21 The Hilbert class polynomial

In the previous lecture we proved that the field of modular functions for $\Gamma_0(N)$ is generated by the functions $j(\tau)$ and $j_N(\tau) := j(N\tau)$, that is, $\mathbb{C}(\Gamma_0(N)) = \mathbb{C}(j, j_N)$, and we showed that $\mathbb{C}(j, j_N)$ is a finite extension of $\mathbb{C}(j)$. We then defined the modular polynomial $\Phi_N(Y)$ as the minimal polynomial of j_N over $\mathbb{C}(j)$ and proved that its coefficients lie in $\mathbb{Z}[j] \subseteq \mathbb{C}(j)$. Replacing j with a formal variable X, we obtain a polynomial $\Phi_N \in \mathbb{Z}[X, Y]$ that gives a canonical defining equation for the modular curve $X_0(N)$.[1]

In this lecture we will use Φ_N to prove that the *Hilbert class polynomial*[2]

$$H_D(X) := H_{\mathcal{O}}(X) := \prod_{j(E)\in \mathrm{Ell}_{\mathcal{O}}(\mathbb{C})} \bigl(X - j(E)\bigr)$$

also has integer coefficients; here $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C}) := \{j(E) : \mathrm{End}(E) \simeq \mathcal{O}\}$ is the set of j-invariants of elliptic curves $E/\mathbb{C}$ with complex multiplication (CM) by the imaginary quadratic order $\mathcal{O}$ with discriminant $D = \mathrm{disc}(\mathcal{O})$. Recall that D uniquely determines $\mathcal{O}$ (and vice versa), by Theorem 18.17, so the notation H_D is unambiguous (both H_D and $H_{\mathcal{O}}$ appear in the literature, we will use the former).

The fact that $H_D \in \mathbb{Z}[x]$ implies that the j-invariant of any elliptic curve $E/\mathbb{C}$ with complex multiplication must be an algebraic integer, meaning that E can actually be defined over a number field (a finite extension of $\mathbb{Q}$). This is a remarkable result. It implies that of the uncountably many isomorphism classes of elliptic curves over $\mathbb{C}$, only countable many have complex multiplication. In order to prove this we will exploit the interpretation of $X_0(N)$ as the "moduli space" of cyclic N-isogenies of elliptic curves; our first task is to explain what this means.

21.1 Isogenies

Recall from §18.5 in Lecture 18 that if $L_1 \subseteq L_2$ are lattices in $\mathbb{C}$, and E_1 and E_2 are the elliptic curves corresponding to the complex tori $\mathbb{C}/L_1$ and $\mathbb{C}/L_2$, then the inclusion $L_1 \subseteq L_2$ induces an isogeny $\phi\colon E_1 \to E_2$ whose kernel is isomorphic to the finite abelian group L_2/L_1. Indeed, we have the commutative diagram

$$\begin{tikzcd}
\mathbb{C}/L_1 \arrow[r, "\iota"] \arrow[d, "\simeq"] & \mathbb{C}/L_2 \arrow[d, "\simeq"] \\
E_1(\mathbb{C}) \arrow[r] & E_2(\mathbb{C})
\end{tikzcd}$$

where the top map ι is induced by the inclusion $L_1 \subseteq L_2$ (lift from $\mathbb{C}/L_1$ to $\mathbb{C}$ then project to $\mathbb{C}/L_2$). If we replace L_2 by the homothetic lattice NL_2, where $N = [L_2 : L_1] = \deg\phi$, the inclusion $NL_2 \subseteq L_1$ induces an isogeny in the reverse direction which, after composing with the isomorphism corresponding to the homethety $L_2 \sim NL_2$, is the dual isogeny $\hat{\phi}\colon E_2 \to E_1$. The composition $\phi \circ \hat{\phi}$ is the multiplication-by-N map on E_2, corresponding to the lattice inclusion $NL_2 \subseteq L_2$, with kernel isomorphic to $L_2/NL_2 \simeq \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$.

[1] The curve $\Phi_N(X, Y) = 0$ is a singular affine curve with the same function field as $X_0(N)$; the desingularization of its projective closure is a smooth projective curve isomorphic to $X_0(N)$.

[2] Some authors use the term *Hilbert class polynomial* only when $\mathcal{O}$ is a maximal order (they then use the term *ring class polynomial* for the general case); we won't make this distinction.

Lecture by Andrew Sutherland

Definition 21.1. If L_1 is a sublattice of L_2 for which the group L_2/L_1 is cyclic, then we say that L_1 is a *cyclic sublattice* of L_2. Similarly, an isogeny $\phi\colon E_1 \to E_2$ is said to be *cyclic* if its kernel is a cyclic group. If ϕ is induced by the lattice inclusion $L_1 \subseteq L_2$ then ϕ is cyclic if and only if L_1 is a cyclic sublattice of L_2.

