

20 The modular equation

In the previous lecture we defined modular curves as quotients of the extended upper half plane under the action of a congruence subgroup (a subgroup of $\mathrm{SL}_2(\mathbb{Z})$ that contains a principal congruence subgroup $\Gamma(N) := \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} \equiv_N \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right\}$ for some $N \in \mathbb{Z}_{>0}$). Of particular interest is the modular curve $X_0(N) := \mathbb{H}^*/\Gamma_0(N)$, where

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : c \equiv 0 \pmod{N} \right\}.$$

This modular curve plays a central role in the theory of elliptic curves. One form of the modularity theorem (a special case of which implies Fermat's last theorem) is that every elliptic curve E/\mathbb{Q} admits a morphism $X_0(N) \rightarrow E$ for some $N \in \mathbb{Z}_{\geq 1}$. It is also a key ingredient for algorithms that use isogenies of elliptic curves over finite fields, including the Schoof-Elkies-Atkin algorithm, an improved version of Schoof's algorithm that is the method of choice for counting points on elliptic curves over a finite fields of large characteristic. Our immediate interest in the modular curve $X_0(N)$ is that we will use it to prove the first main theorem of complex multiplication; among other things, this theorem implies that the j -invariants of elliptic curve E/\mathbb{C} with complex multiplication are algebraic integers.

There are two properties of $X_0(N)$ that make it so useful. The first, which we will prove in this lecture, is that it has a canonical model over \mathbb{Q} with integer coefficients; this allows us to interpret $X_0(N)$ as a curve over any field, including finite fields. The second is that it parameterizes isogenies between elliptic curves (in a sense that we will make precise in the next lecture). In particular, given the j -invariant of an elliptic curve E and an integer N , we can use our explicit model of $X_0(N)$ to determine the j -invariants of all elliptic curves that are related to E by an isogeny whose kernel is a cyclic group of order N .

In order to better understand modular curves, we need to introduce modular functions.

20.1 Modular functions

Modular functions are meromorphic functions on a modular curve. To make this statement precise, we first need to discuss q -expansions. The map $q: \mathbb{H} \rightarrow \mathbb{D}$ defined by

$$q(\tau) = e^{2\pi i \tau} = e^{-2\pi \operatorname{im} \tau} (\cos(2\pi \operatorname{re} \tau) + i \sin(2\pi \operatorname{re} \tau))$$

bijectionally maps each vertical strip $\mathbb{H}_n := \{\tau \in \mathbb{H} : n \leq \operatorname{re} \tau < n+1\}$ (for any $n \in \mathbb{Z}$) to the punctured unit disk $\mathbb{D}_0 := \mathbb{D} - \{0\}$. We also note that

$$\lim_{\operatorname{im} \tau \rightarrow \infty} q(\tau) = 0.$$

If $f: \mathbb{H} \rightarrow \mathbb{C}$ is a meromorphic function that satisfies $f(\tau+1) = f(\tau)$ for all $\tau \in \mathbb{H}$, then we can write f in the form $f(\tau) = f^*(q(\tau))$, where $f^*: \mathbb{D}_0 \rightarrow \mathbb{C}$ is a meromorphic function that we can define by fixing a vertical strip \mathbb{H}_n and putting $f^* := f \circ (q|_{\mathbb{H}_n})^{-1}$.

The q -expansion (or q -series) of $f(\tau)$ is obtained by composing the Laurent-series expansion of f^* at 0 with the function $q(\tau)$:

$$f(\tau) = f^*(q(\tau)) = \sum_{n=-\infty}^{+\infty} a_n q(\tau)^n = \sum_{n=-\infty}^{+\infty} a_n q^n.$$

As on the RHS above, it is customary to simply write q for $q(\tau) = e^{2\pi i \tau}$, as we shall do henceforth; but keep in mind that the symbol q denotes a function of $\tau \in \mathbb{H}$.

If f^* is meromorphic at 0 (meaning that $z^{-k}f^*(z)$ has an analytic continuation to an open neighborhood of $0 \in \mathbb{D}$ for some $k \in \mathbb{Z}_{\geq 0}$) then the q -expansion of f has only finitely many nonzero a_n with $n < 0$ and we can write

$$f(\tau) = \sum_{n=n_0}^{\infty} a_n q^n,$$

with $a_{n_0} \neq 0$, where n_0 is the order of f^* at 0. We then say that f is *meromorphic at ∞* , and call n_0 the *order of f at ∞* .

