Attaching Galois representations to weight 1 modular forms

1 Review of Galois representations for higher weight

Some references I used were:
- these notes by Tom Lovering,
- these lecture notes by Johannes Anschütz,
- these notes by Bas Edixhoven,
- the original paper by Deligne and Serre, [1].

Most of what follows comes from one of these.

Last week, Arun explained how we can attach Galois representations to weight 2 modular forms by studying various cohomology theories for the modular curve $X = X_1(N)$. The key ingredients were:
- The Eichler–Shimura isomorphism,
$$H^0(X, \Omega^1_X) \oplus \overline{H^0(X, \Omega^1_X)} \cong H^1(X, \mathbb{C}), \tag{1}$$
where the left hand side can be interpreted as weight 2 modular forms via $f \rightsquigarrow f(z)dz$ and the right hand side is singular cohomology.
- By fixing an isomorphism $\mathbb{C} \cong \overline{\mathbb{Q}}_\ell$, we can compare singular cohomology with étale cohomology,
$$H^1(X, \mathbb{C}) \cong H^1\left(X, \overline{\mathbb{Q}}_\ell\right) \cong H^1_{\text{ét}}\left(X, \overline{\mathbb{Q}}_\ell\right), \tag{2}$$
where the right hand side has a natural Galois action.
- The Eichler–Shimura *relation*, which gives a relationship between the p-th Hecke operator and the Frobenius at p. To make this precise, we compare the étale cohomology of X with that of its special fibre $X_{\overline{\mathbb{F}}_p}$, where we found that the action of the "geometric" Frobenius was compatible with the action of the Frobenius element at p, $\sigma_p \in G_{\mathbb{Q}} := \mathrm{Gal}\left(\overline{\mathbb{Q}}/\mathbb{Q}\right)$. This eventually tells us that the characteristic polynomial of $\rho_f\left(\sigma_p\right)$ is $T^2 - a_p(f)T + \varepsilon(p)p$, where ε is the Nebentypus character of f, and this implies that $L\left(s, \rho_f\right) = L(s, f)$.

This was proved by Eichler and Shimura, but not phrased in the language of Galois representations or étale cohomology, which came later. However, this was indispensible for Deligne's contruction in weight ≥ 2. [2]

1.1 Geometric modular forms

In this section, we describe modular forms of level $\Gamma_1(N)$ in terms of the geometry of $Y = Y_1(N)$. Let S be a scheme over $\operatorname{Spec} \mathbb{Z}[1/N]$. Since $Y / \operatorname{Spec} \mathbb{Z}[1/N]$ is constructed as a moduli space of elliptic curves with extra data, the general yoga of representable functors gives a *universal elliptic curve* $\pi : \mathcal{E} \to Y$. We define $\omega := \pi_* \Omega^1_{\mathcal{E}/Y}$.

This might seem a bit mysterious, but in practice, it's quite simple: $\mathcal{E}$ is given by an equation
$$\mathcal{E} : y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2, \tag{3}$$
where a_1, a_2, a_3 are holomorphic modular forms of level N with coefficients in $\mathbb{Z}[1/N]$, and these can be computed quite explicitly. The sheaf ω is also relatively easy to understand: at a point

$x = (E, P) \in Y$, the stalk is given by $\omega_x = \Omega^1_E$. We can also extend this to X via the natural inclusion $j : Y \hookrightarrow X$.

We then define a weight k modular form to be a global section in $\Gamma(Y, \omega^k)$. What is the algebraic definition of a cusp form? They are precisely the ones which extend from Y to X, so precisely the elements of $\Gamma(Y, \Omega^1_Y \otimes \omega^{k-2})$. Note that over Y, $\omega^2 \cong \Omega^1_Y$, by comparing sections.

We also see these definitions work equally well when $S = \operatorname{Spec} \mathbb{F}_p$! We take this as our definition of a *mod p modular form*. Then we have the following theorem:

Theorem 1.1: Let $k \geq 2$, and write $M_k(\Gamma_1(N), R) := \Gamma(Y_R, \omega^k)$ for any $\mathbb{Z}[1/N]$-algebra $\mathbb{R}$. For any $p \nmid N$,

$$M_k(\Gamma_1(N), \mathbb{F}_p) \cong M_k(\Gamma_1(N), \mathbb{Z}[1/N]) \otimes \mathbb{F}_p. \quad (4)$$

In other words, every weight ≥ 2 mod p modular form lifts to characterstic 0. This is no longer true when $k = 1$!