As we proved in Corollary 6.11, up to isomorphism, every isogeny is a composition of isogenies of prime degree, which are necessarily cyclic. So we may as well restrict our attention to cyclic isogenies ϕ, which we will show correspond to points on the modular curve $X_0(N)$, with $N = \deg\phi$, and in our proofs we will be content to restrict to the case where N is prime, since we can always decompose ϕ into a composition of isogenies of prime degree. It is thus enough for us to understand cyclic sublattices of prime index.

Lemma 21.2. *Let $L = [1,\tau]$ be a lattice with $\tau \in \mathbb{H}$ and let N be prime. The cyclic sublattices of L of index N are the lattice $[1, N\tau]$ and the lattices $[N, \tau + k]$, for $0 \le k < N$.*

Proof. The lattices $[1, N\tau]$ and $[N, \tau+k]$ are clearly index N sublattices of L, and they must be cyclic sublattices, since N is prime. Conversely, any sublattice $L' \subseteq L$ can be written as $[d, a\tau + k]$, where d is the least positive integer in L' and the index of L' in L is $ad = N$. Since N is prime, either $d = 1$ and $a = N$, in which case $L' = [1, N\tau]$, or $d = N$ and $a = 1$, in which case $L' = [N, \tau + k]$, and we may assume $0 \le k < N$. □

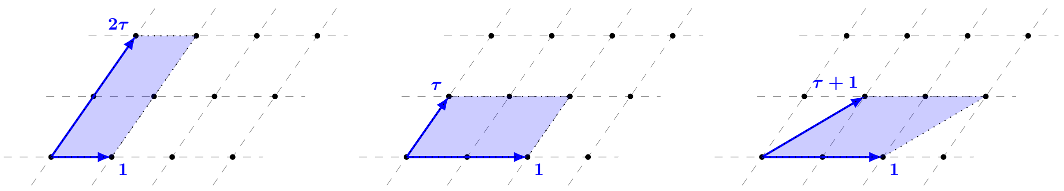

Figure 1: The three cyclic sublattices of $[1,\tau]$ of index 2.

Theorem 21.3. *For all $j_1, j_2 \in \mathbb{C}$, we have $\Phi_N(j_1, j_2) = 0$ if and only if j_1 and j_2 are the j-invariants of elliptic curves over $\mathbb{C}$ over that are related by a cyclic isogeny of degree N.*

Proof for N prime. We will prove the equivalent statement that $\Phi_N(j(L_1), j(L_2)) = 0$ if and only if L_1 is homothetic to a cyclic sublattice of L_2 of index N, equivalently, L_2 is homothetic to a cyclic sublattice of L_1. We may assume without loss of generality that $L_1 = [1, \tau_1]$ and $L_2 = [1, \tau_2]$, where $\tau_1, \tau_2 \in \mathbb{H}$. As in the proof of Theorem 20.17 we have

$$\Phi_N(j(\tau), Y) = (Y - j(N\tau)) \prod_{k=0}^{N-1} (Y - j(N\gamma_k\tau)), \tag{1}$$

where $\gamma_k := ST^k$, and

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

Thus

$$\Phi_N(j(L_1), j(L_2)) = \Phi_N(j([1,\tau_1]), j([1,\tau_2])) = \Phi_N(j(\tau_1), j(\tau_2))$$

is zero if and only if τ_2 is $\mathrm{SL}_2(\mathbb{Z})$-equivalent to $N\tau_1$ or $(\tau_1 + k)/N$, with $0 \le k < N$, hence if and only if L_2 is homothetic to a cyclic sublattice of L_1 of index N, by Lemma 21.2. □

1 Theorem 21.3 applies more generally to any field that can be embedded in $\mathbb{C}$, including all number fields. It can be extended via the Lefschetz principle [8, Thm. VI.6.1] to any field of characteristic zero, and as shown by Igusa [4], it also holds in fields of positive characteristic $p \nmid N$.

Theorem 21.4. *Let k be a field whose characteristic is not a divisor of $N \in \mathbb{Z}_{>1}$. For all $j_1, j_2 \in k$ we have $\Phi_N(j_1, j_2) = 0$ if and only if j_1 and j_2 are the j-invariants of elliptic curves over k that are related by a cyclic isogeny of degree N defined over k.*

Remark 21.5. We could have written the theorem as $\Phi_N(j(E_1), j(E_2)) = 0$ if and only if E_1 and E_2 are related by a cyclic isogeny of degree N, because over $\mathbb{C}$ the j-invariant characterizes elliptic curves up to isomorphism. Over a non-algebraically closed field the theorem remains true as written, but it is not necessarily true that $\Phi_N(j(E_1), j(E_2) = 0$ implies the existence of a cyclic N-isogeny $E_1 \to E_2$; one might need to replace E_1 or E_2 by a twist (a curve with the same j-invariant that is isomorphic over an extension field but not necessarily over the field of definition).

