_-)\big(\operatorname{trace}\pi_+(\nu) - \operatorname{trace}\overline{\pi}_-(\nu)\big)\\
&= (m_+ - m_-)\big(\overline{\mathfrak{G}} - \mathfrak{G}\big)\\
m_+ - m_- &= \frac{\operatorname{trace}\pi(\nu) - \operatorname{trace}\overline{\pi}(\nu)}{\overline{\mathfrak{G}} - \mathfrak{G}}\,.
\end{aligned}$$

3. The modular curves

We are going to apply Proposition 2.1 to the representation of $\mathrm{SL}_2(\mathbb{Z}/p)$ on the space of holomorphic differential forms on $\overline{X}_p$. The first if well known step is to understand the transition from $\mathrm{SL}_2(\mathbb{Z})$ to $\mathrm{SL}_2(\mathbb{Z}/p)$.

The group $\Gamma(p)$ is normal in $\mathrm{SL}_2(\mathbb{Z})$. The action of $\mathrm{SL}_2(\mathbb{Z})$ passes through its quotient by $\Gamma(p)$.

3.1. Lemma. *The sequence*

$$1 \to \Gamma(p) \longrightarrow \mathrm{SL}_2(\mathbb{Z}) \longrightarrow \mathrm{SL}_2(\mathbb{Z}/p) \longrightarrow 1$$

is exact.

Proof. It must be shown that for each γ in $\mathrm{SL}_2(\mathbb{Z}/p)$ there exists g in $\mathrm{SL}_2(\mathbb{Z})$ with image γ modulo p. The Bruhat decomposition of $\mathrm{SL}_2(\mathbb{Z}/p)$ reduces this to three cases:

$$\gamma = \begin{cases} \begin{bmatrix} 1 & x \\ 0 & 1 \end{bmatrix} \\ \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \quad (ab \equiv 1 \bmod p) \\ \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\,. \end{cases}$$

Only the second is non-trivial. But if $ab = 1$ in $\mathbb{Z}/p$, choose $\alpha \equiv a$ modulo p. Since α is invertible modulo p, it is also invertible modulo p^2, so we may choose β such that $\beta \equiv b$ and $\alpha\beta \equiv 1$ modulo p^2, If $\alpha\beta = 1 + kp^2$ then

$$\begin{bmatrix} \alpha & kp \\ p & \beta \end{bmatrix} \longmapsto \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} . \quad \square$$

The usual proof of this is somewhat more elementary. This one has the advantage that it generalises well to $\mathrm{SL}_2(\mathfrak{o}_K)$ where K is an arbitrary number field. This theorem is the simplest case of the strong approximation theorem, valid for arbitrary simply connected, semi-simple groups over number fields.

3.2. Corollary. *The action of* $\mathrm{SL}_2(\mathbb{Z})$ *on* $\overline{X}_p$ *induces one of* $\mathrm{SL}_2(\mathbb{Z}/p)$.

The next step is to recall how to interpret the points of $\overline{X}_p$ as parameters of certain structures. This is quite different for interior points and cusps.

Interior points On $\mathbb{C}$ define the symplectic form

$$u \wedge v = \mathrm{IM}(u \cdot \overline{v}), \quad (x_1 + iy_1) \wedge (x_2 + iy_2) = x_2y_1 - x_1y_2 .$$

We have

$$cu \wedge cv = |c|^2 (u \wedge v) ,$$

so that if (u, v) is a pair with $u \wedge v > 0$ so is (cu, cv). Hence it makes sense to define the space $\mathfrak{H}$ to be that of all pairs (u, v) of points of $\mathbb{C}$ with $u \wedge v > 0$ modulo scalar multiplication by $c \in \mathbb{C}^\times$. Similarly, if (u, v) is a pair with $u \wedge v = 1$ and $|c| = 1$ then also $cu \wedge cv = 1$, and it makes sense to define $\mathfrak{H}_1$ to be the space of all pairs (u, v) with $u \wedge v = 1$ modulo multiplication by c in the unit group

$$\mathbb{S} = \left\{ c \in \mathbb{C} \,\middle|\, |c| = 1 \right\} .$$

The map $z \mapsto (z, 1)$ induces a bijection of the upper half plane $\mathcal{H}$ with $\mathfrak{H}$, and similarly the inclusion of $\mathfrak{H}_1$ into $\mathfrak{H}$ is a bijection. The group $\mathrm{SL}_2(\mathbb{R})$ acts on the left on $\mathfrak{H}$ and $\mathfrak{H}_1$:

$$g = \begin{bmatrix} a & b \\ c & d \end{bmatrix} : (u, v) \longmapsto (au + bv, cu + dv) .$$

The scalar matrices $\pm I$ act trivially in both cases, since $\pm 1 \in \mathbb{S}$. The identifications of $\mathfrak{H}$ and $\mathfrak{H}_1$ with $\mathcal{H}$ are compatible with the classical action on $\mathcal{H}$ by fractional linear transformations.