1.2 Eichler–Shimura in weight > 2

To formulate a version of the Eichler–Shimura isomorphism, the main difficulty is to sort out the right hand side: what local system to put on X to generalise $H^1(X, \mathbb{C})$ for $k = 2$. Apparently, the Hodge–de Rham spectral sequence will tell you that $R^1\pi_*\mathbb{Z}$, whose stalk at $x = (E, P)$ is $H^1(E, \mathbb{Z})$, is the right choice.

Theorem 1.2 (The Eichler–Shimura isomorphism for $k > 2$): There is a Hecke-equivariant isomorphism

$$H^0(Y, \Omega^1_Y \otimes \omega^{k-2}) \oplus \overline{H^0(Y, \Omega^1_Y \otimes \omega^{k-2})} \cong H^1_p(X, \operatorname{Sym}^{k-2} R^1\pi_*\mathbb{Z}), \quad (5)$$

where $H^1_p = \operatorname{Im}(H^1_c \to H^1)$.

A proof of this can be found in Brian Conrad's notes here.

The H^1_p might seem ad-hoc, but[1] it's not obvious which cohomology theory to use on the compactified modular curve, and "parabolic cohomology" H_p turns out to be a special case of the "correct one", which is intersection cohomology. It is probably a good exercise to check that this specialises to Equation 1 when $k = 2$.

The next step is then to compare this with étale cohomology of X with coefficients in the étale local system $R^1\pi_*\overline{\mathbb{Q}}_\ell$, and to prove the Eichler–Shimura relations in this setting, which implies that the characteristic polynomial of $\rho_f(\sigma_p)$ is $T^2 - a_p(f)T + \varepsilon(p)p$, for a weight k modular form f.

2 Weight 1 modular forms

For weight 1 eigenforms, there are several reasons why the above strategy doesn't work:

1. The local system $\operatorname{Sym}^{k-2} R^1\pi_*\mathbb{Z}$ doesn't make sense because $k - 2 < 0$, so it's not clear in what ℓ-adic cohomology group one should look;

[1] According to Will Sawin here

2. The Hecke eigenvalue system associated to a form f can be found in the coherent cohomology groups $H^0(X, \omega)$ as well as $H^1(X, \omega)$. In particular, is related to the reason why there are no simple formula for $\dim S_1(\Gamma_1(N))$, unlike for higher weights.
3. A posteriori, the Galois representations attached to weight 1 modular forms are qualitatively different because their image is finite.

2.1 An example

Let $N = 145 = 29 \cdot 5$, and set $F := \mathbb{Q}\left(\sqrt{29}\right)$. The ideal (5) splits in F, and we pick a prime $\mathfrak{p}$ above 5. The ray class group of F with modulus $\mathfrak{p}$, $\mathrm{Cl}_{\mathfrak{p}}$, is isomorphic to C_4, and hence we can pick a quartic character $\chi : \mathrm{Cl}_{\mathfrak{p}} \to \mathbb{C}^\times$. By identifying $\mathrm{Cl}_{\mathfrak{p}}$ with the Galois group of the ray class field $H_{\mathfrak{p}}$ over F, we then obtain a 2-dimensional Galois representation $\rho := \mathrm{Ind}_{G_{\mathbb{Q}}}^{G_F} \chi : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{C})$.

On the other hand, we can define a theta series θ_χ by the q-expansion

$$\theta_\chi(z) := \sum_{[\mathfrak{a}] \in \mathrm{Cl}_{\mathfrak{p}}} \chi(\mathfrak{a}) \sum_{\alpha \in \mathfrak{a}} q^{\mathrm{Nm}(x) / \mathrm{Nm}(\mathfrak{a})} \quad \text{where } q := e^{2\pi i z}. \tag{6}$$

Using the Poisson summation formula, one can show that this is a modular form of weight 1 and level 145, and a Hecke eigenform at that. This satisfies $L(s, f) = L(s, \rho)$.

This construction – inducing a character of a quadratic field and matching up with a theta series – gives a large number of weight 1 modular forms and their Galois representations, but not all. We call these forms dihedral, and any non-dihedral form is said to be "exotic". The first exotic form in the `lmfdb` is 124.1.i.a.

