18 Riemann surfaces and modular curves

Let $\mathcal{O}$ be an order in an imaginary quadratic field and let $\mathrm{cl}(\mathcal{O})$ be its ideal class group (proper $\mathcal{O}$-ideals up to homethety, or equivalently, invertible fractional $\mathcal{O}$-ideals modulo invertible principal $\mathcal{O}$-ideals). In the previous lecture we showed that the set

$$\mathrm{Ell}_{\mathcal{O}}(\mathbb{C}) := \{j(E) : E/\mathbb{C} \text{ with } \mathrm{End}(E) = \mathcal{O}\}$$

of isomorphism classes of elliptic curves $E/\mathbb{C}$ with complex multiplication by $\mathcal{O}$ is a torsor for the group $\mathrm{cl}(\mathcal{O})$. If $\mathfrak{a}$ and $\mathfrak{b}$ are proper $\mathcal{O}$-ideals and $E_{\mathfrak{b}}$ is the elliptic curve corresponding to the complex torus $\mathbb{C}/\mathfrak{b}$, then $E_{\mathfrak{b}}$ has CM by $\mathcal{O}$ and the $\mathcal{O}$-ideal $\mathfrak{a}$ acts on $E_{\mathfrak{b}}$ via

$$\mathfrak{a}E_{\mathfrak{b}} = E_{\mathfrak{a}^{-1}\mathfrak{b}}.$$

The isogeny $\phi_{\mathfrak{a}}\colon E_{\mathfrak{b}} \to \mathfrak{a}E_{\mathfrak{b}}$ induced by the lattice inclusion $\mathfrak{b} \subseteq \mathfrak{a}^{-1}\mathfrak{b}$ has kernel

$$\ker\phi_{\mathfrak{a}} = E_{\mathfrak{b}}[\mathfrak{a}] := \{P \in E(\mathbb{C}) : \alpha P = 0 \text{ for all } \alpha \in \mathfrak{a} \subseteq \mathcal{O} \simeq \mathrm{End}(E_{\mathfrak{b}})\},$$
$$\#\ker\phi_{\mathfrak{a}} = \deg\phi_{\mathfrak{a}} = \mathrm{N}\mathfrak{a} := [\mathcal{O} : \mathfrak{a}].$$

To make further progress in our development of the theory of complex multiplication, we need a better understanding of the isogenies $\phi_{\mathfrak{a}}$. The key to doing so, both from a theoretical and practical perspective, is to understand the *modular curves* that "parameterize" isogenies of elliptic curves (in a sense that will be made clear in later lectures).

In this lecture our goal is simply to introduce the notion of a modular curve, beginning with the canonical example $X(1)$. Modular curves, and the *modular functions* that comprise their function fields are a major topic in their own right, one to which entire courses are devoted; we shall necessarily only scratch the surface of this rich and beautiful subject. Our presentation is adapted from [2, V.1] and [4, I.2].

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

Recall from Lecture 15 that the modular group $\Gamma := \mathrm{SL}_2(\mathbb{Z})$ acts on the upper half plane $\mathcal{H} := \{\tau \in \mathbb{C} : \mathrm{im}\,\tau > 0\}$ via linear fractional transformations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\tau := \frac{a\tau + b}{c\tau + d}.$$

The quotient $\mathcal{H}/\Gamma$ (the Γ-orbits of $\mathcal{H}$) is known as the *modular curve* $Y(1)$, whose points may be identified with points in the fundamental region

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

You may be wondering why we call $Y(1)$ a curve. Recall from Theorem 15.11 that the j-function defines 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: if we put $f(x, y) = y$ then the zero locus of f is $\{(x, 0) : x \in \mathbb{C}\} \simeq \mathbb{C}$.

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

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

Lecture by Andrew Sutherland

we should also include the set of rational numbers in our extended upper half plane. So let

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

and let Γ act on $\mathcal{H}^*$ by extending its action on $\mathcal{H}$ to $\mathbb{P}^1(\mathbb{Q})$ via

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

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

We can now define the modular curve $X(1) = \mathcal{H}^*/\Gamma$, 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 $Y(1)$ in $\mathbb{P}^2$. It is also a *Riemann surface*, a connected complex manifold of dimension one. Before stating precisely what this means, our first goal is to prove that $X(1)$ is a compact Hausdorff space.

