20.1 The Hilbert class polynomial

Let $\mathcal{O}$ be an order of discriminant D in an imaginary quadratic field K. In Lecture 19 we saw that there is a one-to-one relationship between isomorphism classes of elliptic curves with complex multiplication by $\mathcal{O}$ (the set $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$), and equivalence classes of proper $\mathcal{O}$-ideals (the group $\mathrm{cl}(\mathcal{O})$). The first main theorem of complex multiplication states that the elements of $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$ are algebraic integers that all have the same minimal polynomial

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

known as the *Hilbert class polynomial* (of discriminant D).[1] Moreover, it states that the splitting field $K_{\mathcal{O}}$ of $H_D(X)$ over K has Galois group isomorphic to $\mathrm{cl}(\mathcal{O})$. The roots of $H_D(X)$ are precisely the elements of $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$, and the action of the Galois group $\mathrm{Gal}(K_{\mathcal{O}}/K)$ is precisely the $\mathrm{cl}(\mathcal{O})$-action on $\mathrm{Ell}_{\mathcal{O}}(\mathbb{C})$ that we saw in Lecture 19.

The first main theorem of complex multiplication is one of the central results of what is known as *class field theory*. In order to prove it, and in order to develop efficient algorithms for explicitly computing $H_D(X)$, we need to temporarily divert our attention to the study of *modular curves*. These curves, and the *modular functions* that are defined on them, are a major topic in their own right, one to which entire courses (and research careers) are devoted. We shall only scratch the surface of this subject, focusing on the specific results that we need. Our presentation is adapted from [1, V.1] and [2, I.2].

20.2 The modular curves $X(1)$ and $Y(1)$

Recall the modular group $\Gamma = \mathrm{SL}_2(\mathbb{Z})$, which acts on the upper half plane $\mathbb{H}$ via linear fractional transformations. The quotient $\mathbb{H}/\Gamma$ (the Γ-orbits of $\mathbb{H}$) is known as the *modular curve* $Y(1)$, whose points may be identified with points in the fundamental region

$$\mathcal{F} = \{z \in \mathbb{H} : \mathrm{re}(z) \in [-1/2, 1/2) \text{ and } |z| \geq 1, \text{ with } |z| > 1 \text{ if } \mathrm{re}(z) > 0\}.$$

You may be wondering why we call $Y(1)$ a curve. Recall from Theorem 18.5 that the j-function gives a holomorphic bijection from $\mathcal{F}$ to $\mathbb{C}$, and we shall prove that in fact $Y(1)$ is isomorphic, as a complex manifold, to the complex plane $\mathbb{C}$, which we may view as an affine curve: let $f(X, Y) = Y$ and note that the zero locus of f is just $\{(X, 0) : X \in \mathbb{C}\} \simeq \mathbb{C}$.

The fundamental region $\mathcal{F}$ is not a compact subset of $\mathbb{H}$, since it is unbounded along the positive imaginary axis. To remedy this deficiency, we compactify it by adjoining a point at infinity to $\mathbb{H}$ and including it in $\mathcal{F}$. But we also want $\mathrm{SL}_2(\mathbb{Z})$ to act on our extended upper half plane. Given that

$$\lim_{\mathrm{im}\,\tau\to\infty} \frac{a\tau + b}{c\tau + d} = \frac{a}{c},$$

we need to include the set of rational numbers in our extended upper half plane in order for Γ to act continuously. So let

$$\mathbb{H}^* = \mathbb{H} \cup \mathbb{Q} \cup \{\infty\} = \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q}),$$

[1] Some authors reserve this term for the case $\mathcal{O} = \mathcal{O}_K$, generically referring to $H_D(X)$ as the *ring class polynomial* for the order $\mathcal{O}$.