Remark 21.6. We should note that if $\phi\colon E_1 \to E_2$ is a cyclic N-isogeny, the pair of j-invariants $(j(E_1), j(E_2))$ does *not* uniquely determine ϕ, not even up to isomorphism. For example, suppose $\operatorname{End}(E_1) \simeq \mathcal{O}$ and $\mathfrak{p} \neq \bar{\mathfrak{p}}$ is a proper $\mathcal{O}$-ideal of prime norm p such that $[\mathfrak{p}]$ has order 2 in the class group $\operatorname{cl}(\mathcal{O})$. Then $\mathfrak{p}E_1 \simeq \bar{\mathfrak{p}}E_1$, and the isogenies $\phi_{\mathfrak{p}} : E_1 \to \mathfrak{p}E_1$ and $\phi_{\bar{\mathfrak{p}}} : E_1 \to \bar{\mathfrak{p}}E_1$ have distinct kernels but isomorphic images. These isogenies are not isomorphic (there is no automorphism we can compose with one to get the other, their kernels are distinct). In this situation $\Phi_p(j(E_1), Y)$ will have $j(E_2)$ as a double root.

The existence of the dual isogeny implies that $\Phi_N(j_1, j_2) = 0$ if and only if $\Phi_N(j_2, j_1) = 0$. In fact $\Phi_N(X, Y) = \Phi_N(Y, X)$ is symmetric in the variables X and Y.

Theorem 21.7. $\Phi_N(X, Y) = \Phi_N(Y, X)$

Proof. As in the proof of Theorem 21.3, the function $j(N\gamma_0\tau) = j(\tau/N)$ is a root of $\Phi_N(j, Y) \in \mathbb{C}(j)[Y]$ (this is true whether or not N is prime). We also have the identity $\Phi_N(j(\tau), j(N\tau)) = 0$, which implies $\Phi_N(j(\tau/N), j(\tau)) = 0$, so $j(\tau/N)$ is also a root of $\Phi_N(Y, j) \in \mathbb{C}(j)[Y]$. But $\Phi_N(j, Y)$ is irreducible in $\mathbb{C}(j)[Y]$, since it is the minimal polynomial of j_N over $\mathbb{C}(j)$, so $\Phi_N(j, Y)$ must divide $\Phi_N(Y, j)$ in $\mathbb{C}(j)[Y]$ (otherwise their GCD would properly divide $\Phi_N(j, Y)$). It follows from Theorem 21.3 that $\Phi_N(j, Y)$ and $\Phi_N(Y, j)$ have the same degree, since in both cases, for any lattice $L \subseteq \mathbb{C}$, the number of roots of $\Phi_N(j(L), Y)$ and $\Phi_N(Y, j(L))$ when counted with multiplicity is the number of cyclic sublattices of index N in L (which does not depend on the choice of L).[3] It follows that $\Phi_N(Y, j) = f(j)\Phi_N(j, Y)$ for some $f \in \mathbb{C}(j)$, and plugging in $Y = j$ shows that $f(j) = 1$. □

It follows that for prime N the polynomial $\Phi_N(X, Y)$ has degree $N + 1$ in X and Y.

Example 21.8. For $N = 2$ we have

$$\begin{aligned}\Phi_2(X, Y) = X^3 + Y^3 - X^2Y^2 + 1488(X^2Y + XY^2) - 162000(X^2 + Y^2)\\ + 40773375XY + 8748000000(X + Y) - 157464000000000.\end{aligned}$$

[3] Note that, per Remark 21.6, we cannot assume the j-invariants are distinct, but the cyclic sublattices are distinct; some may have the same j-invariant because distinct sublattices may be homothetic.

As can be seen in this example, the integer coefficients of Φ_N are already large when $N = 2$, and they grow rapidly as N increases. For N prime it is known that the logarithm of the absolute value of the largest coefficient of Φ_N is on the order of $6N \log N + O(N)$, see [2], and it has $O(N^2)$ coefficients. Thus the total number of bits required to write down Φ_N is quasi-cubic in N; in practical terms, Φ_{1009} is about 4 GB, and Φ_{10007} is about 5 TB. This makes it quite challenging to compute these polynomials; you will explore an efficient method for doing so on Problem Set 12.

21.2 Modular curves as moduli spaces

In the same way that the j-function defines a bijection from $Y(1) = \mathbb{H}/\Gamma(1)$ to $\mathbb{C}$ (which we may regard as an affine curve in $\mathbb{C}^2$), the functions $j(\tau)$ and $j_N(\tau)$ define a bijection from $Y_0(N) = \mathbb{H}/\Gamma_0(N)$ to the affine curve $\Phi_N(X, Y) = 0$ via the map

$$\tau \mapsto (j(\tau), j_N(\tau)).$$

If $\{\gamma_k\}$ is a set of right coset representatives for $\Gamma_0(N)$ then for each γ_k we have

$$\gamma_k\tau \mapsto (j(\gamma_k\tau), j_N(\gamma_k\tau)) = (j(\tau), j_N(\gamma_k\tau)),$$

and as in the proof of Theorem 21.3, each of these points corresponds to a cyclic N-isogeny $E \to E'$ with $j(E) = j(\tau)$ and $j(E') = j_N(\gamma_k\tau))$. We can thus view the modular curve $Y_0(N)$, equivalently, the non-cuspidal points on $X_0(N)$, as parameterizing cyclic N-isogenies.