2.2 The Deligne–Serre theorem

Theorem 2.1 (Deligne–Serre [1]): Let $N \geq 1$, $\varepsilon : (\mathbb{Z}/N)^\times \to \mathbb{C}^\times$ be a Dirichlet character such that $\varepsilon(-1) = -1$, and fix a normalised eigenform $f \in S_1(\Gamma_0(N), \varepsilon)$. Then there exists an irreducible representation $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{C})$ unramified away from N such that

$$\det \rho\left(\sigma_p\right) = \varepsilon(p) \quad \text{and} \quad \mathrm{tr}\, \rho\left(\sigma_p\right) = a_p(f), \tag{7}$$

for $p \nmid N$.

In particular, we have $L(s, f) = L(s, \rho)$. This clearly generalises the example in the previous section. From this we deduce:

Corollary 2.2 (Ramanujan conjecture): For any p, $a_p(f)$ is a sum of two roots of unity; in particular, $|a_p(f)| \leq 2$.

This is an example of "Galois information $\Rightarrow$ Automorphic information". In the reverse direction, the Khare–Wintenberger theorem implies that the map $f \rightsquigarrow \rho$ is surjective onto the set of *odd* Artin representations, implying:

Corollary 2.3 (Artin holomorphy conjecture): For any odd irreducible Artin representation $\rho : G_{\mathbb{Q}} \to \mathrm{GL}_2(\mathbb{C})$, the L-function $L(s, \rho)$ is everywhere holomorphic.

2.3 Outline of proof

The proof is completely different from the higher weight counterpart, and rather clever.

1. Let $\ell \nmid N$ be a rational prime. Reduce f modulo ℓ.
2. By multiplying f with a suitable Eisenstein series, obtain a congruence with a mod ℓ eigenform of higher weight.
3. This is an eigenform, but its obvious lift to characteristic 0 is not. So we apply the *Deligne–Serre lifting lemma* to lift instead the associated Hecke eigenvalue system (as opposed to the eigenvector) to characteristic 0.
4. This system corresponds to a new eigenform of weight ≥ 2, to which we can attach a λ-adic representation using Deligne's generalisation of Eichler–Shimura. Here λ is some prime above ℓ in the coefficient field of f.
5. This gives one λ-adic representation for each ℓ; the next step is to show that they are all compatible, which we do by reducing mod λ. Since we know the characteristic polynomial of $\overline{\rho}_\lambda(\sigma_p)$ for all p and these generate $G_\mathbb{Q}$, we can prove that their images are uniformly bounded as we vary ℓ. Using this, we can lift to characteristic 0: the crucial point is that for a finite group G with $\ell \nmid |G|$, there is a bijection between mod-ℓ representations of G and characteristic zero representations.
6. Finally, to show that the resulting representation is irreducible, one uses an analytic estimate: we know $L\left(s, f \otimes \overline{f}\right)$ has a pole at $s = 1$ by a result of Rankin, and if $\rho_f = \chi_1 \oplus \chi_2$, we obtain a contradiction by comparing growths.

2.4 Further details

Let K be the smallest extension of $\mathbb{Q}$ containing $a_n(f)$ for all n as well as the values of ε, and for each rational prime ℓ pick a prime λ of K above ℓ. By k_λ we mean the residue field of K at λ.

The Eisenstein series mentioned in step 2 is

$$E_k(z) = 1 - \frac{4}{B_k}\sum_{n=1}^{\infty} \sigma_{k-1}(n)q^n \quad \text{where} \quad \sigma_{k-1}(n) = \sum_{d|n} d^{k-1}. \tag{8}$$

The Clausen–von Staudt congruences imply that $E_{\ell-1} \equiv 1 \bmod \ell$. Thus, as an element in $M_{p-1}(\Gamma_1(N), \mathbb{F}_\ell)$, $E_{\ell-1}$ has q-expansion identically equal to 1. Note however that $E_{\ell-1}$ is not the constant function, which has weight 0. Then the form $f \cdot E_{\ell-1} \bmod \lambda$ lives in $M_\ell(\Gamma_1(N), k_\lambda)$ and has q-expansion equal to that of $f \bmod \lambda$. One can check that it is an eigenform using explicit formulas for weight ℓ Hecke operators, although $f \cdot E_{\ell-1}$ in characteristic zero is not.

The next step is to show that we can lift the Hecke eigenvalues of $f \cdot E_{\ell-1}$ to characteristic zero.

Theorem 2.4 (Deligne–Serre lifting lemma): Let $k \geq 2$, and let $\overline{g} \in M_k(\Gamma_1(N), k_\lambda)$ be an eigenform which is the reduction mod λ of a modular form $g \in M_k\left(\Gamma_1(N), \mathcal{O}_{K,(\lambda)}\right)$. Then there exists a finite field extension K'/K and an eigenform $g' \in M_k(\Gamma_1(N), K')$ such that $g' \equiv \overline{g} \bmod \lambda'$ for some prime λ' above λ.