Andrew V. Sutherland

and let Γ act on $\mathbb{P}^1(\mathbb{Q})$ via

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} (x : y) = (ax + by : cx + dy).$$

The points in $\mathbb{H}^* \backslash \mathbb{H} = \mathbb{P}^1(\mathbb{Q})$ are called *cusps*; as shown in Problem Set 8, the cusps are all Γ-equivalent. Thus we may extend our fundamental region $\mathcal{F}$ for $\mathbb{H}$ to a fundamental region $\mathcal{F}^*$ for $\mathbb{H}^*$ by including the cusp at infinity: the point $\infty = (1 : 0) \in \mathbb{P}^1(\mathbb{Q})$, which we may view as lying infinitely far up the positive imaginary axis.

We can now define the modular curve $X(1) = \mathbb{H}^*/\Gamma$, which contains all the points in $Y(1)$, plus the cusp at infinity. This is a projective curve, in fact it is the projective closure of $Y(1)$. It is also a compact *Riemann surface*, a connected complex manifold of dimension 1. Before stating precisely what this means, our first goal is to prove that $X(1)$ is a compact Hausdorff space.

To give the extended upper half plane $\mathbb{H}^*$ a topology, we begin with the usual (open) neighborhoods about points $\tau \in \mathbb{H}$ (all open disks about τ that lie in $\mathbb{H}$). For cusps $\tau \in \mathbb{Q}$ we take the union of $\{\tau\}$ with any open disk in $\mathbb{H}$ tangent to τ to be a neighborhood of τ. For the cusp at infinity, any set of the form $\{\infty\} \cup \{\tau \in \mathbb{H} : \operatorname{im} \tau > r\}$ with $r > 0$ is a neighborhood of ∞.

With this topology it is clear that $\mathbb{H}^*$ is a Hausdorff space (any two points can be separated by neighborhoods). It does not immediately follow that $X(1) = \mathbb{H}^*/\Gamma$ is a Hausdorff space; a quotient of a Hausdorff space need not be Hausdorff. To prove that $X(1)$ is Hausdorff we first derive two lemmas that will be useful in what follows.

Lemma 20.1. *For any compact sets A and B in $\mathbb{H}$ the set $S = \{\gamma : \gamma A \cap B \neq \emptyset\}$ is finite.*

Proof. Let $m = \min\{\operatorname{im} \tau : \tau \in A\}$ and $M = \max\{|\operatorname{re} \tau| : \tau \in A\}$, and define

$$r = \max\{\operatorname{im} \tau_A / \operatorname{im} \tau_B : \tau_A \in A, \tau_B \in B\}.$$

Recall that for any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ we have $\operatorname{im} \gamma\tau = \operatorname{im} \tau / |c\tau + d|^2$. If γ sends $\tau_A \in A$ to $\tau_B \in B$, then $|c\tau_A + d|^2 = \operatorname{im} \tau_A / \operatorname{im} \tau_B \leq r$. This implies $(cm)^2 \leq r$ and $(cM + d)^2 \leq r$, which gives upper bounds on $|c|$ and $|d|$ for any $\gamma \in S$. Thus the number of pairs (c, d) arising among $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in S$ is finite. Let us now fix one such pair and define

$$s = \max\{|\tau_B||c\tau_A + d| : \tau_A \in A, \tau_B \in B\}.$$

For any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ we have $|\gamma\tau| = |a\tau + b|/|c\tau + d|$. If γ sends $\tau_A \in A$ to $\tau_B \in B$, then $|a\tau_A + b| = |\tau_B||c\tau_A + d| \leq s$. As above, this gives upper bounds on $|a|$ and $|b|$, proving that the number of pairs (a, b) arising among $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in S$ is finite. So S is finite. $\square$

Lemma 20.2. *For $\tau_1, \tau_2 \in \mathbb{H}^*$ there exist neighborhoods U_1 of τ_1 and U_2 of τ_2 such that*

$$\gamma U_1 \cap U_2 \neq \emptyset \quad \Longleftrightarrow \quad \gamma\tau_1 = \tau_2,$$

for all $\gamma \in \Gamma$. In particular, every $\tau \in \mathbb{H}^*$ has a neighborhood containing no points $\gamma\tau \neq \tau$.