As noted above such an isogeny is not always uniquely determined by a pair of j-invariants (these correspond to singular points on the curve $\Phi_N(X, Y) = 0$), but a cyclic N-isogeny $\phi\colon E \to E'$ is uniquely determined by the pair $(E, \langle P\rangle)$, where P is any generator for $\ker \phi$ (so P is a point of order N). Recall from Theorem 6.10 that every finite subgroup of points on an elliptic curve determines a separable isogeny that is unique up to isomorphism. Every pair $(E, \langle P\rangle)$ thus corresponds to a non-cuspidal point of $X_0(N)$; two pairs $(E, \langle P\rangle)$ and $(E', \langle P'\rangle)$ correspond to the same point if and only if there exists an isomorphism $\varphi\colon E \xrightarrow{\sim} E'$ such that $\varphi(\langle P\rangle) = \langle P'\rangle$.

With this interpretation the modular curve $X_0(N)$ can be viewed as the "moduli space" of cyclic N-isogenies of elliptic curves, each identified by a pair $(E, \langle P\rangle)$, up to the isomorphism defined above. We won't formally define the notion of a moduli space in this course, but this can be done, and it provides an alternative definition of $X_0(N)$. The key point from our perspective is that this moduli interpretation is valid over any field, not just $\mathbb{C}$. The modular curves $X_0(N)$ play a key role in many algorithms that work with elliptic curves over finite fields, including the Schoof-Elkies-Atkin (SEA) point-counting algorithm (a faster version of Schoof's algorithm), and fast algorithms to compute Hilbert class polynomials, which are the key to the CM method that we will discuss in the next lecture.

Other modular curves also have characterizations as moduli spaces. We have already seen that the modular curve $X(1)$ is the moduli space of isomorphism classes of elliptic curves, and for $N > 1$ the modular curve $X(N)$ is the moduli space of triples (E, P_1, P_2), where $\{P_1, P_2\}$ is a basis for the N-torsion subgroup of E, and the modular curve $X_1(N)$ is the moduli space of pairs (E, P), where P is a point of order N on E. Note that in each case one considers triples or pairs only up to a suitable isomorphism, as with $X_0(N)$ above.

21.3 The Hilbert class polynomial

We now turn our attention to the Hilbert class polynomial. Recall that for each imaginary quadratic order $\mathcal{O}$, we have the set

$$\mathrm{Ell}_{\mathcal{O}}(\mathbb{C}) := \{j(E) \in \mathbb{C} : \mathrm{End}(E) \simeq \mathcal{O}\}$$

of equivalence classes of elliptic curves with complex multiplication (CM) by $\mathcal{O}$, and the ideal class group $\mathrm{cl}(\mathcal{O})$ acts on $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$ via isogenies, as we now recall. Every elliptic curve $E/\mathbb{C}$ with CM by $\mathcal{O}$ is of the form $E_{\mathfrak{b}}$ corresponding to the torus $\mathbb{C}/\mathfrak{b}$, where $\mathfrak{b}$ is a proper $\mathcal{O}$-ideal for which $j(\mathfrak{b}) = j(E)$ (note that $j(\mathfrak{b}) = j(E)$ depends only on the class $[\mathfrak{b}]$ in $\mathrm{cl}(\mathcal{O})$). If $[\mathfrak{a}]$ is an element of $\mathrm{cl}(\mathcal{O})$, then $\mathfrak{a}$ acts on $E_{\mathfrak{b}}$ by the isogeny

$$\phi_{\mathfrak{a}} \colon E_{\mathfrak{b}} \to E_{\mathfrak{a}^{-1}\mathfrak{b}}$$

of degree $\mathrm{N}\mathfrak{a}$ induced by the lattice inclusion $\mathfrak{b} \subseteq \mathfrak{a}^{-1}\mathfrak{b}$. As with $E_{\mathfrak{b}}$, the isomorphism class of $E_{\mathfrak{a}^{-1}\mathfrak{b}}$ depends only on the class $[\mathfrak{a}^{-1}\mathfrak{b}]$ in $\mathrm{cl}(\mathcal{O})$, and we proved that this action is free and transitive, meaning that $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$ is a $\mathrm{cl}(\mathcal{O})$-torsor. This implies that the set $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$ is finite, with cardinality equal to the class number $h(\mathcal{O}) := \#\,\mathrm{cl}(\mathcal{O})$.

We may uniquely identify $\mathcal{O}$ by its discriminant D (by Theorem 18.17), and the Hilbert class polynomial

$$H_D(X) = \prod_{j(E)\in \mathrm{Ell}_{\mathcal{O}}(\mathbb{C})} (X - j(E))$$

is the monic polynomial whose roots are the distinct j-invariants of all elliptic curves with CM by $\mathcal{O}$. We now want to use the fact that $\Phi_N \in \mathbb{Z}[X,Y]$ to prove that $H_D \in \mathbb{Z}[X]$. To do this we need the following lemma.