More generally, if f satisfies $f(\tau + N) = f(\tau)$ for all $\tau \in \mathbb{H}$, then we can write f as

$$f(\tau) = f^*(q(\tau)^{1/N}) = \sum_{n=-\infty}^{\infty} a_n q^{n/N}, \quad (1)$$

and we say that f is meromorphic at ∞ if f^* is meromorphic at 0.

If Γ is a congruence subgroup of level N , then for any Γ -invariant function f we have $f(\tau + N) = f(\tau)$ (for $\gamma = \begin{pmatrix} 1 & N \\ 0 & 1 \end{pmatrix} \in \Gamma$ we have $\gamma\tau = \tau + N$), so f can be written as in (1), and it makes sense to say that f is (or is not) meromorphic at ∞ .

Definition 20.1. Let $f : \mathbb{H} \rightarrow \mathbb{C}$ be a meromorphic function that is Γ -invariant for some congruence subgroup Γ . The function $f(\tau)$ is said to be *meromorphic at the cusps* if for every $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ the function $f(\gamma\tau)$ is meromorphic at ∞ .

It follows immediately from the definition that if $f(\tau)$ is meromorphic at the cusps, then for any $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ the function $f(\gamma\tau)$ is also meromorphic at the cusps. In terms of the extended upper half-plane \mathbb{H}^* , notice that for any $\gamma \in \mathrm{SL}_2(\mathbb{Z})$,

$$\lim_{\mathrm{im} \tau \rightarrow \infty} \gamma\tau \in \mathbb{P}^1(\mathbb{Q}),$$

and recall that $\mathbb{P}^1(\mathbb{Q})$ is the $\mathrm{SL}_2(\mathbb{Z})$ -orbit of $\infty \in \mathbb{H}^*$, whose elements are called *cusps*. To say that $f(\gamma\tau)$ is meromorphic at ∞ is to say that $f(\tau)$ is meromorphic at $\gamma\infty$. To check whether f is meromorphic at the cusps, it suffices to consider a set of Γ -inequivalent cusp representatives $\gamma_1\infty, \gamma_2\infty, \dots, \gamma_n\infty$, one for each Γ -orbit of $\mathbb{P}^1(\mathbb{Q})$; this is a finite set because the congruence subgroup Γ has finite index in $\mathrm{SL}_2(\mathbb{Z})$.

If f is a Γ -invariant meromorphic function, then for any $\gamma \in \Gamma$ we must have

$$\lim_{\mathrm{im} \tau \rightarrow \infty} f(\gamma\tau) = \lim_{\mathrm{im} \tau \rightarrow \infty} f(\tau)$$

whenever either limit exists, and if neither limit exists then f must still have the same order at ∞ and $\gamma\infty$. Thus if f is meromorphic at the cusps it determines a meromorphic function $g : X_\Gamma \rightarrow \mathbb{C}$ on the modular curve $X_\Gamma := \mathbb{H}^*/\Gamma$ (as a Riemann surface). Conversely, every meromorphic function $g : X_\Gamma \rightarrow \mathbb{C}$ determines a Γ -invariant meromorphic function $f : \mathbb{H} \rightarrow \mathbb{C}$ that is meromorphic at the cusps via $f := g \circ \pi$, where π is the quotient map $\mathbb{H} \rightarrow \mathbb{H}/\Gamma$.

Definition 20.2. Let Γ be a congruence subgroup. A *modular function* for Γ is a Γ -invariant meromorphic function $f : \mathbb{H} \rightarrow \mathbb{C}$ that is meromorphic at the cusps; equivalently, it is a meromorphic function $g : X_\Gamma \rightarrow \mathbb{C}$ (as explained above).

Sums, products, and quotients of modular functions for Γ are modular functions for Γ , as are constant functions, thus the set of all modular functions for Γ forms a field $\mathbb{C}(\Gamma)$ that we view as a transcendental extension of \mathbb{C} . As we will shortly prove for $X_0(N)$, modular curves X_Γ are not only Riemann surfaces, they are algebraic curves over \mathbb{C} ; the field $\mathbb{C}(\Gamma)$ of modular functions for Γ is isomorphic to the function field $\mathbb{C}(X_\Gamma)$ of X_Γ/\mathbb{C} .