Proof. We first note that if $\gamma\tau_1 = \tau_2$, then $\gamma U_1 \cap U_2 \neq \emptyset$ for all neighborhoods U_1 of τ_1 and U_2 of τ_2, so we only need to prove the forward implication in the statement of the lemma.

We first consider $\tau_1, \tau_2 \in \mathbb{H}$, with compact neighborhoods C_1 and C_2, respectively, and let $S = \{\gamma : \gamma C_1 \cap C_2 \neq \emptyset \text{ and } \gamma\tau_1 \neq \tau_2\}$. If S is empty then let $U_1 \subset C_1$ be a neighborhood of τ_1 and let $U_2 \subset C_2$ be a neighborhood of τ_2. Otherwise, pick $\gamma \in S$, pick a neighborhood U_1 of τ_1 such that $\tau_2 \notin \gamma U_1$, pick a neighborhood U_2 of τ_2 such that $\gamma U_1 \cap U_2 = \emptyset$, and replace C_1 and C_2 by the closures of U_1 an U_2, respectively, yielding a smaller set S. Note that the existence of U_1 and U_2 is guaranteed by the continuity of the function $f(\tau) = \gamma\tau = (a\tau + b)/(c\tau + d)$. By Lemma 20.1, S is finite, so we eventually have $S = \emptyset$ and neighborhoods U_1 and U_2 that satisfy the lemma.

We now consider $\tau_1 \in \mathbb{H}$ and $\tau_2 = \infty$. Let U_1 be a neighborhood of τ_1 with $\overline{U}_1 \subset \mathbb{H}$. The set $\{|c\tau + d| : \tau \in U_1, c, d \in \mathbb{Z}$ not both $0\}$ is bounded below, and $\{\operatorname{im}\gamma\tau : \gamma \in \Gamma, \tau \in U_1\}$ is bounded above, say by r, since $\operatorname{im}\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\tau = \operatorname{im}\tau/|c\tau + d|^2$. If we let $U_2 = \{\tau : \operatorname{im}\tau > r\}$ be our neighborhood of $\tau_2 = \infty$, then $\gamma U_1 \cap U_2 = \emptyset$ for all $\gamma \in \Gamma$ and the lemma holds. This argument extends to all the cusps in $\mathbb{H}^*$, since every cusp is Γ-equivalent to ∞, and we can easily reverse the roles of τ_1 and τ_2, since if $\gamma U_1 \cap U_2 = \emptyset$ then $U_1 \cap \gamma^{-1} U_2 = \emptyset$. □

Theorem 20.3. *$X(1)$ is a connected compact Hausdorff space.*

Proof. It is clear that $\mathbb{H}$ is connected, hence its closure $\mathbb{H}^*$ is connected, and the quotient of a connected space is connected. So $X(1)$ is connected.

To show that $X(1)$ is compact, we show that every open cover has a finite subcover. Let $\{U_i\}$ be an open cover of $X(1)$ and let $\pi\colon \mathbb{H}^* \to X(1)$ be the quotient map. Then $\{\pi^{-1}(U_i)\}$ is an open cover of $\mathbb{H}^*$, so it contains an open set V_0 containing the point ∞. Let $\{V_1, \ldots, V_n\}$ be a finite subset of $\{\pi^{-1}(U_i)\}$ covering the compact set $\overline{\mathcal{F}} \backslash V_0$ (note that V_0 contains a neighborhood $\{z : \operatorname{im} z > r\}$ of ∞). Then $\{V_0, \ldots, V_n\}$ is a finite cover of $\mathcal{F}^*$, and $\{\pi(V_0), \ldots, \pi(V_n)\}$ is a finite subset of $\{U_i\}$ covering $X(1)$.

