MODULAR AND SHIMURA CURVES

JIM STANKEWICZ

1. Lecture 1: Riemann surfaces

As promised, in the first half of the course, we will cover modular and Shimura curves over the complex numbers. As these are curves over the complex numbers, they form surfaces over the real numbers. These will in fact lie inside a collection of orientable surfaces which we will call Riemann surfaces. Throughout these notes we will denote the complex numbers by $\mathbf{C}$. In general, we will attempt to use the bold font for objects which are *unique up to unique isomorphism* in whichever category is appropriate. For instance, the integers will be denoted $\mathbf{Z}$ and are unique in the category of rings. The rational numbers are denoted $\mathbf{Q}$ and are unique in the category of fields. The real numbers are denoted $\mathbf{R}$ and are unique in the category of topological fields. The complex numbers are unique, but not up to unique isomorphism in the category of topological fields (there is still complex conjugation floating around). To fix this, we will enrich our category by using $\mathbf{C}$ to denote the complex numbers *with a fixed choice of* $\sqrt{-1}$ which we call i. This choice also forces an orientation on any Riemann surface which we informally call a real surface with complex-analytic functions.

Definition 1. *Let $\mathbb{D}$ denote the complex unit disc, i.e. $\mathbb{D} = \{x + iy : x, y \in \mathbf{R}, x^2 + y^2 < 1\}$.*

Definition 2. *A Riemann Surface is a connected Hausdorff topological space X which admits a collection of charts $\{U_\alpha, z_\alpha\}$ such that*

- *$\{U_\alpha\}$ is an open cover of X,*
- *for each α, $z_\alpha : U_\alpha \to \mathbb{D}$ is a homeomorphism, and*
- *$z_\beta \circ z_\alpha^{-1}$ and $z_\alpha \circ z_\beta^{-1}$ are analytic maps $\mathbb{D} \to \mathbb{D}$.*

In practice we don't often use this sort of definition, but it can be stated with relative ease and demonstrates that in some sense we want Riemann surfaces to be complex manifolds. It also allows us to see a few classical examples explicitly as Riemann surfaces.

Example 3. *For each $\alpha \in \mathbf{C}$, let U_α be the disc of radius 1 centered at α and let z_α be the map $z \mapsto z + \alpha$ so that $z_\alpha^{-1} : z \mapsto z - \alpha$. It follows*

that $z_\beta \circ z_\alpha^{-1}(z) = z + \beta - \alpha$ *and* $z_\alpha \circ z_\beta^{-1} = z + \alpha - \beta$*, which are each clearly analytic maps.*

Example 4. *For each* $\alpha \in \mathbf{C}^\times$*, let* $U_\alpha = \{z \in \mathbf{C} : |z-\alpha| < |\alpha|\}$ *and let* z_α *be the map* $z \mapsto \frac{z}{|\alpha|} + \alpha$*. Therefore* z_α^{-1} *is the map* $z \mapsto |\alpha|(z-\alpha)$*. Once again,* $z_\beta \circ z_\alpha^{-1}$ *and* $z_\alpha \circ z_\beta^{-1}$ *are analytic maps* $\mathbb{D} \to \mathbb{D}$*.*

Example 5. *Let* $S^2 = \{(x,y,z) : x,y,z \in \mathbf{R}, x^2+y^2+z^2 = 1\}$*. Let* $U_{z-} = \{(x,y,z) : z > 0\}$ *and* z_{z-} *be the map* $(x,y,z) \mapsto \frac{x}{1-z} + i\frac{y}{1-z}$*. Let* $U_{y-} = \{(x,y,z) : y < 0\}$ *and* z_{y+} *be the map* $(x,y,z) \mapsto \frac{x}{1+y} + i\frac{z}{1+y}$*. Similarly define* U_α *and* z_α *for* $\alpha = z+, x+, x-,$ *and* $y-$*. Note that* $z_{z-}^{-1} : x+iy \mapsto \left(\frac{2x}{1+x^2+y^2}, \frac{2y}{1+x^2+y^2}, \frac{-1+x^2+y^2}{1+x^2+y^2}\right)$ *and* $z_{y+}^{-1} : x+iy \mapsto \left(\frac{2x}{1+x^2+y^2}, \frac{1-x^2-y^2}{1+x^2+y^2}, \frac{2y}{1+x^2+y^2}\right)$*. Note then that*

$$z_{z-} \circ z_{y+}^{-1}(x+iy) = \frac{2x}{1+x^2+y^2} + i\frac{1-x^2-y^2}{1+x^2+y^2} = \frac{-i(x+iy+i)}{x+iy-i}$$

, which is analytic away from $(x,y) = (0,1)$*.*

Now let's see how this definition gives us complex-analytic functions on Riemann surfaces.

Definition 6. *A complex-analytic (or holomorphic or just* analytic*) function on a Riemann surface* X *with charts* $\{U_\alpha, z_\alpha\}$ *is a function* $f : X \to \mathbf{C}$ *such that* $f \circ z_\alpha^{-1} : \mathbb{D} \to \mathbf{C}$ *is a complex-analytic function.*

This definition reveals the original reason that Riemann surfaces were developed: analytic continuation! We will however use this definition to complete our definition of the category of Riemann surfaces. We have defined the objects, now we will define the morphisms as the analytic maps between Riemann surfaces X and Y.

Definition 7. *We say that a continuous map of Riemann surfaces* $f : X \to Y$ *is* analytic *if for all analytic maps* $\phi : Y \to \mathbf{C}$*, all open sets* $U \subset Y$ *and for all* U_α *intersecting* $f^{-1}(U)$*, the map* $z\alpha^{-1}(f^{-1}(U) \cap U_\alpha) \to \mathbf{C}$ *induced by* ϕ *is analytic. That is to say that for all analytic functions* $\phi : Y \to \mathbf{C}$*,* $\phi \circ f : X \to \mathbf{C}$ *is analytic.*

Now we recall a bit of algebraic topology.

Definition 8. *Let* (X, τ) *be a connected topological space. We say that* X *is simply connected if for all pairs of points* $a, b \in X$*, any two*

continuous paths between a and b in X are homotopic. Equivalently, for any basepoint $x \in X$, $\pi_1(X, x)$ is the trivial group.

As we know, any connected Riemann surface X is both locally path connected and semi-locally simply connected, and thus possesses a universal cover $p : \widetilde{X} \to X$. By the local homeomorphism property of p (and perhaps after modifying our atlas (U_α, z_α)), we have $p^{-1}(U_\alpha) = \coprod_\beta V_{\alpha,\beta}$ and thus an atlas $(V_{\alpha,\beta}, z_\alpha \circ p)$ for $\widetilde{X}$.

Example 9. *The standard example of a universal cover from topology is $\mathbf{R} \to S^1 = \{z \in \mathbf{C} : |z| = 1\}$ by $t \mapsto e^{2\pi i t}$. There is a Riemann surface analogue of this: the analytic map $\mathbf{C} \to \mathbf{C}^\times$ by $z \mapsto e^{2\pi i z}$.*

Example 10. *The identity map $\mathbf{C} \to \mathbf{C}$ is a covering map, and it is unique up to homeomorphism. This is indeed always the case with covering spaces.*

We will thus study Riemann surfaces, starting with their universal covers, i.e., the simply connected Riemann surfaces

We begin in earnest with the following theorem.

Theorem 11 (Uniformization)**.** *If X is a simply connected Riemann surface, then there is an analytic isomorphism between one of the following Riemann surfaces:*

- *The unit disc $\mathbb{D}$, or*
- *The complex plane $\mathbf{C}$, or*
- *The complex projective line $\mathbb{P}^1(\mathbf{C}) \cong S^2$.*

Proving this theorem is hard, but it is easy to see that at the very least these are all distinct. Topologically, we note that S^2 is compact while the other two are not. To see that $\mathbb{D}$ and $\mathbf{C}$ do not admit an analytic isomorphism, consider that any analytic map $\mathbf{C} \to \mathbb{D}$ must be constant by Liouville's thoerem. Therefore no isomorphism is possible.

A sketch for how this goes is to use the theory of harmonic functions to embed X into $\mathbb{P}^1(\mathbf{C})$, at which point we can invoke the Riemann Mapping Theorem on simply connected subregions of $\mathbf{C}$.

We now investigate what Riemann surfaces has each of these as a universal cover.

Theorem 12. *Let X be a Riemann surface.*

- *If $p : \mathbb{P}^1(\mathbf{C}) \to X$ is a universal cover, $X \cong \mathbb{P}^1(\mathbf{C})$.*
- *If $p : \mathbf{C} \to X$ is a universal cover, $X \cong \mathbf{C}^\times$ or X is isomorphic to a closed orientable surface of genus one.*

Proof. The case of $\mathbb{P}^1$ is easy because any deck transformation must be an automorphism of $\mathbb{P}^1$ and thus a linear fractional transformation,

and therefore must have at least two fixed points if it does not fix ∞. Therefore our only deck transformation must be the identity and therefore $X \cong \mathbb{P}^1$.

The case of $\mathbf{C}$ is similar because the linear fractional transformations which fix infinity are the degree one polynomials $az + b$. If there are no fixed points, then there must be no $z \in \mathbf{C}$ such that $az + b = z$. Therefore we must have $a = 1$. This is to say that the only deck transformations must be translations. If our group of deck transformations is isomorphic to $\mathbf{Z}$, then the exponential map provides an isomorphism $X \cong \mathbf{C}^\times$. If our group of deck transformations is isomorphic to $\mathbf{Z}^2$, there is a homeomorphism between X and a genus one surface. If the group of deck transformations is any larger, then X admits a covering map by a genus one surface Σ. If the covering map is finite degree d, $\chi(X) = d\chi(\Sigma)$ by the properties of covering maps. Therefore $\chi(X) = \chi(\Sigma) = 0$ and X is a genus one surface.

If $p : \Sigma \to X$ is a covering map, recall that Σ is compact and X is Hausdorff. Let $x \in X$, and let $x \in U$ such that $\{U_\alpha\}$ is a system of covering sets in Σ such that $p : U_\alpha \cong U$. Complete $\{U_\alpha\}$ to a cover of Σ, pick a finite subcover and let $\alpha_1, \ldots, \alpha_n$ be the α's appearing in that subcover. Since $U_\alpha \cap U_\beta = \emptyset$ for all α, β, $p : \Sigma \to X$ is finite.

□

We can see now that most of our attention should be focused on $\mathbb{D}$, which covers all other Riemann surfaces.

Remark 13. *There is a similar dichotomy throughout all of algebraic geometry. There is the* Fano *case, analogous to being covered by* $\mathbb{P}^1$*. There is the case of Calabi-Yau (and friends), analogous to the case of being covered by* $\mathbf{C}$*, and finally there is the case of "general type," corresponding to being covered by* $\mathbb{D}$*.*

2. Lecture 2: Universal covers and Group actions

Let's consider a few explicit universal covering maps. We have already seen that $\mathbb{P}^1(\mathbf{C})$ and $\mathbf{C}$ form their own universal covers.

Example 14. *The universal cover of* $\mathbf{C}^\times$ *is* $\mathbf{C}$*, given by the map* $\exp\colon \mathbf{C} \to \mathbf{C}^\times$

This should be no surprise, because when we restrict this map to the set $i\mathbf{R} = \{0 + iy : y \in \mathbf{R}\}$, we get the most famous universal cover $\mathbf{R} \to S^1$ and $\mathbf{C}^\times$ is readily homotopy equivalent to S^1. Let's use this to recognize another important covering space.

Example 15. *The universal cover of the punctured unit disc* $\mathbb{D}^\bullet = \mathbb{D} - \{0\}$ *is the left half-plane* $\{x + iy : x, y \in \mathbf{R}, x < 0\}$*, or after multiplying by* $-i$*, the upper half-plane* $\mathcal{H}$*.*

Note that the map $\exp i\colon \mathcal{H} \to \mathbb{D}^\bullet$ is a function $f(z)$ satisfying $f(z + 2\pi n) = f(z)$ for all $n \in \mathbf{Z}$. Clearly by the theory of covering spaces, we may then realize $\mathbb{D}^\bullet$ as a quotient $\mathbf{Z} \setminus \mathcal{H}$. In fact, this example will allow us to go beyond the theory of covering spaces. To gird ourselves for this journey, we recall some properties of group actions.

Let G be a group and let X be a topological space.

Definition 16. *We say that* $\rho : G \times X \to X$ *is a* group action *when*

- *for all* $g \in G$*, the map* $\rho_g : X \to X$ *defined by* $\rho_g(x) = \rho(g, x)$ *is a continuous map,*
- *for all* $g, h \in G$*,* $\rho(h, \rho(g, x)) = \rho(hg, x)$*, and*
- *for all* $x \in X$*,* $\rho(1, x) = x$*.*

Definition 17. *The* orbit *in* X *of a point* $x \in X$ *under* G *is the set* $Gx = \{\rho(g, x) : g \in G\}$*. The* quotient $G \setminus X$ *is the set of* G*-orbits and admits a map* $p : X \to G \setminus X$ *called the* quotient map*. We give* $G \setminus X$ *the weakest topology such that the quotient map is continuous, namely that* $U \subset G \setminus X$ *is open if and only if* $p^{-1}U$ *is open in* X*.*

We can in fact define a quotient space and quotient topology for any equivalence relation. When we take a quotient by a group action however, we have an additional property.

Lemma 18. *The map* $p : X \to G \setminus X$ *is an open map.*

Proof. Let $V \subset X$ be open, so that $p(V) = \{Gx : x \in V\}$, which is open if and only if its preimage is open. Note however that

$$\begin{aligned} p^{-1}(p(V)) &= \{\rho(g, x) : g \in G, x \in V\} \\ &= \bigcup_{g \in G} \{\rho_g(x) : x \in V\} \\ &= \bigcup_g \rho_g(V). \end{aligned}$$

Since ρ_g is a homeomorphism for all g (its inverse is $\rho_{g^{-1}}$), $\rho_g(V)$ is always open, and therefore the union of all of the translates is open. □

Technically, this is a left action, and a right action can be defined analogously. Let us introduce a few common adjectives of topological groups.

Definition 19. *We say that a group action is* free *if for all* $x \in X$*,* $\mathrm{Stab}_G(x) = \{g \in G : \rho_g(x) = x\} = \{1\}$*.*

We recall here that if $p : \tilde{X} \to X$ is a covering map and $\alpha : \tilde{X} \to \tilde{X}$ is a map such that $p \circ \alpha = p$ then we call α a deck transformation of the cover p. Each deck transformation must indeed be a homeomorphism by the properties of covering maps and indeed the set of all deck transformations forms a group under composition and acts freely under the map $\mathrm{Deck}(\tilde{X}/X) \times \tilde{X} \to \tilde{X}$ by $(\alpha, x) \mapsto \alpha(x)$. Note that if p is a universal cover, then the group of deck transformations is naturally isomorphic to $\pi_1(X, x)$ for any $x \in X$.

Suppose then that G is a group acting freely on a topological space X. Does G have to be a deck transformation action? Yes, provided that the action of G keeps points far enough away from one another. We make the following definition.

Lemma 20 (Hatcher, Proposition 1.40)**.** *Suppose that an action of a group G on a path-connected space Y satisfies the following condition: Each $y \in Y$ has a neighborhood U such that for all $g_1, g_2 \in G$, $g_1(U) \cap g_2(U) \neq \emptyset$ implies that $g_1 = g_2$. Then*

1. *The quotient map $p : Y \to G \setminus Y$ is a covering map*
2. *G is the group of deck transformations of p*
3. *p embeds $\pi_1(Y)$ as a normal subgroup of $\pi_1(Y/G)$ and in fact $G \cong \pi_1(G \setminus Y)/p_*(\pi_1(Y))$*

We shall thus incorporate some more topology.

Definition 21. *A topological group G is a group endowed with a topology such that*

$$m : G \times G \to G$$

and

$$()^{-1} : G \to G$$

are continuous.

