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 $\Gamma(N)$ for some integer $N \geq 1$). Of particular interest is the curve $X_0(N) := \mathbb{H}^*/\Gamma_0(N)$, where

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

The modular curve $X_0(N)$ plays a central role in the theory of elliptic curves. From a theoretical perspective, it lies at the heart of the modularity conjecture, a special case of which was used to prove Fermat's last theorem. From a practical perspective, it is a key ingredient for algorithms that work with isogenies of elliptic curves over finite fields, including the Schoof-Elkies-Atkin algorithm, an enhanced version of Schoof's algorithm that is now the standard algorithm for point-counting on elliptic curves over a finite fields.

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 $\mathbb{Z}$, which allows us to use it over any field (including finite fields). The second is that it parameterizes isogenies between elliptic curves; in particular, given the j-invariant of an elliptic curve E and an integer N, we can use $X_0(N)$ to find the j-invariants of all elliptic curves related to E by a cyclic isogeny of degree N (we will define exactly what this means in the next lecture). Both of these properties will play a key role in our proof that the Hilbert class polynomial $H_D(X)$ has integer coefficients, which implies that the j-invariants of elliptic curves $E/\mathbb{C}$ with complex multiplication are algebraic integers, and has many other theoretical and practical applications.

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

20.1 Modular functions

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

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

bijectively maps each horizontal strip $\{\tau : n \leq \operatorname{im} \tau > n+1\}$ of the upper half plane $\mathbb{H}$ to the punctured unit disk $\mathbb{D} - \{0\}$. We also note that

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

If $f\colon \mathbb{H} \to \mathbb{C}$ 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^* is meromorphic on the punctured unit disk. The *q-series* or *q-exansion* for $f(\tau)$ is the Laurent-series expansion of f^* at 0 composed with $q(\tau)$:

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

where we typically just write q for $q(\tau)$ (as we will henceforth). If f^* is meromorphic at 0 then this series 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 $a_{n_0} \neq 0$. We then say that f is *meromorphic at* ∞, and call n_0 the *order of* f *at* ∞; note that n_0 is also the order of f^* at zero.

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

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

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 $f(\tau + N) = f(\tau)$ (consider $\gamma = \left(\begin{smallmatrix} 1 & N \\ 0 & 1 \end{smallmatrix}\right)$), so f can be written in the form (1), and the same is true of the function $f(\gamma\tau)$, for any fixed $\gamma \in \Gamma$.

Definition 20.1. Let Γ be a congruence subgroup and let $f : \mathbb{H} \to \mathbb{C}$ be a Γ-invariant meromorphic function. 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 ∞.

In terms of the extended upper half-plane $\mathbb{H}^*$, notice that for any $\gamma \in \mathrm{SL}_2(\mathbb{Z})$,

$$\lim_{\mathrm{im}\,\tau \to \infty} \gamma\tau \in \mathbb{H}^* \backslash \mathbb{H} = \mathbb{P}^1(\mathbb{Q}).$$

Thus to say that $f(\gamma\tau)$ is meromorphic at ∞ is the same thing as saying that $f(\tau)$ is meromorphic at the cusp $\gamma\infty$. Note that since f is Γ-invariant, in order to check whether or not f is meromorphic at the cusps, it suffices to consider a set of cusp representatives $\gamma_0\infty, \gamma_1\infty, \ldots, \gamma_k\infty$ for Γ; this set is finite because Γ has finite index in $\mathrm{SL}_2(\mathbb{Z})$.

Definition 20.2. Let Γ be a congruence subgroup. A *modular function* for Γ is a meromorphic function $g\colon \mathbb{H}^*/\Gamma \to \mathbb{C}$, equivalently, a Γ-invariant meromorphic function $f\colon \mathbb{H} \to \mathbb{C}$ that is meromorphic at the cusps.

Sums, products, and quotients of modular functions Γ are also modular functions for Γ, as are constant functions, thus the set of all modular functions for Γ is a field that is a transcendental extension of $\mathbb{C}$. Notice that if $f(\tau)$ is a modular function for a congruence subgroup Γ, then $f(\tau)$ is also a modular function for every congruence subgroup $\Gamma' \subseteq \Gamma$: clearly $f(\tau)$ is Γ'-invariant since $\Gamma' \subseteq \Gamma$, and the property of being meromorphic at the cusps does not depend on Γ'.

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 $\mathrm{SL}_2(\mathbb{Z})$-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 note the following lemma, part of which was used in Problem Set 8.