To show that $X(1)$ is Hausdorff, let $x_1, x_2 \in X(1)$ be distinct, and choose $\tau_1, \tau_2 \in \mathcal{F}^*$ so that $\pi(\tau_1) = x_1$ and $\pi(\tau_2) = x_2$. By Lemma 20.2, there exist neighborhoods U_1 and U_2 of τ_1 and τ_2 respectively, such that $\gamma U_1 \cap U_2 = \emptyset$ for all $\gamma \in \Gamma$. It follows that $\pi(U_1)$ and $\pi(U_2)$ are disjoint neighborhoods of x_1 and x_2, thus $X(1)$ is Hausdorff. □

We note that Lemmas 20.1 and 20.2 and Thoerem 20.3 all hold if we replace Γ by any finite-index subgroup of Γ; the proofs are essentially the same, the only change is an additional argument in the proof of Lemma 20.2 to handle inequivalent cusps.

20.3 Riemann surfaces

Definition 20.4. A *complex structure* on a topological space X is an open cover $\{U_i\}$ of X together with a set of compatible homeomorphisms[2] $\psi_i\colon U_i \to \mathbb{C}$ with open images. Homeomorphisms ψ_i and ψ_j are *compatible* if the *transition map*

$$\psi_j \circ \psi_i^{-1}\colon\ \psi_i(U_i \cap U_j) \to \psi_j(U_i \cap U_j)$$

is holomorphic (vacuously true when $U_i \cap U_j = \emptyset$).

The homeomorphisms ψ_i are called *local parameters*, or *charts*, and the set $\{\psi_i\}$ is called an *atlas*. Each of the charts ψ_i allows us to view a local piece of X as a region of the complex plane; the transition map allows us to move smoothly from one region to another. Note that the transition maps are necessarily homeomorphisms; the requirement that they also be holomorphic is the key feature that differentiates complex manifolds from real manifolds.

[2]Recall that a homeomorphism is a bicontinuous function, a continuous function with a continuous inverse.

Definition 20.5. A *Riemann surface* is a connected Hausdorff space with a complex structure (equivalently, a connected complex manifold of dimension 1).[3]

We have already seen examples of Riemann surfaces: the torus $\mathbb{C}/L$ corresponding to an elliptic curve $E/\mathbb{C}$ is a Riemann surface. In order to show that $X(1)$ is a Riemann surface, we need to give it a complex structure. The only difficulty that arises when doing so occurs at points in $\mathbb{H}^*$ that possess extra symmetries under the action of Γ. We may restrict our attention to the fundamental region $\mathcal{F}^*$, and in this region there are three points that we need to worry about: $i, \rho = e^{2\pi i/3}$, and ∞. We require the following lemma.

Lemma 20.6. *For $\tau \in \mathcal{F}^*$, let G_τ denote the stabilizer of τ in $\Gamma = \mathrm{SL}_2(\mathbb{Z})$. Let $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$. Then*

$$G_\tau = \begin{cases} \langle S \rangle \simeq \mathbb{Z}/4\mathbb{Z} & \text{if } \tau = i; \\ \langle ST \rangle \simeq \mathbb{Z}/6\mathbb{Z} & \text{if } \tau = \rho \\ \langle \pm T \rangle \simeq \mathbb{Z} & \text{if } \tau = \infty; \\ \{\pm 1\} \simeq \mathbb{Z}/2\mathbb{Z} & \text{otherwise.} \end{cases}$$

Proof. See problem 4 on Problem Set 8. □

We now define a complex structure for $X(1)$. Let $\pi \colon \mathbb{H}^* \to X(1)$ be the quotient map, and for each point $x \in X(1)$ let τ_x be the unique point in the fundamental region $\mathcal{F}^*$ for which $\pi(\tau_x) = x$, and let $G_x = G_{\tau_x}$ be the stabilizer of τ_x. For each $\tau_x \in \mathcal{F}^*$, we can pick a neighborhood U_x such that $\gamma U_x \cap U_x = \emptyset$ for all $\gamma \notin G_x$, by Lemma 20.2. The sets $\pi(U_x)$ form an open cover of $X(1)$. For $x \neq \infty$, we can map U_x to an open subset of the unit disk $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$ via the homeomorphism $\delta_x \colon \mathbb{H} \to \mathbb{D}$ defined by

$$\delta_x(\tau) = \frac{\tau - \tau_x}{\tau - \overline{\tau}_x}. \tag{1}$$

To visualize the map δ_x, note that it sends τ_x to the origin, and if we extend its domain to $\overline{\mathbb{H}}$, it maps the real line to the unit circle minus the point 1 and sends ∞ to 1.

