The Shimura construction and the modularity theorem

Gregor Bruns

14.05.2014

1 Introduction

We know already the modularity condition that the coefficients a_p, p a prime, in the Fourier expansion of a modular form $f = \sum_{n\geq 0} a_n q^n$ are related to the number of points on the elliptic curve modulo p. Set

$$b_p = (p+1) - \#E(\mathbb{F}_p)$$

Then modularity is exactly $a_p = b_p$ for all primes p.

We will give a geometric definition of modularity and formulate the modularity theorem in these terms.

2 Leftover things

2.1 Motivation for the Tate module

Let A, B be abelian varieties and ℓ a prime. The Tate module $T_\ell(A)$ is defined to be

$$T_\ell(A) = \varprojlim_n A[\ell^n]$$

with the maps being $A[\ell^{n+1}] \to A[\ell^n]$ given by multiplication by ℓ. $T_\ell(A)$ has a natural $\mathbb{Z}_\ell$-module structure. In our case, since the Hecke algebra $\mathbb{T}$ acts on each $E[\ell^n]$, there is an induced action of $\mathbb{T} \otimes \mathbb{Z}_\ell$ on $T_\ell(E)$.

For $\ell \neq \mathrm{char}(k)$ there is an injection

$$\mathrm{Hom}(A, B) \to \mathrm{Hom}_{\mathbb{Z}_\ell}(T_\ell(A), T_\ell(B))$$

which is Galois-equivariant.

The Tate module gives a lattice in $T_\ell(A) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$ (which has dimension $2g$ as a $\mathbb{Q}_\ell$-vector space).

We have two actions on the Tate module $T_\ell(J_\Gamma)$, one by $G_\mathbb{Q}$ and one by $\mathbb{T}_{\mathbb{Z}_\ell}$. A similar thing will be possible for cohomology groups. These actions commute, hence we can decompose them into simultaneous eigenspaces for the Hecke operators and the Galois actions. It turns out we get a decomposition in a direct sum of two-dimensional vector spaces spanned by eigenforms. Each of them has a $G_\mathbb{Q}$-action, so is a two-dimensional Galois representation. Therefore, choosing an eigenform selects one of those spaces and therefore gives an associated Galois representation (more on that in later talks).

Theorem 2.1 (Tate's isogeny theorem)**.** *Two abelian varieties over a finite field are isogeneous if and only if their Tate modules are isomorphic as Galois representations.*

Indeed for abelian varieties the cohomology (the Tate module is poor man's cohomology, in fact equal with coefficients in $\mathbb{Z}_\ell$) tells us almost everything about the variety itself. This is why motivic theory is concerned with abelian varieties.

2.2 The Eichler-Shimura relation

Geometrically, the T_p are about isogenies. Modulo p there is a canonical isogeny, the Frobenius isogeny (raising everything to the p-th power). So T_p intuitively should have something to do with F.

As a correspondence on $X_0(N)$, T_p has the following definition:

$$(E, C) \mapsto \sum_{\phi\colon E \to E'} (E', \phi(C))$$

Here E is an elliptic curve and C a subgroup of order N on it. The sum runs over all degree p isogenies $\phi\colon E \to E'$. In characteristic p, any p-isogeny is either the Frobenius F or its dual isogeny $F^\vee$ (the Verschiebung). Remember the construction of $X_0(N)$ via

$$E_j\colon y^2 + xy = x^3 - \frac{36}{j - 1728}x - \frac{1}{j - 1728}$$

Frobenius sends an elliptic curve E with invariant j to $E^{(p)}$ with invariant j^p. This shows that the correspondence

$$(E, C) \mapsto (E^{(p)}, F(C))$$

on $X_0(N)$ is actually itself the Frobenius on $X_0(N)$. The same for the dual. Since these are the only isogenies we get

$$T_p = F + F^\vee \pmod p$$

Theorem 2.2. *Let $p \nmid N$. Then the relation*

$$T_p = F + \langle p \rangle F'$$

between the endomorphisms T_p, F (Frobenius) and F' (Verschiebung) on $J_{\Gamma/\mathbb{F}_p}$ holds.