Lemma 21.9. *If N is prime then the leading coefficient of $\Phi_N(X,X)$ is -1.*

Proof. Replacing Y with $j(\tau)$ in equation (1) for $\Phi_N(Y)$ yields

$$\Phi_N(j(\tau), j(\tau)) = \big(j(\tau) - j(N\tau)\big) \prod_{k=0}^{N-1} \big(j(\tau) - j(\tfrac{\tau+k}{N})\big).$$

Recall from the proof of Theorem 20.17 that we have the q-expansions

$$j(N\tau) = \frac{1}{q^N} + \cdots\,,$$
$$j(\frac{\tau+k}{N}) = \frac{\zeta_N^{-k}}{q^{1/N}} + \cdots\,,$$

where $q := e^{2\pi i\tau}$, $\zeta_N := e^{2\pi i/N}$, and ellipses denotes terms involving larger powers of q. Thus

$$j(\tau) - j(N\tau) = -\frac{1}{q^N} + \frac{1}{q} + \cdots\,,$$
$$j(\tau) - j(\frac{\tau+k}{N}) = \frac{1}{q} - \frac{\zeta_N^{-k}}{q^{1/N}} + \cdots\,,$$

which implies that the q-expansion of $f(\tau) = \Phi_N(j(\tau), j(\tau))$ begins $-\frac{1}{q^{2N}} + \cdots$. Since $f(\tau)$ is a polynomial in $j(\tau) = \frac{1}{q} + \cdots$, the leading term of $\Phi_N(X,X)$ must be $-X^{2N}$. □

Remark 21.10. Lemma 21.9 does not hold in general; in particular, when N is square $\Phi_N(X, X)$ is not even primitive (its coefficients have a non-trivial common divisor).

Before proving $H_D \in \mathbb{Z}[X]$, we record the following classical result, which was proved for maximal orders by Dirichlet and later generalized by Weber; see [3, p. 190]. Today this is typically cited as a consequence of the Chebotarev[4] density theorem, but since the proof of the Chebotarev density theorem actually uses class field theory, a small part of which we are about to prove, we should note that the result we need was proved earlier.

Theorem 21.11. *Let $\mathcal{O}$ be an imaginary quadratic order. Every ideal class in* $\operatorname{cl}(\mathcal{O})$ *contains infinitely many ideals of prime norm.*

Proof. This follows from Theorems 7.7 and 9.12 in [3]. □

Theorem 21.12. *The coefficients of the Hilbert class polynomial $H_D(X)$ are integers.*

Proof. Let $\mathcal{O}$ be the imaginary quadratic order of discriminant D, let $E/\mathbb{C}$ be an elliptic curve with CM by $\mathcal{O}$, and let $\mathfrak{p}$ be a principal $\mathcal{O}$-ideal of prime norm p (by Theorem 21.11 there are infinitely many choices for $\mathfrak{p}$). Then $[\mathfrak{p}]$ is the identity element of $\operatorname{cl}(\mathcal{O})$, so $\mathfrak{p}$ acts trivially on $\operatorname{Ell}_{\mathcal{O}}(\mathbb{C})$. Thus $\mathfrak{p}E \simeq E$, which implies that, after composing with an isomorphism if necessary, we have a p-isogeny from E to itself, equivalently, an endomorphism of degree p. Such an isogeny is necessarily cyclic, since it has prime degree, so we must have $\Phi_p(j(E), j(E)) = 0$. Thus $j(E)$ is the root of the polynomial $-\Phi_p(X, X)$, which is monic, by Lemma 21.9, and has integer coefficients, by Theorem 20.17. The j-invariant $j(E)$ is thus an algebraic integer, and the elliptic curve E can be defined by a Weierstrass equation $y^2 = x^3 + Ax + B$ whose coefficients lie in the number field $\mathbb{Q}(j(E))$, by Theorem 14.12.

The absolute Galois group $\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on the set of elliptic curves defined over number fields via its action on the Weierstrass coefficients A and B: for each field automorphism $\sigma \in \operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ the curve E^σ is defined by the equation $y^2 = x^3 + \sigma(A)x + \sigma(B)$. Similarly, σ acts on isogenies via its action on the coefficients of the rational map defining the isogeny. If $\phi\colon E \to E$ is an endomorphism, then so is $\phi^\sigma\colon E^\sigma \to E^\sigma$, and for any $\phi, \psi \in \operatorname{End}(E)$ we have $(\phi + \psi)^\sigma = \phi^\sigma + \psi^\sigma$ and $(\phi \circ \psi)^\sigma = \phi^\sigma \circ \psi^\sigma$. Thus each $\sigma \in \operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ induces a ring homomorphism

$$\operatorname{End}(E) \xrightarrow{\sigma} \operatorname{End}(E^\sigma).$$

Applying σ^{-1} to E^σ induces an inverse homomorphism, we thus have a ring isomorphism $\operatorname{End}(E) \simeq \operatorname{End}(E^\sigma)$, which implies that E^σ also has CM by $\mathcal{O}$.