To define ψ_x we need to map $\pi(U_x)$ into $\mathbb{D}$. For $\tau_x \neq i, \rho, \infty$ we have $G_x = \{\pm 1\}$, which fixes every point in U_x, not just τ_x. In this case the restriction of π to U_x is injective, we have $U_x/\Gamma = U_x/G_x = U_x$, so we can simply define $\psi_x = \delta_x \circ \pi^{-1}$.

When $|G_x| > 2$, the restriction of π to U_x is no longer injective (it is at τ_x, but not at points near τ_x), so we cannot use $\psi_x = \delta_x \circ \pi^{-1}$. We instead define $\psi_x(z) = \delta_x(\pi^{-1}(z))^n$, where $n = |G_x|/2$ is the size of the Γ-orbits in $U_x \backslash \{\tau_x\}$. Note that when $G_x = \{\pm 1\}$ we have $n = 1$ and this is the same as defining $\psi_x = \delta_x \circ \pi^{-1}$. To prove that this actually works, we will need the following lemma.

Lemma 20.7. *Let $\tau_x \in \mathbb{H}$, with $\delta_x(\tau)$ as in (1), and let $\varphi \colon \mathbb{H} \to \mathbb{H}$ be a holomorphic function fixing τ_x whose n-fold composition with itself is the identity, with n minimal. Then for some primitive nth root of unity ζ, we have $\delta_x(\varphi(\tau)) = \zeta \delta_x(\tau)$ for all $\tau \in \mathbb{H}$.*

Proof. The map $f = \delta_x \circ \varphi \circ \delta_x^{-1}$ is a holomorphic bijection (conformal map) from $\mathbb{D}$ to $\mathbb{D}$ that fixes 0. Every such function is a rotation $f(z) = \zeta z$ with $|\zeta| = 1$, by [3, Cor. 8.2.3]. Since the n-fold composition of f with itself is the identity map, with n minimal, ζ must be a primitive nth root of unity. □

[3]Strictly speaking, a Riemann surface is also required to be second-countable, meaning that it admits a countable basis of open sets. This is a technical condition that is easily satisfied by all the Riemann surfaces we shall consider (e.g., use neighorboods with rational coordinates and rational radii).

What about $x = \infty$? We have $G_\infty = \langle \pm T \rangle$, so the intersection of the Γ-orbit of any point $\tau \in U_\infty \backslash \{\infty\}$ with U_∞ is the set $\{\tau + m : m \in \mathbb{Z}\}$. Define

$$\delta_\infty(z) = \begin{cases} e^{2\pi i z} & \text{if } z \neq \infty, \\ 0 & \text{if } z = \infty, \end{cases}$$

and let $\psi_\infty = \delta_\infty \circ \pi^{-1}$. Then $\delta_\infty(\tau + m) = \delta_\infty(\tau)$ for all $\tau \in U_\infty \backslash \{\infty\}$ and $m \in \mathbb{Z}$.

The following commutative diagrams summarizes the charts ψ_x:

$$\begin{tikzcd} U_x \arrow[r, "\pi"] \arrow[d, "\delta_x"'] & U_x/G_x \arrow[d, "\psi_x"] \\ \mathbb{D} \arrow[r, "z^n"] & \mathbb{D} \end{tikzcd} \qquad \begin{tikzcd} U_x \arrow[r, "\pi"] \arrow[dr, "\delta_x"'] & U_x/G_x \arrow[d, "\psi_x"] \\ & \mathbb{D} \end{tikzcd}$$

$x \neq \infty,\ \psi_x(\tau) = \frac{\tau - \tau_x}{\tau - \bar{\tau}_x}$
$n = |G_x|/2$