Definition 22. *If G is a topological group then we say that $\rho : G \times X \to X$ is a* continuous group action *if*

- *ρ is a continuous map for the product topology.*
- *for all $g, h \in G$, $\rho(h, \rho(g, x)) = \rho(hg, x)$, and*
- *for all $x \in X$, $\rho(1, x) = x$.*

We note that if G is a topological group and ρ is a continuous group action, then it is automatically a group action. Indeed, since ρ is continuous, ρ_g decomposes as

$$X \cong \{g\} \times X \hookrightarrow G \times X \xrightarrow{\rho} X,$$

and is thus continuous. One very natural action which arises in this way is the action of a subgroup H of a topological group G on G. In this case, ρ is simply the composition

$$H \times G \hookrightarrow G \times G \stackrel{m}{\to} G.$$

Definition 23. *A topological space X is* locally compact *if for each $x \in X$, there is an open set U containing x whose closure is compact.*

It is easy to show that a closed subspace of a locally compact space is again locally compact.

Definition 24. *Let $\mathrm{SL}_2(\mathbf{R})$ denote the 2×2 real matrices of determinant 1. Give $\mathrm{SL}_2(\mathbf{R})$ the subspace topology from $M_2(\mathbf{R}) \cong \mathbf{R}^4$.*

Let $\mathrm{SO}_2(\mathbf{R})$ denote the subgroup of matrices of the form

$$\kappa_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

In the above, note that the determinant map is polynomial and thus continuous. Since $\mathrm{SL}_2(\mathbf{R})$ is the preimage in $M_2(\mathbf{R})$ of the closed set $\{1\}$ under the determinant map, it is a closed subspace of $M_2(\mathbf{R})$ and is thus locally compact. Note also that $\kappa_\theta \kappa_\tau = \kappa_{\theta+\tau}$ and therefore $\mathrm{SO}_2(\mathbf{R})$ admits a natural identification with $\mathbf{R}/2\pi\mathbf{Z} \cong S^1$.

Lemma 25. *Let $\mathrm{SO}_2(\mathbf{R})$ act by translation on $\mathrm{SL}_2(\mathbf{R})$. The quotient $\mathrm{SL}_2(\mathbf{R})/\mathrm{SO}_2(\mathbf{R})$ is homeomorphic to the upper half-plane $\mathcal{H}$.*

Proof. Starting with an element $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{R})$, multiply on the right by an element of $\mathrm{SO}_2(\mathbf{R})$ to reduce to matrices of the form $\begin{pmatrix} \alpha & -\beta \\ \gamma & 0 \end{pmatrix}$ and send such a matrix to the complex number $\dfrac{\alpha + i\beta}{\gamma}$. It is left to the reader to find the appropriate element of $\mathrm{SO}_2(\mathbf{R})$, check that such a number is in the upper half-plane, check that this map defines a homeomorphism, and that in fact $\dfrac{\alpha + i\beta}{\gamma} = \dfrac{ai + b}{ci + d}$. Typically, the way to prove this is by constructing an action of $\mathrm{SL}_2(\mathbf{R})$ on $\mathcal{H}$ by linear fractional transformations and noting that the stabilizer of i is $\mathrm{SO}_2(\mathbf{R})$. Here we can see how linear fractional transformations spring forth from the natural multiplication action. □

We should note that if we are given a topological space X and an action of a random topological group G, then the topology of X/G will heavily depend on the topology of G in a perhaps quite complicated way. This is one of the topics studied by experts in G-spaces, G-bundles, and equivariant cohomology. This is difficult, so we restrict to the cases where the topology on X is the only topological input.

Definition 26. *We say that an action ρ of a group G is* discontinuous *if when we give G the discrete topology, ρ is a continuous group action.*

Perhaps a less common adjective is that of properness. Still it's important enough to list here.

Definition 27. *We say that a map of topological spaces $f : X \to Y$ is* proper *if for all closed sets C of X, $f(C)$ is closed and for all compact sets K of Y, $f^{-1}(K)$ is compact.*

Definition 28. *We say that a continuous group action $\rho : G \times X \to X$ is* proper *if the map $G \times X \stackrel{\rho \times p_2}{\to} X \times X$ is proper. Here p_2 denotes the projection of the second factor $p_2 : G \times X \to X$.*

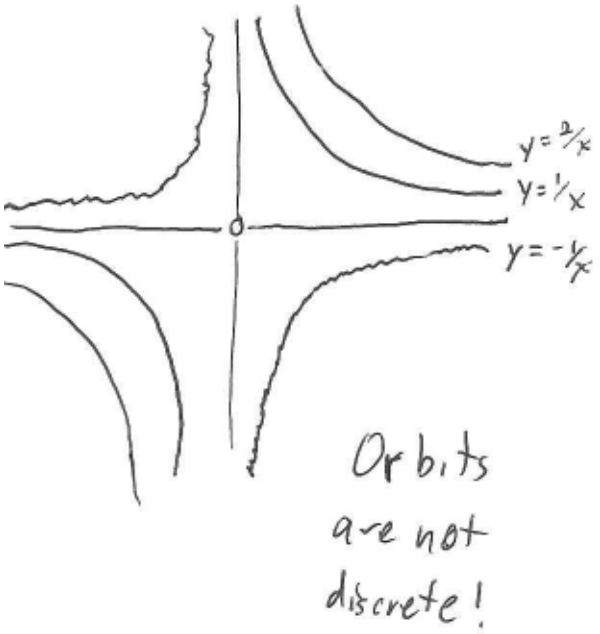
Definition 29. *We say that a continuous group action $\rho : G \times X \to X$ is* properly discontinous *if it is discontinuous and proper.*

We note that simply requiring the covering space condition is not good enough to produce a nice quotient.

Example 30 (Thurston, Geometry and Topology of 3-manifolds, Section 8.2)**.** *The action ρ of $\mathbf{Z}$ on $\mathbf{R}^2 - \{(0,0)\}$ by*

$$\rho(n, (x_0, y_0)) = (2^n x_0, 2^{-n} y_0)$$

is free and for all (x_0, y_0) there exists a U containing (x_0, y_0) such that $nU \cap U = \emptyset$ unless $n = 0$.



We now note one way that properly discontinous actions are special.

Lemma 31. *If a torsion-free group G acts properly discontinouosly on a space X, then the action is free.*

Proof. Let $a : G \times X \to X \times X$ be the action map. Since X is Hausdorff, $\{(x, x)\}$ is a compact subset of $X \times X$ and thus $a^{-1}(x, x)$ is a compact subset of $G \times X$. But then if $p_1 : G \times X \to G$ is the projection map, then $p_1(a^{-1}(x, x))$ is the stabilizer in G of X. Since the continuous image of a compact set is compact and G is discrete, the stabilizer must be a finite subgroup of G. Consider however that all finite groups are torsion, while the torsion subgroup of G is exactly $\{1\}$ by torsion-freeness. Therefore, the stabilizer in G of x must consist of only the identity, so the action of G on X is free. □

Now we give the reason we have introduced this terminology.

Lemma 32. *If X is Hausdorff and $\rho : G \times X \to X$ is a proper discontinuous group action, then $G \setminus X$ is Hausdorff.*

Proof. This proof was gleefully stolen from Brian Conrad's Differential Geometry notes. A space Y is Hausdorff if and only if the image of the diagonal map $Y \to Y \times Y$ is closed. Let S be the image of the diagonal map $(G \setminus X) \to (G \setminus X) \times (G \setminus X)$. Since the quotient map $X \to G \setminus X$ is open, so is the map $X \times X \to (G \setminus X) \times (G \setminus X)$ in the product topology. Therefore S is closed if and only if its preimage in $X \times X$ is closed.

What then is the preimage of S? It is the set $\{(x, x') : Gx = Gx'\}$, or in fact the image of the map $G \times X \to X \times X$ by $(g, x) \mapsto (gx, x)$. By properness, this is a closed map, showing that S is closed and completing the proof.

□

Theorem 33. *Let G be a locally compact Hausdorff group, K a compact subgroup of G, and H a closed subgroup of G. There is a naturally induced action of G on G/K by multiplication and thus of any subgroup of H on G/K. The action of H is properly discontinuous if and only if H is discrete in G.*

Proof. We sketch the proof here.

First, we may show easily that the multiplication action of G on itself is proper, even if G is not locally compact or Hausdorff. Please notice that this only means that the action map $A : G \times G \to G \times G$ by $(x, y) \mapsto (xy, y)$ is proper, not necessarily the multiplication map $m : G \times G \to G$. In fact, the multiplication map is only guaranteed to be proper if G is a compact group. A hint for this is to prove that the

action map is not only proper, but holds many other nice properties as well.

Second, suppose there is an action $\rho : G \times X \to X$ which is proper, $\sigma : G \times Y \to Y$ is an action, and $f : X \to Y$ is a proper map such that $\sigma(g, f(x)) = \rho(g, x)$ (or f is G-equivariant). It follows that σ is proper, so the action of G on G/K is proper.

Finally, if $H \to G$ is a closed embedding, then $H \times X \to G \times X$ is also a closed embedding. Since closed embeddings are proper maps and compositions of proper maps are proper, the proof is complete. □

Corollary 34. *A subgroup Γ of $\mathrm{SL}_2(\mathbf{R})$ acts properly discontinously on $\mathcal{H}$ if and only if Γ is discrete.*

We should note that although $\mathcal{H} \cong \mathrm{SL}_2(\mathbf{R})/\mathrm{SO}_2(\mathbf{R})$, this does not mean that an element γ of $\mathrm{SO}_2(\mathbf{R})$ acts trivially on $\mathcal{H}$. For instance, the transformation $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is a nonidentity element of $\mathrm{SO}_2(\mathbf{R})$ but acts nontrivially. The point is that elements of $\mathrm{SO}_2(\mathbf{R})$ fix the point i. Likewise, if T sends i to a point w then the transformation TST^{-1} acts nontrivially on $\mathcal{H}$ but fixes w. As a different example, the transformation $\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$ fixes the point $\dfrac{1 + i\sqrt{3}}{2}$ in $\mathcal{H}$ but acts nontrivially. In general, the only transformations which act trivially on $\mathcal{H}$ are the scalar diagonal matrices and the topic of fixed points will be taken up in earnest later. For now, we have put off the following definition for long enough.

Definition 35. *We say that a subgroup of $\mathrm{SL}_2(\mathbf{R})$ is Fuchsian if it is discrete.*

We now state the theorem which is the main goal of this lecture.

Theorem 36. *If Γ is a finitely generated Fuchsian group, then $\Gamma \setminus \mathcal{H}$ is a Riemann surface.*

The hypothesis of finite generation is certainly an annoying one, but as we will see next time, it comes for free with a certain more natural assumption that we can make on our group.

We proceed as follows. First we show that properly discontinous actions are almost free. Then we show that if $p : \tilde{X} \to X$ is a covering space and $\tilde{X}$ is a Riemann surface, then X is a Riemann surface. Finally we show that if $p : X \to Y$ is an analytic quotient by a finite group and X is a Riemann surface, then so is Y. Let us begin with a theorem we will not prove.

Theorem 37 (Selberg's Lemma)**.** *Let K be a field of characteristic zero, $n > 0$ and $\Gamma < \mathrm{GL}_n(K)$ a finitely generated subgroup. Then there is a torsion-free finite-index normal subgroup Γ_1 of Γ.*

You should be chuckling at the word "Lemma" here.

Sketch: Let L be a finitely generated field of characteristic zero and $\{x_i\}$ a finite generating set. It was shown by Cassels that there exist infinitely many primes p such that there exists an embedding $\iota : L \hookrightarrow \mathbf{Q}_p$ with $\iota(x_i) \in \mathbf{Z}_p$. Pick a generating set $\{g_j\}$ for Γ which is stable under inversion. Let $\{x_i\} \subset K$ be the coordinates of the g_j's and let $L = \mathbf{Q}(\{x_i\})$. By abuse of notation, let $\iota : \mathrm{GL}_n(L) \hookrightarrow \mathrm{GL}_n(\mathbf{Q}_p)$ be an embedding as above. Note that $\iota(\Gamma) < \mathrm{GL}_n(\mathbf{Z}_p)$. Use the fact that $\ker(\mathrm{GL}_n(\mathbf{Z}_p) \to \mathrm{GL}_n(\mathbf{Z}/(p)))$ is a finite-index normal subgroup, which is torsion-free for $p > 2$. □

Lemma 38. *If $p : \tilde{X} \to X$ is a covering map and $\tilde{X}$ is a Riemann surface, then so is X.*

Proof. Let $\{(U_\alpha, z_\alpha)\}$ be an atlas for $\tilde{X}$ and let V be an open set of X whose preimages V_β are mutually disjoint. Then on V_β and thus on $W_{\alpha\beta} = U_\alpha \cap V_\beta$, p is a homeomorphism. However, via z_α, $W_{\alpha\beta}$ is identified with an open subset of $\mathbb{D}$. If it is a simply-connected open subset then we are done by the Riemann mapping theorem. In case we are not, recall that a base for the topology on $\mathbb{D}$ is given by discs in $\mathbb{D}$ with rational length and rational centers. Cover $W_{\alpha\beta}$ by these discs $D_{\alpha\beta\gamma}$ and let $\phi_{\alpha\beta\gamma}$ scale and translate these discs (considered now in $\mathbf{C}$) to $\mathbb{D}$ so that $\{(p(D_{\alpha\beta\gamma}), z_\alpha \phi^{-1}_{\alpha\beta\gamma} p^{-1})\}$ is an atlas for $\tilde{X}$. □

Lemma 39. *Let X be a Riemann surface and G be a finite group acting on X by complex analytic automorphisms. Then $Y = G \setminus X$ can be given the structure of a Riemann surface in such a way that the quotient map is holomorphic.*

Proof. (For now I'm attributing this to Robert Varley, although this is certainly a well-known and classical result)

Without loss of generality, assume that Y is connected, so that X has finitely many components. If g does not act trivially on any component, then the fixed points of g are discrete. To see this, assume not, i.e. that x is a limit point of some fixed points of g. Pick a neighborhood U of x such that $U \cong \mathbb{D}$, and therefore there is a sequence of fixed points of g tending to x inside of U. Therefore g defines a holomorphic map from U to some open subset of $\mathbf{C}$ which is the identity on a sequence of points tending to x. By the identity theorem of complex analysis, g must be the identity in some neighborhood of x, and therefore on the

whole connected component of x. This is to say that g acts trivially on a component of X.

Let R be a finite set of fixed points of some $g \in G$ and set $B = p(R)$. Since Y is Hausdorff, it suffices to treat the case $B = \{y_0\}$ and $R = \{\rho(g, x_0) : g \in G\}$.

Let $\{1, g_2, \ldots, g_r\}$ be coset representatives for $H = \mathrm{Stab}_G(x_0)$ so that $R = \{x_0, g_2 x_0, \ldots, g_r x_0\}$. Take an open neighborhood U of x_0 admitting a holomorphic isomorphism $\mathcal{H} \to U$ sending i to x_0 and such that $U, g_2 U, \ldots, g_r U$ are mutually disjoint. Since p is an open mapping, $V = p(U)$ is an open neighborhood of y_0 in Y. Now if we consider the induced action of the finite group H on $\mathcal{H}$, we find that it stabilies i and thus $H \subset \mathrm{Stab}_{\mathrm{SL}_2(\mathbf{R})}(i) \cong \mathrm{SO}_2(\mathbf{R})$.

However, every finite subgroup of $\mathrm{SO}_2(\mathbf{R}) \cong \mathbf{R}/2\pi\mathbf{Z}$ is cyclic, so H has a generator. Now identify $\mathcal{H}$ with $\mathbb{D}$ via the Cayley transformation mapping i to 0. Let $e = \#H$ and let ζ denote a primitive e-th root of unity so that the action of H on $\mathbb{D}$ is generated by the map $z \mapsto \zeta z$. Identify $H \setminus \mathbb{D}$ with $\mathbb{D}$ by the H-invariant map $z \mapsto z^e$, and use this copy of $\mathbb{D}$ to define our complex mapping on V. □

3. Lecture 3: Metrics Volumes, and Measures

In our first lecture, we found that if we want to study interesting Riemann surfaces, we should take quotients of $\mathcal{H}$. In a sense, we smelled the porridge at that point. In our second lecture, we found that if we wanted to take a quotient of the upper-half plane by a subgroup of $\mathrm{SL}_2(\mathbf{R})$ and still get a Riemann surface, we should focus our attention on Fuchsian groups, or risk getting a non-Hausdorff space at the end. Since Fuchsian in this context means only discrete, or "small" in some sense, we ignore the groups which are "too big," or in other words, "that porridge is too hot."