The j-invariant of E is a rational function $1728 \cdot 4A^3/(4A^3 + 1728B^2)$ of A and B, so $j(E^\sigma) = j(E)^\sigma$, and we have shown that $j(E^\sigma) \in \operatorname{Ell}_{\mathcal{O}}(\mathbb{C})$. It follows that $\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on the set $\operatorname{Ell}_{\mathcal{O}}(\mathbb{C})$, which are precisely the roots of $H_D(X)$. The coefficients of $H_D(X)$ are all symmetric polynomials in the roots, hence they are fixed by $\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and therefore lie in the fixed field $\mathbb{Q}$. Every root of $H_D(X)$ is a root of $\Phi_p(X, X)$, thus $H_D(X)$ divides $\Phi_p(X, X)$ in $\mathbb{Q}[X]$. But $\Phi_p(X, X)$ has integer coefficients, and it is monic (hence primitive), so by Gauss's lemma [1, §12.3], its factors in $\mathbb{Q}[X]$ are the same as its factors in $\mathbb{Z}[X]$; therefore $H_D \in \mathbb{Z}[X]$ as claimed. □

Corollary 21.13. *Let $E/\mathbb{C}$ be an elliptic curve with complex multiplication. Then $j(E)$ is an algebraic integer.*

[4] Many different transliterations of Chebotarev's Russian name appear in the literature, including Chebotaryov, Čebotarev, Chebotarëv, Čhebotarëv, Tchebotarev, and Tschebotaröw; none is universally accepted.

From the proof of Theorem 21.12, we now have two groups acting on the roots of $H_D(X)$: the class group $\mathrm{cl}(\mathcal{O})$ and the Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. In the latter case there is no need to consider the entire Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, we can always restrict our attention to any Galois subfield $L \subseteq \overline{\mathbb{Q}}$ that contains the splitting field L of $H_D(X)$, since the action of any $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on the roots of $H_D(X)$ is determined by its restriction to $\mathrm{Gal}(L/\mathbb{Q})$. We then have two finite group actions, and it is reasonable to ask whether they are in some sense compatible.

In order to obtain compatible actions we do not want to work with the splitting field L of $H_D(X)$ over $\mathbb{Q}$, since $\mathrm{Gal}(L/\mathbb{Q})$, may contain automorphisms that don't fix the order $\mathcal{O}$. but if we instead let L be the splitting field of $H_D(X)$ over $K := \mathbb{Q}\sqrt{D})$, the Galois group $\mathrm{Gal}(L/K)$ fixes $\mathcal{O}$, and we will show that its action on $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$ is compatible with that of the class group $\mathrm{cl}(\mathcal{O})$. In fact, $\mathrm{Gal}(L/K) \simeq \mathrm{cl}(\mathcal{O})$. This isomorphism is part of the *First Main Theorem of Complex Multiplication*, and our next goal is to prove it.

So let $\mathcal{O}$ be the imaginary quadratic order of discriminant D, and let us fix an elliptic curve E_1 with CM by $\mathcal{O}$. Each $\sigma \in \mathrm{Gal}(L/K)$ can be viewed as the restriction to L of an element of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ that fixes K, thus as in the proof of Theorem 21.12, the elliptic curve E_1^σ also has CM by $\mathcal{O}$. Therefore $E_1^\sigma \simeq \mathfrak{a}E_1$ for some proper $\mathcal{O}$-ideal $\mathfrak{a}$, since $\mathrm{cl}(\mathcal{O})$ acts transitively on $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$. If $E_2 \simeq \mathfrak{b}E_1$ is any other elliptic curve with CM by $\mathcal{O}$, we then have

$$E_2^\sigma \simeq (\mathfrak{b}E_1)^\sigma = \mathfrak{b}^\sigma E_1^\sigma = \mathfrak{b}E_1^\sigma \simeq \mathfrak{b}\mathfrak{a}E_1 = \mathfrak{a}\mathfrak{b}E_1 \simeq \mathfrak{a}E_2. \tag{2}$$

The innocent looking identity $(\mathfrak{b}E_1)^\sigma = \mathfrak{b}^\sigma E_1^\sigma$ used in (2) is not immediate, it requires a somewhat lengthy argument involving a diagram chase that we omit; see [9, Prop. II.2.5] for a proof. The second identity *is* immediate, because $\mathfrak{b} \subset K$ and $\sigma \in \mathrm{Gal}(L/K)$ fixes K; but note that this would not be true if we had instead used $\sigma \in \mathrm{Gal}(L/\mathbb{Q})$.

Since our choice of E_2 was arbitrary, it follows from (2) that the action of σ on $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$ is the same as the action of $\mathfrak{a}$ on $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$. Because $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$ is a $\mathrm{cl}(\mathcal{O})$-torsor, the map that sends each $\sigma \in \mathrm{Gal}(\overline{K}/K)$ to the unique class $[\mathfrak{a}] \in \mathrm{cl}(\mathcal{O})$ for which $E_1^\sigma = \mathfrak{a}E_1$ defines a group homomorphism

$$\Psi\colon \mathrm{Gal}(L/K) \to \mathrm{cl}(\mathcal{O}).$$

This homomorphism is injective because, by definition of the splitting field, the only element of $\mathrm{Gal}(L/K)$ that acts trivially on the roots of $H_D(X)$ is the identity element, and the same is true of $\mathrm{cl}(\mathcal{O})$. We summarize this discussion with the following theorem.