Each pair (u, v) with $u \wedge v > 0$ determines a lattice $L_{u,v} = \mathbb{Z}u + \mathbb{Z}v$. If g lies in $\mathrm{SL}_2(\mathbb{Z})$ then (gu, gv) determines the same lattice. If $u \wedge v = 1$ then the volume of $\mathbb{C}/L$ is 1—it is a **unit lattice**.

3.1.1. Lemma. *If* L *is a unit lattice in* $\mathbb{C}$ *there exists a basis* (u, v) *for* L *such that* $u \wedge v = 1$*, which is unique modulo* $\mathrm{SL}_2(\mathbb{Z})$.

Proof. If u, v make up a basis, then the familiar volume formula tells us that $|u \wedge v| = 1$, so if necessary we can swap u and v to obtain $u \wedge v = 1$. The last claim follows from the equation $g(u) \wedge g(v) = \det(g)\,(u \wedge v)$. $\square$

3.1.2. Corollary. *The map* $(u, v) \mapsto L_{u,v}$ *induces a bijection of the quotient* $\mathrm{SL}_2(\mathbb{Z})\backslash\mathfrak{H}_1$ *with the space of unit lattices modulo* $\mathbb{S}$.

I want now to describe X_p similarly. It parametrizes **level structures**. If L is a unit lattice in $\mathbb{C}$ then the symplectic form on $\mathbb{C}$ induces an integral symplectic form on L, hence also a symplectic form on L/pL with values in $\mathbb{Z}/p$. Given (u, v) in $\mathfrak{H}_1$, let $\overline{u}$, $\overline{v}$ be the images of u, v in L/pL.

3.1.3. Proposition. *The map*

$$(u, v) \longmapsto L_{u,v}, \quad (\overline{u}, \overline{v})$$

induces a bijection of $\Gamma(p)\backslash\mathfrak{H}_1$ *with the set of unit lattices* L *together with a symplectic basis of* L/pL *modulo scalar multiplication by* ± 1*, all modulo* $\mathbb{S}$.

Cusps If Γ is an arbitrary arithmetic subgroup of $\mathrm{SL}_2(\mathbb{Q})$, the cusps of $\Gamma\backslash\mathcal{H}$ are in bijection with the Γ-orbits on $\mathbb{P}^1(\mathbb{Q})$. A point of $\mathbb{P}^1(\mathbb{Q})$ is either ∞ or a fraction a/c. The expression a/c for a reduced fraction is unique up to multiplication of numerator and denominator by ± 1, so the map from relatively prime pairs (a, c)—called **primitive points**—in $\mathbb{Z}^2$ to $\mathbb{P}^1(\mathbb{Q})$ is a double covering. (Primitive points of any free module over a ring can be characterized as part of a babis.) For example, if $\Gamma = \mathrm{SL}_2(\mathbb{Z})$ there is just one cusp since this Γ acts transitively on primitive points (a, c). The Euclidean algorithm shows this explicitly—given a, c with a, c relatively prime, we can find b, d with $ad - bc = 1$, and hence

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} : (1, 0) \sim \infty \longmapsto (a, c) \sim a/c\,.$$

Reduction modulo p maps primitive points of $(\mathbb{Z})^2$ to primitive poinst of $(\mathbb{Z}/p)^2$.

3.2.1. Proposition. *Reduction modulo p induces a bijection of the orbits of $\Gamma(p)$ among primitive points (a, c) with primitive points of $(\mathbb{Z}/p)^2$.*

Proof. Let $\Gamma = \Gamma(p)$. I first ask, what is the Γ-orbit of $(1, 0)$? If γ lies in Γ and γ takes $(1, 0)$ to (a, c) then of course $a \equiv 1$, $c \equiv 0$ modulo p. Conversely, suppose $a \equiv 1$, $c \equiv 0$ modulo p with a relatively prime to c. Then a is also prime to c^2, so we can find k, m with $ka - mc^2 = 1$. If

$$\gamma = \begin{bmatrix} a & lc \\ c & k \end{bmatrix}$$

then $\gamma \equiv I$ modulo p and $\gamma(1, 0) = (a, c)$. Thus the orbit of $(1, 0)$ with respect to Γ is the set of all primitive (a, c) with $a \equiv 1$ and $c \equiv 0$ modulo p.