$x = \infty,\ \psi_x(\tau) = e^{2\pi i \tau}$

We are now ready to prove that $X(1)$ is a compact Riemann surface. Theorem 20.3 states that $X(1)$ is a connected compact Haussdorff space, so we just need to prove that we have a complex structure on $X(1)$. This means verifying that the maps $\psi_x \colon \pi(U_x) \to \mathbb{D}$ are well-defined (we must have $\psi(\pi(\gamma\tau)) = \psi(\pi(\tau))$ for all $\tau \in U_x$ and $\gamma \in G_x$), that they are homeomorphisms, and that the transition maps are holomorphic.

Theorem 20.8. *The open cover $\{U_x\}$ and atlas $\{\psi_x\}$ define a complex structure on $X(1)$.*

Proof. As above, let $x = \pi(\tau_x)$ with $\tau_x \in \mathcal{F}^*$. We first verify that the maps ψ_x are well-defined and homeomorphisms.

Let us assume $x \neq \infty$. By Lemma 20.6, the stabilizer G_x of τ_x is cyclic of order $2n$, and $\gamma^n = \pm 1$ acts trivially for all $\gamma \in G_x$. Applying Lemma 20.7 to the function $\varphi(\tau) = \gamma\tau$, we have $\delta_x(\gamma z) = \zeta \delta_x(z)$ for all $z \in U_x$, where ζ is a primitive nth root of unity. Thus

$$\psi_x(\pi(\gamma z)) = \delta_x(\gamma z)^n = \zeta^n \delta_x(z)^n = \delta_x(z)^n = \psi_x(\pi(z))$$

for all $z \in U_x$. It follows that ψ_x is well defined on U_x/G_x. To show that ψ_x is a homeomorphism, it suffices to show that it is holomorphic and injective, by the open mapping theorem [3, Thm. 5.5.4]. It is clearly holomorphic, since $\delta_x(\tau)$ is a rational function with no poles in U_x. To prove injectivity, assume $\psi_x(\pi(\tau_1)) = \psi_x(\pi(\tau_2))$. Then for some integer k

$$\begin{aligned} \delta_x(\tau_1)^n &= \delta_x(\tau_2)^n \\ \delta_x(\tau_1) &= \zeta^k \delta_x(\tau_2) = \delta_x(\gamma^k \tau_2) \\ \tau_1 &= \gamma^k \tau_2 \\ \pi(\tau_1) &= \pi(\tau_2). \end{aligned}$$

Thus ψ_x is an injective and therefore a homeomorphism.

For $x = \infty$, the point $\tau = \infty \in \mathbb{H}^*$ is the unique point in U_∞ for which $\pi(\tau) = \infty$, and $\psi_x(\tau) = 0$ if and only if $\tau = \infty$. So ψ_∞ is well defined at ∞. For $\tau \in U_\infty \backslash \{\infty\}$, we have

$$\psi_\infty(\pi(\tau + m)) = \delta_\infty(\tau + m) = e^{2\pi i(\tau + m)} = e^{2\pi i \tau} = \delta_\infty(\tau) = \psi_\infty(\pi(\tau))$$

for all $m \in \mathbb{Z}$, thus ψ_∞ is well defined. The map ψ_∞ is clearly continuous, and it has a continuous inverse

$$\psi_\infty^{-1}(z) = \begin{cases} \pi\left(\frac{1}{2\pi i}\log z\right) & \text{if } z \neq 0, \\ \infty & \text{otherwise,} \end{cases}$$

thus it is a homeomorphism.