In this lecture, we will see that simply being Fuchsian is not restrictive enough to avoid pathologies. It is possible for our Fuchsian groups to be "too cold," or more accurately, "too small," so that their quotients are too big to be of regular use to us. For us, the porridge which is "just right," will be those Fuchsian groups of finite covolume. Come with me and avoid the bears!

Definition 40. *The complex upper half-plane $\mathcal{H}$ can be equipped with a hyperbolic metric*

$$\mathrm{d}s = \frac{\sqrt{\mathrm{d}x^2 + \mathrm{d}y^2}}{y}.$$

This metric induces the hyperbolic area or volume $\dfrac{\mathrm{d}x\mathrm{d}y}{y^2}$.

Theorem 41 (Katok, Theorem 1.2.6(i)). *For $z \neq w \in \mathcal{H}$, the hyperbolic distance between the two is given by*

$$\delta(z,w) = \ln\left(\frac{|\, z-\bar{w}\,| + |\, z-w\,|}{|\, z-\bar{w}\,| - |\, z-w\,|}\right).$$

The geodesics for this metric are the vertical lines and circles centered on the real axis.

While the study of such metrics is far from our main concern, this metric carries a deep connection to Fuchsian groups as we see in the following. We recall first that the matrix $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ acts trivially on $\mathcal{H}$. There is thus no harm in working with the coindex two quotient $\mathrm{PSL}_2(\mathbf{R})$, and indeed reductions and lifts of discrete subgroups are again discrete subgroups.

Definition 42. *A subgroup Γ of $\mathrm{PSL}_2(\mathbf{R})$ is Fuchsian if it is discrete.*

Theorem 43 (Katok, Theorem 1.3.1). *The group of orientation preserving isometries of $\mathcal{H}$ with the hyperbolic metric is given by $\mathrm{PSL}_2(\mathbf{R})$. This is an index 2 subgroup of the full group of isometries of $\mathcal{H}$, and the nontrivial coset includes the mapping $z \mapsto -\overline{z}$.*

Of course we know that matrices in $\mathrm{PSL}_2(\mathbf{R})$ induce conformal mappings on $\mathcal{H}$, so we even have the following.

Lemma 44. *Let $\{U_\alpha, z_\alpha\}$ be a set of holomorphic charts of $\mathcal{H}$ (or a set V thereof) such that each U_α is a circle in $\mathcal{H}$. If $T \in \mathrm{PSL}_2(\mathbf{R})$ then $\{T(U_\alpha), z_\alpha \circ T^{-1}\}$ is another such set of holomorphic charts of $\mathcal{H}$ (or of $T(V)$).*

This lemma is how we will take our Fuchsian group quotients $\Gamma \setminus \mathcal{H}$, for now only locally compact Hausdorff topological spaces, and give each of them the structure of a Riemann surface. We will now formalize the process of finding subsets V of $\mathcal{H}$ whose points will be coset representatives of $\Gamma \setminus \mathcal{H}$.

Definition 45. *We say that a Fundamental domain for a Fuchsian group Γ is a subset F of $\mathcal{H}$ which is*

- *The closure in $\mathcal{H}$ of an open subset F^o of $\mathcal{H}$,*
- *satisfying $\bigcup_{\gamma\in\Gamma} \gamma F = \mathcal{H}$, and*
- *satisfying $\gamma(F^o) \cap F^o = \emptyset$ for all nonidentity $\gamma \in \Gamma$.*

Lemma 46 (Katok, Theorem 3.1.1). *If Γ is a Fuchsian group, F_1, F_2 are fundamental domains for Γ, the area of F_1 is finite, and $\partial F_1, \partial F_2$ have area zero then the area of F_2 is equal to the area of F_1.*

Proof.

$$F_1 \supset F_1 \cap \left(\bigcup_{T \in \Gamma} T(F_2^o) \right) = \bigcup_{T \in \Gamma} (F_1 \cap T(F_2^o))$$

Since F_2 is a fundamental domain, the elements of $\{F_1 \cap T(F_2^o)\}$ are mutually disjoint. Let μ denote the hyperbolic area, and thus

$$\mu(F_1) \geq \sum_{T \in \Gamma} \mu\left(F_1 \cap T(F_2^o)\right) = \sum_{T \in \Gamma} \left(T^{-1}(F_1) \cap F_2^o\right) = \sum_{T \in \Gamma} \left(T(F_1) \cap F_2^o\right).$$

Since F_1 is a fundamental domain,

$$\bigcup_{T \in \Gamma} (T(F_1)) = \mathcal{H}$$

and thus

$$\bigcup_{T \in \Gamma} (T(F_1) \cap F_2^o) = F_2^o$$

. Therefore $\mu(F_1) \geq \mu(F_2^o) = \mu(F_2)$. But then in the above argument, we can start with F_2 and obtain $\mu(F_2) \geq \mu(F_1)$. □

Therefore if Γ has a fundamental domain with finite area and measure zero boundary, there is a well-defined area associated to Γ. Note however that if $\Gamma < \mathrm{PSL}_2(\mathbf{R})$ is Fuchsian, then it has a finite index normal torsionfree subgroup, and any nonidentity element of Γ acts nontrivially on $\mathcal{H}$. Therefore, the fixed points of Γ are discrete in $\mathcal{H}$ and so in any compact subset of $\mathcal{H}$ we may find a point P which is not fixed by any nonidentity element of Γ.

Definition 47. *The Dirichlet domain for Γ centered at P is the set*

$$D_P(\Gamma) = \{z \in \mathcal{H} : \delta(P, z) \leq \delta(P, T(z)) \forall T \in \Gamma\}.$$

4. Lecture 4:Siegel's Theorem

Example 48. *For the cyclic Fuchsian group $\langle T \rangle$ with $T : z \mapsto 2z$ we compute the Dirichlet domain centered at $P = i$. It is easy to compute that the midpoint between i and $2i$ is $\sqrt{2}i$ and the midpoint of i and $1/2i$ is $\frac{1}{\sqrt{2}}i$. In fact, the circles centered at zero of radii $\sqrt{2}$ and $\frac{1}{\sqrt{2}}$ form the boundary of this Dirichlet domain. In general, if $r \in \mathbf{R}_{>0}$ then the set of points $\{z : \delta(z, i) = \delta(z, r^2 i)\}$ is the circle of radius r centered at zero.*

To see this, recall that our formula for δ gives that $\cosh(\delta(z, w))$ for $z, w \in \mathcal{H}$ must be (via a somewhat hard calculation)

$$1 + \frac{|z - w|^2}{2\Im(z)\Im(w)}.$$

Therefore the set of $z = x + iy$ such that $\delta(z, i) = \delta(z, r^2 i)$ is the set of z such that

$$\frac{|z-i|^2}{2y} = \frac{|z-r^2 i|^2}{2r^2 y},$$

which is equal after expanding to $x^2 + y^2 = r^2$.

We note the structure of a Dirichlet domain here. If $T \in \Gamma$ is not the identity, then let $L_P(T)$ denote the geodesic between P and $T(P)$, and on that geodesic, let Q denote the midpoint between P and $T(P)$. Let ψ denote the linear fractional transformation sending i to P, 0 to Q and $-i$ to $T(P)$. Let $H_P(T) = \psi^{-1}(\mathcal{H})$ and because $P \in H_P(T)$, $D_P(\Gamma)^o \subset H_P(T)$. In fact,

$$D_P(\Gamma) = \overline{\bigcap_{T \in \Gamma - \{0\}} H_P(T)}.$$

Lemma 49 (Katok, Theorem 3.2.2)**.** *For P not fixed by any nonidentity element of Γ, $D_P(\Gamma)$ is a connected, convex (and thus boundary measure zero) fundamental domain for Γ.*

Proof. Let $z \in \mathcal{H}$, so that by proper discontinuity Γz is discrete in $\mathcal{H}$. Therefore we can find some $z_0 \in \Gamma z$ which is closest to P in the hyperbolic metric. By definition, $\delta(z_0, P) \leq \delta(T(z_0), P)$ for all $T \in \Gamma$ and so $z_0 \in D_P(\Gamma)$. In fact, as we will see, if $z_0 \in D_P(\Gamma)^o$ then for all nonidentity $T \in \Gamma \subset \mathrm{PSL}_2(\mathbf{R})$, $\delta(z_0, P) < \delta(T(z_0), P)$.

If in fact $z \in \mathcal{H}$ and $T \in \Gamma$ such that $\delta(z, P) = \delta(T(z), P)$ then $\delta(z, P) = \delta(z, T^{-1}P)$. Therefore, z is on the boundary of $H_P(T^{-1})$. If additionally $z \in D_P(\Gamma)$, then it must lie on the boundary of $D_P(\Gamma)$. Therefore if $z_1, z_2 \in D_P(\Gamma)^0$ then z_1 and z_2 cannot be in the same orbit. □

Definition 50. *We say that Γ is a Fuchsian group of finite covolume if it has a fundamental domain of finite volume in $\mathcal{H}$.*

We use this terminology because in fact, the measure induced by $\dfrac{\mathrm{d}x\mathrm{d}y}{y^2}$ descends down to $\Gamma \setminus \mathcal{H}$. For a proof of this, see section 2.5 of Shimura's book. Therefore, $\Gamma \setminus \mathcal{H}$ has a natural compact closure. Finite covolume Fuchsian groups are important for many reasons. As we will see, they are finitely generated!

To show this first notice that we have essentially shown that a Dirichlet domain is essentially polygonal, with geodesic sides perpendicular to the lines $L_p(T)$ for $T \in \Gamma$. Call these lines $B_P(T)$ as they are perpendicular bisectors of $L_P(T)$. Clearly T moves the line $L_P(T)$ to $L_{T(P)}(T)$

and so on. In fact, the elements of Γ can pair one side of a fundamental domain to another.

Lemma 51 (Katok, Theorem 3.5.4)**.** *Let $\{T_i\}$ be the subset of Γ consisting of those elements which pair the sides of a fixed fundamental domain F. Then $\{T_i\}$ is a set of generators for Γ.*

Proof. Let Σ denote the subgroup generated by the side-pairing elements T_i. Let $S_1 \in \Sigma$ and let $S_2 \in \Gamma$ such that $S_1(F)$ shares a side with $S_2(F)$. Then F shares a side with $S_1^{-1}S_2(F)$ and thus $S_1^{-1}S_2 = T_k$ for some k by definition. Therefore $S_2 = S_1T_k$ and since S_1 is already a product of T_i's, S_2 is a product of T_i's and therefore $S_2 \in \Gamma$.

In fact, if $S \in \Sigma$ and $T \in \Gamma$ such that $T(F) \cap S(F) \neq \emptyset$ then $T \in \Gamma$. We have already taken care of the case of a side, now we take care of the case of a vertex v shared by $T(F)$ and $S(F)$. By Proper discontinuity, for v a vertex of $S(F)$ there are only finitely many $T \in \Gamma$ such that $v \in T(F)$. Then F intersects $S^{-1}T(F)$ in $u = S^{-1}v$. The above argument applied possibly more than once yields $T \in \Sigma$.

It follows that $\bigcup_{S\in\Sigma} S(F)$ and $\bigcup_{T\in\Gamma-\Sigma} T(F)$ are disjoint sets, and since F is a fundamental domain, the union of these two sets is $\mathcal{H}$. It follows that the first of these is nonempty and both are closed, so by the connectedness of $\mathcal{H}$, $\Gamma - \Sigma$ is empty and thus $\Gamma = \Sigma$.

We note that both are closed, because if $z \in \mathcal{H}$ is a limit point of a sequence $\{z_j\}$ in a union of $T(F)$'s, then any compact neighborhood meets only finitely many of the $T(F)$'s and any finite union of $T(F)$'s is closed, thus containing all of their limit points. □

As a consequence, if there is a fundamental domain with *finitely many* sides, then Γ is finitely generated. We now show that this is frequently the case.

Theorem 52 (Siegel's Theorem, Katok, Theorem 4.1.1)**.** *If Γ has finite covolume, then there is a fundamental domain for Γ with finitely many sides.*

Proof. The key is to show that if $\{\omega\}$ is the set of angles of a fundamental domain F, $\sum_\omega(\pi - \omega) \leq 2\pi + \mu(F)$. □

Example 53. *Taking area into account, we see that if Γ is a Fuchsian group of finite covolume, we have a natural compactification of $\Gamma \setminus \mathcal{H}$. We will explore how to explicitly form that quotient using the theory of cusps.*

5. Lecture 5: Quaternion Algebras

Throughout this lecture, let F be a field of characteristic not equal to two. In the case of characteristic two, we have other definitions for quaternion algebras, but we'll never really work outside of number fields or other fields of characteristic zero.

Definition 54. *Let $a, b \in F^\times$. Define $\left(\frac{a,b}{F}\right)$ to be the associative F-algebra whose elements are of the form $x + iy + jz + kw$ such that $x, y, z, w \in F$, $i^2 = a$, $j^2 = b$, and $ij = k = -ji$. A quaternion algebra over F is one of the form $\left(\frac{a,b}{F}\right)$ for some $a, b \in F^\times$.*

Example 55. *It is known that the unique noncommutative associative $\mathbf{R}$-algebra without zero divisors is Hamilton's quaternions $\left(\frac{-1,-1}{\mathbf{R}}\right)$.*

Note that there could be many choices of $a, b \in F^\times$ which give the same F-algebra up to isomorphism. It can be shown that independently of that choice of a and b, there is a canonical involution - or order two automorphism - of $\left(\frac{a,b}{F}\right)$. This takes the form

$$\overline{x + iy + jz + kw} = x - iy - jz - kw.$$

Definition 56. *If $\alpha = x+iy+jz+kw$ lies inside a quaternion algebra $\left(\frac{a,b}{F}\right)$, we define the reduced trace of α to be $t(\alpha) = \alpha + \overline{\alpha} = 2x$ and the reduced norm of α to be $n(\alpha) = \alpha\overline{\alpha} = x^2 - ay^2 - bz^2 + abw^2$.*

Note that the reduced trace is additive and the reduced norm is multiplicative.

Lemma 57. *If $a \in F^{\times 2}$, then $\left(\frac{a,b}{F}\right) \cong M_2(F)$. If $\left(\frac{a,b}{F}\right) \ncong M_2(F)$ then $\left(\frac{a,b}{F}\right)$ is a division algebra, i.e. every nonzero element has an inverse.*

Proof. The isomorphism is an easy one, and explicitly described as

$$\begin{aligned} 1 &\mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ i &\mapsto \begin{pmatrix} \sqrt{a} & 0 \\ 0 & -\sqrt{a} \end{pmatrix} \\ j &\mapsto \begin{pmatrix} 0 & 1 \\ b & 0 \end{pmatrix} \\ k &\mapsto \begin{pmatrix} 0 & \sqrt{a} \\ -b\sqrt{a} & 0 \end{pmatrix}. \end{aligned}$$

For the statement on division algebras, note that $\left(\frac{a,b}{F}\right)$ is division if and only if for all $\alpha \neq 0$, $n(\alpha) \neq 0$. Clearly if $\left(\frac{a,b}{F}\right) \ncong M_2(F)$ then $F(i)$ is a quadratic field extension of F. On the quadratic extension $F(i)$, the canonical involution is simply Galois conjugation, so the reduced norm restricts to the standard multiplicative field norm $N_{F(i)/F}\colon F(\beta) \to F$. The exact same proof shows that $F(i)$ is a division algebra if and only if the field norm is nondegenerate.

Let $\alpha = x + iy + jz + kw$ be such that $0 = n(\alpha) = x^2 - ay^2 - bz^2 + abw^2 = N_{F(i)/F}(x+iy) - bN_{F(i)/F}(z+iw)$. If $N_{F(i)/F}(z+iw) \neq 0$, then

$$b = \frac{N_{F(i)/F}(x+iy)}{N_{F(i)/F}(z+iw)} = N_{F(i)/F}(\frac{x+iy}{z+iw})$$

because the field norm is nondegenerate and thus $z+iw \neq 0$. Since $F(i)$ is a field, we can express that fraction as $q+ir$ for $q, r \in F$. Therefore $b = q^2 - ar^2$ and we can give an isomorphism $\left(\frac{a,b}{F}\right) \to M_2(F)$ by

$$\begin{aligned} 1 &\mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ i &\mapsto \begin{pmatrix} 0 & 1 \\ a & 0 \end{pmatrix} \\ j &\mapsto \begin{pmatrix} q & -r \\ ar & -q \end{pmatrix} \\ k &\mapsto \begin{pmatrix} ra & -q \\ qa & -ra \end{pmatrix}. \end{aligned}$$

This is impossible though, because $M_2(F) \ncong \left(\frac{a,b}{F}\right)$. It follows that $N_{F(i)/F}(z+iw) = 0$. If $N_{F(i)/F}(z+iw) = 0$ then $z = w = 0$, $N_{F(i)/F}(x+iy) = 0$ and $x = y = 0$, so $\alpha = 0$.