We extend the topology of $\mathcal{H}$ to a topology on $\mathcal{H}^*$ by taking as a basis of open neighborhoods:

- $\tau \in \mathcal{H}$: all open disks about τ that lie in $\mathcal{H}$;
- $\tau \in \mathbb{Q}$: all sets $\{\tau\} \cup D$, where $D \subseteq \mathcal{H}$ is an open disk tangent to the real line at τ;
- $\tau = \infty$: all sets of the form $\{\tau \in \mathcal{H} : \operatorname{im} \tau > r\}$ for any $r > 0$;

The topology of $\mathcal{H}^*$ is generated by these open neighborhoods under unions and finite intersections; note that the induced subspace topology on $\mathcal{H}$ is just its standard topology.

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

Lemma 18.1. *For any compact sets $A, B \subseteq \mathcal{H}$ the set $S = \{\gamma \in \Gamma : \gamma A \cap B \neq \emptyset\}$ is finite.*

Proof. Recall that for any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ we have

$$\operatorname{im} \gamma\tau = \operatorname{im} \frac{a\tau + b}{c\tau + d} = \operatorname{im} \frac{(a\tau + b)(c\bar{\tau} + d)}{|c\tau + d|^2} = \frac{(ad - bc) \operatorname{im} \tau}{|c\tau + d|^2} = \frac{\operatorname{im} \tau}{|c\tau + d|^2}.$$

Now define

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

If $\gamma\tau_A = \tau_B$ for some $\tau_A \in A$ and $\tau_B \in B$, then $|c\tau_A + d|^2 = \operatorname{im} \tau_A / \operatorname{im} \tau_B \leq r$, which implies 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 $\gamma\tau_A = \tau_B$ for some $\tau_A \in A$ and $\tau_B \in B$, then $|a\tau_A + b| = |\tau_B||c\tau_A + d| \leq s$, which gives upper bounds on $|a|$ and $|b|$ as above. The number of pairs (a, b) arising among $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in S$ is thus finite, hence S is finite. $\square$

Lemma 18.2. *For any $\tau_1, \tau_2 \in \mathcal{H}^*$ there exist open neighborhoods U_1, U_2 of τ_1, 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, each $\tau \in \mathcal{H}^*$ has an open neighborhood U in which it is the sole representative of its Γ-orbit and $\gamma U \cap \gamma' U = \emptyset$ for all $\gamma, \gamma' \in \Gamma$ such that $\gamma\tau \neq \gamma'\tau$.

Proof. We first note that if $\gamma\tau_1 = \tau_2$, then $\gamma U_1 \cap U_2 \neq \emptyset$ for all open neighborhoods U_1, U_2 of τ_1, τ_2, so we only need to consider γ for which $\gamma\tau_1 \neq \tau_2$.

We first consider $\tau_1, \tau_2 \in \mathcal{H}$ and let $C_1, C_2 \subseteq \mathcal{H}$ be closed disks about them. Let $S(C_1, C_2) := \{\gamma : \gamma C_1 \cap C_2 \neq \emptyset \text{ and } \gamma\tau_1 \neq \tau_2\}$. If S is nonempty, pick $\gamma \in S$, and let U_3 and U_2' be disjoint open neighborhoods of $\gamma\tau_1$ and τ_2 respectively (they exist because $\mathcal{H}$ is Hausdorff). Then $\gamma^{-1}U_3$ is an open neighborhood of τ_1 (since γ acts continuously), and it contains a closed disk $C_1' \subseteq C_1$ about τ_1, and the open set U_2' similarly contains a closed disk $C_2' \subseteq C_2$ about τ_2. We then have $S(C_1', C_2') \subsetneq S(C_1, C_2)$, since by construction, $\gamma \notin S(C_1', C_2')$. By Lemma 18.1, S is finite, so if we continue in this fashion we will eventually have $S(C_1, C_2) = \emptyset$, at which point we may take U_1, U_2 to be the interiors of C_1, C_2.

We now consider $\tau_1 \in \mathcal{H}$ and $\tau_2 = \infty$. Let U_1 be a neighborhood of τ_1 with $\overline{U}_1 \subseteq \mathcal{H}$. The set $\{|c\tau + d| : \tau \in U_1, c, d \in \mathbb{Z} \text{ not both } 0\}$ 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 $\mathcal{H}^*$, 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 $U_1 \cap \gamma^{-1}U_2 = \emptyset$.

