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 $j_N(\tau) := j(N\tau)$, and we showed that $\mathbb{C}(j, j_N)$ is a finite extension of $\mathbb{C}(j)$. We then defined the *classical modular polynomial* $\Phi_N(Y)$ as the minimal polynomial of j_N over $\mathbb{C}(j)$, and we proved that its coefficients are integer polynomials in j. Replacing j with a new variable X, we can view $\Phi_N\mathbb{Z}[X, Y]$ as an integer polynomial in two variables.

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

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

also has integer coefficients; here $D = \mathrm{disc}(\mathcal{O})$ and $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C}) := \{j(E) : \mathrm{End}(E) \simeq \mathcal{O}\}$ is the finite set of j-invariants of elliptic curves $E/\mathbb{C}$ with complex multiplication (CM) by $\mathcal{O}$. This implies that each $j(E) \in \mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$ is an algebraic integer, meaning that any elliptic curve $E/\mathbb{C}$ with complex multiplication can actually be defined over a finite extension of $\mathbb{Q}$ (a number field). This fact is the key to relating the theory of elliptic curves over the complex numbers to elliptic curves over finite fields.

21.1 Isogenies

Recall from Lecture 18 that if L_1 is a sublattice of L_2, and $E_1 \simeq \mathbb{C}/L_1$ and $E_2 \simeq \mathbb{C}/L_2$ are the corresponding elliptic curves, then there is 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

$$\begin{array}{ccc} \mathbb{C}/L_1 & \longrightarrow^{\iota} & \mathbb{C}/L_2 \\ \downarrow\wr & & \downarrow\wr \\ E_1(\mathbb{C}) & \longrightarrow^{\phi} & E_2(\mathbb{C}) \end{array}$$

where the top map is induced by the inclusion $L_1 \subseteq L_2$ (lift from $\mathbb{C}/L_1$ to $\mathbb{C}$ then quotient by the finer lattice L_2). The relationship between $E_1(\mathbb{C})$ and $E_2(\mathbb{C})$ is symmetric, since if we replace L_2 by the homothetic lattice NL_2, where $N = [L_2 : L_1] = \deg\phi$, then NL_2 is a sublattice of L_1 and we obtain the dual isogeny $\hat{\phi}\colon E_2 \to E_1$ (the elliptic curves corresponding to $\mathbb{C}/L_2$ and $\mathbb{C}/NL_2$ are both isomorphic to E_2). The composition $\phi \circ \hat{\phi}$ is the multiplication-by-N map on E_2, induced by the lattice inclusion $NL_2 \subseteq L_2$ and has kernel $L_2/NL_2 \simeq \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$, and $\hat{\phi} \circ \phi$ is the multiplication-by-N map on E_1.

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 called a *cyclic isogeny* if its kernel is a cyclic group. If ϕ is induced by the lattice inclusion $L_1 \subseteq L_2$ then ϕ is a cyclic isogeny if and only if L_1 is a cyclic sublattice.

Cyclic isogenies are of particular interest because they are effectively parameterized by the modular polynomial Φ_N; we will prove this only for prime N, but it holds in general. We begin by describing the cyclic sublattices of prime index in a given lattice.

Lemma 21.2. *Let $L = [1, \tau]$ be a lattice with $\tau \in \mathbb{H}$. The cyclic sublattices of L with prime 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 equal to ad. If $[L : L'] = N$ is prime, then 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 $0 \le k < N$. □

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 $\mathbb{C}$ that are related by a cyclic isogeny of degree N.*

Proof for N prime. We will prove the equivalent statement that $\Phi_N\big(j(L_1), j(L_2)\big) = 0$ if and only if L_2 is homothetic to a cyclic sublattice of L_1 with index N. We may assume without loss of generality that $L_1 = [1, \tau_1]$ and $L_2 = [1, \tau_2]$, where $\tau_1, \tau_2 \in \mathbb{H}$. With $\gamma_k = ST^k$ as in the proof of Theorem 20.13, we have