□

We note a few consequences to this proof.

Porism 58. *If $a, b \in F^\times$, then $\left(\frac{a,b}{F}\right) \cong M_2(F)$ if and only if there exist $x, y \in F$ such that $b = x^2 - ay^2$.*

Proof. If $a \notin F^{\times 2}$, the proof above explicitly computes this result. If not, set $1 = x - y\sqrt{a}$ and $b = x + y\sqrt{a}$, thus $x = 1 + y\sqrt{a}$, $y = \frac{b-x}{\sqrt{a}} = \frac{b-1}{\sqrt{a}} - y$ or $y = \frac{b-1}{2\sqrt{a}}$. Therefore $x = \frac{b+1}{2}$, $y = \frac{b-1}{2\sqrt{a}}$ work as values. □

Note also that if $\left(\frac{a,b}{F}\right) \cong M_2(F)$ then under the described isomorphism, the canonical involution is transformed into the classical adjoint

$$\begin{pmatrix} x & y \\ z & w \end{pmatrix} \mapsto \begin{pmatrix} w & -y \\ -z & x \end{pmatrix}.$$

Therefore,

$$t\begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} x+w & 0 \\ 0 & x+w \end{pmatrix} = x + w = \operatorname{tr}\begin{pmatrix} x & y \\ z & w \end{pmatrix},$$

and

$$n\begin{pmatrix} x & y \\ z & w \end{pmatrix} = \begin{pmatrix} xw-yz & 0 \\ 0 & xw-yz \end{pmatrix} = xw - yz = \det\begin{pmatrix} x & y \\ z & w \end{pmatrix}.$$

We now address the relation of quaternion algebras to Fuchsian groups. If F admits an embedding $\iota : F \hookrightarrow \mathbf{R}$ under which $\iota(a)$ (respectively $\iota(b)$) is positive, then $\sqrt{\iota(a)} + 1$ (resp. $\sqrt{\iota(b)} + 1$) is a nonzero element of norm zero in $\left(\frac{\iota(a),\iota(b)}{\mathbf{R}}\right)$. Therefore ι defines an embedding $\iota : \left(\frac{a,b}{F}\right) \hookrightarrow \left(\frac{\iota(a),\iota(b)}{\mathbf{R}}\right) \cong M_2(\mathbf{R})$.

We now restrict our attention to the case $F = \mathbf{Q}$, which has a unique embedding into $\mathbf{R}$. If we do not do this, we can run into technical difficulties such as the following.

Example 59. *Let $F = \mathbf{Q}(\sqrt{2})$ and let $\iota_1(\sqrt{2}) \approx 1.414\ldots$, $\iota_2(\sqrt{2}) \approx -1.414\ldots$. It is perfectly reasonable to consider $B = \left(\frac{\sqrt{2}-1,-1}{\mathbf{Q}(\sqrt{2})}\right)$. Note however that ι_1 takes B into $M_2(\mathbf{R})$ while ι_2 takes it into Hamilton's quaternions. There are also fields such as $\mathbf{Q}(\sqrt{-1})$ which admit no real embeddings, or fields like $\mathbf{Q}(\sqrt[3]{2})$ which don't admit the full number of embeddings. Each of these phenomena pose a different technical problem from the perspective of Fuchsian groups.*

We note now that there is no embedding of $\left(\frac{a,b}{\mathbf{Q}}\right)$ into $M_2(\mathbf{R})$ precisely when $a, b < 0$. If A is a subring of a quaternion algebra, let A^1 denote the elements of norm one. If we let $a, b \in \mathbf{Q}^\times$, $a > 0$ then $\left(\frac{a,b}{\mathbf{Q}}\right)^1 \hookrightarrow \mathrm{SL}_2(\mathbf{R})$.

Of course, if A is a quaternion algebra over $\mathbf{Q}$, then there A is essentially $\mathbf{Q}^4$ inside of $\mathbf{R}^4$ and A^1 is essentially $\mathbf{Q}^3$ inside of $\mathbf{R}^3$ - certainly not discrete! We introduce the following workaround.

Definition 60. *Let A be a quaternion algebra over $\mathbf{Q}$. We say that $\alpha \in A$ is integral if $t(\alpha), n(\alpha) \in \mathbf{Z}$. A $\mathbf{Z}$-order(or just an order) $\mathcal{O} \subset A$ is a subring which admits generators $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ as a $\mathbf{Z}$-algebra which are integral and $\mathbf{Q}$-linearly independent. A maximal order is an order which is maximal with respect to inclusion.*

Example 61. *Inside of the rational Hamilton quaternions $\left(\frac{-1,-1}{\mathbf{Q}}\right)$, $\mathcal{O} = \mathbf{Z} \oplus i\mathbf{Z} \oplus j\mathbf{Z} \oplus k\mathbf{Z}$ is an order. It is however, not a maximal order. Hurwitz found that if we take*

$$\alpha = \frac{1+i+j+k}{2},$$

then $t(\alpha) = 1$ and $n(\alpha) = 2$. He also found that the $\mathbf{Z}$-module $\mathbf{Z} \oplus i\mathbf{Z} \oplus j\mathbf{Z} \oplus \alpha\mathbf{Z}$ is a ring, and in fact a maximal order.

To a student of algebraic number theory, one might think it natural to consider the set of all integral elements of a quaternion algebra B. Unfortunately, non-commutativity gets in the way and this set is usually not a subring.

Example 62. *Consider the following elements of the quaternion algebra $M_2(\mathbf{Q})$:*

$$A = \begin{pmatrix} 1/2 & -3 \\ 1/4 & 1/2 \end{pmatrix}, B = \begin{pmatrix} 0 & -5 \\ 1/5 & 0 \end{pmatrix}.$$

Note that both A and B and integral but neither $A+B$ nor AB are integral. Therefore, while there are orders which contain A and orders which contain B, no order can contain both. In fact this would follow from either of $A+B$ or AB not being integral.

Now we investigate the properties of the trace map $t : \alpha \mapsto \alpha + \overline{\alpha}$. We can use it to define a bilinear form on our quaternion algebra A over F. If $\alpha, \beta, \gamma, \delta \in A$, then

$$t((\alpha+\gamma)\beta) = t(\alpha\beta) + t(\gamma\beta)$$

and

$$t(\alpha(\beta+\delta)) = t(\alpha\beta) + t(\alpha\delta).$$

Consider $A = \left(\frac{a,b}{F}\right)$ for some $a, b \in F^\times$. Either over F or $F(\sqrt{a})$,

$$t(\alpha\beta) = \mathrm{tr}(\alpha\beta) = \mathrm{tr}(\beta\alpha) = t(\beta\alpha).$$

Therefore the trace pairing is in fact a symmetric bilinear pairing, so we can think of it as a sort of inner product. With that in mind, we consider the following generalization of the Gram matrix and its determinant.

Definition 63. *Let $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ be a basis for an order $\mathcal{O}$ as a $\mathbf{Z}$-module. Consider the matrix M whose entry in the i-th row and j-th column is $t(\alpha_i\overline{\alpha_j})$. If $\beta_1, \ldots, \beta_4$ is another basis for $\mathcal{O}$ as a $\mathbf{Z}$-module, and N is the matrix whose entry in the i-th row and j-th column is $t(\beta_i\overline{\beta_j})$. Let J be the matrix expressing $\{\beta_i\}$ in terms of $\{\alpha_i\}$, so that $\det(M) = \det(J)^2 \det(N)$. Since these are each bases, $J \in \mathrm{GL}_4(\mathbf{Z})$ and we can define the* non-reduced discriminant *of $\mathcal{O}$ as the determinant of M.*

Example 64. *In the above, if we let $A = \left(\frac{-1,-1}{\mathbf{Q}}\right)$ and $\mathcal{O} = \mathbf{Z} \oplus i\mathbf{Z} \oplus j\mathbf{Z} \oplus k\mathbf{Z}$ then the matrix M produced above is*

$$\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}.$$

Therefore the non-reduced discriminant of $\mathcal{O}$ is 16.
For the Hurwitz quaternions the matrix is

$$\begin{pmatrix} 2 & 0 & 0 & 1 \\ 0 & 2 & 0 & 1 \\ 0 & 0 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix},$$

therefore the non-reduced discriminant is 4.

Lemma 65. *If B is a quaternion algebra over $\mathbf{Q}$, the non-reduced discriminant is a square integer. Let $\mathrm{disc}(\mathcal{O})$ be the square root of the non-reduced discriminant. If $\mathrm{disc}(\mathcal{O})$ is square-free then $\mathcal{O}$ is maximal.*

Theorem 66. *If $\mathcal{O}_1$, $\mathcal{O}_2$ are maximal orders in a quaternion algebra B over $\mathbf{Q}$ which embeds into $M_2(\mathbf{R})$, then there is some $\phi \in B$ such that $\mathcal{O}_1 = \phi^{-1}\mathcal{O}_2\phi$.*

Example 67. *It is easy to show that $M_2(\mathbf{Z})$ is a maximal order in $M_2(\mathbf{Q})$. Another maximal order is of the form*

$$\begin{pmatrix} \mathbf{Z} & 1/p\mathbf{Z} \\ p\mathbf{Z} & \mathbf{Z} \end{pmatrix} = \begin{pmatrix} 1/p & 0 \\ 0 & 1 \end{pmatrix} M_2(\mathbf{Z}) \begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}.$$

Theorem 68. *If $\mathcal{O}$ is a maximal order in a quaternion algebra $B \hookrightarrow M_2(\mathbf{R})$ over $\mathbf{Q}$, then $\mathcal{O}^1$ is a Fuchsian group. If $\mathrm{disc}(\mathcal{O}) = \prod_{i=1}^n p_i$ then the covolume of $\mathcal{O}^1$ is $\prod_{i=1}^n (p_i - 1)$.*

Proof. Since left-translation is a homeomorphism, it suffices to find a neighborhood of the identity in $\mathrm{SL}_2(\mathbf{R})$ whose intersection with $\mathcal{O}$ is the identity. Consider the set U of matrices $\begin{pmatrix} x & y \\ z & w \end{pmatrix}$ such that $|x-1|, |w-1| < 1/2$ and $|y|, |z| < 1/2$. Clearly for $M_2(\mathbf{Z})$ this works. Assume now that B is a division algebra, so $B \cong \left(\frac{a,b}{\mathbf{Q}}\right)$ such that $a \notin \mathbf{Q}^{\times 2}$. It suffices to prove that $\mathbf{Z} \oplus i\mathbf{Z} \oplus j\mathbf{Z} \oplus k\mathbf{Z}$ is discrete, as this is at worst finite-index in a maximal order.

Recall that if $\alpha x + iy + jz + kw \in \mathcal{O}^1$ then it embeds into $\mathrm{SL}_2(\mathbf{R})$ as

$$\begin{pmatrix} x + y\sqrt{a} & z + w\sqrt{a} \\ b(z - w\sqrt{a}) & x - y\sqrt{a} \end{pmatrix}.$$

Since $|t(\alpha)| = 2|x-1| < 1$, we have $x = 1$. Since $|b| > 1$, $|z - w\sqrt{a}| < \dfrac{1}{2|b|} < 1/2$. Therefore $|2z| < 1$ and thus $z = 0$. Therefore $|w\sqrt{a}|, |y\sqrt{a}| < 1/2$ and thus $y = w = 0$.

Computing the volume involves studying the zeta function of $\mathcal{O}$ and its residue. □

It turns out that outside of the case $M_2(\mathbf{Z})$, the Fuchsian groups coming from quaternion algebras are not just finite volume, but cocompact.

6. Lecture 6: Traces and the Geometry of Shimura curves

We give a basic definition of a Shimura curve. We may take a somewhat more expansive definition later.

Definition 69. *If Γ is a finite covolume Fuchsian Group, let $Y(\Gamma) = \Gamma \setminus \mathcal{H}$.*

Theorem 70. *If Γ is a finitely generated Fuchsian group and $g \in SL_2(\mathbf{R})$, then there is a holomorphic isomorphism between $\Gamma \setminus \mathcal{H}$ and $(g\Gamma g^{-1}) \setminus \mathcal{H}$.*

Proof. Since the two groups are isomorphic, $g\Gamma g^{-1}$ is finitely generated if and only if Γ is, so their quotients are Riemann surfaces. The map $\mathcal{H} \to \mathcal{H}$ by $\tau \mapsto g\tau$ is clearly holomorphic. If we post-compose with the map $\mathcal{H} \to (g\Gamma g^{-1}) \setminus \mathcal{H}$, consider two points τ_1 and τ_2 which map to the same point of $(g\Gamma g^{-1}) \setminus \mathcal{H}$. Thus $(g\Gamma g^{-1})g\tau_1 = (g\Gamma g^{-1})g\tau_2$, so there exists some $\gamma \in \Gamma$ such that $\tau_2 = \gamma\tau_1$. Therefore there is a holomorphic map $\Gamma \setminus \mathcal{H} \to (g\Gamma g^{-1}) \setminus \mathcal{H}$ with the obvious inverse $(g\Gamma g^{-1}) \setminus \mathcal{H} \to (g^{-1}(g\Gamma g^{-1})g) \setminus \mathcal{H} = \Gamma \setminus \mathcal{H}$. □

Corollary 71. *If $\mathcal{O}_1$, $\mathcal{O}_2$ are maximal orders in a quaternion algebra $\psi : B \hookrightarrow M_2(\mathbf{R})$ then $Y(\mathcal{O}_1^1) \cong Y(\mathcal{O}_2^1)$.*

Proof. If so, there is some $\phi \in B$ such that $\mathcal{O}_2 = \phi\mathcal{O}_1\phi^{-1}$. Therefore $\psi(\mathcal{O}_2^1) = \psi(\phi)\psi(\mathcal{O}_1^1)\psi(\phi)^{-1}$. If $\psi(\phi) \in \mathrm{SL}_2(\mathbf{R})$, we are done.

If $n(\phi) < 0$, recall that we may write B as $\left(\frac{a,b}{\mathbf{Q}}\right)$ with $a, b \in \mathbf{Z}$, $a > 0$. Therefore $n(i) = -a < 0$ and so $n(\phi i) > 0$. If $i \in \mathcal{O}_1$, then we may replace ϕ with ϕi because

$$\phi i\mathcal{O}_1(\phi i)^{-1} = \phi(i\mathcal{O}_1 i^{-1})\phi^{-1} = \phi\mathcal{O}_1\phi^{-1} = \mathcal{O}_2.$$

If $i \notin \mathcal{O}_1$, then we know already that $t(i) = 0$ so i is integral and lies in some maximal order $\mathcal{O}_3 = \alpha\mathcal{O}_1\alpha^{-1}$. In this case, we may replace ϕ with $\phi\alpha i\alpha^{-1}$. Therefore without loss of generality, $n(\phi) = \det(\psi(\phi)) > 0$.

If $n(\phi) > 0$ but is not equal to one, replace $\psi(\phi)$ with $\frac{1}{\sqrt{n(\phi)}}\psi(\phi) \in \mathrm{SL}_2(\mathbf{R})$. □

Let $\mathcal{O}$ be a maximal order in a quaternion algebra $B/\mathbf{Q}$. Since all maximal orders in B are conjugate to $\mathcal{O}$, the quotient $Y(\mathcal{O}^1)$ is a natural invariant of B.

Definition 72. *Let $B/\mathbf{Q}$ be a quaternion algebra. Define the Shimura curve Y_B to be $Y(\mathcal{O}^1)$ for any maximal order $\mathcal{O}$ of B.*

To get a better idea of the geometry of these quotients by Fuchsian groups, especially those with finite volume, we chop up $\mathrm{PSL}_2(\mathbf{R})$ into three categories. Note first that on $\mathrm{PSL}_2(\mathbf{R})$, the map $\alpha \mapsto |\operatorname{tr}(\alpha)|$ is well-defined. We see a bit more evidence of the following trend: geometric information is more easily accessible via $\mathrm{PSL}_2(\mathbf{R})$ while algebraic information is more easily accessible via $\mathrm{SL}_2(\mathbf{R})$.