Finally, if $\tau_1 = \tau_2 = \infty$ we let $U_1 = U_2 = \{\tau \in \mathcal{H} : \operatorname{im}\tau > 1\} \cup \{\infty\}$: for $\operatorname{im}\tau > 1$ either $\operatorname{im}\gamma\tau = \operatorname{im}\tau$, in which case $\gamma = \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right)$ fixes ∞, or $\operatorname{im}\gamma\tau = \operatorname{im}\tau/|c\tau + d|^2 < 1$.

To prove the last statement in the lemma, take $\tau_1 = \tau_2 = \tau$ and $U = U_1 \cap U_2$. □

Theorem 18.3. *$X(1)$ is a connected compact Hausdorff space.*

Proof. It is clear that $\mathcal{H}$ is connected, hence its closure $\mathcal{H}^*$ 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 $\pi\colon \mathcal{H}^* \to X(1)$ be the quotient map. Then $\{\pi^{-1}(U_i)\}$ is an open cover of $\mathcal{H}^*$ and 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 $\overline{\mathcal{F}} - V_0$ (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 subcover of $\{U_i\}$.

To show that $X(1)$ is Hausdorff, let $x_1, x_2 \in X(1)$ be distinct, and choose τ_1, τ_2 so that $\pi(\tau_1) = x_1$ and $\pi(\tau_2) = x_2$. Then $\tau_2 \neq \gamma\tau_1$ for all $\gamma \in \Gamma$ (since $x_1 \neq x_2$), so by Lemma 18.2, there are neighborhoods U_1 and U_2 of τ_1 and τ_2 respectively for which $\gamma U_1 \cap U_2 = \emptyset$ for all $\gamma \in \Gamma$. Thus $\pi(U_1)$ and $\pi(U_2)$ are disjoint neighborhoods of x_1 and x_2. □

We note that Lemmas 18.1 and 18.2 and Theorem 18.3 all hold if we replace Γ by any finite-index subgroup of Γ; the proofs are essentially the same, the only difference is an additional argument in the proof of Lemma 18.2 to handle inequivalent cusps.

18.2 Riemann surfaces

Definition 18.4. A *complex structure* on a topological space X is an open cover $\{U_i\}$ of X together with a set of compatible homeomorphisms[1] $\psi_i\colon U_i \to \mathbb{C}$ with open images.

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

Homeomorphisms ψ_i and ψ_j are compatible if whenever $U_i \cap U_j \neq \emptyset$ 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.

The homeomorphisms ψ_i are called *charts* (or *local parameters*), and the collection $\{\psi_i\}$ is called an *atlas*. Each chart ψ_i allows us to view a local neighborhood U_i of X as a region of the complex plane, and the transition maps allow us to move smoothly from one region to another. Note that transition maps are automatically homeomorphisms; the requirement that they be holomorphic is a stronger condition (this is what differentiates complex manifolds from real manifolds).

Definition 18.5. A *Riemann surface* is a connected Hausdorff space with a complex structure (equivalently, it is a connected complex manifold of dimension one).[2]

Example 18.6. The torus $\mathbb{C}/L$ corresponding to an elliptic curve $E/\mathbb{C}$ is a Riemann surface. To give $\mathbb{C}/L$ a complex structure let $\pi\colon \mathbb{C} \to \mathbb{C}/L$ be the quotient map, let $r > 0$ be less than half the length of the shortest vector in L, and for each $z \in \mathbb{C}$ in a fundamental region for L, let $U_z \subseteq \mathbb{C}$ be the open disc or radius r centered at z. The restriction of π to each U_z is injective (by our choice of r) and defines a homeomorphism. We may thus take $\{\pi(U_z)\}$ as our open cover and the inverse maps $\pi^{-1}\colon \pi(U_z) \to U_z$ as our charts. The transition maps are all the identity map, hence holomorphic.