Remark 20.3. In fact, every compact Riemann surface corresponds to a smooth projective (algebraic) curve over \mathbb{C} that is uniquely determined up to isomorphism. Conversely, if X/\mathbb{C} is a smooth projective curve then the set $X(\mathbb{C})$ can be given a topology and a complex structure that makes it a compact Riemann surface S . The function field of X and the field of meromorphic functions on S are both finite extensions of a purely transcendental extension of \mathbb{C} (of transcendence degree one), and the two fields are isomorphic. We will make this isomorphism completely explicit for $X(1)$ and $X_0(N)$.

Remark 20.4. If f is a modular function for a congruence subgroup Γ , then it is also a modular function for any congruence subgroup $\Gamma' \subseteq \Gamma$, since Γ -invariance obviously implies Γ' -invariance, and the property of being meromorphic at the cusps does not depend on Γ' . Thus for all congruence subgroups Γ and Γ' we have

$$\Gamma' \subseteq \Gamma \implies \mathbb{C}(\Gamma) \subseteq \mathbb{C}(\Gamma'),$$

and the corresponding inclusion of function fields $\mathbb{C}(X_\Gamma) \subseteq \mathbb{C}(X_{\Gamma'})$ induces a morphism $X_{\Gamma'} \rightarrow X_\Gamma$ of algebraic curves, a fact that has many useful applications.

20.2 Modular Functions for $\Gamma(1)$

We first consider the modular functions for $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$. In Lecture 16 we proved that the j -function is $\Gamma(1)$ -invariant and holomorphic (hence meromorphic) on \mathbb{H} . To show that the $j(\tau)$ is a modular function for $\Gamma(1)$ we just need to show that it is meromorphic at the cusps. The cusps are all $\Gamma(1)$ -equivalent, so it suffices to show that the $j(\tau)$ is meromorphic at ∞ , which we do by computing its q -expansion. We first record the following lemma, which was used in Problem Set 8.

Lemma 20.5. Let $\sigma_k(n) = \sum_{d|n} d^k$, and let $q = e^{2\pi i\tau}$. We have

$$\begin{aligned} g_2(\tau) &= \frac{4\pi^4}{3} \left(1 + 240 \sum_{n=1}^{\infty} \sigma_3(n) q^n \right), \\ g_3(\tau) &= \frac{8\pi^6}{27} \left(1 - 504 \sum_{n=1}^{\infty} \sigma_5(n) q^n \right), \\ \Delta(\tau) &= g_2(\tau)^3 - 27g_3(\tau)^2 = (2\pi)^{12} q \prod_{n=1}^{\infty} (1 - q^n)^{24}. \end{aligned}$$

Proof. See Washington [1, pp. 273-274]. □

Corollary 20.6. With $q = e^{2\pi i\tau}$ we have

$$j(\tau) = \frac{1}{q} + 744 + \sum_{n=1}^{\infty} a_n q^n,$$

where the a_n are integers.

Proof. Applying Lemma 20.5 yields

$$\begin{aligned} g_2(\tau)^3 &= \frac{64}{27}\pi^{12}(1 + 240q + 2160q^2 + \cdots)^3 = \frac{64}{27}\pi^{12}(1 + 720q + 179280q^2 + \cdots), \\ 27g_3(\tau)^2 &= \frac{64}{27}\pi^{12}(1 - 504q - 16632q^2 - \cdots)^2 = \frac{64}{27}\pi^{12}(1 - 1008q + 220752q^2 + \cdots), \\ \Delta(\tau) &= \frac{64}{27}\pi^{12}(1728q - 41472q^2 + \cdots) = \frac{64}{27}\pi^{12}1728q(1 - 24q + 252q^2 + \cdots), \end{aligned}$$

and we then have

$$j(\tau) = \frac{1728g_2(\tau)^3}{\Delta(\tau)} = \frac{1}{q} + 744 + \sum_{n=1}^{\infty} a_n q^n,$$

with $a_n \in \mathbb{Z}$, since $1 - 24q + 252q^2 + \cdots$ is an element of $1 + \mathbb{Z}[[x]]$, hence invertible. \square

Remark 20.7. The proof of Corollary 20.6 explains the factor 1728 that appears in the definition of the j -function: it is the least positive integer that ensures that the q -expansion of $j(\tau)$ has integral coefficients.