Let $f' \in M(\Gamma_1(N), K')$ denote the lift of $f \cdot E_{p-1}$ to characteristic 0, and let $\rho_{\lambda'} : G_\mathbb{Q} \to \mathrm{GL}_2(K'_{\lambda'})$ be the associated λ'-adic representation. Up to conjugating $\rho_{\lambda'}$, we can assume the image lands in $\mathrm{GL}_2\left(\mathcal{O}_{K'_{\lambda'}}\right)$. Now it makes sense to reduce $\rho_{\lambda'}$ mod λ', giving $\overline{\rho}_{\lambda'}$ valued in $\mathrm{GL}_2(\mathcal{O}_{K'}/\lambda')$. In fact, since we know that the characteristic polynomial of $\overline{\rho}_{\lambda'}(\sigma_p)$ is $T^2 - a_p(f)T + \varepsilon(p) \in k_\lambda[T]$ for any $p \nmid N\ell$, the Chebotarov density theorem implies that $\overline{\rho}_{\lambda'}$ actually takes values in k_λ, at least after replacing with the semisimplification.

So now we have a collection of k_λ-valued representations $\overline{\rho}_{\lambda'}$, indexed by primes $\ell \geq 5$. We restrict to ℓ such that λ is totally split in K, which implies that $\mathbb{F}_\lambda = \mathbb{F}_\ell$. The next step is to show that since we know the characteristic polynomial of the Frobenii, we can bound the size of $G_\ell := \mathrm{Im}(\overline{\rho}_{\lambda'}) \subset \mathrm{GL}_2(\mathbb{F}_\ell)$ uniformly as we vary ℓ. This allows us to lift $\overline{\rho}_\lambda$ to characteristic 0 – the crux of the argument here is that when $\ell \nmid |G|$, then reduction mod λ gives a bijection of (isomorphism classes of) finite group representations.

So, suppose $|G_\ell| \leq A$, and fix $\ell, \ell' > A$. We obtain two characteristic zero lifts ρ_ℓ and $\rho_{\ell'}$ with

$$\det(T - \rho_\ell(\sigma_p)) = T^2 - a_p(f)T + \varepsilon(p) = \det(T - \rho_{\ell'}(\sigma_p)), \tag{9}$$

for all $p \neq \ell$, and so we conclude that $\rho_\ell \cong \rho_{\ell'}$ and that this representation is unramified away from N. In fact, we should be a bit careful here and argue that we have a representation ρ such that $\det(T - \rho(\sigma_p))$ is congruent to $T^2 - a_p(f)T + \varepsilon(p)$ modulo infinitely many primes, and so they have to be equal.

Finally, to show that the representation is irreducible, we argue by contradiction. A result of Rankin implies that $L(s, f \otimes \overline{f}) = \sum_{n=1}^{\infty} |a_n|^2 \cdot n^{-s}$ has a simple pole at $s = 1$, and from this one obtains

$$\sum_{p \nmid N} \frac{|a_p|^2}{p^s} \leq \log\left(\frac{1}{s-1}\right) + O(1) \text{ as } s \downarrow 1. \tag{10}$$

On the other hand, if $\rho \cong \chi_1 \oplus \chi_2$, then $a_p = \chi_1(p) + \chi_2(p)$ and

$$\begin{aligned}\sum_{p \nmid N} \frac{|a_p|^2}{p^s} &= 2\sum_{p \nmid N} \frac{1}{p^s} + \sum_{p \nmid N} \frac{\chi_1(p)\overline{\chi_2(p)}}{p^s} + \sum_{p \nmid N} \frac{\chi_2(p)\overline{\chi_1(p)}}{p^s} \\ &= 2\log\left(\frac{1}{s-1}\right) + O(1),\end{aligned} \tag{11}$$

as $\chi_1 \neq \chi_2$ since $\chi_1\chi_2(-1) = -1$. This gives a contradiction, and irreducibility follows.

Bibliography

[1] P. Deligne and J.-P. Serre, "Formes modulaires de poids 1", *Annales scientifiques de l'École Normale Supérieure*, vol. 7, no. 4, pp. 507–530, 1974, doi: 10.24033/asens.1277.

[2] P. Deligne, "Formes modulaires et représentations l-adiques", *Séminaire Bourbaki*, vol. 11, pp. 139–172, 1969, [Online]. Available: http://eudml.org/doc/109756