It is clear that $\mathbb{C}/L$ is a connected Hausdorff space, hence a Riemann surface, in fact a compact Riemann surface. We can compute its genus by triangulating a fundamental parallelogram and computing its Euler characteristic. Recall Euler's formula

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

where V counts vertices, E counts edges, F counts faces, and g is the genus. If $L = [\omega_1, \omega_2]$, we may triangulate the parallelogram $\overline{\mathcal{F}_0}$ by drawing a diagonal from ω_1 to ω_2. We then have $V = 1$ (every lattice point is equivalent to 0), $E = 3$ (edges on the opposite side of the parallelogram are equivalent, so 2 edges on the border plus the diagonal), and $F = 2$ (two triangles, one on each side of the diagonal). We thus have

$$1 - 3 + 2 = 2 - 2g,$$

and $g = 1$, as expected.

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 $\mathcal{H}^*$ 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 only three points that we need to worry about, the points $i, \rho := e^{2\pi i/3}$, and ∞. We require the following lemma.

Lemma 18.7. *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*

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

[2]Some texts require Riemann surfaces to be second-countable (admit a countable basis of open sets), but in fact this requirement is automatically satisfied; this is a celebrated theorem of Radó.

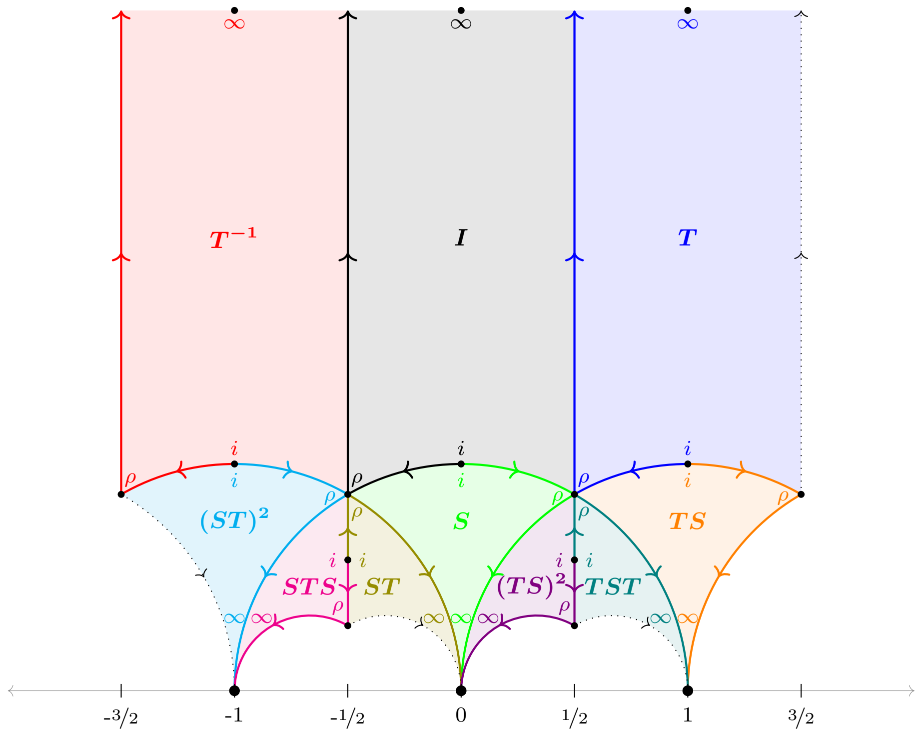

Figure 1: $\mathcal{H}^*/\Gamma$

Proof. See Problem Set 8, or stare at Figure 1 and note $-I$ acts trivially and $T\infty = \infty$. $\square$

18.3 The modular curve $X(1)$ as a Riemann surface

We now put a complex structure on $X(1)$. Let $\pi\colon \mathcal{H}^* \to X(1)$ 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 18.2. 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 $\mathcal{D} := \{z \in \mathbb{C} : |z| < 1\}$ via the homeomorphism $\delta_x\colon \mathcal{H} \to \mathcal{D}$ 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 $\overline{\mathcal{H}} \subseteq \mathbb{C}$, it maps the real line to the unit circle minus the point 1 and sends ∞ to 1. Note that $\operatorname{im}\tau > 0$ and $\operatorname{im}\overline{\tau}_x < 0$, so $\delta_x(\tau)$ is defined and nonzero for all $\tau \in \mathcal{H}$.