The corollary implies that the j -function is a modular function for $\Gamma(1)$, with a simple pole at ∞ . We proved in Theorem 18.5 that the j -function defines a holomorphic bijection from $Y(1) = \mathbb{H}/\Gamma(1)$ to \mathbb{C} . If we extend the domain of j to \mathbb{H}^* by defining $j(\infty) = \infty$, then the j -function defines an isomorphism from $X(1)$ to the Riemann sphere $\mathcal{S} := \mathbb{P}^1(\mathbb{C})$ that is holomorphic everywhere except for a simple pole at ∞ . In fact, if we fix $j(\rho) = 0$, $j(i) = 1728$, and $j(\infty) = \infty$, then the j -function is uniquely determined by this property (as noted above, we put $j(i) = 1728$ to obtain an integral q -expansion). It is for this reason that the j -function is sometimes referred to as *the* modular function. Indeed, every modular function for $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$ can be written in terms of the j -function.

Theorem 20.8. *Every modular function for $\Gamma(1)$ is a rational function of $j(\tau)$; in other words $\mathbb{C}(\Gamma(1)) = \mathbb{C}(j)$.*

Proof. As noted above, the j -function is a modular function for $\Gamma(1)$, so $\mathbb{C}(j) \subseteq \mathbb{C}(\Gamma(1))$. If $g: X(1) \rightarrow \mathbb{C}$ is a modular function for $\Gamma(1)$ then $f := g \circ j^{-1}: \mathcal{S} \rightarrow \mathbb{C}$ is meromorphic, and Lemma 20.9 below implies that f is a rational function. Thus $g = f \circ j \in \mathbb{C}(j)$. \square

Lemma 20.9. *Every meromorphic function $f: \mathcal{S} \rightarrow \mathbb{C}$ on the Riemann sphere $\mathcal{S} := \mathbb{P}^1(\mathbb{C})$ is a rational function.*

Proof. Let $f: \mathcal{S} \rightarrow \mathbb{C}$ be a nonzero meromorphic function. We may assume without loss of generality that f has no zeros or poles at $\infty := (1 : 0)$, since we can always apply a linear fractional transformation $\gamma \in \mathrm{SL}_2(\mathbb{C})$ to move a point where f does not have a pole or a zero to ∞ and replace f by $f \circ \gamma$ (note that γ and γ^{-1} are rational functions, and if $f \circ \gamma$ is a rational function, so is $f = f \circ \gamma \circ \gamma^{-1}$).

Let $\{p_i\}$ be the set of poles of $f(z)$, with orders $m_i := -\mathrm{ord}_{p_i}(f)$, and let $\{q_j\}$ be the set of zeros of f , with orders $n_j := \mathrm{ord}_{q_j}(f)$. We claim that

$$\sum_i m_i = \sum_j n_j.$$

To see this, triangulate \mathcal{S} so that all the poles and zeros of $f(z)$ lie in the interior of a triangle. It follows from Cauchy's argument principle (Theorem 15.17) that the contour integral

$$\int_{\Delta} \frac{f'(z)}{f(z)} dz$$

about each triangle (oriented counter clockwise) is the difference between the number of zeros and poles that $f(z)$ in its interior. The sum of these integrals must be zero, since each edge in the triangulation is traversed twice, once in each direction.

The function $h: \mathcal{S} \rightarrow \mathbb{C}$ defined by

$$h(z) = f(z) \cdot \frac{\prod_i (z - p_i)^{m_i}}{\prod_j (z - q_j)^{n_j}}$$

has no zeros or poles on \mathcal{S} . It follows from Liouville's theorem (Theorem 15.30) that h is a constant function, and therefore $f(z)$ is a rational function of z . \square

Corollary 20.10. *Every modular function $f(\tau)$ for $\Gamma(1)$ that is holomorphic on \mathbb{H} is a polynomial in $j(\tau)$.*

Proof. Theorem 20.8 implies that f can be written as a rational function of j , so

$$f(\tau) = c \frac{\prod_i (j(\tau) - \alpha_i)}{\prod_k (j(\tau) - \beta_k)},$$

for some $c, \alpha_i, \beta_j \in \mathbb{C}$. Now the restriction of j to any fundamental region for $\Gamma(1)$ is a bijection, so $f(\tau)$ must have a pole at $j^{-1}(\beta_k)$ for each β_k . But $f(\tau)$ is holomorphic and therefore has no poles, so the set $\{\beta_j\}$ is empty and $f(\tau)$ is a polynomial in $j(\tau)$. \square

We proved in the previous lecture that the j -function $j: X(1) \xrightarrow{\sim} \mathcal{S}$ determines an isomorphism of Riemann surfaces. As an algebraic curve over \mathbb{C} , the function field of $X(1) \simeq \mathcal{S} = \mathbb{P}^1(\mathbb{C})$ is the rational function field $\mathbb{C}(t)$, and we have just shown that the field of modular functions for $\Gamma(1)$ is the field $\mathbb{C}(j)$ of rational functions of j . Thus, as claimed in Remark 20.3, the function field $\mathbb{C}(X(1)) = \mathbb{C}(t)$ and the field of modular functions $\mathbb{C}(\Gamma(1)) = \mathbb{C}(j)$ are isomorphic, with the isomorphism given by $t \mapsto j$. More generally, for every congruence subgroup Γ , the field $\mathbb{C}(X_\Gamma) \simeq \mathbb{C}(\Gamma)$ is a finite extension of $\mathbb{C}(t) \simeq \mathbb{C}(j)$.