Definition 73. *Let $x \in \mathrm{PSL}_2(\mathbf{R})$. If $|\operatorname{tr}(x)| = 2$, we say that x is parabolic. If $|\operatorname{tr}(x)| \leq 2$, we say that x is elliptic. If $|\operatorname{tr}(x)| \geq 2$, we say that x is hyperbolic.*

- If $x \in \mathrm{SL}_2(\mathbf{R})$ reduces to a parabolic element, it is conjugate to a matrix of the form $\pm \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ for some $\lambda \in \mathbf{R}$
- If $x \in \mathrm{SL}_2(\mathbf{R})$ reduces to an elliptic element, it is conjugate to a matrix of the form $\begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$ for some $\theta \in \mathbf{R}$
- If $x \in \mathrm{SL}_2(\mathbf{R})$ reduces to a hyperbolic element, it is conjugate to a matrix of the form $\begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$ for $\lambda \in \mathbf{R}^\times$

Let us simply refer to elements of $\mathrm{SL}_2(\mathbf{R})$ as elliptic, hyperbolic, or parabolic as is appropriate. We will shortly see why the above conjugacy statements are true. In the mean time, let's see some consequences.

Lemma 74. *If $\alpha \in \mathrm{SL}_2(\mathbf{R})$ is not elliptic, then $\langle \alpha \rangle \cong \mathbf{Z}$. If $\alpha \in \mathrm{SL}_2(\mathbf{R})$ is torsion then α is elliptic. Let $\Gamma < \mathrm{SL}_2(\mathbf{R})$ be a Fuchsian group. If $\alpha \in \Gamma$ is elliptic, then $\langle \alpha \rangle$ is finite and thus torsion.*

Proof. The first statement is clear and the second statement is the contrapositive of the first. Now note that $\langle \alpha \rangle$ is discrete in $\mathrm{SL}_2(\mathbf{R})$, however it also lies inside some conjugate of $\mathrm{SO}_2(\mathbf{R})$, which is compact. □

Definition 75. *If $\Gamma < \mathrm{PSL}_2(\mathbf{R})$ is a Fuchsian group and α is a nondivisible elliptic element of Γ, we say that $\langle \alpha \rangle$ is an elliptic cycle of length $|\alpha|$ in Γ.*

Equivalently, $\langle \alpha \rangle$ is a maximal cyclic elliptic subgroup of Γ. We can therefore see that if Γ does not contain any elliptic cycles, then $\mathcal{H} \to \Gamma \setminus \mathcal{H}$ is a covering map. In general if $\alpha \neq \pm 1$ is elliptic and $z \in \mathcal{H}$ is a fixed point of α, then near z, the map $\mathcal{H} \to \Gamma \setminus \mathcal{H}$ is of the form $w \mapsto w^{|\alpha|}$. Now we begin our investigation in earnest.

Lemma 76. *Let* $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{R})$*, acting on* $\mathbb{P}^1(\mathbf{C})$ *by linear fractional transformations. Let* $z_1, z_2 \in \mathbb{P}^1(\mathbf{C})$ *(possibly equal) be the fixed points of* α*.*

- *If* α *is parabolic,* $z_1 = z_2 \in \mathbf{R} \cup \infty$*.*
- *If* α *is hyperbolic,* $z_1 \neq z_2 \in \mathbf{R} \cup \infty$*.*
- *If* α *is elliptic,* $z_1 \neq z_2 \in \mathbf{C} - \mathbf{R}$ *and precisely one lies in* $\mathcal{H}$*.*

Proof. Suppose that z is a fixed point of $\alpha \neq 1$, i.e., $az + b = cz^2 + dz$ and thus $cz^2 + (d-a)z - b = 0$. By the quadratic formula, z_1 and z_2 must be

$$\begin{aligned} \frac{(a-d) \pm \sqrt{(d-a)^2 - 4(c)(-b)}}{2c} &= \frac{(a-d) \pm \sqrt{a^2 - 2ad + d^2 + 4bc}}{2c} \\ &= \frac{(a-d) \pm \sqrt{(a+d)^2 - 4ad + 4bc}}{2c} \\ &= \frac{(a-d) \pm \sqrt{\mathrm{tr}(\alpha)^2 - 4(ad-bc)}}{2c} \\ &= \frac{(a-d) \pm \sqrt{\mathrm{tr}(\alpha)^2 - 4}}{2c}. \end{aligned}$$

The result follows. □

We now obtain our result on conjugation. Let α be elliptic and let z be its fixed point in $\mathcal{H}$. Let R be the linear fractional transformation taking z to i. Then $R\alpha R^{-1}$ fixes i and therefore lies inside $\mathrm{SO}_2(\mathbf{R})$.

If α is parabolic, let x be the fixed point and let R send x to ∞, so that $\beta = R\alpha R^{-1}$ fixes ∞. Therefore β is of the form $\pm \begin{pmatrix} \kappa & \mu \\ 0 & 1/\kappa \end{pmatrix}$. Since β is parabolic, $\kappa = \pm 1$.

If α is hyperbolic, let $a \neq b$ be the fixed points of α. Let R send a to 0 and b to ∞. Namely, $R(z) = \dfrac{1}{a-b}\dfrac{z-a}{z-b}$. Then $R\alpha R^{-1}$ fixes 0 and ∞, and is thus of the form $\begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$.

Maximal parabolic and elliptic subgroups are therefore in bijection with their fixed points in $\mathcal{H} \cup \mathbf{R} \cup \infty$. In fact, we may think about parabolic elements as being limits of elliptic elements, or as "elliptic elements with infinite order." To this end, consider the following Theorem.

Theorem 77 (Katok, Theorem 4.2.1)**.** *If a Fuchsian group* Γ *has a compact Dirichlet domain* F*, then* Γ *contains no parabolic elements.*

Proof. Define a function

$$\eta(z) = \inf\{\delta(z, T(z)) : T \in \Gamma - \{1\}, |\operatorname{tr}(T)| \geq 2\}.$$

Note that we have just thrown out all elliptic elements because we want to avoid $z = T(z)$. Note that $\eta(z)$ is a continuous function of $z \in \mathcal{H}$ and $\eta(z) > 0$. Since F is compact,

$$\eta = \min\{\eta(z) : z \in F\},$$

exists and is positive. If $z \in \mathcal{H}$, let $S \in \Gamma$ such that $w = S(z) \in F$. Therefore if $\gamma \in \Gamma - \{1\}$ is not elliptic,

$$\delta(z, \gamma(z)) = \delta(S(z), S(\gamma(z))) = \delta(w, S\gamma S^{-1}w) \geq \eta,$$

and thus $\inf\{\delta(z, \gamma(z)) : z \in \mathcal{H}\} = \eta > 0$.

Now if Γ contains a nonidentity parabolic element π, then let z_π be the unique fixed point of π in $\partial\mathcal{H} = \mathbf{R} \cup \infty$. For any $R \in \mathrm{PSL}_2(\mathbf{R})$, $R(F)$ is a compact fundamental domain for $\Gamma' = R\Gamma R^{-1}$. Pick R sending z_π to ∞ so that $R\pi R^{-1}$ fixes ∞ and is thus of the form $\pm \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ for some $\lambda \in \mathbf{R}$. Therefore $R\pi R^{-1} \in \Gamma'$ is the map $z \mapsto z + \lambda$. Note however that

$$\delta(x + iy, x + \lambda + iy) \leq \int_0^\lambda \frac{\mathrm{d}x}{y} = \frac{\lambda}{y},$$

so as $\Im(z) \to \infty$, we have $\delta(z, z + \lambda) \to 0$. If there were a compact fundamental domain F, there would be a fixed lower bound η given by F. It follows that if Γ contains a nonidentity parabolic element, there can be no compact fundamental domain. □

Therefore if $\Gamma \setminus \mathcal{H}$ is compact and $\mathcal{H} \to \Gamma \setminus \mathcal{H}$ is a covering map, then Γ consists entirely of hyperbolic elements.

Example 78. *Let* $A = \left(\dfrac{-2, 13}{\mathbf{Q}}\right)$ *and let*

$$\mathcal{O} = \mathbf{Z} \oplus \frac{1+j}{2}\mathbf{Z} \oplus i\frac{1+j}{2}\mathbf{Z} \oplus k\mathbf{Z},$$

a maximal order. The Fuchsian group $\mathcal{O}^1$ *has generators* $\{-11/2 + 3/2j, 5/2 + 3/2i - 1/2j - 1/2k, 3/2 - i - 1/2j, 4 - 3/2i - j - 1/2k\}$. *The matrix representatives for these are*

$$\left\{ \begin{pmatrix} \dfrac{3\sqrt{13} - 11}{2} & 0 \\ 0 & \dfrac{-3\sqrt{13} - 11}{2} \end{pmatrix}, \begin{pmatrix} \dfrac{\sqrt{13} + 5}{2} & \sqrt{13} - 3 \\ \dfrac{\sqrt{13} + 3}{2} & \dfrac{-\sqrt{13} + 5}{2} \end{pmatrix} \right\},$$

$$\left\{ \begin{pmatrix} \frac{\sqrt{13}+3}{2} & 2 \\ -1 & \frac{-\sqrt{13}+3}{2} \end{pmatrix}, \begin{pmatrix} \sqrt{13}+4 & \sqrt{13}+3 \\ \frac{\sqrt{13}-3}{2} & -\sqrt{13}+4 \end{pmatrix} \right\}.$$

This group is purely hyperbolic, and the hyperbolic octagon formed by the Dirichlet domain centered at i forms a compact fundamental domain for the genus two curve $\psi(\mathcal{O}^1) \setminus \mathcal{H}$.

Definition 79. *If Γ is a Fuchsian group, let $\Lambda(\Gamma)$ be the limiting set of Γ, i.e., the elements z of $\partial\mathcal{H}$ such that there exists some nonidentity parabolic element g such that $gz = z$.*

Example 80. *If $\Gamma = \mathrm{SL}_2(\mathbf{Z}) = \langle T, S \rangle$ where $T : z \mapsto z + 1$ and $S : z \mapsto -1/z$, then $\Lambda(\Gamma) = \mathbf{Q} \cup \infty$. We at least have ∞ in this set because $T\infty = \infty$. We also have every element $a/b \in \mathbf{Q}$ where the gcd of a and b is one. To see this, let $x, y \in \mathbf{Z}$ such that $ay - bx = 1$. Let R be the transformation $\begin{pmatrix} a & x \\ b & y \end{pmatrix}$, sending ∞ to a/b. It follows that $RTR^{-1}(a/b) = a/b$, and since the trace of a matrix is invariant under conjugation, RTR^{-1} is also parabolic. On the other hand, if $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z})$ is parabolic, then the fixed point of α is $\dfrac{a-d}{2c} \in \mathbf{Q}$.*

This is a standard sort of argument as we see below.

Lemma 81. *The group Γ preserves the set $\Lambda(\Gamma)$.*

Proof. Let $z \in \Lambda(\Gamma)$ and let $g \in \Gamma$ be parabolic such that $gz = z$. If $h \in \Gamma$, then $hz \in \Lambda(\Gamma)$ because $hgh^{-1}(hz) = hgz = hz$. Since the trace of a matrix is invariant under conjugation, hgh^{-1} is parabolic. Finally $hz \in \partial\mathcal{H}$ because any element h of $\mathrm{SL}_2(\mathbf{R})$ preserves the set $\partial\mathcal{H} = \mathbf{R} \cup \infty$. □

Definition 82. *We define the cusps of Γ as $\mathrm{cusp}(\Gamma) = \Gamma \setminus \Lambda(\Gamma)$.*

Therefore our natural compactification of $Y(\Gamma)$ is the set $Y(\Gamma) \cup \mathrm{cusp}(\Gamma)$. Of course, we don't just want a set, we want a complex analytic space. To give that, we study quotients of a new topological space $\mathcal{H}^\Gamma = \mathcal{H} \cup \Lambda(\Gamma)$.

7. Lecture 7: The horocycle topology and cusps

As in the previous lecture, we want to form the natural compactification of $Y(\Gamma)$. We will do so by taking quotients of $\mathcal{H}^\Gamma = \mathcal{H} \cup \Lambda(\Gamma)$ where $\Lambda(\Gamma)$ is the set of fixed points of parabolic elements of Γ. We would like to give $\mathcal{H}^\Gamma$ the structure of a topological space.

Example 83 (The Toy Model)**.** *Let* $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ *and let* $\Gamma = \langle T \rangle$*. The fixed point of* T *is* ∞ *so* $\Lambda(\Gamma) = \{\infty\}$*. It follows that* $\mathcal{H}^\Gamma = \mathcal{H} \cup \{\infty\}$*. We already have a topology on* $\mathcal{H}$ *so to give* $\mathcal{H}^\Gamma$ *a Hausdorff topology, it suffices to find, for each* $z \in \mathcal{H}$*, a neighborhood* U_z *on* ∞ *which does not contain* z*. We will call this a* neighborhood basis *of* ∞*.*

The natural choice of neighborhood we will take is $U_z = \{x + iy : x \in \mathbf{R}, y > \Im(z)\} \cup \{\infty\}$*. One might ask why this is natural. To tend to* ∞*, at least one of* $|x|$ *or* y *must go to* ∞*. However,* x *only matters up to addition or subtraction by integers once we take a quotient. For this reason, most people refer to* ∞ *as* $i\infty$*.*

Incidentally, the quotient map $\mathcal{H} \to \Gamma \setminus \mathcal{H} \cong \mathbb{D}^*$ *is the map* $z \mapsto e^{2\pi i z}$*. If we take the limit of this function as* $\Im(z) \to \infty$*, we get zero. Therefore, this fits into the following diagram.*

$$\begin{array}{ccc} \mathcal{H} & \to & \Gamma \setminus \mathcal{H} \cong \mathbb{D}^* \\ \downarrow & & \downarrow \\ \mathcal{H}^\Gamma & \to & \Gamma \setminus \mathcal{H}^\Gamma \cong \mathbb{D} \end{array}$$

It is a nice exercise to check that if a Fuchsian group Γ has $\#\Lambda(\Gamma) > 1$, then $\Lambda(\Gamma)$ is infinite. We note that if this is the case and $a \neq \infty$ is in $\Lambda(\Gamma)$, then to get a neighborhood basis for a, we need only pull back the neighborhood basis at ∞ to a via the map $\phi_a : z \mapsto \dfrac{-1}{z - a}$. We thus make the following definition.

Definition 84. *Let* Γ *be a Fuchsian group with limit set* $\Lambda(\Gamma)$*. We define the* extended upper half plane $\mathcal{H}^\Gamma$ *as the set* $\mathcal{H} \cup \Lambda(\Gamma)$ *given the* horocycle topology*. This is the topology generated by the topology on* $\mathcal{H}$ *along with the sets* U_z *if* $\infty \in \Lambda(\Gamma)$ *and if* $a \in \Lambda(\Gamma)$*, the sets* $\phi_a^{-1}(U_z)$*, which form open circles tangent to* $\mathbf{R}$ *at* a *along with the point* a*.*

Despite the name and construction, the extended upper half-plane is quite different from $\mathcal{H}$ as a topological space.

Lemma 85. *The topological space* $\Gamma \setminus \mathcal{H}$ *is compact if and only if* $\Lambda(\Gamma)$ *is empty if and only if* $\mathcal{H}^\Gamma$ *is locally compact.*

Proof. The first equivalence is easy to see. For the second one, notice first that if $\Lambda(\Gamma)$ is empty then $\mathcal{H}^\Gamma = \mathcal{H}$, which is locally compact. On the other hand, if it is not empty, we may conjugate Γ and assume without loss of generality that $\infty \in \Lambda(\Gamma)$. If $\mathcal{H}^\Gamma$ is locally compact, then there is a compact neighborhood K of ∞. Since K is a neighborhood, there exists some z such that $U_z \subset K$. Since K is closed, $\overline{U_z} \subset K$. Therefore $\partial\overline{U_z} = \mathbf{R} + i\Im(z)$ is a closed subset of K. Therefore $\mathbf{R} + i\Im z$ is compact. This cannot be, hence there is no such neighborhood K. $\square$

So perhaps most terrifyingly of all, this means that in any situation where you'd want to introduce it, $\mathcal{H}^\Gamma$ is not locally compact and is thus not a Riemann surface. While this sounds bad, we don't *need* $\mathcal{H}^\Gamma$ to be Riemann surfaces. We need *quotients* of $\mathcal{H}^\Gamma$ to be Riemann surfaces. We get around this by reducing to the case of the toy model. At least in the case of finite-covolume Fuchsian groups, which have only finitely cusps, we can see how to do this.