Definition 2.3. The vector space version of the Tate module is

$$\mathcal{V} = T_\ell(J_\Gamma) \otimes_{\mathbb{Z}_\ell} \mathbb{Q}_\ell$$

This is a $\mathbb{T}_{\mathbb{Q}_\ell}$-module.

Lemma 2.4. $\mathcal{V}$ is a free $\mathbb{T}_{\mathbb{Q}_\ell}$-module of rank 2.

Theorem 2.5. *The Frobenius F as an endomorphism on the $\mathbb{T}_{\mathbb{Q}_\ell}$-module $\mathcal{V}$ has characteristic polynomial*

$$F^2 - T_p X + \langle p \rangle p$$

Proof. Multiplying the Eichler-Shimura relation by F and using $FF' = p$ we get

$$F^2 - T_p F + \langle p \rangle p = 0$$

This is certainly a good candidate for the characteristic polynomial and we can check by computing the trace of F, i.e. we have to show that $\mathrm{tr}(F) = T_p$. For this we define a pairing on $T_\ell(J_\Gamma)$ which is too lengthy here. Sorry. □

2.3 The Weil pairing

We need this to prove things about the Tate module that help us figure out equivalences of modularity statements. Let E be an elliptic curve.

Definition 2.6 (Weil pairing)**.** Let $P, Q \in E(K)[n]$. Choose a function in $K(E)$ having the following divisor:

$$\mathrm{div}(f) = \sum_{k=0}^{n-1} (P + kQ) - \sum_{k=0}^{n-1} kQ$$

Of course this divisor has degree 0, f having a zero at each point $P + kQ$ and a simple pole at each kQ. Then f is unique up to scaling. Let g be the translate of f by Q. Then g has the same divisor as f ($nQ = 0$) and therefore f/g is a constant, in fact an n-th root of unity (translating n times must give the identity). Define

$$e_n(P, Q) = \frac{f}{g} \in \mu_n$$

and call this the Weil pairing of P and Q.

Lemma 2.7 (Properties of the Weil pairing)**.** The pairing e_n enjoys the following properties:

1. $e_n(P_1 + P_2, Q) = e_n(P_1, Q) e_n(P_2, Q)$
2. $e_n(P, P) = 1$
3. e_n is nondegenerate

4. the e_{ℓ^n} can be combined into a bilinear, alternating, nondegenerate pairing

$$e\colon T_\ell(E) \times T_\ell(E) \to T_\ell(\mu)$$

such that for an isogeny $\phi\colon E_1 \to E_2$, ϕ and $\widehat{\phi}$ are adjoints for e.

Lemma 2.8. Let $\phi \in \mathrm{End}(E)$ and $\phi_\ell\colon T_\ell(E) \to T_\ell(E)$ be the induced map on the Tate module. Then

$$\det(\phi_\ell) = \deg(\phi)$$

Proof. Choose a $\mathbb{Z}_\ell$-basis $\{v_1, v_2\}$ for $T_\ell(E)$ and let the matrix of ϕ_ℓ for this basis be

$$\phi_\ell = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Then we compute

$$\begin{aligned} e(v_1, v_2)^{\deg(\phi)} &= e([\deg(\phi)]v_1, v_2) \\ &= e(\widehat{\phi}_\ell \phi_\ell v_1, v_2) \\ &= e(\phi_\ell v_1, \phi_\ell v_2) \\ &= e(av_1 + cv_2, bv_1 + dv_2) \\ &= e(v_1, v_2)^{ad-bc} \\ &= e(v_1, v_2)^{\det(\phi_\ell)} \end{aligned}$$

Nondegeneracy of e gives $\det(\phi_\ell) = \deg(\phi_\ell)$. □

3 Recap: Eigenvalues

Remember that an eigenform is a simultaneous eigenfunction for all operators in the Hecke algebra $\mathbb{T}$, i.e. for all T_n and $\langle d \rangle$. This means that for an eigenform f there is, for each $T \in \mathbb{T}$, an eigenvalue $\lambda_f(T) \in \mathbb{C}$ such that

$$Tf = \lambda_f(T) f$$

Since the Hecke algebra is commutative, we get an algebra homomorphism

$$\lambda_f\colon \mathbb{T} \to \mathbb{C}$$

associated to an eigenform f. The most important fact about the Hecke operators T_n is that they extract the Fourier coefficient a_n out of their eigenfunctions, i.e. $\lambda_f(T_n) = a_n$ if $f = \sum a_n q^n$.