Theorem 20.11. *Let Γ be a congruence subgroup. The field $\mathbb{C}(\Gamma)$ of modular functions for Γ is a finite extension of $\mathbb{C}(j)$ of degree at most $n := [\Gamma(1) : \Gamma]$.*

Proof. Let γ_1 be the identity in $\Gamma(1)$ and let $\{\gamma_1, \dots, \gamma_n\} \subseteq \Gamma(1)$ be a set of right coset representatives for Γ as a subgroup of $\Gamma(1)$ (so $\Gamma(1) = \Gamma\gamma_1 \sqcup \dots \sqcup \Gamma\gamma_n$).

Let $f \in \mathbb{C}(\Gamma)$ and for $1 \leq i \leq n$ define $f_i(\tau) := f(\gamma_i\tau)$. For any $\gamma'_i \in \Gamma\gamma_i$ the functions $f(\gamma'_i\tau)$ and $f(\gamma_i\tau)$ are the same, since f is Γ -invariant. For any $\gamma \in \Gamma(1)$, the set of functions $\{f(\gamma_i\gamma\tau)\}$ is therefore equal to the set of functions $\{f(\gamma_i\tau)\}$, since multiplication on the right by γ permutes the cosets $\{\Gamma\gamma_i\}$. Any symmetric polynomial in the functions f_i is thus $\Gamma(1)$ -invariant, and meromorphic at the cusps (since f , and therefore each f_i , is), hence an element of $\mathbb{C}(j)$, by Theorem 20.8. Now let

$$P(Y) = \prod_{i \in \{1, \dots, n\}} (Y - f_i).$$

Then $f = f_1$ is a root of P (since γ_1 is the identity), and the coefficients of $P(Y)$ lie in $\mathbb{C}(j)$, since they are all symmetric polynomials in the f_i .

It follows that every $f \in \mathbb{C}(\Gamma)$ is the root of a monic polynomial in $\mathbb{C}(j)[Y]$ of degree n ; this implies that $\mathbb{C}(\Gamma)/\mathbb{C}(j)$ is an algebraic extension, and it is separable, since we are in characteristic zero. We now claim that $\mathbb{C}(\Gamma)$ is finitely generated: if not we could pick functions $g_1, \dots, g_{n+1} \in \mathbb{C}(\Gamma)$ such that

$$\mathbb{C}(j) \subsetneq \mathbb{C}(j)(g_1) \subsetneq \mathbb{C}(j)(g_1, g_2) \subsetneq \dots \subsetneq \mathbb{C}(j)(g_1, \dots, g_{n+1}).$$

But then $\mathbb{C}(j)(g_1, \dots, g_{n+1})$ is a finite separable extension of $\mathbb{C}(j)$ of degree at least $n+1$, and the primitive element theorem implies it is generated by some $g \in \mathbb{C}(\Gamma)$ whose minimal polynomial must have degree greater than n , which is a contradiction. The same argument then shows that $[\mathbb{C}(\Gamma) : \mathbb{C}(j)] \leq n$. \square

Remark 20.12. If $-I \in \Gamma$ then in fact $[\mathbb{C}(\Gamma(1)) : \mathbb{C}(\Gamma)] = [\Gamma(1) : \Gamma]$; we will prove this for $\Gamma = \Gamma_0(N)$ in the next section. In general $[\mathbb{C}(\Gamma(1)) : \mathbb{C}(\Gamma)] = [\bar{\Gamma}(1) : \bar{\Gamma}]$, where $\bar{\Gamma}$ denotes the image of Γ in $\mathrm{PSL}_2(\mathbb{Z}) := \mathrm{SL}_2(\mathbb{Z})/\{\pm I\}$.

20.2.1 Modular functions for $\Gamma_0(N)$

We now consider modular functions for the congruence subgroup $\Gamma_0(N)$.