To define ψ_x we need to map $\pi(U_x)$ into $\mathcal{D}$. 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 $U_x - \{\tau_x\}$. 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 18.8. *Let $\tau_x \in \mathcal{H}$, with $\delta_x(\tau)$ as in (1), and let $\varphi\colon \mathcal{H} \to \mathcal{H}$ 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 $\tau \in \mathcal{H}$.*

Proof. The map $f = \delta_x \circ \varphi \circ \delta_x^{-1}$ is a holomorphic bijection (conformal map) from $\mathcal{D}$ to $\mathcal{D}$ that fixes 0. Every such function is a rotation $f(z) = \zeta z$ with $|\zeta| = 1$, by [5, 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. □

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

The following commutative diagrams summarize the charts ψ_x:

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

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

We are now ready to prove that $X(1)$ is a compact Riemann surface. Theorem 18.3 states that $X(1)$ is a connected compact Hausdorff space, so we just need to prove that we have a complex structure on $X(1)$. This means verifying that the maps $\psi_x\colon \pi(U_x) \to \mathcal{D}$ 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 18.9. *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 homeomorphisms.

We first consider $x \neq \infty$. By Lemma 18.7, 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 18.8 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 [5, 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 $\tau = \infty \in \mathcal{H}^*$ 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 $\tau \in U_\infty - \{\infty\}$, 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 $z \in \psi_x(\pi(U_x) \cap \pi(U_y)) \subseteq \mathcal{D}$ 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 $\mathcal{D}$, so it suffices to show that it is a power series in z^{n_x}; this will imply that $\delta_y^{n_y} \circ \delta_x^{-1}(z^{1/n_z})$ is defined by a power series in z, hence holomorphic. 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 18.8. 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 $q = e^{2\pi iz}$ (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. □

Theorem 18.10. *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 18.3 and 18.9. 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 $g = 0$. □

Theorem 18.10 implies that $X(1)$ is homeomorphic to the Riemann sphere $S = \mathbb{P}^1(\mathbb{C})$, since up to homeomorphism, 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. As we saw in Lecture 17, $Y(1)$ is homeomorphic to the complex plane $\mathbb{C}$ via the j-function.

18.4 Modular curves

We also wish to consider modular curves defined as quotients $\mathcal{H}^*/\Gamma$ for various finite index subgroups Γ of $\mathrm{SL}_2(\mathbb{Z})$ that have desirable arithmetic properties.

Definition 18.11. The *principal congruence subgroup* $\Gamma(N)$ is defined by

$$\Gamma(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) : \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right) \bmod N \right\}.$$

A *congruence subgroup* (of level N) is any subgroup of $\mathrm{SL}_2(\mathbb{Z})$ that contains $\Gamma(N)$. A *modular curve* is a quotient of $\mathcal{H}^*$ or $\mathcal{H}$ by a congruence subgroup.

Remark 18.12. Every congruence subgroup is a finite index subgroup of $\mathrm{SL}_2(\mathbb{Z})$. The converse does not hold; in fact, most finite index subgroups of $\mathrm{SL}_2(\mathbb{Z})$ are not congruence subgroups, although it is surprisingly difficult to write down explicit examples (you will have the opportunity to explore this question in Problem Set 10).

There are two families of congruence subgroups of particular interest:

$$\Gamma_1(N) := \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) : \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 1 & * \\ 0 & 1 \end{smallmatrix}\right) \bmod N \right\};$$
$$\Gamma_0(N) := \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z}) : \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right) \bmod N \right\};$$

Note that $\Gamma(1) = \Gamma_1(1) = \Gamma_0(1) = \mathrm{SL}_2(\mathbb{Z})$. We now define the modular curves

$$X(N) := \mathcal{H}^*/\Gamma(N), \qquad X_1(N) := \mathcal{H}^*/\Gamma_1(N), \qquad X_0(N) := \mathcal{H}^*/\Gamma_0(N),$$

and similarly define

$$Y(N) := \mathcal{H}/\Gamma(N), \qquad Y_1(N) := \mathcal{H}/\Gamma_1(N), \qquad Y_0(N) := \mathcal{H}/\Gamma_0(N).$$

Following the same strategy we used for $X(1)$, one can show that these are all compact Riemann surfaces (the only difference in the proof is that in general a fundamental region may contain multiple cusps, we only had to consider the cusp ∞).

References

[1] Renzo Cavalieri and Eric Miles, *Riemann surfaces and algebraic curves*, Cambridge University Press, 2016.

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

[3] Rick Miranda, *Algebraic curves and Riemann surfaces*, American Mathematical Society, 1995.

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

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