I next ask, what are the other Γ-orbits? If $p_1 = (a_1, c_1)$ and $p_2 = (a_2, c_2)$ are in the same Γ-orbit then clearly $(a_2, c_2) \equiv \pm(a_1, c_1)$ modulo p. Conversely, suppose this condition to hold. Say $(a_1, c_1) = \alpha(1, 0)$. Because of the congruence conditions on p_1 and p_2, the first part of our argument implies that $\alpha^{-1}(p_2)$ will be in the orbit of $(1, 0)$, say $\alpha^{-1}(p_2) = \gamma(\infty)$ with $\gamma \equiv I$ modulo p. Then $\alpha\gamma\alpha^{-1}(p_1) = p_2$. ∎

3.2.2. Corollary. *The map*

$$a/c \longmapsto (a, c)$$

induces a bijection of the cusps of $\Gamma(p)$ with the primitive points of $(\mathbb{Z}/p)^2$ modulo ± 1.

4. Hecke's theorem

Hecke's really interesting observation was that the difference between the representations $\pi_\pm$ has global arithmetic significance. They do not occur with the same multiplicity in the representation of $\mathrm{SL}_2(\mathbb{Z}/p)$ on differential forms on $\overline{X}_p$. Instead, the difference in multiplicities is accounted for by the cusp forms contributed by Grössencharaktere associated (through binary θ-functions) to $\mathbb{Q}(\sqrt{-p})$. These had been constructed in [Hecke:1928]. Hecke proved this by a clever application of Riemann-Roch, but it can be more cleanly proved by applying the Atiyah-Bott trace formula (which was in fact motivated by an earlier result of Eichler about algebraic curves, brought to the attention of Atiyah and Bott by Shimura at the Woods Hole conference of 1965).

Let π be the representation of $G = \mathrm{PSL}_2(\mathbb{Z}/p)$ at hand. Let $h(-p)$ be the ideal class number of the quadratic imaginary quadratic extension $\mathbb{Q}(\sqrt{-p})$. Suppose that as a representation of G

$$\pi = \sum m_{\pi_k}\, \pi_k\,.$$

4.1. Proposition. (Hecke) *The difference $m_{\pi_+} - m_{\pi_-}$ is equal to $h(-p)$.*

The class number will arise through Dirichlet's formula (found for example as Theorem 1 of §5.4 in [Borevich-Shafarevich:1966])

$$h(-p) = -\frac{1}{p}\sum_{1}^{p-1} m\left(\frac{m}{p}\right) .$$

Recall that

$$\nu = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} .$$

According to Proposition 2.1

$$m_+ - m_- = \frac{\operatorname{trace}\pi(\nu) - \operatorname{trace}\overline{\pi}(\nu)}{\overline{\mathfrak{G}} - \mathfrak{G}} .$$

So we must calculate

$$\tau = \operatorname{trace}\pi(\nu) - \operatorname{trace}\overline{\pi}(\nu) ,$$

for which I'll use the formula of Eichler-Atiyah-Bott. More precisely, I use the formulation on p. 12 of [Atiyah-Bott:1964]: *If X is a compact Riemann surface and $f\colon X \to X$ a holomorphic map, let α be the trace of f^* on $H^0(X,\Omega)$. If the fixed points of f are isolated, then*

$$1 - \overline{\alpha} = \sum_{f(x)=x} \frac{1}{1 - f'(x)} ,$$

where $f'(x)$ is the map of the holomorphic tangent space induced by f at x. In our situation the group has been made to act by the inverse of f^* and f is of finite order, so $\overline{\alpha}$ itself will be the trace we are looking at. In other words

$$\tau - \overline{\tau} = -\sum_{\nu(x)=x} \left(\frac{1}{1-\zeta_x} - \frac{1}{1-\overline{\zeta}_x}\right) ,$$

where ν acts as multiplication by ζ_x around the fixed point x.

Thus we must classify the points of $\overline{X}_p$ fixed by ν. *What points are fixed by ν? How does ν act on the tangent space of those points?*

4.2. Proposition. *The fixed points of ν acting on $\overline{X}_p$ are the cusps associated to the points $(x,0)$ for x in $(\mathbb{Z}/p)^\times$.*

Since $(\pm x, 0)$ give rise to the same cusp, there are $(p-1)/2$ of them.

Proof. It is fairly straightforward to see that if γ in $\mathrm{SL}_2(\mathbb{Z}/p)$ fixes an interior point of $\overline{X}_p$ then the point is the transform of i or of a sixth root of unity, and γ has order divisible by 2 or 3. Since ν has order $p > 3$ this possibility is excluded. Any fixed point must therefore be a cusp. There is one simple possibility, the cusp at infinity, where ν acts locally by multiplication by $\zeta = e^{2\pi i/p}$. Which of the other cusps are fixed by ν?

Any cusp fixed by ν has to correspond to a non-zero point of $(\mathbb{Z}/p)^2$ fixed by the image of ν modulo p, so must be of the form a/c with $c \equiv 0$ modulo p. Conversely, consider $(x,0)$ for x in $\mathbb{Z}$ with $x \not\equiv 0$ modulo p. Then (x,p) is a primitive point of $(\mathbb{Z})^2$ with image $(x,0)$ modulo p.