Lemma 20.3. *Let $\sigma_k(n) = \sum_{d|n} d^k$, and let $q = e^{2\pi i\tau}$. We have*

$$g_2(\tau) = \frac{4\pi^4}{3}\left(1 + 240\sum_{n=1}^{\infty}\sigma_3(n)q^n\right),$$

$$g_3(\tau) = \frac{8\pi^6}{27}\left(1 - 504\sum_{n=1}^{\infty}\sigma_5(n)q^n\right),$$

$$\Delta(\tau) = g_2^3(\tau) - 27g_3^2(\tau) = (2\pi)^{12}q\sum_{n=1}^{\infty}(1-q^n)^{24}.$$

Proof. See Washington [4, pp. 273-274]. □

Corollary 20.4. *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. We have

$$g_2^3(\tau) = \frac{64}{27}\pi^{12}(1 + 240q + O(q^2))^3 = \frac{64}{27}\pi^{12}(1 + 720q + O(q^2)),$$
$$\Delta(\tau) = \frac{64}{27}\pi^{12}(3^3\cdot 2^6)q(1 - 24q + O(q^2)),$$

where each $O(q^2)$ denotes sums of higher order terms with integer coefficients. Thus

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

with $a_n \in \mathbb{Z}$, as desired. □

Remark 20.5. The proof of Corollary 20.4 explains the factor $1728 = 3^3 \cdot 2^6$ 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, fixing $j(i) = 1728$ ensures 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 expressed in terms of the j-function.

Theorem 20.6. *Every modular function for $\Gamma(1)$ is a rational function of $j(\tau)$. Equivalently, $\mathbb{C}(j)$ is the field of modular functions for $\Gamma(1)$.*

Proof. Let $g\colon X(1) \to \mathbb{C}$ be a modular function for $\Gamma(1)$. Then $f = g \circ j^{-1}\colon \mathcal{S} \to \mathbb{C}$ is meromorphic. By Lemma 20.7 below, this implies that f is a rational function. Therefore $g = f \circ j \in \mathbb{C}(j)$, as desired. □

Lemma 20.7. *If $f\colon \mathcal{S} \to \mathbb{C}$ is meromorphic, then $f(z)$ is a rational function.*

Proof. We may assume without loss of generality that f has no zeros or poles at ∞ (the north pole of $\mathcal{S}$). If this is not the case, we may replace $f(z)$ by $f(z+c)$ with an appropriate constant $c \in \mathbb{C}$; in terms of $\mathbb{P}^1(\mathbb{C})$ this corresponds to applying the linear fractional transformation $\left(\begin{smallmatrix} 1 & c \\ 0 & 1 \end{smallmatrix}\right)$ which sends affine projective points $(z:1)$ to $(z+c:1)$ and moves the point $(1:0)$ at infinity to $(c:0)$. Note that if $f(z)$ is a rational function in z, so is $f(z+c)$.

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 15.16) that the counter integral

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

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 $h\colon \mathcal{S} \to \mathbb{C}$ 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 that h is a constant function, and therefore $f(z)$ is a rational function of z. □

Corollary 20.8. *Every modular function $f(\tau)$ for $\Gamma(1)$ that is holomorphic on $\mathbb{H}$ is a polynomial in $j(\tau)$.*

Proof. Theorem 20.6 implies that f is a rational function in j, which we may write as

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

for some $c, \alpha_i, \beta_j \in \mathbb{C}$. Now $j\colon \mathcal{F} \to \mathbb{C}$ is a bijection, so $f(\tau)$ must have a pole at $j^{-1}(\beta_k) \in \mathcal{F}$ 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 $j(\tau)$. □

20.2.1 Modular functions for $\Gamma_0(N)$

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

Theorem 20.9. *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 $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$. We have

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

where

$$\gamma' = \begin{pmatrix} a & bN \\ c/N & d \end{pmatrix}.$$

We now note that $\gamma' \in \mathrm{SL}_2(\mathbb{Z})$, since $\det(\gamma') = \det(\gamma) = 1$ and $c \equiv 0 \pmod{N}$ implies that c/N is an integer. And $j(\tau)$ is $\mathrm{SL}_2(\mathbb{Z})$-invariant, so

$$j_N(\gamma\tau) = j(\gamma' N\tau) = j(N\tau) = j_N(\tau),$$

thus $j_N(\tau)$ is $\Gamma_0(N)$-invariant. □

Theorem 20.10. *$\mathbb{C}(j, j_N)$ is the field of modular functions for $\Gamma_0(N)$.*

Cox gives a very concrete proof of this result in [1, Thm. 11.9]; here we give a simpler, but somewhat more abstract proof that is adapted from Milne [2, Thm. V.2.3].