Theorem 21.14. *Let $\mathcal{O}$ be an imaginary quadratic order of discriminant D and let L be the splitting field of $H_D(X)$ over $K := \mathbb{Q}(\sqrt{D})$. The map $\Psi : \mathrm{Gal}(L/K) \to \mathrm{cl}(D)$ that sends each $\sigma \in \mathrm{Gal}(L/K)$ to the unique $\alpha_\sigma \in \mathrm{cl}(\mathcal{O})$ for which $j(E)^\sigma = \alpha_\sigma j(E)$ for all $j(E) \in \mathrm{Ell}_{\mathcal{O}}(E)$ is an injective group homomorphism.*

We thus have an embedding of $\mathrm{Gal}(L/K)$ in $\mathrm{cl}(\mathcal{O})$ that is compatible with the actions of both groups on $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$. It remains only to prove that Ψ is surjective, which is equivalent to proving that $H_D(X)$ is irreducible over K. To do this we need to introduce the Artin map (named after Emil Artin), which allows us to associate to each $\mathcal{O}$-ideal $\mathfrak{p}$ of prime norm satisfying certain constraints an automorphism $\sigma_{\mathfrak{p}} \in \mathrm{Gal}(L/K)$ whose action on $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$ corresponds to the action of $[\mathfrak{p}]$. In order to define the Artin map we need to briefly delve into a bit of algebraic number theory. We will restrict our attention to the absolute minimum that we need. Those who would like to know more may wish to consult one of [6, 7] (and/or take 18.785 in the fall); those who do not may treat the Artin map as a black box.

21.4 The Artin map

Let L be a finite Galois extension of a number field K. The nonzero prime ideals $\mathfrak{p}$ in the ring of integers $\mathcal{O}_K$ are called "primes of K".[5] The $\mathcal{O}_L$-ideal $\mathfrak{p}\mathcal{O}_L$ is typically not a prime ideal, but it can be uniquely factored as

$$\mathfrak{p}\mathcal{O}_L = \mathfrak{q}_1 \cdots \mathfrak{q}_n$$

where the $\mathfrak{q}_i$ are not-necessarily-distinct primes of L (prime ideals of $\mathcal{O}_L$) that are characterized by the property $\mathfrak{q}_i \cap \mathcal{O}_K = \mathfrak{p}$. The primes $\mathfrak{q}_i$ are said to "lie above" the prime $\mathfrak{p}$, and it is standard to write $\mathfrak{q}_i|\mathfrak{p}$ as shorthand for $\mathfrak{q}_i|\mathfrak{p}\mathcal{O}_L$ and use $\{\mathfrak{q}|\mathfrak{p}\}$ to denote the set $\{\mathfrak{q}_1, \ldots, \mathfrak{q}_n\}$.

We should note that the ring $\mathcal{O}_L$ is typically *not* a unique factorization domain, but it is a *Dedekind domain*, and this implies unique factorization of ideals.[6]

When the $\mathfrak{q}_i$ are distinct, we say that $\mathfrak{p}$ is *unramified* in L, which is true of all but finitely many primes $\mathfrak{p}$. If we apply an automorphism $\sigma \in \mathrm{Gal}(L/K)$ to both sides of the equation above, the LHS must remain the same: σ fixes every element of $\mathfrak{p} \subseteq K$, and it maps algebraic integers to algebraic integers, so it preserves the set $\mathcal{O}_L$. For the RHS, it is clear that σ must map $\mathcal{O}_L$-ideals to $\mathcal{O}_L$-ideals, and since the $\mathfrak{q}_i$ are all prime ideals, σ must permute them. Thus the Galois group $\mathrm{Gal}(L/K)$ acts on the set $\{\mathfrak{q}_1, \ldots, \mathfrak{q}_n\} = \{\mathfrak{q}|\mathfrak{p}\}$; one can show that this action is transitive, but it is typically not faithful.