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

where

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

Thus

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

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

Remark 21.4. 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 still holds 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 21.5. 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. As an example, suppose $\mathrm{End}(E_1) \simeq \mathcal{O}$ and $\mathfrak{p}$ is an unramified proper $\mathcal{O}$-ideal of prime norm p such that $[\mathfrak{p}]$ has order 2 in the class group $\mathrm{cl}(\mathcal{O})$. Then $\mathfrak{p}E_1 \simeq \bar{\mathfrak{p}}E_1$, and there are two distinct p-isogenies from E_1 to $E_2 = \mathfrak{p}E_1$.[1] These isogenies are not isomorphic (there is no automorphism we can compose with one to get the other). In this situation the polynomial $\Phi_p(j(E_1), Y)$ will have $j(E_2)$ as a double root.

Corollary 21.6. $\Phi_N(X, Y) = \Phi_N(Y, X)$

[1] Recall that if $E_1 \simeq \mathbb{C}/L$ then $\mathfrak{p}E_1$ denotes the elliptic curve $E_2 \simeq \mathbb{C}/\mathfrak{p}^{-1}L$, see Lecture 19.

Proof. The function j_N is a root of $\Phi_N(j,Y)$ in $\mathbb{C}(j,j_N)$, and it is also a root of the polynomial $\Phi_N(Y,j)$; this follows from the existence of the dual isogeny. By definition, the polynomial $\Phi_N(Y) = \Phi_N(j,Y)$ is irreducible over $\mathbb{C}(j)$, hence over $\mathbb{Q}(j)$, which means that $\Phi_N(j,Y)$ must divide $\Phi_N(Y,j)$ in $\mathbb{Q}[j,Y]$. But the theorem and its proof imply that $\Phi_N(Y,j)$ and $\Phi_N(j,Y)$ must have the same degree. When N is prime, for example, there are exactly $N+1$ cyclic sub-lattices L' of index N in any lattice L, and L is a cyclic sublattice of index N of exactly $N+1$ lattices, namely, the lattices $\frac{1}{N}L'$. Thus the number of roots of the two polynomials is the same when counted with multiplicity (per the remark above, we do not assume that these lattices all have distinct j-invariants)

It follows that $\Phi_N(j,Y)$ and $\Phi_N(Y,j)$ can differ only by a nonzero scalar multiple λ. If we plug in the j-function for Y we then have $\Phi_N(j,j) = \lambda\Phi_N(j,j)$, and if λ is not 1 this implies $\Phi_N(j,j) = 0$; but this is impossible because $\Phi_N(j,Y)$ is irreducible over $\mathbb{Q}(j)$ and cannot have j as a root. □

It follows from the corollary that when N is prime $\Phi_N(X,Y)$ has degree $N+1$ in both X and Y.

Example 21.7. For $N = 2$ we have

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

As can be seen in the 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 $6N \log N + O(N)$ [2], and as we have seen 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 the size of Φ_1009 is about 4 gigabytes, and Φ_{10007} is about 5 terabytes. 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 embedded in $\mathbb{C}^2$), functions $j(\tau)$ and $j_N(\tau)$ define a bijection from $Y_0(N) = \mathbb{H}/\Gamma_0(N)$ to the affine curve defined by $\Phi_N(X,Y) = 0$ via the map

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

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

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

thus there is a point on the curve $\Phi_N(X,Y) = 0$ corresponding teach cyclic N-isogeny $E \to E'$ with $j(E) = j(\tau)$. Thus we can view the modular curve $Y_0(N)$ (equivalently, the non-cuspidal points on $X_0(N)$) as parameterizing cyclic isogenies of degree N. As noted above such an isogeny is not always uniquely determined by a pair of j-invariants, but each is uniquely determined by a pair $(E, \langle P\rangle)$, where P is a point of order N on $E(\mathbb{C})$ and $\langle P\rangle$ is the cyclic subgroup it generates. Recall from Theorem 6.8 that every finite subgroup of points on an elliptic curve determines a separable isogeny that is unique up to isomorphism, thus there is a one-to-one correspondence between pairs $(E, \langle P\rangle)$ and the

non-cuspidal points of $X_0(N)$; note that this point depends only on the group $\langle P\rangle$, not the choice of the generator P.