Proof. Let $\{\gamma_1, \cdots, \gamma_m\} \subset \Gamma(1)$ be a set of right coset representatives for $\Gamma_0(N)$ as a subgroup of $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$; this means that the cosets $\Gamma_0(N)\gamma_1, \cdots, \Gamma_0(N)\gamma_m$ are distinct and cover $\Gamma(1)$. Without loss of generality, we may assume $\gamma_1 = I$ is the identity. Let K_N denote the field of modular functions for $\Gamma_0(N)$. By the previous theorem, $j_N \in K_N$, and clearly $j \in K_N$, since j is a modular function for $\Gamma(1)$ and therefore for $\Gamma_0(N) \subseteq \Gamma(1)$. Thus K_N is an extension the field $\mathbb{C}(j, j_N)$, we just need to show that it is a trivial extension, i.e. that $[K_N : \mathbb{C}(j, j_N)] = 1$.

We first bound the degree of K_N as an extension of the subfield $\mathbb{C}(j)$. Consider any function $f \in K_N$, and for $1 \le i \le m$ define $f_i(\tau) := f(\gamma_i\tau)$. Since $f(\tau)$ is $\Gamma_0(N)$-invariant, the function $f_i(\tau)$ does not depend on the choice of the right-coset representative γ_i (for any $\gamma_i' \in \Gamma_0(N)\gamma_i$ the functions $f(\gamma_i'\tau)$ and $f(\gamma_i\tau)$ are the same). This implies that for any $\gamma \in \Gamma(1)$, the set of functions $\{f(\gamma_i\gamma\tau)\}$ is equal to the set of functions $\{f(\gamma_i\tau)\}$, since right-multiplication by γ permutes the right cosets $\{\Gamma_0(N)\gamma_i\}$. Thus any symmetric polynomial in the functions f_i is $\Gamma(1)$-invariant, and therefore a rational function of $j(\tau)$, by Theorem 20.6. Now let

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

Then $f = f_1$ is a root of P (since $\gamma_1 = I$), and the coefficients of $P(Y)$ lie in $\mathbb{C}(j)$, since they are all symmetric polynomials in the f_i. Thus every $f \in K_N$ is the root of a monic polynomial over $\mathbb{C}(j)$ of degree m; this implies that $K_N/\mathbb{C}(j)$ is an algebraic extension, and it is separable, since we are in characteristic zero. We claim that K_N is also finitely generated: if not we could pick functions $g_1, \ldots, g_{m+1} \in K_N$ such that

$$\mathbb{C}(j) \subsetneq \mathbb{C}(j)(g_1) \subsetneq \mathbb{C}(j)(g_1, g_2) \subsetneq \cdots \subsetneq \mathbb{C}(j)(g_1, \ldots, g_{m+1}).$$

But then $\mathbb{C}(j)(g_1, \ldots, g_{m+1})$ is a finite separable extension of $\mathbb{C}(j)$ of degree at least $m+1$, and the primitive element theorem implies it is generated by some function g whose minimal

polynomial most have degree greater than m, which is a contradiction. The same argument then shows that $[K_N : \mathbb{C}(j)] \le m$.

Now let $F \in \mathbb{C}(j)[Y]$ be the minimal polynomial of f over $\mathbb{C}(j)$, which necessarily divides $P(Y)$, but may have lower degree. We can regard $F(j(\tau), f(\tau))$ as a function of τ, which must be the zero function. If we then replace τ by $\gamma_i\tau$, for every $\tau \in \mathbb{H}$ we have

$$F(j(\gamma_i\tau), f(\gamma_i\tau)) = F(j(\tau), f(\gamma_i\tau)) = F(j(\tau), f_i(\tau)) = 0,$$

where we have used the fact that the j-function is $\Gamma(1)$-invariant. Thus the functions f_i all have the same minimal polynomial F as f, which implies that $P = F^n$ for some $n \ge 1$. We have $n = 1$ if and only if the f_i are distinct, and if this is the case then we must have $K_N = \mathbb{C}(j, f)$, since $[K_N : \mathbb{C}(j)] \le m$ and $[\mathbb{C}(j, f) : \mathbb{C}(j)] = m$.

Now consider $f = j_N$. By the argument above, to prove $K_N = \mathbb{C}(j, j_N)$ we just need to show that the functions $f_i(\tau) = j_N(\gamma_i\tau) = j(N\gamma_i\tau)$ are distinct functions of τ as i varies.

