Cusp forms and the Eichler-Shimura relation

September 9, 2013

In the last lecture we observed that the family of modular curves $X_0(N)$ has a model over the rationals. In this lecture we use this fact to attach Galois representations to cusp forms of weight 2. The goal is to prove the following theorem over the next few lectures.

Theorem 0.0.1. *Let f be a normalized cuspidal eigenform of weight 2 for $\Gamma_0(N)$, with Fourier expansion $f = \sum_{n\geq 1} a_n q^n$. The a_n lie in a number field K. Let λ be a prime of K, and let K_λ be the completion of K at λ. There exists an irreducible representation*

$$\rho\colon \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}_2(K_\lambda)$$

having the following property. For every prime p prime to λ and N, ρ is unramified at p, so that $\rho(\mathrm{Frob}_p)$ is well-defined, and the characteristic polynomial of $\rho(\mathrm{Frob}_p)$ is $X^2 - a_p X + p$.

We only need the following facts about modular curves:

1. For each p prime to N, there is a Hecke correspondence T_p on $X_0(N)$,
2. The T_p induce linear operators on $H^0(X_0(N), \Omega^1_{X_0(N)/\mathbb{Q}})$ which can be simultaneously diagonalized,
3. We have the relationship

$$T_p = \mathrm{Frob}_p + \mathrm{Frob}_p^\vee$$

in the endomorphism algebra of $\mathrm{Jac}\, X_0(N)_{\mathbb{F}_p}$.

1 Modular forms and cusp forms

The classical theory of modular forms is due to Hecke. It has to do with the action of discrete subgroups of $\mathrm{SL}_2(\mathbb{R})$ (known as Fuchsian groups) on the upper half plane $\mathcal{H}$. We start with the *principal congruence subgroup*

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \;\middle|\; a, d \equiv 1 \pmod{N},\ b, c \equiv 0 \pmod{N} \right\}$$

and say that a subgroup $\Gamma \subset \mathrm{SL}_2(\mathbb{Z})$ is a *congruence subgroup* if it contains $\Gamma(N)$ for some N. Two important examples are $\Gamma_0(N)$ (resp., $\Gamma_1(N)$), which are those subgroups of $\mathrm{SL}_2(\mathbb{Z})$ where $c \equiv 0 \pmod{N}$ (resp., $c \equiv 0 \pmod{N}$ and $a \equiv b \equiv 1 \pmod{N}$).

Let Γ be a congruence subgroup, and let k be an integer. A *modular form* of weight k for Γ is a complex-valued function f on the upper half-plane $\mathcal{H}$ with the following properties:

1. For all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$,
$$f(\gamma z) = (cz + d)^k f(z).$$
2. f is holomorphic on $\mathcal{H}$.
3. f is holomorphic at the cusps of $\Gamma \backslash \mathcal{H}$.

For any $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{R})$ with positive determinant, we write

$$f|_{g,k}(z) = (cz + d)^{-k} (\det g)^{k/2} f(gz);$$

this actually defines an action of $\mathrm{GL}_2^+(\mathbb{R})$ on functions on $\mathcal{H}$. (The center acts trivially, so that in fact we have an action of $\mathrm{PGL}_2^+(\mathbb{R})$.) The first condition can be restated as $f|_{\gamma,k} = f$ for all $\gamma \in \Gamma$.

The last point requires some explanation. A *cusp* of Γ is an equivalence class in an element of $\mathbb{R} \cup \{\infty\}$ which is fixed by a parabolic subgroup of Γ (a subgroup which has exactly one fixed point on the Riemann sphere). For instance, ∞ is always fixed by the intersection of $\begin{pmatrix} 1 & \mathbb{Z} \\ & 1 \end{pmatrix}$ with Γ, so ∞ is always a cusp of a congruence subgroup. We now define what it means for