4 The Jacobian of $X(N)$

The Hecke operators T_p act on the Jacobian as follows: $J = V/\Lambda$ where $V = S_2(\Gamma)^\vee$ and $\Lambda = H_1(X(\Gamma), \mathbb{Z})$ is embedded as a lattice in V via

$$\gamma \mapsto \left(\omega_f \mapsto \int_\gamma \omega_f\right)$$

If $\phi \in V$, $\phi\colon S_2(\Gamma) \to \mathbb{C}$, then

$$T\phi\colon S_2(\Gamma) \to \mathbb{C}, \quad (T\phi)(f) = \phi(Tf)$$

The action is best described via geometric correspondences.

Let C be a curve. The points of the Jacobian parametrize just the rational divisor classes of degree 0, i.e. $J_C(k) \cong \mathrm{Pic}^0_k(C)$. We will assume C has a rational point, i.e. $C(k) \neq \emptyset$. This is always true for modular curves. Fix such a point $P \in C(k)$. Then we can (on points) embed our curve in the Jacobian:

$$C(k) \to J_C(k), \quad Q \mapsto [Q] - [P]$$

The underlying algebraic morphism $i_C\colon C \to J_C$ satisfies the following universal property:

> If A is an abelian variety and $g\colon C \to A$ is a morphism sending P to $0 \in A$, then there is a unique homomorphism $\psi\colon J \to A$ of abelian varieties making the diagram commute:

$$\begin{array}{ccc} C & \xrightarrow{i_C} & J_C \\ & \searrow^{f} & \downarrow{\scriptstyle \psi} \\ & & A \end{array}$$

Jacobians are always principally polarized, meaning there is a canonical isomorphism $J_C \to J_C^\vee$ where $J_C^\vee$ should be thought of again as $\mathrm{Pic}^0(J_C)$.

5 The Shimura construction

Remember that $\mathbb{T}^0 \subseteq \mathbb{T}$ was the Hecke subalgebra generated by all T_n where $n \nmid N$ and all $\langle d \rangle$.

Theorem 5.1. *Let $f \in S_2^{\mathrm{new}}(\Gamma)$ be an eigenvector for all the operators in $\mathbb{T}^0$. Then f is an eigenform for the whole algebra $\mathbb{T}$ and therefore is unique up to scaling. More generally, if f is a newform of level $N_f \mid N$ then*

$$S_f = \{g \in S_2(\Gamma) \mid Tg = \lambda_f(T)g \text{ for all } T \in \mathbb{T}^0\}$$

is stable under $\mathbb{T}$. It has a basis consisting of the $f(az)$ where a ranges over the divisors of N/N_f. We have

$$S_2(\Gamma) = \bigoplus_f S_f$$

where the sum is taken over all newforms f of all levels $N_f \mid N$.

We start with an eigenform $f = \sum a_n q^n$ for Γ. Let K_f be the field extension of $\mathbb{Q}$ generated by all the a_n, which is a finite extension by some previous theorem. By λ_f we denote the associated eigenvalue algebra homomorphism $\mathbb{T}_{\mathbb{Q}} \to K_f$. We are going to associate to f an abelian variety A_f (defined over $\mathbb{Q}$) of dimension $[K_f : \mathbb{Q}]$. So if actually the Fourier coefficients are all rational, then A_f will be an elliptic curve.