Suppose not. Then $j(N\gamma_i\tau) = j(N\gamma_k\tau)$ for some $i \ne k$ and $\tau \in \mathbb{H}$ that we can choose to have stabilizer $\pm I$ (distinct meromorphic functions cannot agree on any open set where both are defined so we can easily avoid $\Gamma(1)$-translates of $e^{\pi i}$ and $e^{2\pi/3}$). 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 $\beta N\gamma_j\tau$ lie in $\mathcal{F}$. The j-function is injective on $\mathcal{F}$, so

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

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

Now let $\gamma = \alpha^{-1}\beta = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. 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 $\gamma_i\gamma_k^{-1}$, so b/N is an integer, and $cN \equiv 0 \bmod N$, so in fact $\gamma_i\gamma_k^{-1} \in \Gamma_0(N)$. But then γ_i and γ_k lie in the same right coset of $\Gamma_0(N)$, which is a contradiction. □

20.3 The modular polynomial

Definition 20.11. The *modular polynomial* Φ_N is the minimal polynomial of j_N over $\mathbb{C}(j)$.

As in the proof of Theorem 20.10, we may write $\Phi_N \in \mathbb{C}(j)[Y]$ as

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

where the γ_i are 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 20.10 they are $\Gamma(1)$-invariant; and they are holomorphic on $\mathbb{H}$, so they are polynomials in j, by Corollary 20.8. 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 an element of $\mathbb{C}[X, Y]$ 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 only prove this for prime N, which is all that is needed in many practical applications (such as the SEA algorithm), and suffices to prove the main theorem of complex multiplication.[1]

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

Lemma 20.12. *For prime N we can write the right cosets of $\Gamma_0(N)$ in $\Gamma(1)$ as*

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

where $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$.

Proof. We first show that the union of these cosets is $\Gamma(1)$. Let $\gamma = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) \in \Gamma$. If $C \equiv 0 \bmod N$, then $\gamma \in \Gamma_0(N)$ lies in the first coset above. Otherwise, we note that

$$ST^k = \begin{pmatrix} 0 & -1 \\ 1 & k \end{pmatrix} \quad \text{and} \quad (ST^k)^{-1} = \begin{pmatrix} k & 1 \\ -1 & 0 \end{pmatrix},$$

and for $C \not\equiv 0 \bmod N$, we may pick k such that $kC \equiv D \bmod N$, since N is prime. Then

$$\gamma_0 := \gamma(ST^k)^{-1} = \begin{pmatrix} kA - B & A \\ kC - D & C \end{pmatrix} \in \Gamma_0(N),$$

and $\gamma = \gamma_0(ST^k) \in \Gamma_0(N)ST^k$.

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 $0 \le k < N$, or (b) $\gamma_1 ST^j = \gamma_2 ST^k$ with $0 \le j < k < N$. Let $\gamma_2 = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. 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 $d \equiv 0 \bmod N$. But then $\det \gamma_2 = ad - bc \equiv 0 \bmod N$, 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 $-c - dm \equiv 0 \bmod N$, and since $c \equiv 0 \bmod N$ and $m \not\equiv 0 \bmod N$, we must have $d \equiv 0 \bmod N$, which again implies $\det \gamma_2 = ad - bc \equiv 0 \bmod N$, a contradiction. □

Theorem 20.13. $\Phi_N \in \mathbb{Z}[X,Y]$.

Proof (for N prime). Let $\gamma_k = ST^k$. By Lemma 20.12 we have

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

Let $f(\tau)$ be a coefficient of $\Phi_N(Y)$. Then $f(\tau)$ is holomorphic function on $\mathbb{H}$, since $j(\tau)$ is, $f(\tau)$ is $\Gamma(1)$-invariant, since, as in the proof of Theorem 20.10, it is symmetric polynomial

[1] 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.

in $j_N(\tau)$ and the functions $j_N(\gamma_k\tau)$, corresponding to a 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 $\Gamma(1)$ and therefore a polynomial in $j(\tau)$, by Corollary 20.8. By Lemma 20.14 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 $f(\tau)$ has have 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

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

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

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

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 Galois group $\mathrm{Gal}(\mathbb{Q}(\zeta_N)/\mathbb{Q})$ 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, which is a symmetric polynomial in these functions, are fixed by $\mathrm{Gal}(\mathbb{Q}(\zeta_N))/\mathbb{Q})$ and therefore lie in $\mathbb{Q}$; thus $f \in \mathbb{Q}((q^{1/N}))$.

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 $f(\tau) \in \mathbb{Z}((q^{1/N}))$.

Finally, we recall that $f(\tau)$ is a polynomial in $j(\tau)$, so its q-expansion can have only integral powers of q; therefore $f(\tau) \in \mathbb{Z}((q))$, as desired. □

Lemma 20.14 (Hasse q-expansion principal)**.** *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 20.8, 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. □

References

[1] David A. Cox, *Primes of the form $x^2 + ny^2$: Fermat, class field theory, and complex multiplication*, second edition, Wiley, 2013.

[2] J.S. Milne, *Elliptic curves*, BookSurge Publishers, 2006.

[3] J.H. Silverman, *Advanced topics in the the arithmetic of elliptic curves*, Springer, 1994.

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