f to be holomorphic at ∞. The subgroup $\Gamma_\infty \subset \Gamma$ which fixes ∞ is of the form $\begin{pmatrix} 1 & \mathbb{Z}w \\ & 1 \end{pmatrix}$ for some $w \geq 1$, known as the width of the cusp at ∞. We have $f(z+w) = f(z)$ for $z \in \mathcal{H}$, so that it makes sense to define $\hat{f}(q)$ on the domain $0 < |q| < 1$ by

$$\hat{f}(e^{2\pi i z/w}) = f(z)$$

Then the condition of holomorphicity at ∞ is the condition that $\hat{f}$ extend to a holomorphic function on $|q| < 1$. For general cusps c, one can always find a $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ which translates ∞ onto c. Then the condition that f be holomorphic at c is understood to be the same as the condition that $f|_{\gamma,k}$ (which is attempting to be a modular form for $\gamma^{-1}\Gamma\gamma$) be holomorphic at ∞. One must check that this definition does not depend on the choice of γ.

Modular forms are sections of certain line bundles on $X(\Gamma)$. Last time we established that $X(\Gamma)$ is an algebraic curve, which suggests that modular forms can be defined in some purely algebraic manner. Indeed this is the case. For instance, if $f(z)$ is a cusp form of weight 2 for Γ, then the 1-form $f(z)dz$ on $\mathcal{H}$ is easily checked to be invariant under Γ, so that $f(z)dz$ descends to a holomorphic 1-form on $\Gamma\backslash\mathcal{H}$. Further, the condition that $f(z)$ is cuspidal means that $f(z)dz$ will be holomorphic at the cusps (exercise). Thus $f(z)dz$ is holomorphic on the compact Riemann surface $X(\Gamma)(\mathbb{C})$. By GAGA, this means that there exists a corresponding differential form ω in $H^0(X(\Gamma), \Omega^1_{X(\Gamma)/\mathbb{C}})$. In fact there is an isomorphism:

$$S_2(\Gamma) \cong H^0(X(\Gamma), \Omega^1_{X(\Gamma)/\mathbb{C}}).$$

If it so happens that $X(\Gamma)$ has a $\mathbb{Q}$-rational model, such as when $\Gamma = \Gamma_0(N)$, then the complex vector space $S_2(\Gamma)$ has a $\mathbb{Q}$-rational model as well, namely $H^0(X(\Gamma), \Omega^1_{X(\Gamma)/\mathbb{Q}})$. The same is true for the spaces $M_k(\Gamma)$ and $S_k(\Gamma)$ for higher weights k.

We very often consider the Taylor series expansion of $\hat{f}(q)$ around $q = 0$, which converges for all $z \in \mathcal{H}$:

$$f(z) = \sum_{n \geq 0} a_n e^{2\pi i n z/w}.$$

Note that in the cases of $\Gamma_0(N)$ and $\Gamma_1(N)$, the width of ∞ is $w = 1$.

Write $M_k(\Gamma)$ for the complex vector space of modular forms of weight k on Γ. Write $S_k(\Gamma)$ for the space of modular forms which are *cusp forms*,

which means they are 0 at every cusp of Γ. It turns out that $M_k(\Gamma)$ is a finite-dimensional vector space, so that $S_k(\Gamma)$ is a finite-dimensional Hilbert space.

There is a bilinear pairing $S_k(\Gamma) \times S_k(\Gamma) \to \mathbb{C}$ known as the Petersson inner product:

$$(f, g)_{k,\Gamma} = \int_{\mathcal{F}} f(z)\overline{g(z)}y^k \frac{dxdy}{y^2},$$

where $\mathcal{F}$ is a fundamental domain for $\Gamma\backslash\mathcal{H}$. (Check that the integrand really is Γ-invariant!)