Let I_f be the ideal $\ker(\lambda_f) \cap \mathbb{T}_{\mathbb{Z}}$ in $\mathbb{T}_{\mathbb{Z}}$. The operators in I_f act on J_Γ, we can therefore consider the image $I_f(J_\Gamma)$. It is a connected subgroup of J_Γ and therefore an abelian subvariety.

Definition 5.2. The abelian variety A_f associated to f is

$$A_f = J_\Gamma / I_f(J_\Gamma)$$

Remark 5.3. A_f is defined over $\mathbb{Q}$ (since J_Γ and $I_f(J_\Gamma)$ are) and depends only on the Galois orbit $[f]$ of f (because $\lambda_{f^\sigma} = \lambda_f^\sigma$ and its kernel doesn't change).

Lemma 5.4. The Jacobian $J_0(N)$ splits via an isogeny as

$$J_0(N) \simeq \prod_{[f]} A_f^{m_f}$$

where the product runs over all Galois orbits of newforms of some level $N_f | N$ and $m_f = \sigma_0(N/N_f)$ is the number of divisors of N/N_f.

To describe the complex points as a torus we set

$$V_f = \langle g \mid g \in [f] \rangle^\vee$$

which is a vector space of dimension $[K_f : \mathbb{Q}]$. We can also restrict the lattice $H_1(X_0(N), \mathbb{Z})$ of $S_2(\Gamma_0(N))^\vee$ to V_f, which gives a subgroup Λ_f there. Then the homomorphism

$$J_0(N) \to V_f / \Lambda_f$$

induces an isomorphism $A_f \to V_f/\Lambda_f$. To check this in detail is not very pretty.

We will now connect this construction with the Fourier coefficient definition of modularity.

Lemma 5.5. Let f be an eigenform of level N and $p \nmid N$ a prime. Then A_f has good reduction at p. The reason is that X_Γ has good reduction at p which implies J_Γ has good reduction at p.

Let p be a prime not dividing N. We want to study the number $N_{f,p}$ of points on A_f over $\mathbb{F}_p$.

Proposition 5.6. *We have*

$$N_{f,p} = \mathrm{Norm}_{K_f/\mathbb{Q}}(\lambda_f(1 - a_p(f) + \langle p\rangle p))$$

Here $a_p(f) = T_p(f)$, the p-th Fourier coefficient of f.

Proof. We only show the case $K_f = \mathbb{Q}$ in which case we can drop the norm and $A_f = E_f$ is an elliptic curve.

The fixed points of the Frobenius $\phi\colon (x, y) \mapsto (x^p, y^p)$ are exactly the points defined over $\mathbb{F}_p$. Reason: The elements of $\mathbb{F}_p$ are exactly the solutions to $X^p = X$ in $\overline{\mathbb{F}}_p$; all p points $x \in \mathbb{F}_p$ satisfy this and the equation, being of degree p, has exactly p solutions. Therefore

$$P \in E_f(\mathbb{F}_p) \Leftrightarrow P \in \ker(1 - F)$$

Now since F is not separable, $(1 - F)$ is and therefore $(1 - F)$ is unramified. In particular, $\#\ker(1 - F) = \#(1 - F)^{-1}(O_E) = \deg(1 - F)$. But by an above lemma we know $\deg(1 - F) = \det(1 - F)$. We have established

$$N_{f,p} = \det(1 - F)$$

We already know that

$$X^2 - T_pX + \langle p\rangle p = 0$$

is the characteristic polynomial for F. An algebraic manipulation shows that the characteristic polynomial of $(1 - F)$ is

$$X^2 + (T_p - 2)X + (\langle p\rangle p - T_p + 1) = 0$$

and we deduce $\det(1 - F) = \langle p\rangle p - T_p + 1$ (over the $\mathbb{T}_{\mathbb{Q}}$-module $\mathcal{V}$). On $T_\ell(E_f)$ this acts by λ_f and we get

$$\det(1 - F) = \lambda_f(1 - T_p + \langle p\rangle p) = 1 - a_p + p$$

since we are working with $\Gamma_0(N)$ and $\langle p\rangle$ has eigenvalue 1. □

Corollary 5.7. Let f be an eigenform with rational coefficients so we get an elliptic curve E_f. Remember the definition $b_p = (p + 1) - \#E_f(\mathbb{F}_p)$. The proposition then says $a_p = b_p$ which is the modularity statement for E_f. We have shown: The elliptic curves which we get by the Shimura construction are modular.