We now show that the transition maps are holomorphic. Let us first consider U_x, U_y with $x, y \neq \infty$. For any $z \in \psi_x(\pi(U_x) \cap \pi(U_y)) \subseteq \mathbb{D}$ we have

$$\psi_y \circ \psi_x^{-1}(z) = \psi_y \circ \pi \circ \pi^{-1} \circ \psi_x^{-1}(z) = (\psi_y \circ \pi) \circ (\psi_x \circ \pi)^{-1}(z) = \delta_y^{n_y} \circ \delta_x^{-1}(z^{1/n_x}),$$

where $n_x = |G_x|/2$ and $n_y = |G_y|/2$. The map $\delta_y^{n_y} \circ \delta_x^{-1}$ is holomorphic on $\mathbb{D}$, so it suffices to show that it is a power series in z^{n_x}. Let ζ be an n_xth root of unity such that $\delta_x(\gamma z) = \zeta\delta_x(z)$, where γ generates G_x, as in Lemma 20.7. Note that $\pi \circ \gamma = \pi$ for any $\gamma \in \Gamma$, so we have

$$\delta_y^{n_y} \circ \delta_x^{-1}(\zeta z) = (\psi_y \circ \pi) \circ (\gamma \circ \delta_x^{-1}(z)) = \psi_y \circ \pi \circ \delta_x^{-1}(z) = \delta_y^{n_y} \circ \delta_x^{-1}(z).$$

It follows that $\delta_y^{n_y} \circ \delta_x^{-1}$ is a power series in z^{n_x}, since it maps ζz and z to the same point.

For $x \neq \infty$ and $y = \infty$ we have

$$\begin{aligned} \psi_\infty \circ \psi_x^{-1}(z) &= \psi_y \circ \pi \circ \pi^{-1} \circ \psi_x^{-1}(z) = (\psi_y \circ \pi) \circ (\psi_x \circ \pi)^{-1}(z) \\ &= \delta_\infty \circ \delta_x^{-1}(z^{1/n_x}) = \exp\left(2\pi i\ \delta_x^{-1}(z^{1/n_x})\right), \end{aligned}$$

where $\delta_\infty \circ \delta_x^{-1}$ is holomorphic. and the same argument used above shows that it is actually a power series in z^{n_x}.

For the case $x = \infty$ and $y \neq \infty$, we have

$$\delta_y^{n_y}(z+1) = \psi_y \circ \pi \circ Tz = \psi_y \circ \pi(z) = \delta_y^{n_y}(z),$$

so $\delta_y^{n_y}$ is a holomorphic function in the variable $q = e^{2\pi i z}$ (note $z \in U_\infty \cap U_y$ is bounded). Thus the transition map

$$\psi_y \circ \psi_\infty^{-1}(z) = \delta_y^{n_y}\left(\frac{1}{2\pi i}\log z\right)$$

is holomorphic. The case $x = y = \infty$ is trivial, since $\psi_\infty \circ \psi_\infty^{-1}$ is the identity map. $\square$

Theorem 20.9. *The modular curve $X(1)$ is a compact Riemann surface of genus 0.*

Proof. That $X(1)$ is a compact Riemann surface follows immediately from Theorems 20.3 and 20.8. To show that it has genus 0, we triangulate $X(1)$ by connecting the points i, ρ, and ∞, partitioning the surface into two triangles. Applying Euler's formula

$$V - E + F = 2 - 2g$$

with $V = 3, E = 3$, and $F = 2$, we see that $g = 0$. $\square$

Theorem 20.9 implies that $X(1)$ is homeomorphic to the Riemann sphere $S = \mathbb{P}^1(\mathbb{C})$, since, up to isomorphism, S is the unique compact Riemann surface of genus 0. The modular curve $Y(1)$ is also a Riemann surface of genus 0, but it is not compact; it is homeomorphic to the complex plane $\mathbb{C}$.

We also wish to consider modular curves defined as quotients $\mathbb{H}^*/\Gamma$ for certain finite index subgroups Γ of $\mathrm{SL}_2(\mathbb{Z})$ that we will define in the next lecture. The methods we used to prove that $X(1)$ is a compact Riemann surface apply to all of the modular curves we shall consider, the only additional complication is that there will generally by more than one Γ-equivalence class of cusps to consider.

References

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

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

[3] Elias M. Stein and Rami Shakarchi, *Complex analysis*, Princeton University Press, 2003.