It is generally far easier to produce examples of elements of $M_k(\Gamma)$ than it is to produce cusp forms in $S_k(\Gamma)$. For k even, we have the *Eisenstein series*

$$E_k(z) = \frac{1}{2\zeta(k)} \sum_{(c,d)} \frac{1}{(cz+d)^k},$$

where the sum is over all pairs of integers $(c, d) \neq (0, 0)$. E_k is very easily seen to ba modular form for $\mathrm{SL}_2(\mathbb{Z})$ of weight k. A little manipulation shows that its Fourier expansion is

$$E_k(z) = 1 + \frac{2}{\zeta(1-k)} \sum_{n \geq 1} \sigma_{k-1}(n) q^n,\ q = e^{2\pi i z}.$$

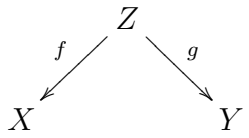
In can be shown that the graded ring $\oplus_{k \text{ even}} M_k(\mathrm{SL}_2(\mathbb{Z}))$ is generated over $\mathbb{C}$ by the elements E_4 and E_6. As a result $M_{12}(\mathrm{SL}_2(\mathbb{Z}))$ is spanned by E_4^3 and E_6^2, and

$$\Delta = \frac{1}{1728}(E_4^3 - E_6^2) \in S_{12}(\mathrm{SL}_2(\mathbb{Z}))$$

spans the (one-dimensional) space of cusp forms for $\mathrm{SL}_2(\mathbb{Z})$.

2 Hecke correspondences

If X and Y are two projective curves, then a *correspondence* between X and Y is a diagram



where Z is a third projective curve and the arrows represent finite morphisms. (There is a more general definition for varieties X and Y of higher dimension.) If $h\colon X \to Y$ is a finite morphism, then h gives a correspondence as above by setting $Z = X$, but not all correspondences will arise this way. But a correspondence can act like a morphism, in that it induces maps between various vector spaces attached to X and Y. For example, if ω is a differential form on Y, we may obtain a differential form on X by first pulling back ω through g and then pushing forward through f. Or if D is a divisor on X, we can get a divisor on Y by taking the preimage of D in Z (this will increase the degree by a factor of $\deg(f)$) and then pushing this into Y.

Of particular use is a correspondence T_p between $X_0(N)$ and $X_0(N)$, where p is a prime not dividing N. The role of Z will be played by $X_0(Np)$. This curve (or at least an affine part of it) classifies triples (E, C_N, C_p), where E is an elliptic curve, C_N is a cyclic subgroup of E of order N, and C_p is a cyclic subgroup of order p. There is an obvious map $X_0(Np) \to X_0(N)$ defined by forgetting C_p, but there is also another one where (E, C_N, C_p) gets sent to $(E/C_p, (C_N + C_p)/C_p)$. These two maps give the correspondence T_p.

The Hecke correspondence T_p induces an operator on $M_k(\Gamma_0(N))$. On the level of functions on the upper half plane, this works out to

$$T_p f(z) = p^{k-1} f(pz) + \frac{1}{p} \sum_{a=0}^{p-1} f\left(\frac{z+a}{p}\right)$$

On the level of q-expansions, this is

$$T_p f(z) = p^{k-1} \sum_{n \geq 1} a_n q^{pn} + \sum_{n \geq 1} a_{pn} q^n \tag{1}$$

The operators T_n for n composite can be defined the same way, although the formulas will be more complicated. One has the rules

$$\begin{aligned} T_m T_n &= T_{mn}, \ \gcd(m,n) = 1 \\ T_p T_{p^n} &= T_{p^{n+1}} + p^{k-1} T_{p^{n-1}} \end{aligned}$$

In particular all the T_m commute with one another. One also checks that

$$(T_n f, g)_{k,\Gamma} = (f, T_n g)_{k,\Gamma}$$

so that the T_n are a collection of commuting self-adjoint operators on a finite-dimensional Hilbert space $S_k(\Gamma_0(N))$. Therefore they can be simultaneously diagonalized.

An important observation is that $S_k(\Gamma_0(N))$ has a rational model which is preserved by the Hecke operators. This means that the eigenvalues of the T_n are algebraic numbers (in fact they are algebraic integers).