Theorem 20.13. *The function $j_N(\tau) := j(N\tau)$ is a modular function for $\Gamma_0(N)$.*

Proof. The function $j_N(\tau)$ is obviously meromorphic (in fact holomorphic) on \mathbb{H} , since $j(\tau)$ is, and it is meromorphic at the cusps for the same reason (note that τ is a cusp if and only if $N\tau$ is). We just need to show that $j_N(\tau)$ is $\Gamma_0(N)$ -invariant.

Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$. Then $c \equiv 0 \pmod{N}$ and

$$j_N(\gamma\tau) = j(N\gamma\tau) = j\left(\frac{N(a\tau + b)}{c\tau + d}\right) = j\left(\frac{aN\tau + bN}{\frac{c}{N}N\tau + d}\right) = j(\gamma'N\tau) = j(N\tau) = j_N(\tau),$$

where

$$\gamma' = \begin{pmatrix} a & bN \\ c/N & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}),$$

since c/N is an integer and $\det(\gamma') = \det(\gamma) = 1$. Thus $j_N(\tau)$ is $\Gamma_0(N)$ -invariant. \square

Theorem 20.14. *The field of modular functions for $\Gamma_0(N)$ is an extension of $\mathbb{C}(j)$ of degree $n := [\Gamma(1) : \Gamma_0(N)]$ generated by $j_N(\tau)$.*

Proof. By the previous theorem, we have $j_N \in \mathbb{C}(\Gamma_0(N))$, and from Theorem 20.11 we know that $\mathbb{C}(\Gamma_0(N))$ is a finite extension of $\mathbb{C}(j)$ of degree at most n , so it suffices to show that the minimal polynomial of j_N over $\mathbb{C}(j)$ has degree at least n .

As in the proof of Theorem 20.11, let us fix a set of right coset representatives $\{\gamma_1, \dots, \gamma_n\}$ for $\Gamma_0(N) \subseteq \Gamma(1)$, and let $P \in \mathbb{C}(j)[Y]$ be the minimal polynomial of j_N over $\mathbb{C}(j)$. We may view $P(j(\tau), j_N(\tau))$ as a function of τ , which must be the zero function. If we replace τ by $\gamma_i\tau$ then for each γ_i we have

$$0 = P(j(\gamma_i\tau), j_N(\gamma_i\tau)) = P(j(\tau), j_N(\gamma_i\tau)),$$

so the function $j_N(\gamma_i\tau)$ is also a root of $P(Y)$.

To prove that $\deg P \geq n$ it suffices to show that the n functions $j_N(\gamma_i\tau)$ are distinct. Suppose not. Then $j(N\gamma_i\tau) = j(N\gamma_k\tau)$ for some $i \neq k$ and $\tau \in \mathbb{H}$ that we can choose to have stabilizer $\pm I$. Fix a fundamental region \mathcal{F} for $\mathbb{H}/\Gamma(1)$ and pick $\alpha, \beta \in \Gamma(1)$ so that $\alpha N\gamma_i\tau$ and $\beta N\gamma_k\tau$ lie in \mathcal{F} . The j -function is injective on \mathcal{F} , so

$$j(\alpha N\gamma_i\tau) = j(\beta N\gamma_k\tau) \iff \alpha N\gamma_i\tau = \pm \beta N\gamma_k\tau \iff \alpha N\gamma_i = \pm \beta N\gamma_k,$$

where we may view N as the matrix $\begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix}$, since $N\tau = \frac{N\tau+0}{0\tau+1}$.

Now let $\gamma = \alpha^{-1}\beta = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. We have

$$\begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix} \gamma_i = \pm \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix} \gamma_k,$$

and therefore

$$\gamma_i \gamma_k^{-1} = \pm \begin{pmatrix} 1/N & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix} = \pm \begin{pmatrix} a & b/N \\ cN & d \end{pmatrix}.$$

We have $\gamma_i \gamma_k^{-1} \in \mathrm{SL}_2(\mathbb{Z})$, so b/N is an integer, and $cN \equiv 0 \pmod{N}$, so $\gamma_i \gamma_k^{-1} \in \Gamma_0(N)$. But then γ_i and γ_k lie in the same right coset of $\Gamma_0(N)$, which is a contradiction. \square

20.3 The modular polynomial

Definition 20.15. The *modular polynomial* Φ_N is the minimal polynomial of j_N over $\mathbb{C}(j)$.