Example 86. *Let* $\Gamma(2)$ *be the kernel of the map* $\mathrm{SL}_2(\mathbf{Z}) \to \mathrm{SL}_2(\mathbf{Z}/2\mathbf{Z})$. *The element* $\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ *is a parabolic element which fixes* 0. *It is easy, but not necessary, to show that* 0 *and* ∞ *are inequivalent cusps for* $\Gamma(2)$. *Let* $\Gamma'(2) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \Gamma(2) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. *We need only consider that* $Y(\Gamma(2)) \cong Y(\Gamma'(2))$, *and*

$$\gamma = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}.$$

This is a parabolic matrix fixing ∞. *Let* F_γ *be a fundamental domain for the group* $\langle\gamma\rangle$. *If* F *is a fundamental domain for* $\Gamma'(2)$, *we can find some* $z \in \mathcal{H}$ *such that* $U_z \cap F = U_z \cap F_\gamma$. *Since* $\langle\gamma\rangle \setminus F_\gamma \cong \mathbb{D}^*$, $\langle\gamma\rangle \backslash U_z \cap F \cong \mathbb{D}^*$. *Therefore* $\langle\gamma\rangle \backslash (U_z \cap F) \cup \{i\infty\} \cong \mathbb{D}$, *just as in the toy model. Therefore* $Y(\Gamma(2)) \cup \{0\}$ *has the structure of a Riemann surface near* 0. *It follows that* $\Gamma(2) \setminus \mathcal{H}^{\Gamma(2)}$ *has the structure of a Riemann surface and is indeed compact.*

Definition 87. *Let* Γ *be a finite covolume Fuchsian group. Let* $X(\Gamma) = \Gamma \setminus \mathcal{H}^\Gamma$.

By applying the above procedure at a finite number of cusps we obtain the following theorem.

Theorem 88. *If* Γ *is a finite covolume Fuchsian group,* $X(\Gamma)$ *is a compact Riemann surface.*

Proof. Let $F = D_P(\Gamma)$ be a fundamental domain for Γ. For $\mathrm{cusp}(\Gamma) = \{a_1, \ldots, a_n\}$ let $\gamma_1, \ldots, \gamma_n$ be the corresponding maximal parabolic stabilizers and let $z_1, \ldots, z_n \in \mathcal{H}$ be such that $\phi_{a_i}^{-1}(U_{z_i}) \cap F = \phi_{a_i}^{-1}(U_{z_i}) \cap F_{\gamma_i}$. Let $V_i = \langle\gamma_i\rangle \setminus (\phi_{a_i}^{-1}(U_{z_i}) \cap F) \cong \mathbb{D}$. Clearly $X(\Gamma)$ is compact if and only if the closed set $X(\Gamma) - \bigcup_i V_i$ is compact. However, a fundamental domain for this latter set is $G = F - \bigcup_i \phi_{a_i}^{-1}(U_{z_i})$, which is closed obviously. However, since there are only finitely many z_i, there is a maximum value m of $\Im(z_i)$. Since the hyperbolic metric is invariant under elements of $\mathrm{SL}_2(\mathbf{R})$ such as ϕ_a, $G \subset \{z \in \mathcal{H} : \delta(p, z) \leq m\}$. Therefore G is closed and bounded and thus compact. □

Since $X(\Gamma)$ is a compact Riemann surface, it is a closed orientable surface of some genus g. By abuse of notation, we make the following definition.

Definition 89. *If Γ is a finite covolume Fuchsian group, define $g(\Gamma)$ to be the genus of $X(\Gamma)$. Let $\{\alpha_1, \ldots, \alpha_r\}$ be generators of disjoint maximal elliptic cycles in Γ and $\{\alpha_{r+1}, \ldots, \alpha_{r+s}\}$ be generators of maximal parabolic subgroups of Γ. Let e_i be such that $e_i = |\alpha_i|$ if $i \leq r$ and $e_i = \infty$ if $i > r$. We define the signature of Γ to be*

$$\sigma(\Gamma) = (g; e_1, \ldots, e_{r+s}).$$

8. Lecture 8: Signature and the fundamental equation

Last time we defined the signature of a Fuchsian group. This time we aim to prove the following.

Theorem 90 (The Fundamental equation)**.** *Let Γ be a Fuchsian group of finite covolume and let F be a fundamental domain for Γ. Let $\sigma(\Gamma) = (g; e_1, \ldots, e_{r+s})$ be the signature of Γ. Then*

$$\begin{aligned} area(F) &= 2\pi \left[(2g-2) + \sum_{i=1}^{r+s} \left(1 - \frac{1}{e_i}\right) \right] \\ &= 2\pi \left[(2g-2) + s + \sum_{i=1}^{r} \left(1 - \frac{1}{e_i}\right) \right]. \end{aligned}$$

We shall now proceed to lemma this thing to death. First we start with a general result on hyperbolic polygons.

Lemma 91. *Let F be a hyperbolic polygon (possibly with vertices on $\partial\mathcal{H}$, in which the angle at that vertex is necessarily zero) with n sides and vertex angles $\{\vartheta_j\}$. Let $A = \sum_j \vartheta_j$. We then have*

$$\text{area}(F) = \pi(n-2) - A.$$

Proof. Choose and interior point P for F and draw lines between P and the vertices of F. Therefore we have chopped F into n triangles T_i. Let $\{\theta_i, \alpha_i, \beta_i\}$ be the angles of T_i with θ_i the angle at P. Therefore $\sum_i \theta_i = 2\pi$ and $\sum_i (\alpha_i + \beta_i) = \sum_j \vartheta_j = A$. It follows that

$$\begin{aligned}\text{area}(F) &= \sum_i \text{area}(T_i) \\ &= \sum_i (\pi - \theta_i - \alpha_i - \beta_i) \\ &= \sum_i \pi - \sum_i \theta_i - \sum_i (\alpha_i + \beta_i) \\ &= n\pi - 2\pi - A.\end{aligned}$$

□

We use this to show an easy variant.

Corollary 92. *If Γ is purely hyperbolic of finite covolume then there is a compact fundamental region F and area$(F) = 4\pi(g(\Gamma) - 1)$.*

Proof. Note that if Γ is purely hyperbolic of finite covolume, then $\mathcal{H} \to Y(\Gamma) = X(\Gamma)$ is a covering map and Γ has a compact fundamental region F. Since the quotient map is the universal covering map of a genus $g(\Gamma)$ surface, F is a $4g(\Gamma)$-sided compact polygon, all of whose vertices are identified in $X(\Gamma)$. Therefore we need only show that the angle sum of F is 2π.

Label the vertices $\{v_j\}$ with angles ϑ_j. Note that if $S, T \in \Gamma$ send v_j to v_1 then $S^{-1}Tv_j = v_j$. Since Γ is a group of deck transformations, it acts freely on $\mathcal{H}$ and thus $S^{-1}T = 1$ and $S = T$. Therefore there are unique $\gamma_1 = 1, \gamma_2, \ldots, \gamma_{4g(\Gamma)}$ such that $\gamma_j v_j = v_1$. Therefore if U is an open neighborhood of v_1, every element of U lies in some $\gamma_j F$, which is unique except for the boundary. Since each γ_j preserves angles, $\sum_j \vartheta_j = 2\pi$. □

Now what if Γ is not purely hyperbolic?

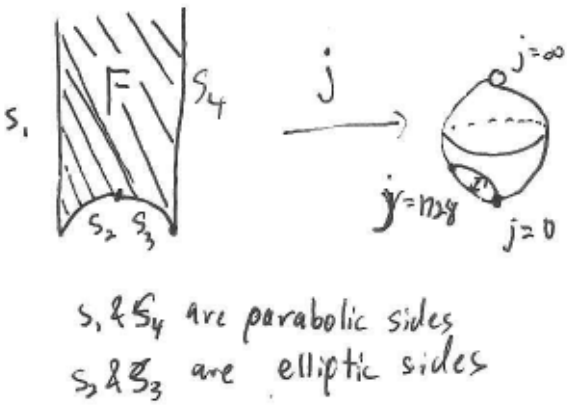
Definition 93. *Let F be a fundamental domain with no interior fixed points for $\Gamma < \mathrm{PSL}_2(\mathbf{R})$, a finite covolume Fuchsian group. If a point v of F is fixed by some elliptic element besides ± 1, we say that v is an* elliptic vertex. *Let $v_1, \ldots, v_r$ be the non-Γ-equivalent elliptic vertices and let $\alpha_1, \ldots, \alpha_r$ be such that $\langle \alpha_i \rangle = \mathrm{Stab}_\Gamma(v_i)$. If a vertex v of F is fixed by some parabolic element besides ± 1, we say that v is a* parabolic vertex. *Let s denote the number of non-Γ-equivalent parabolic vertices of F, and correspondingly let $\alpha_{r+1}, \ldots, \alpha_{r+s}$ be generators of their stabilizers. If v is neither elliptic nor parabolic, we say that v is a* hyperbolic vertex. *Let h denote the number of non-Γ-equivalent hyperbolic vertices of F.*

Lemma 94. *If Γ is a finite covolume Fuchsian group, let F be an n-sided fundamental domain for Γ and let r, s, h be as above. The Euler characteristic of $X(\Gamma)$ is*

$$2 - 2g(\Gamma) = (r + s + h) - \frac{n}{2} + 1.$$

Proof. Since the quotient $\mathcal{H}^\Gamma \to X(\Gamma)$ *pairs* sides of F and identifies Γ-congruent vertices of F (including ones on $\partial\mathcal{H}$, F gives an explicit representation of the underlying topological space $X(\Gamma)$ as a CW-complex. This complex has a single face F^o, $n/2$ sides, and $(r+s+h)$ vertices. □

Example 95. *One might read the above and be confused as to how this works with the classical fundamental domain of* $\mathrm{PSL}_2(\mathbf{Z})$*. It is natural to view this as a 3-sided hyperbolic polygon. It is! But as a fundamental domain, elliptic fixed points must be vertices. In particular, i is an elliptic fixed point of order 2, and therefore the angle at i must be $\frac{2\pi}{2} = \pi$. Therefore, at fixed points of order 2 in* $\mathrm{PSL}_2(\mathbf{R})$*, we have to chop up a seemingly contiguous side a little further.*



Continuing this thread, we consider angles of a fundamental domain.

Lemma 96. *Let Γ have finite covolume and let F, h, r, s, and $\alpha_1, \ldots \alpha_{r+s}$ be as above. Let $m_i = |\alpha_i|$, so that if α_i is parabolic, $m_i = \infty$ and $\frac{1}{m_i} = 0$. Then the angle sum of F is*

$$A = 2\pi \left(h + \sum_{i=1}^{r+s} \frac{1}{m_i} \right).$$

Proof. For hyperbolic vertices, this was covered in Corollary 92. For parabolic vertices, the angle must be zero. Therefore we need only consider Γ-equivalent elliptic vertices $v_1, \ldots v_t$ such that ϑ_j is the angle

at v_j. For each j select $T_j \in \Gamma$ sending v_j to v_1. Each element sending v_j to v_1 will be of the form HT_j where $H = \mathrm{Stab}_\Gamma(v_1)$. Note that $\#H = m_i$ and therefore there are m_i elements sending v_j to v_1. Therefore the sets $\bigcup_{h\in H}\bigcup_j hT_j(F)$ cover a small enough neighborhood of v_1 and thus $m_i \sum_{j=1}^{t} \vartheta_j = 2\pi$.

Therefore the sum of all the angles at elliptic vertices is $\sum_{i=1}^{s} \frac{2\pi}{m_i}$, the sum of all the angles at parabolic vertices is $0 = \sum_{i=r+1} r + s\frac{2\pi}{m_i}$, and the sum of all the angles at hyperbolic vertices is $2\pi h$. □

Now we complete the proof of the fundamental equation.

Proof of the Fundamental equation. Let Γ, F, r, s, h, m_i and n be as above. Fix $g = g(\Gamma)$.

We have learned from the Gauss-Bonnet theorem that the area of F is

$$(n-2)\pi - A.$$

We have learned from the Euler characteristic that

$$n = 2(2g - 2 + 1 + r + s + h).$$

We have just calculated that

$$A = 2\pi\left(h + \sum_{i=1}^{r+s} \frac{1}{m_i}\right).$$

Therefore,

$$\begin{aligned}
\mathrm{area}(F) &= (2(2g-2+1+r+s+h)-2)\pi - 2\pi\left(h + \sum_{i=1}^{r+s}\frac{1}{m_i}\right) \\
&= 2\pi\left(2g-2+1+r+s+h-1-h-\sum_{i=1}^{r+s}\frac{1}{m_i}\right) \\
&= 2\pi\left[2g-2+\sum_{i=1}^{r+s}\left(1-\frac{1}{m_i}\right)\right].
\end{aligned}$$

□

We note an easy and yet fundamental fact which can be easily deduced from the fundamental equation.

Corollary 97 (Riemann-Hurwitz)**.** *Let X_1, X_2 be compact Riemann Surfaces and let $\pi : X_1 \to X_2$ be a holomorphic map. Then*

$$\chi(X_1) = \deg(\pi)\chi(X_2) - \sum_{P \in X_1} (e_P - 1),$$

where in a small enough neighborhood of P, π looks like $z \mapsto z^{e_P}$.

This theorem holds in very great generality and will take the place of the fundamental equation in the algebraic setting.

So why study Shimura curves in the first place? They are objects whose algebra and geometry exactly coincide. To see this, let B be a quaternion algebra over $\mathbf{Q}$, let $\mathcal{O}$ be a maximal order in B and let $\psi : B \hookrightarrow M_2(\mathbf{R})$. We consider the elliptic points of $Y_B = Y(\psi(\mathcal{O}^1))$.

Definition 98. *Each $\alpha \in B$ satisfies its own characteristic polynomial*

$$\chi_\alpha(X) = (X - \alpha)(X - \overline{\alpha}) = X^2 - t(\alpha)X + n(\alpha) \in \mathbf{Q}[X].$$

Note that if $\alpha \in \mathcal{O}$ then by definition $t(\alpha), n(\alpha) \in \mathbf{Z}$. Suppose now that $\alpha \in \mathcal{O}^1$ is such that $\psi(\alpha)$ is an elliptic element. Recall that to be elliptic means that either $|t(\alpha)| < 2$ or $\alpha = \pm 1$. Then α satisfies a polynomial of the form

$$X^2 - tX + 1 : \qquad t \in \{-1, 0, 1\}. \tag{1}$$

If $t = 0$ then $\alpha^2 = -1$ and therefore in $\mathrm{PSL}_2(\mathbf{R})$, α gives an elliptic element of order 2. If $t = \pm 1$, note that if α is a root of $X^2 - X + 1$, then

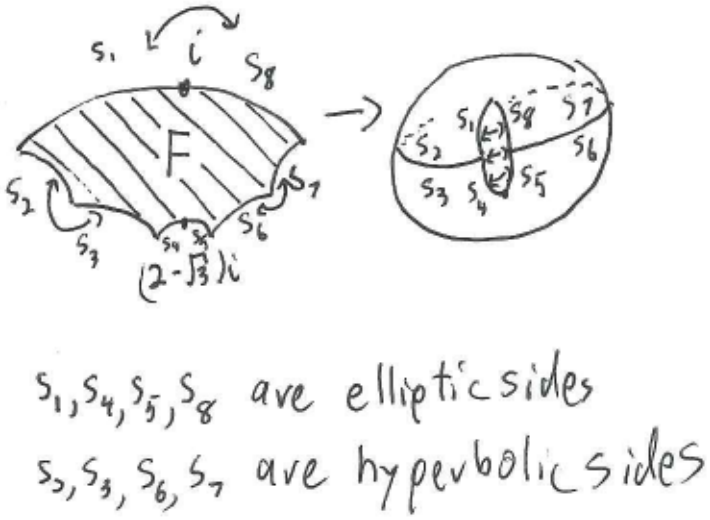
$$(-\alpha)^2 + (-\alpha) + 1 = \alpha^2 - \alpha + 1 = 0.$$

Therefore, roots of these polynomials are equivalent in $\mathrm{PSL}_2(\mathbf{R})$. Moreover, since $(X - 1)(X^2 + X + 1) = X^3 - 1$, roots of $X^2 \pm X + 1$ give rise to elliptic elements of order three. That's it!