For each $\mathfrak{q}|\mathfrak{p}$, the stabilizer of $\mathfrak{q}$ under this action is a subgroup

$$D_\mathfrak{q} := \{\sigma \in \mathrm{Gal}(L/K) : \mathfrak{q}^\sigma = \mathfrak{q}\} \subseteq \mathrm{Gal}(L/K)$$

known as the *decomposition group* of $\mathfrak{q}$. Each $\sigma \in D_\mathfrak{q}$ fixes $\mathfrak{q}$ and therefore induces an automorphism $\bar\sigma$ of the quotient $\mathbb{F}_\mathfrak{q} := \mathcal{O}_L/\mathfrak{q}$ defined by $\bar\sigma(\bar x) = \overline{\sigma(x)}$, where $x \mapsto \bar x$ is the quotient map $\mathcal{O}_L \to \mathcal{O}_L/\mathfrak{q}$. Note that the quotient is a field (in a Dedekind domain every nonzero prime ideal is maximal), and $\mathfrak{q}$ has finite index $\mathrm{N}\mathfrak{q} := [\mathcal{O}_L : \mathfrak{q}]$ in $\mathcal{O}_L$, so it is a finite field whose cardinality $\mathrm{N}\mathfrak{q}$(which must be a prime power. The image of $\mathcal{O}_K$ under the quotient map $\mathcal{O}_L \to \mathcal{O}_L/\mathfrak{q} = \mathbb{F}_q$ is $\mathcal{O}_K/(\mathfrak{q} \cap \mathcal{O}_K) = \mathcal{O}_K/p = \mathbb{F}_\mathfrak{p}$, thus the finite field $\mathbb{F}_\mathfrak{p}$ is a subfield of $\mathbb{F}_\mathfrak{q}$ (and necessarily has the same characteristic). It follows that $\bar\sigma \in \mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p})$, and we have a group homomorphism

$$\begin{aligned} D_q &\to \mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p}) \\ \sigma &\mapsto \bar\sigma. \end{aligned}$$

This homomorphism is surjective [7, Prop. I.9.4], and when $\mathfrak{p}$ is unramified it is also injective [7, Prop. I.9.5], and therefore an isomorphism, which we now assume.

The group $\mathrm{Gal}(\mathbb{F}_\mathfrak{q}/\mathbb{F}_\mathfrak{p})$ is cyclic, generated by the Frobenius automorphism $x \to x^{\mathrm{N}\mathfrak{p}}$, where $\mathrm{N}\mathfrak{p} = [\mathcal{O}_K : \mathfrak{p}] = \#\mathbb{F}_\mathfrak{p}$. The unique $\sigma_\mathfrak{q} \in D_\mathfrak{q}$ for which $\bar\sigma_\mathfrak{q}$ is the Frobenius automorphism is called the *Frobenius element* of $\mathrm{Gal}(L/K)$ at $\mathfrak{q}$. In general the Frobenius element $\sigma_\mathfrak{q}$ depends on our choice of $\mathfrak{q}$, but the $\sigma_\mathfrak{q}$ for $\mathfrak{q}|\mathfrak{p}$ are all conjugate, since if $\tau(\mathfrak{q}_i) = \mathfrak{q}_j$ then we must have $\sigma_{\mathfrak{q}_j} = \tau^{-1}\sigma_{\mathfrak{q}_i}\tau$.

In the case we are interested in, $\mathrm{Gal}(L/K) \hookrightarrow \mathrm{cl}(\mathcal{O})$ is abelian, so conjugacy implies equality, and the $\sigma_\mathfrak{q}$ are all the same. Thus when $\mathrm{Gal}(L/K)$ is abelian, each prime $\mathfrak{p}$ of K

[5] This is an abuse of terminology: as a ring, K does not have any nonzero prime ideals (it is a field).

[6] There are several equivalent definitions of Dedekind domains: it is an integral domain with unique factorization of ideals, and it also an integral domain in which every nonzero fractional ideal is invertible. We have seen that the latter applies to rings of integers in number fields (at least for imaginary quadratic fields), so the former must as well (this equivalence is a standard result from commutative algebra).

determines a unique Frobenius element that we denote $\sigma_{\mathfrak{p}}$. The map

$$\mathfrak{p} \mapsto \sigma_{\mathfrak{p}}$$

is known as the *Artin map* (it extends multiplicatively to all $\mathcal{O}_K$-ideals, but this is not relevant to us). The automorphism $\sigma_{\mathfrak{p}}$ is uniquely characterized by the fact that

$$\sigma_{\mathfrak{p}}(x) \equiv x^{\mathrm{N}\mathfrak{p}} \bmod \mathfrak{q}, \tag{3}$$

for all $x \in \mathcal{O}_L$ and primes $\mathfrak{q}|\mathfrak{p}$.

In the next lecture we will use the Artin map to prove that $\Psi\colon \mathrm{Gal}(L/K) \to \mathrm{cl}(\mathcal{O})$ is surjective, hence an isomorphism.

References

[1] Michael Artin, *Algebra*, second edition, Pearson, 2011.

[2] Paula Cohen, *On the coefficients of the transformation polynomials for the elliptic modular function*, Mathematical Proceedings of the Cambridge Philosophical Society **95** (1984), 389–402.

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

[4] Jun-Ichi Igusa, *Kroneckerian Model of Fields of Elliptic Modular Functions*, American Journal of Mathematics **81** (1959), 561–577.

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

[6] J. S. Milne, *Algebraic number theory*, course notes, 2014.

[7] Jürgen Neukirch, *Algebraic number theory*, Springer, 1999.

[8] Joseph H. Silverman, *The arithmetic of elliptic curves*, second edition, Springer, 2009.

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

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