I am going to show that ν fixes the image of (x,p) in $\overline{X}_p$. For this, it suffices to find γ in $\mathrm{SL}_2(\mathbb{Z})$ such that $\gamma(x,p) = (x,p)$, $\gamma \equiv \nu$ modulo p.

Choose y, k such that $xy - kp^2 = 1$, so now

$$\alpha = \begin{bmatrix} x & kp \\ p & y \end{bmatrix}$$

takes $(1,0)$ to (x,p). Let

$$\mu = \begin{bmatrix} 1 & y^2 \\ 0 & 1 \end{bmatrix}.$$

Since

$$\alpha \equiv \begin{bmatrix} x & 0 \\ 0 & 1/x \end{bmatrix}$$

modulo p, $y \equiv 1/x$ modulo p, and

$$\begin{bmatrix} x & 0 \\ 0 & 1/x \end{bmatrix} \begin{bmatrix} 1 & y^2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/x & 0 \\ 0 & x \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$$

I set $\gamma = \alpha\mu\alpha^{-1}$.

4.3. Proposition. *The action of ν on the tangent space associated to $(x,0)$ is multiplication by $e^{2\pi i y^2/p}$, where $xy \equiv 1$ modulo p.*

Proof. Follows immediately from the definition of γ in the proof of the preceding Proposition.

The fixed point formula tells us what $1 - \operatorname{trace} \pi(\nu)$ is, and then also $1 - \operatorname{trace} \overline{\pi}(\nu)$. Taking the difference, we get

$$\operatorname{trace} \pi(\nu) - \operatorname{trace} \overline{\pi}(\nu) = \sum_{m \bmod p} \left(\frac{m}{p}\right) \frac{1}{\zeta^m - 1}$$

where $\zeta = e^{2\pi i/p}$.

We want to rewrite this expression. Start with the algebraic formula:

$$z^p - 1 = (z-1)(z^{p-1} + z^{p-2} + \cdots + z + 1)$$
$$\frac{1}{z-1} = \frac{z^{p-1} + z^{p-2} + \cdots + z + 1}{z^p - 1}$$

If we set z equal to a p-th root of unity, the right hand side must be evaluated by l'Hôpital's rule:

$$\frac{1}{z-1} = \frac{\sum_1^{p-1} k z^{k-1}}{p z^{p-1}} = \frac{1}{p} \sum k z^k .$$

The trace on $\pi - \overline{\pi}$ is therefore

$$\begin{aligned} \frac{1}{p} \sum_{m=1}^{p-1} \left(\frac{m}{p}\right) \left(\sum_{k=1}^{p-1} k \zeta^{km} \right) &= \frac{1}{p} \sum_k k \left(\sum \left(\frac{m}{p}\right) \zeta^{km} \right) \\ &= \frac{1}{p} \sum_k k \left(\frac{k}{p}\right) \left(\sum \left(\frac{n}{p}\right) \zeta^n \right) \\ &= \frac{\mathfrak{G} - \overline{\mathfrak{G}}}{p} \sum_k k \left(\frac{k}{p}\right) \end{aligned}$$

To get the difference $m_+ - m_-$, we divide by $\overline{\mathfrak{G}} - \mathfrak{G}$ to give us $h(-p)$.

Which of the two representations $\pi_\pm$ occurs more often than the other is implicitly related, as Hecke himself observed, to the sign of Gauss sums. This sort of problem—i.e. relations with class field theory—is ubiquitous in this business. Hecke didn't tie all these facts together systematically, but this was done by Labesse and Langlands around 1971, who generalized Hecke's observation to all representations on all quotients $\mathrm{SL}_2(F)\backslash SL_2(\mathbb{A}_F)$.

5. References

1. M. Atiyah and R. Bott, **Notes on the Lefschetz fixed point theorem for elliptic complexes**, notes from the Harvard University Mathematics Department, 1964.

2. Z. I. Borevich and I. R. Shafarevich, **Number theory**, Academic Press, 1966,

3. E. Hecke, 'Bestimmungen der Perioden gewisser Integrale durch die Theorie der Klassenkörper', *Mathematisches Zeitschrift* **28** (1928), 708–727.

4. ——, Uber das verhalten der Integrale 1. Gattung bei Abbildungen, insbesondere in der Theorie der elliptischen Modulfunktionen', *Abhandlungen aus einem Seminar der Hamburgische Universität* **8** (1930), 271–281.

5. J.-P. Labesse and R. P. Langlands, 'L-indistinguishability for $\mathrm{SL}(2)$', *Can. Jour. of Math.* **XXXI** (1979), 726–785.