Corollary 99. *If $\mathcal{O}$ is a maximal order in a quaternion algebra $B/\mathbf{Q}$ then let Γ be the reduction of $\psi(\mathcal{O}^1)$ in $\mathrm{PSL}_2(\mathbf{R})$ so that $\Gamma \setminus \mathcal{H}^\Gamma = X_B$. Then we have*

$$g(X_B) = 1 + \frac{\phi(\mathrm{disc}(\mathcal{O}))}{12} - \frac{e_2}{4} - \frac{e_3}{3} - \frac{\#\,\mathrm{cusp}(\Gamma)}{2},$$

where e_2 is the number of non-conjugate elliptic elements of order 2 in Γ and e_3 is the number of non-conjugate elliptic elements of order 3 in Γ.



To make this a precise number, we will need to be able to count the number of embeddings of $\mathbf{Z}[X]/(X^2+1)$ and $\mathbf{Z}[X]/(X^2+X+1)$ into $\mathcal{O}$ up to conjugacy.

First let's get an idea for when a ring $\mathbf{Z}[X]/(f(X))$ can embed into $\mathcal{O}$, given that $f(X)$ is monic of degree 2. It is easy to see that such rings depend, up to isomorphism, only upon $\mathrm{disc}(f)$. Moreover, they are either of the form $\mathbf{Z}[\sqrt{d}]$ or $\mathbf{Z}\left[\dfrac{1+\sqrt{d}}{2}\right]$ if $d \equiv 1 \bmod 4$.

Example 100. *If R is any quadratic ring, there exists an embedding $R \hookrightarrow M_2(\mathbf{Z})$. Clearly if $R = \mathbf{Z}[\sqrt{d}]$ then our previous discussion of quaternions leads us to the embedding*

$$\mathbf{Z}[\sqrt{d}] \hookrightarrow M_2(\mathbf{Z}): \qquad \sqrt{d} \mapsto \begin{pmatrix} 0 & 1 \\ d & 0 \end{pmatrix}.$$

If $d = 4m+1$, then the minimal polynomial for $\dfrac{1+\sqrt{d}}{2}$ is X^2-X-m. Therefore we have an embedding

$$\mathbf{Z}\left[\frac{1+\sqrt{d}}{2}\right] \hookrightarrow M_2(\mathbf{Z}): \qquad \frac{1+\sqrt{d}}{2} \mapsto \begin{pmatrix} 1 & 1 \\ m & 0 \end{pmatrix}.$$

Note that in $M_2(\mathbf{Z})$ if $M = \begin{pmatrix} 1 & 1 \\ m & 0 \end{pmatrix}$ then

$$M^2 - M = \begin{pmatrix} m+1 & 1 \\ m & m \end{pmatrix} - \begin{pmatrix} 1 & 1 \\ m & 0 \end{pmatrix} = \begin{pmatrix} m & 0 \\ 0 & m \end{pmatrix}.$$

We note that if we formally take $x = \frac{1+\sqrt{r}}{2}$ then $2x - 1 = \sqrt{r}$. If $r = 4m + 1$ and we take $M = \begin{pmatrix} 1 & 1 \\ m & 0 \end{pmatrix}$, then we saw that we could take M as a copy of our formal x in $M_2(\mathbf{Z})$. What can we say about $2M - 1 = \begin{pmatrix} 1 & 2 \\ 2m & -1 \end{pmatrix}$? Could it be $M_2(\mathbf{Z})$-conjugate to our other square root of r as given by $\begin{pmatrix} 0 & 1 \\ r & 0 \end{pmatrix}$?

Let $I = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbf{Z})$ so that $I^{-1} = \pm \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. We will see that the negative sign does not matter, because

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 2m & -1 \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} * & * \\ 2dc - 2c^2 + 2d^2m & * \end{pmatrix}.$$

Since $r = 4m + 1$, $2 \nmid r$ and this is impossible.

The reason these embeddings are different is that the embedding given by $\sqrt{r} \mapsto \begin{pmatrix} 0 & 1 \\ r & 0 \end{pmatrix}$ is optimal for $\mathbf{Z}[\sqrt{r}]$ while if $r = 4m + 1$,the embedding $\sqrt{r} \mapsto \begin{pmatrix} 1 & 2 \\ 2m & -1 \end{pmatrix}$ is optimal for $\mathbf{Z}\left[\frac{1+\sqrt{r}}{2}\right]$.

It turns out that the only way that we can get non-conjugate optimal embeddings of a quadratic ring R is if the class number of R is greater than one.

Definition 101. *To any binary quadratic form $q(x, y) \in \mathbf{Z}[x, y]$, we can naturally associate the discriminant of $f(x) = q(x, 1)$, which is an isomorphism invariant of q. Let $\Delta \in \mathbf{Z}$ be such that $\Delta \equiv 0, 1 \bmod 4$, so that Δ can be the discriminant of a binary quadratic form. Define the class number $h(\Delta)$ to be the number of non-isomorphic binary quadratic forms of discriminant Δ.*

Gauss discovered a natural connection between the ideals of the quadratic ring R_Δ and the binary quadratic forms of discriminant Δ.

Example 102. *The ring $\mathbf{Z}[\sqrt{-5}]$ has discriminant -20. There are two isomorphism classes of quadratic forms $q(x, y)$ of discriminant -20, represented by x^2+5y^2 and $2x^2+2xy+3y^2$. We see how this is reflected in embeddings of $\mathbf{Z}[\sqrt{-5}]$ into $M_2(\mathbf{Z})$.*

Consider the matrices $\begin{pmatrix} 0 & 1 \\ -5 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & -2 \\ 3 & -1 \end{pmatrix}$. Both square to $\begin{pmatrix} -5 & 0 \\ 0 & -5 \end{pmatrix}$. Suppose that these two matrices were conjugate.

Consider that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -5 & 0 \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} * & a^2+5b^2 \\ * & * \end{pmatrix}.$$

Clearly then, these two matrices are $\mathrm{GL}_2(\mathbf{Z})$*-conjugate if and only if* $a^2+5b^2 = \pm 2$*. Clearly though this is impossible because if so, then there exists some* $a \in \mathbb{F}_5$ *such that* $a^2 = \pm 2$*, while the only squares mod 5 are* $0, \pm 1$*.*

Similarly, the other isomorphism class shows up otherwise as

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & -2 \\ 3 & -1 \end{pmatrix}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} * & * \\ 2c^2+2cd+3d^2 & * \end{pmatrix}.$$

Hopefully now, we have given evidence to suggest the following theorem.

Theorem 103. *If* $\Delta \equiv 0, 1 \bmod 4$ *then the number of non-conjugate optimal embeddings of the unique (up to isomorphism) quadratic ring* R_Δ *of discriminant* Δ *into* $M_2(\mathbf{Z})$ *is* $h(\Delta)$*.*

It is easy to verify that $R_{-3} = \mathbf{Z}[X]/(X^2+X+1)$ and $R_{-4} = \mathbf{Z}[X]/(X^2+1)$ each have class number 1 and thus verify that

$$g(SL_2(\mathbf{Z})) = 1 + \frac{1}{12} - \frac{1}{4} - \frac{1}{3} - \frac{1}{2} = 0.$$

As an exercise, we may also use this theorem to verify that the class number of $\mathbf{Z}[X]/(X^2)$ is one. We will also understand embeddings into division algebras. For now though, we now recall the definition of the Legendre symbol.

Definition 104. *For an odd prime* p*, define the* Legendre symbol $\left(\frac{\Delta}{p}\right)$ *to be* 0 *if* $p \mid \Delta$*,* 1 *if* Δ *is a nonzero square modulo* p*, and* -1 *if* Δ *is not a square modulo* p*.*

If p is an odd prime and $R = R_\Delta$ then R/pR is respectively in the split, inert, or ramified cases if $\left(\frac{\Delta}{p}\right) = 1, -1$, or 0. We extend the Legendre Symbol now to $p = 2$.

Definition 105. *If* p *is a prime and* $\Delta \in \mathbf{Z}$ *then if* p *is odd then* $\left(\frac{\Delta}{p}\right)$ *is the Legendre symbol. If* $p = 2$ *and* Δ *is even then* $\left(\frac{\Delta}{2}\right) = 0$*. If* $p = 2$ *and* Δ *is odd then* $\left(\frac{\Delta}{2}\right) = (-1)^{\frac{\Delta^2-1}{8}}$*.*

As a corollary of a more general theorem on embeddings, we can state the following lemma.

Lemma 106. *Let $B/\mathbf{Q}$ be a quaternion algebra, $B \hookrightarrow M_2(\mathbf{R})$, and $\mathcal{O}$ a maximal order. Let e_2 be the number of conjugacy classes of embeddings $\mathbf{Z}[X]/(X^2+1) \hookrightarrow \mathcal{O}$ and e_3 the number of conjugacy classes of embeddings $\mathbf{Z}[X]/(X^2+X+1) \hookrightarrow \mathcal{O}$. Then all such embeddings are optimal and*

$$e_2 = \prod_{p|\mathrm{disc}(\mathcal{O})} \left(1 - \left(\frac{-4}{p}\right)\right), e_3 = \prod_{p|\mathrm{disc}(\mathcal{O})} \left(1 - \left(\frac{-3}{p}\right)\right).$$

We can now easily and numerically compute the genus of a Shimura curve X_B.

Theorem 107. *Let $B/\mathbf{Q}$ be a quaternion algebra, $B \hookrightarrow M_2(\mathbf{R})$, and $\mathcal{O}$ a maximal order. We have*

$$\begin{aligned} g(X_B) &= 1 + \frac{\phi(\mathrm{disc}(\mathcal{O})}{12} - \prod_{p|\mathrm{disc}(\mathcal{O})} \left(1 - \left(\frac{-4}{p}\right)\right) \\ &- \prod_{p|\mathrm{disc}(\mathcal{O})} \left(1 - \left(\frac{-3}{p}\right)\right) - \frac{\#\,\mathrm{cusp}(\mathcal{O}^1)}{2}. \end{aligned}$$

9. Lecture 9: Embeddings, cont'd, Involutions on Curves, Normalizers and Automorphisms

So where do the numbers in that genus formula come from anyhow?

Suppose that R is a quadratic ring, i.e. $R = \mathbf{Z}[X]/(f)$ where $f \in \mathbf{Z}[X]$ is monic of degree 2. Let $K = \mathbf{Q}[X]/(f)$ and let R_0 be the maximal quadratic ring in K containing R. Let $\Delta = \mathrm{disc}(R)$, $\Delta_0 = \mathrm{disc}(R_0)$.

Definition 108. *If $\Delta \neq 0$, let $f(\Delta) = \sqrt{\frac{\Delta}{\Delta_0}}$, else $f(\Delta) = 0$.*

Example 109. *If $R = \mathbf{Z}[\sqrt{-45}]$ then $K = \mathbf{Z}[\sqrt{-5}]$ and $R_0 = \mathbf{Z}[\sqrt{-5}]$. Note that $\Delta = -180$ and $\Delta_0 = -20$, so $f(\Delta) = 3$.*

Example 110. *If $R = \mathbf{Z}[\sqrt{-3}]$ then $K = \mathbf{Z}[\sqrt{-3}]$ and so $R_0 = \mathbf{Z}\left[\frac{1+\sqrt{-3}}{2}\right]$. Note that $\Delta = -12$ and $\Delta_0 = -3$ so $f(\Delta) = 2$.*

Note that if $\mathcal{O}$ is an order in a quaternion algebra $B/\mathbf{Q}$ and $\phi : R \hookrightarrow \mathcal{O}$ is an embedding of a quadratic ring, then there is an embedding $\phi^0 : K \hookrightarrow B$. Note that there's nothing specific to quaternions here,

just that R and $\mathcal{O}$ are torsion-free $\mathbf{Z}$-algebras. We note that $(\phi^0)^{-1}(\mathcal{O})$ is a quadratic ring inside K containing $\mathcal{O}$.

Definition 111. *We say that ϕ is optimal when $(\phi^0)^{-1}(\mathcal{O}) = R$.*

The idea here is that Optimal embeddings $R \hookrightarrow \mathcal{O}$ should be in bijection with Optimal Embeddings " modulo p for all primes p." The relevant tool here is the Hasse-Minkowski theorem for quadratic forms. If R is a quadratic ring, we know many things about R/pR. It must be a two-dimensional $\mathbb{F}_p$-algebra, and must be of one of the following three forms.

- $\mathbb{F}_p \oplus \mathbb{F}_p$ (the split algebra)
- $\mathbb{F}_{p^2}$ (the inert algebra)
- $\mathbb{F}_p[T]/(T^2)$ (the ramified algebra)

Example 112. *Let $R_{-4} = \mathbf{Z}[X]/(X^2+1)$. Note that $R_{-4}/2R_{-4} = \mathbb{F}_2[T]/(T^2+1) = \mathbb{F}_2[T]/(T+1)^2 = \mathbb{F}_2[T+1]/(T+1)^2$ and hence we are in the ramified case.*

Let $R_{-3} = \mathbf{Z}[X]/(X^2+X+1)$. Note that $R_{-3}/2R_{-3} = \mathbb{F}_2[T]/(T^2+T+1) \cong \mathbb{F}_4$ and hence we are in the ramified case.

Let $R_{-7} = \mathbf{Z}[X]/(X^2-X+2)$. Note that $R_{-7}/2R_{-7} = \mathbb{F}_2[T]/(T^2-T+2) = \mathbb{F}_2[T]/(T(T+1))$ and hence we are in the split case.

We also have the following structure theorem for maximal orders in quaternion algebras.

Theorem 113. *If $\mathcal{O}$ is a maximal order in a quaternion algebra B then*

1. *If $p \nmid \operatorname{disc}(\mathcal{O})$ then $\mathcal{O}/p\mathcal{O} \cong M_2(\mathbb{F}_p)$.*
2. *If $p \mid \operatorname{disc}(\mathcal{O})$ then*

$$\mathcal{O}/p\mathcal{O} \cong \left\{ \begin{pmatrix} x & y \\ 0 & x^p \end{pmatrix} \in M_2(\mathbb{F}_{p^2}) \right\}.$$

Lemma 114. *Suppose $\mathcal{O}$ is a maximal order in a quaternion algebra B, $R = R_\Delta$, and $\operatorname{disc}(\mathcal{O}) = \prod_{i=1}^n p_i$. The number of non-conjugate embeddings $R/pR \hookrightarrow \mathcal{O}/p\mathcal{O}$ is*

- *0 in the split case,*
- *1 in the ramified case, and*
- *2 in the inert case.*

Before proving this, recall the following definition.

Definition 115. *An idempotent of a ring R is an element e such that $e^2 = e$. A* nontrivial idempotent *is an idempotent besides 0 or 1.*

Proof. For the split case, note that $\mathbb{F}_p \oplus \mathbb{F}_p$ has nontrivial idempotents such as $(1,0)$ and $(0,1)$. Meanwhile, if $\begin{pmatrix} x & y \\ 0 & x^p \end{pmatrix}$ is idempotent, we have that $x^2 = x$ and $y(x + x^p) = y$. It is easy to show then that there are only trivial idempotents in $\mathcal{O}/p\mathcal{O}$, i.e. $x = 0, 1$ and $y = 0$.

For the ramified case, we have the natural embedding $x + yT \mapsto \begin{pmatrix} x & y \\ 0 & x^p \end{pmatrix} = \begin{pmatrix} x & y \\ 0 & x \end{pmatrix}$. All other embeddings are conjugate to this one as it has to be the reduction of an embedding $\mathbf{Z}[X]/(X^2) \hookrightarrow M_2(\mathbf{Z})$.

For the inert case, we have the natural embedding $x \mapsto \begin{pmatrix} x & 0 \\ 0 & x^p \end{pmatrix}$ and the Galois conjugate map $\begin{pmatrix} x^p & 0 \\ 0 & x \end{pmatrix}$. □

Now to really be able to calculate the number of embeddings up to conjugacy, we need to make a slight correction to the Kronecker symbol $\left(\frac{\cdot}{p}\right)$.