It follows from the proof of Theorem 20.14, we may write $\Phi_N \in \mathbb{C}(j)[Y]$ as

$$\Phi_N(Y) = \prod_{i=1}^n (Y - j_N(\gamma_i\tau)),$$

where $\{\gamma_1, \dots, \gamma_n\}$ is a set of right coset representatives for $\Gamma_0(N)$. The coefficients of $\Phi_N(Y)$ are symmetric polynomials in $j_N(\gamma_i\tau)$, so as in the proof of Theorem 20.11 they are $\Gamma(1)$ -invariant. They are holomorphic on \mathbb{H} , so they are polynomials in j , by Corollary 20.10. Thus $\Phi_N \in \mathbb{C}[j, Y]$. If we replace every occurrence of j in Φ_N with a new variable X we obtain a polynomial in $\mathbb{C}[X, Y]$ that we write as $\Phi_N(X, Y)$.

Our next task is to prove that the coefficients of $\Phi_N(X, Y)$ are actually integers, not just complex numbers. To simplify the presentation, we will prove this only prove for prime N , which is all that is needed in most practical applications (such as the SEA algorithm), and suffices to prove the main theorem of complex multiplication. The proof for composite N is essentially the same, but explicitly writing down a set of right coset representatives γ_i and computing the q -expansions of the functions $j_N(\gamma_i\tau)$ is more complicated.

We begin by fixing a specific set of right coset representatives for $\Gamma_0(N)$.

Lemma 20.16. For prime N we can write the right cosets of $\Gamma_0(N)$ in $\Gamma(1)$ as

$$\{\Gamma_0(N)\} \cup \{\Gamma_0(N)ST^k : 0 \leq k < N\},$$

where $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.

Proof. We first show that these cosets cover $\Gamma(1)$. Let $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$. If $c \equiv 0 \pmod{N}$, then $\gamma \in \Gamma_0(N)$ lies in the first coset. Otherwise, pick $k \in [0, N-1]$ so that $kc \equiv d \pmod{N}$ (c is nonzero modulo the prime N , so this is possible), and let

$$\gamma_0 := \begin{pmatrix} ka - b & a \\ kc - d & c \end{pmatrix} \in \Gamma_0(N).$$

Then

$$\gamma_0 ST^k = \begin{pmatrix} ka - b & a \\ kc - d & c \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & k \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \gamma,$$

lies in $\Gamma_0(N)ST^k$.

We now show the cosets are distinct. Suppose not. Then there must exist $\gamma_1, \gamma_2 \in \Gamma_0(N)$ such that either (a) $\gamma_1 = \gamma_2 ST^k$ for some $0 \leq k < N$, or (b) $\gamma_1 ST^j = \gamma_2 ST^k$ with $0 \leq j < k < N$. Let $\gamma_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. In case (a) we have

$$\gamma_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & k \end{pmatrix} = \begin{pmatrix} b & bk - a \\ d & dk - c \end{pmatrix} \in \Gamma_0(N),$$

which implies $d \equiv 0 \pmod{N}$ and $\det \gamma_2 = ad - bc \equiv 0 \pmod{N}$, a contradiction. In case (b), with $m = k - j$ we have

$$\gamma_1 = \gamma_2 ST^m S^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & m \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} a - bm & b \\ c - dm & d \end{pmatrix} \in \Gamma_0(N).$$

Thus $c - dm \equiv 0 \pmod{N}$, and since $c \equiv 0 \pmod{N}$ and $m \not\equiv 0 \pmod{N}$, we must have $d \equiv 0 \pmod{N}$, which again implies $\det \gamma_2 = ad - bc \equiv 0 \pmod{N}$, a contradiction. \square

Theorem 20.17. $\Phi_N \in \mathbb{Z}[X, Y]$.

Proof (for N prime). Let $\gamma_k := ST^k$. By Lemma 20.16 we have

$$\Phi_N(Y) = (Y - j_N(\tau)) \prod_{k=0}^{N-1} (Y - j_N(\gamma_k \tau)).$$

Let $f(\tau)$ be a coefficient of $\Phi_N(Y)$. Then $f(\tau)$ is a holomorphic function on \mathbb{H} , since $j(\tau)$ is, and $f(\tau)$ is $\Gamma(1)$ -invariant, since it is a symmetric polynomial in $j_N(\tau)$ and the functions $j_N(\gamma_k \tau)$, corresponding to a complete set of right coset representatives for $\Gamma_0(N)$; and $f(\tau)$ is meromorphic at the cusps, since it is a polynomial in functions that are meromorphic at the cusps. Thus $f(\tau)$ is a modular function for $\Gamma(1)$ holomorphic on \mathbb{H} and therefore a polynomial in $j(\tau)$, by Corollary 20.10. By Lemma 20.18 below, if we can show that the q -expansion of $f(\tau)$ has integer coefficients, then it will follow that $f(\tau)$ is an integer polynomial in $j(\tau)$ and therefore $\Phi_N \in \mathbb{Z}[X, Y]$.