6 The modularity theorem

Proposition 6.1. *Let E be an elliptic curve with arithmetic conductor N. Then the following are equivalent:*

1. *E is isogeneous over $\mathbb{Q}$ to E_f for some newform f on some congruence subgroup Γ.*
2. *E is isogeneous over $\mathbb{Q}$ to E_f for some newform f on $\Gamma_0(N)$.*
3. *There is a non-constant morphism $\phi\colon X_0(N) \to E$, defined over $\mathbb{Q}$.*

If E satisfies these conditions, then it is called modular.

Theorem 6.2 (Modularity theorem)**.** *All elliptic curves are modular.*

Proof. We will not show the equivalence of the first two statements. For the implication $2 \Rightarrow 3$ consider an isogeny $E_f \to E$. Composing with $J_0(N) \to E_f$ gives a surjective map $J_0(N) \to E$. Compose this now with the embedding $X_0(N) \to J_0(N)$ to get a map $\phi\colon X_0(N) \to E$.

Lemma 6.3. The map ϕ is non-constant and therefore surjective.

Proof. We can work over $\mathbb{C}$. Set $C = X_0(N)$. Let $C \to J_C$ be given by a complex point $O \in C(\mathbb{C})$. Let $D \in J_C$ be a divisor of degree 0, i.e. $D = \sum_{i=1}^n P_i - \sum_{i=1}^n Q_i$. Then also

$$D = \sum_{i=1}^{n} (P_i - O) - \sum_{i=1}^{n} (Q_i - O)$$

that is every point in J_C can be written as a linear combination of points in the embedded image of C. If taking the quotient J_C/I would kill C, it would have to kill the whole J_C and the image would be 0-dimensional. □

Conversely, for $3 \Rightarrow 2$, assume a non-constant map $\phi\colon X_0(N) \to E$. E is an abelian variety, therefore by the universal property of the Jacobian, ϕ factors over a surjective map $\phi_*\colon J_0(N) \to E$. We know that $J_0(N)$ is isogeneous to a product of abelian varieties A_f where f runs over all newforms of level $N_f | N$. Therefore there is a surjective map $A_f \to E$ for some A_f. Now all A_f are simple, therefore this map has to be an isogeny and A_f an elliptic curve. □

We have seen that "geometric" modularity implies modularity in terms of the Fourier coefficients. How do we see the converse? Assume E is an elliptic curve such that there is a cusp form $f = \sum a_n q^n$ such that $b_p(E) = a_p$ for the p where E has good reduction.

We then have, by our previous considerations, that for all these p the number of $\mathbb{F}_p$-points on E and E_f are equal:

$$\#E(\mathbb{F}_p) = \#E_f(\mathbb{F}_p)$$

This implies by Tate's isogeny theorem that there is, for all those p, an isogeny $E \to E_f$. Does this lift to an isogeny over $\mathbb{Q}$?

In fact this follows from Falting's isogeny theorem (the same as Tate's, only over an arbitrary number field). We have to show that the Tate modules $T_\ell(E)$ and $T_\ell(E_f)$ are isomorphic as Galois representations. But we will learn later (next talk even) that a Galois representation for $G_{\mathbb{Q}}$ is determined by the action of the Frobenius for all but finitely many primes.

7 Relation to L-functions

There is this cool theorem. Modularity is important, outside of Fermat's last theorem.

Proposition 7.1. *If $E \simeq E_f$ is modular then*

$$L(E/\mathbb{Q}, s) = L(f, s)$$

In particular, $L(E/\mathbb{Q}, s)$ has an analytic continuation to the whole of $\mathbb{C}$ and a functional equation.

This has connections to Birch-Swinnerton-Dyer etc.