One then says that the modular curve $X_0(N)$ corresponds to the "moduli space" of cyclic N-isogenies of elliptic curves, each identified by a pair $(E, \langle P\rangle)$, up to isomorphism. 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 $\mathbb{C}$, and the modular curves $X_0(N)$ actually 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.

The other modular curves we have defined 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 in general 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. The modular curve $X_1(N)$ is the moduli space of pairs (E, P), where P is a point of order N on E.[2]

21.3 The Hilbert class polynomial

We now turn our attention to the Hilbert class polynomial introduced in Lecture 18. 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

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

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 Lemma 17.12), and the Hilbert class polynomial

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

is then defined as the monic polynomial whose roots are precisely the j-invariants of the 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 21.8. *If N is prime then the leading coefficient of $\Phi_N(X, X)$ is -1.*

[2] One needs to define a suitable notion of isomorphism in each case, for example, we don't distinguish isomorphic elliptic curves, but we do distinguish different choices of the point P or the points P_1 and P_2.

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

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

Recall from the proof of Theorem 20.13 that we have the q-expansions

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

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

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

which implies that the q-expansion of $f(\tau) = \Phi_N\big(j(\tau),j(\tau)\big)$ is $-\frac{1}{q^{2N}} + \cdots$. 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 21.9. Lemma 21.8 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[3] 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), it is important to note that that the result we need was known earlier.

Theorem 21.10. *Let $\mathcal{O}$ be an imaginary quadratic order. Every ideal class in* $\mathrm{cl}(\mathcal{O})$ *contains infinitely many ideals of prime norm.*

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

Theorem 21.11. *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 wih CM by $\mathcal{O}$, and let $\mathfrak{p}$ be a principal $\mathcal{O}$-ideal of prime norm p (the existence of $\mathfrak{p}$ is guaranteed by Theorem 21.10). Then $[\mathfrak{p}]$ is the identity element of $\mathrm{cl}(\mathcal{O})$, so $\mathfrak{p}$ acts trivially on $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$. 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 $\Phi_p\big(j(E),j(E)\big) = 0$. Thus $j(E)$ is the root of the polynomial $-\Phi_p(X,X)$, which has integer coefficients and is also monic, by Lemma 21.8. Therefore $j(E)$ is 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 $\mathbb{Q}(j(E))$, a finite extension of $\mathbb{Q}$.

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

The absolute Galois group $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ acts on elliptic curves defined over number fields via its action on the Weierstrass coefficients A and B: for each field automorphism $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ the curve E^σ is defined by the equation $y^2 = x^3 + \sigma(A)x + \sigma(B)$. Similarly, σ acts on isogenies between such curves 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 $\phi, \psi \in \mathrm{End}(E)$ we have $(\phi+\psi)^\sigma = \phi^\sigma + \psi^\sigma$ and $(\phi\circ\psi)^\sigma = \phi^\sigma \circ \psi^\sigma$. Thus each $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ induces a ring homomorphism

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

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

The j-invariant of E is a rational function of the Weierstrass coefficients A and B, so $j(E^\sigma) = j(E)^\sigma$, and we have shown that $j(E^\sigma) \in \mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$. It follows that each element of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ permutes the elements of $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$, which are the roots of $H_D(X)$. The coefficients of $H_D(X)$ are all symmetric polynomials in the roots, hence they are fixed by $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ 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 primitive by Lemma 21.8, 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 $H_D \in \mathbb{Z}[X]$. □

Corollary 21.12. *Let $E/\mathbb{C}$ be an elliptic curve with complex multiplication. Then $j(E)$ is an algebraic integer.*