We first show that the q -expansion of $f(\tau)$ has rational coefficients. We have

$$j_N(\tau) = j(N\tau) = \frac{1}{q^N} + 744 + \sum_{n=1}^{\infty} a_n q^{nN},$$

where the a_n are integers, thus $j_N \in \mathbb{Z}((q))$. For $j_N(\gamma_k \tau)$, we have

$$\begin{aligned} j_N(\gamma_k \tau) &= j(N\gamma_k \tau) = j\left(\begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix} ST^k \tau\right) \\ &= j\left(S \begin{pmatrix} 1 & 0 \\ 0 & N \end{pmatrix} \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \tau\right) = j\left(\begin{pmatrix} 1 & 0 \\ 0 & N \end{pmatrix} \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix} \tau\right) = j\left(\frac{\tau + k}{N}\right), \end{aligned}$$

where we are able to drop the S because $j(\tau)$ is Γ -invariant. If we let $\zeta_N = e^{\frac{2\pi i}{N}}$, then

$$q^{((\tau+k)/N)} = e^{2\pi i(\frac{\tau+k}{N})} = e^{2\pi i\frac{k}{N}} q^{1/N} = \zeta_N^k q^{1/N},$$

and

$$j_N(\gamma_k \tau) = \frac{\zeta_N^{-k}}{q^{1/N}} + \sum_{n=0}^{\infty} a_n \zeta_N^{kn} q^{n/N},$$

thus $j_N(\gamma_k \tau) \in \mathbb{Q}(\zeta_N)((q^{1/N}))$. The action of the Galois group $\text{Gal}(\mathbb{Q}(\zeta_N)/\mathbb{Q})$ on the coefficients of the q -expansions of each $j_N(\gamma_k \tau)$ induces a permutation of the set $\{j_N(\gamma_k \tau)\}$ and fixes $j_N(\tau)$. It follows that the coefficients of the q -expansion of f are fixed by $\text{Gal}(\mathbb{Q}(\zeta_N)/\mathbb{Q})$ and must lie in \mathbb{Q} . Thus $f \in \mathbb{Q}((q^{1/N}))$, and $f(\tau)$ is a polynomial in $j(\tau)$, so its q -expansion contains only integral powers of q and $f \in \mathbb{Q}((q))$.

We now note that the coefficients of the q -expansion of $f(\tau)$ are algebraic integers, since the coefficients of the q -expansions of $j_N(\tau)$ and the $j_N(\gamma_k)$ are algebraic integers, as is any polynomial combination of them. This implies $f(\tau) \in \mathbb{Z}((q))$. \square

Lemma 20.18 (Hasse q -expansion principle). *Let $f(\tau)$ be a modular function for $\Gamma(1)$ that is holomorphic on \mathbb{H} and whose q -expansion has coefficients that lie in an additive subgroup A of \mathbb{C} . Then $f(\tau) = P(j(\tau))$, for some polynomial $P \in A[X]$.*

Proof. By Corollary 20.10, we know that $f(\tau) = P(j(\tau))$ for some $P \in \mathbb{C}[X]$, we just need to show that $P \in A[X]$. We proceed by induction on $d = \deg P$. The lemma clearly holds for $d = 0$, so assume $d > 0$. The q -expansion of the j -function begins with q^{-1} , so the q -expansion of $f(\tau)$ must have the form $\sum_{n=-d}^{\infty} a_n q^n$, with $a_n \in A$ and $a_{-d} \neq 0$. Let $P_1(X) = P(X) - a_{-d}X^d$, and let $f_1(\tau) = P_1(j(\tau)) = f(\tau) - a_{-d}j(\tau)^d$. The q -expansion of the function $f_1(\tau)$ has coefficients in A , and by the inductive hypothesis, so does $P_1(X)$, and therefore $P(X) = P_1(X) + a_{-d}X^d$ also has coefficients in A . \square

References

- [1] Lawrence C. Washington, *Elliptic curves: number theory and cryptography*, second edition, Chapman & Hall/CRC, 2008.