From the proof of Theorem 21.11, 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 may as well restrict our attention to automorphisms of 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 L. We then have two finite group actions, and it is reasonable to ask whether they are in some sense compatible. In order for this to be true, we do not want to work with $\mathrm{Gal}(L/\mathbb{Q})$, since this Galois group may contain automorphisms that don't fix the order $\mathcal{O}$. But if we restrict our attention to the subgroup $\mathrm{Gal}(L/K)$ of automorphisms that fix $K = \mathbb{Q}(\sqrt{D})$ (and hence the order $\mathcal{O}$) then the group actions are indeed compatible. 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 $\mathcal{O}$. As in the proof of Theoerem 21.11, for each $\sigma \in \mathrm{Gal}(\bar{K}/K)$, the elliptic curve E_1^σ also has CM by $\mathcal{O}$, and therefore $E_1^\sigma \simeq \mathfrak{a}E_1$ for some proper $\mathcal{O}$-ideal $\mathfrak{a}$ (because $\mathrm{cl}(\mathcal{O})$ 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 [7, 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 $\sigma \in \mathrm{Gal}(\bar{K}/K)$ 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 21.13. *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 $\alpha \in \mathrm{cl}(\mathcal{O})$ for which $j(E)^\sigma = \alpha j(E)$ for all $j(E) \in \mathrm{Ell}_{\mathcal{O}}(E)$ is an injective group homomorphism.*

We have have an embedding of $\mathrm{Gal}(L/K)$ in $\mathrm{cl}(\mathcal{O})$ that is compatible with both group actions 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 will allow 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 [5] and/or [6] (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".[4] 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. Note: 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.[5]

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 $\mathfrak{p}$; henceforth we assume $\mathfrak{p}$ is unramified. 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 $\{\mathfrak{q}_i\}$, and one can show that this action is transitive, but it is typically not faithful.

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

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

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

[5] There are several equivalent definitions of Dedekind domains: one is an integral domain with unique factorization of ideals, and another is 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).

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 $\mathcal{O}_L/\mathfrak{q}$. This quotient is a field (in a Dedekind domain every nonzero prime ideal is maximal), and $\mathfrak{q}$ has finite index $N\mathfrak{q} := [\mathcal{O}_L : \mathfrak{q}]$, so it in fact a finite field $\mathbb{F}_\mathfrak{q} := \mathcal{O}_L/\mathfrak{q}$ of cardinality $N\mathfrak{q}$ (which must be a prime power). The image of $\mathcal{O}_K$ under the quotient map $\mathcal{O}_L \to \mathbb{F}_q$ is $\mathcal{O}_K/(\mathfrak{q} \cap \mathcal{O}_K)$. The intersection $\mathfrak{q} \cap \mathcal{O}_K$ clearly contains $\mathfrak{p}$, and it is not equal to $\mathcal{O}_K$ (because it does not contain 1), so it must be equal to $\mathfrak{p}$ (because $\mathfrak{p}$ is maximal); thus $\mathbb{F}_\mathfrak{p} := \mathcal{O}_K/\mathfrak{p}$ is a subfield of $\mathbb{F}_\mathbb{Q}$. 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 [6, Prop. I.9.4], and our assumption that $\mathfrak{p}$ is unramified means that it is also injective [6, Prop. I.9.5], and therefore an isomorphism.

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 various $\sigma_\mathfrak{q}$ are all conjugate: if $\tau(\mathfrak{q}) = \mathfrak{q}'$ then $\sigma_{\mathfrak{q}'} = \tau^{-1}\sigma_\mathfrak{q}\tau$.

In the situation we are interested in, $\mathrm{Gal}(L/K) \hookrightarrow \mathrm{cl}(\mathcal{O})$ is abelian, so the $\sigma_\mathfrak{q}$ must all be equal. Thus when $\mathrm{Gal}(L/K)$ is abelian, each prime $\mathfrak{p}$ of K 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 $\mathfrak{q}$ that divide $\mathfrak{p}\mathcal{O}_L$.

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

[5] J.S. Milne, *Algebraic Number Theory*, course notes, 2014.

[6] Jürgen Neukirch, *Algebraic Number Theory*, Springer, 1999.

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

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