Example 116. *Looking just mod p, one might think that* $\mathbf{Z}[\sqrt{-45}]$ *embeds into a maximal order* $\mathcal{O}$ *with discriminant* 6 *(say inside* $\left(\frac{-1,3}{\mathbf{Q}}\right)$*). After all* $\left(\frac{-180}{2}\right) = 0$ *and* $\left(\frac{-180}{3}\right) = 0$ *so there are embeddings mod p for all p. The trick is that Hasse Minkowski is not really about mod p behavior, but about mod* p^n *behavior for all n. In fact* $\mathbf{Z}[\sqrt{-45}]$ *does not embed, and this is suggested (but not proved) by the fact that* $\mathbf{Z}[\sqrt{-5}]$ *does not embed. Note that even modulo 3,* $\mathbf{Z}[\sqrt{-5}]$ *does not embed because*

$$\left(\frac{-20}{3}\right) = \left(\frac{-1}{3}\right)\left(\frac{5}{3}\right) = \left(\frac{-1}{3}\right)^2 = 1.$$

Definition 117. *Let the Eichler symbol be defined as*

$$\left\{\frac{\Delta}{p}\right\} = \begin{cases} \left(\frac{\Delta}{p}\right) & p \nmid f(\Delta) \\ 1 & p \mid f(\Delta). \end{cases}$$

This is the final piece of terminology we need to state the following very imporant theorem.

Theorem 118 (Eichler's Embedding Theorem)**.** *Let* $B/\mathbf{Q}$ *be a quaternion algebra,* $B \hookrightarrow M_2(\mathbf{R})$*, and* $\mathcal{O}$ *a maximal order of* B*. Then the*

number of conjugacy classes of optimal embeddings $R_\Delta \hookrightarrow \mathcal{O}$ is

$$h(\Delta) \prod_{p|\mathrm{disc}(\mathcal{O})} \left(1 - \left\{\frac{\Delta}{p}\right\}\right).$$

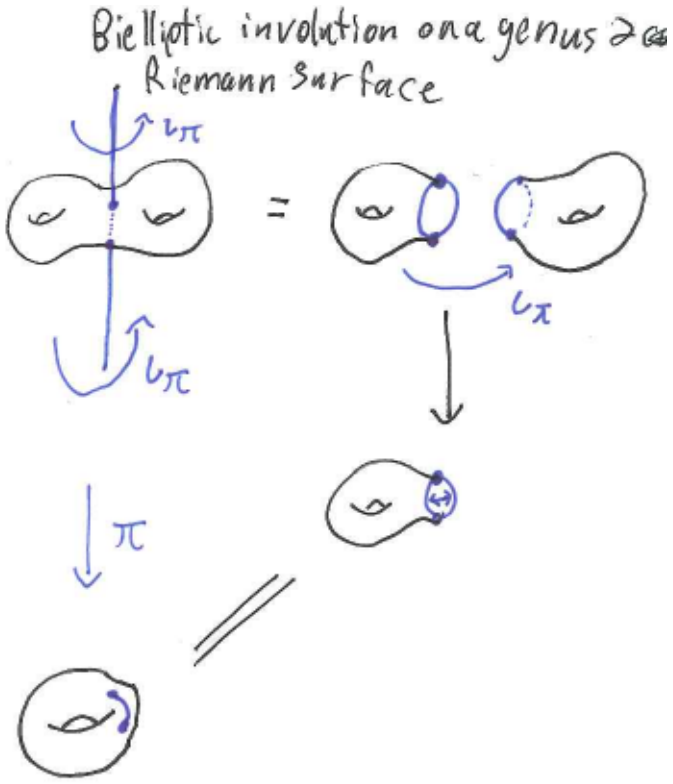
We are concerned with these embedding numbers because they are the number of fixed points of certain important automorphisms of Shimura curves. Let us briefly consider involutions on compact Riemann surfaces.

Definition 119. *If X is a compact Riemann surface, an involution $\iota : X \to X$ is a nonidentity holomorphic map such that ι^2 is the identity.*

Lemma 120. *There is a one-to-one correspondence between involutions $\iota : X \to X$ and degree two holomorphic maps $\pi : X \to Y$ where Y is a Riemann surface.*

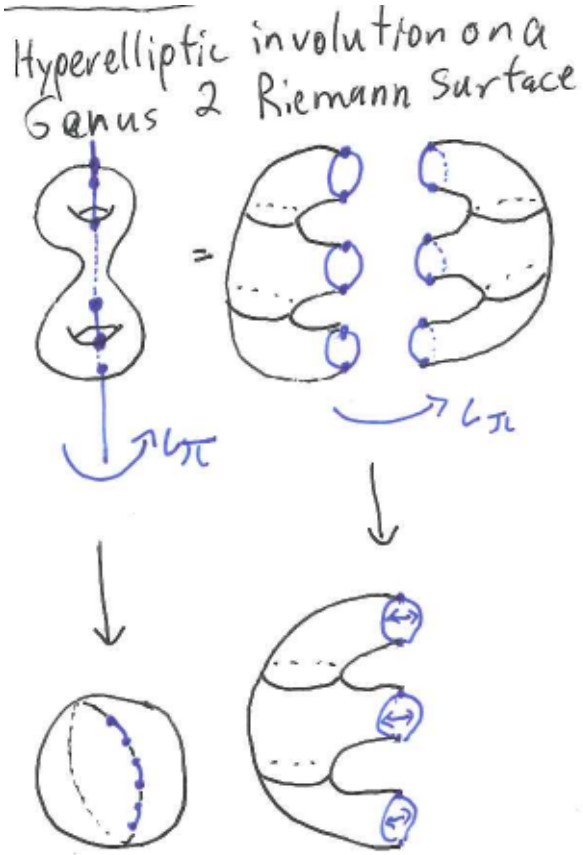
Proof. To an involution ι, we have the action of the finite group $\langle \iota \rangle$ on X. Since this group is finite, it acts properly discontinuously on X and the quotient map $\pi_\iota : X \to X/\langle \iota \rangle$ is a holomorphic map, giving a Riemann surface structure to its target. Conversely, if $\pi : X \to Y$ has degree 2, we can form an involution π_ι taking x in x to the " other" preimage of $\pi(x)$ in X. We put "other" in quotation marks because x may be a ramification point of π, in which case it is a fixed point of ι_π. □

Definition 121. *We say that X is a bielliptic Riemann Surface if there is a $2:1$ map $\pi : X \to Y$ such that the genus of Y is 1.*



Definition 122. *We say that X is a hyperelliptic Riemann Surface if there is a $2:1$ map $\pi X \to \mathbb{P}^1(\mathbf{C})$.*

We note that a hyperelliptic involution must have $2g(X)+2$ fixed points.



We focus on involutions, because the curves X_B will tend to have lots of involutions in some sense.

Definition 123. *Suppose* $\Gamma < \mathrm{PSL}_2(\mathbf{R})$. *Define* $N\Gamma = N_{\mathrm{PSL}_2(\mathbf{R})}(\Gamma) = \{\phi \in \mathrm{PSL}_2(\mathbf{R}) : \phi\Gamma\phi^{-1} = \Gamma\}$.

Theorem 124 (Katok, Theorem 2.3.8)**.** *If* $\Gamma < \mathrm{PSL}_2(\mathbf{R})$ *is Fuchsian and non-abelian, then* $N\Gamma$ *is a Fuchsian group.*

Idea. First you show that if $\alpha, \beta \in \mathrm{PSL}_2(\mathbf{R})$, then $\alpha\beta = \beta\alpha$ if and only if the fixed point sets of α and β are equal. If $N\Gamma$ is not Fuchsian, then you can find $T_i \in N\Gamma$ such that $T_i \to 1$ and each T_i is distinct. Therefore if $\gamma \in \Gamma$ then $T_i\gamma T_i^{-1} \to \gamma$. Pick $\gamma' \in \Gamma$ with a different fixed point set than γ. Since Γ is discrete, there are only finitely many i such that $T_i\gamma T_i^{-1} \neq \gamma$. Therefore, for all but finitely many i, the fixed point sets of γ and T_i are the same. On the other hand, $T_i\gamma' T_i^{-1} \neq \gamma'$ for only finitely many i. Therefore, for infinitely many i, γ, T_i and γ' all have the same fixed point set, which is a contradiction. □

Corollary 125. *If* $\mathrm{covol}(\Gamma) < \infty$ *then* $N\Gamma/\Gamma$ *is finite.*

Proof. $\mathrm{covol}(N\Gamma)[N\Gamma : \Gamma] = \mathrm{covol}(\Gamma)$ □

Definition 126. *If* $\mathrm{covol}(\Gamma) < \infty$*, we say that* $N\Gamma/\Gamma$ *is the* modular automorphism group *of* $X(\Gamma)$*.*

If we give this group such a flashy name, then there ought to be a reason to do so. In particular, this group had better act on $X(\Gamma)$.

Theorem 127. *Let* $\Gamma < \mathrm{PSL}_2(\mathbf{R})$ *be a finite covolume Fuchsian group. Then there is a natural map* $N\Gamma \to \mathrm{Aut}(X(\Gamma))$ *with kernel* Γ*.*

Proof. First, we have a natural inclusion $\mathrm{Aut}(Y(\Gamma)) \hookrightarrow \mathrm{Aut}(X(\Gamma))$. If we have an automorphism $Y(\Gamma) \to Y(\Gamma)$ then this map has removable singularities at each cusp. Consider the map $\mathcal{H} \to \mathcal{H}$ by $\tau \mapsto \phi\tau$ with $\phi \in N\Gamma$. When we map to the quotient $\phi\Gamma\phi^{-1} \setminus \mathcal{H} = \Gamma \setminus \mathcal{H}$, we are left with $(\phi\Gamma\phi^{-1})\phi\tau = \phi\Gamma\tau = \Gamma\phi\tau$. The preimage of $\phi\Gamma\phi^{-1}$ is Γ and thus we have an automorphism $Y(\Gamma) \to Y(\Gamma)$ by $\Gamma\tau \mapsto \phi\Gamma\tau$. Clearly this map is trivial if and only if $\phi \in \Gamma$. □

We note that in many cases, the modular automorphism group of $X(\Gamma)$ is the full automorphism group.

Lemma 128. *If* Γ *is purely hyperbolic, then* $\mathrm{Aut}(X(\Gamma)) = N\Gamma/\Gamma$*.*

Proof. Kontogeorgis-Rotger Proposition 1.2. □

Now we return to the case of Shimura curves. In the following, let $B/\mathbf{Q}$ be a quaternion algebra,$\psi : B \hookrightarrow M_2(\mathbf{R})$ and let $\mathcal{O}$ be a maximal order. Let $\Gamma = \pm\psi(\mathcal{O}^1) < \mathrm{PSL}_2(\mathbf{R})$, let $\tilde{\Gamma} = \psi(\mathcal{O}^1) < \mathrm{SL}_2(\mathbf{R})$, and let $N\tilde{\Gamma} = N_{\mathrm{SL}_2(\mathbf{R})}(\tilde{\Gamma})$. The quotient maps $N\tilde{\Gamma} \to N\Gamma \to N\Gamma/\Gamma$ induce an isomorphism $N\tilde{\Gamma}/\tilde{\Gamma} \cong N\Gamma/\Gamma$ because $\pm 1 \in \tilde{\Gamma} \subset N\tilde{\Gamma}$.

Moreover, if $\phi \in N\tilde{\Gamma}$ then $\phi\psi(\mathcal{O}^1)\phi^{-1} = \psi(\mathcal{O}^1)$. If we extend this to B, we get $\phi\psi(B)\phi^{-1} = \psi(B)$ in $M_2(\mathbf{R})$. Therefore ϕ induces an automorphism of B and by the Skolem-Noether Theorem, there is some $\alpha \in B^\times$ such that $\phi = \psi(\alpha)$.

Theorem 129 (Skolem-Noether)**.** *If* $\epsilon : B \to B$ *is an automorphism, there is some* $\alpha_\epsilon \in B^\times$ *such that* $\epsilon(b) = \alpha_\epsilon b \alpha_\epsilon^{-1}$ *for all* $b \in B$*.*

With this in mind, we are really looking at the normalizer in B of $\mathcal{O}$. To this end, consider the following.

Definition 130. *If* $\mathcal{O}$ *is a maximal order in an indefinite quaternion algebra* $B/\mathbf{Q}$*, set*

$$N\mathcal{O} = \{b \in B^\times : b\mathcal{O}b^{-1} = \mathcal{O}, n(b) > 0\}.$$

In a theorem due to Michon, $N\mathcal{O}$ has a very definite structure.

Theorem 131 (Michon)**.** *Let $\mathcal{O}$ be a maximal order of an indefinite quaternion algebra B. For all $p \mid \mathrm{disc}(\mathcal{O})$ there exist $\beta_p \in \mathcal{O}$ such that $n(\beta_p) = p$ and if $b \in N\mathcal{O}$, there exist $\epsilon_p \in \{0,1\}$, $r \in \mathbf{Q}^\times$ and $\gamma \in \mathcal{O}^\times$ (and thus $\mathcal{O}^1$) such that*

$$b = \left(\prod_{p|\mathrm{disc}(\mathcal{O})} \beta_p^{\epsilon_p} \right) \gamma r,$$

and while the order of the β_p, may change r or γ, it will not change the ϵ_p.

Tools and references. This theorem is really a statement about the "two-sided fractional ideals" of $\mathcal{O}$. The existence of the β_p may be proved using the Eichler-Selberg trace formula, which is a topic that may deserve its own course. This proof is claimed in Michon's "Courbes de Shimura Hyperelliptiques," although the proof is passed off to the book of Vignerás. □

Corollary 132. *There is an isomorphism*

$$N\mathcal{O}/\mathcal{O}^1\mathbf{Q}^\times \cong N\Gamma/\Gamma \cong (\mathbf{Z}/2\mathbf{Z})^{\#\{p|\mathrm{disc}(\mathcal{O}),p \text{ prime.}\}}.$$

Proof. This isomorphism is induced by the map $N\mathcal{O} \to N\tilde{\Gamma}$ by $b \mapsto \dfrac{\psi(b)}{\sqrt{n(b)}}$. Note that if $b \in \mathbf{Q}^\times$, then $n(b) = b^2$ and $\sqrt{n(b)} = |b| \in \mathbf{R}$. Likewise, if b has diagonal image under ψ, then $b \in \mathbf{Q}$. By Skolem-Noether, we thus have an isomorphism between $N\mathcal{O}/\mathbf{Q}^\times$ and $N\tilde{\Gamma}/\{\pm 1\}$. The result follows. □

10. Lecture 10: Complex Tori and Abelian varieties

Last time, we showed that if $B/\mathbf{Q}$ is a quaternion algebra, $\psi : B \hookrightarrow M_2(\mathbf{R})$, $\mathcal{O}$ is a maximal order, and $\Gamma = \pm\psi(\mathcal{O}^1)$, then

$$N\Gamma/\Gamma \cong (\mathbf{Z}/2\mathbf{Z})^{\#\{p|\mathrm{disc}(\mathcal{O}),p \text{ prime.}\}} \hookrightarrow \mathrm{Aut}(X(\Gamma)).$$

Definition 133. *To an element $(\ldots, \epsilon_p, \ldots) \in (\mathbf{Z}/2\mathbf{Z})^{\#\{p|\mathrm{disc}(\mathcal{O}),p \text{ prime.}\}}$ where $\epsilon_p \in \{0,1\}$, let $m = \prod_{\substack{p|\mathrm{disc}(\mathcal{O}) \\ \epsilon_p = 1}} p$. Denote the corresponding involution of $N\Gamma/\Gamma$ by w_m and call this an* Atkin-Lehner Involution.

These are important objects in the study of Shimura curves because they give rise to Hecke Operators. One might still ask though, why one would want to study Shimura curves at all. The reason we will give today is that Shimura curves parametrize certain special objects called abelian varieties with a certain structure. We will further see that

under this correspondence, fixed points of w_m give abelian varieties with still more structure given by the ring $\mathbf{Z}[\sqrt{-m}]$.

Definition 134. *A topological torus is a topological space homeomorphic to $(S^1)^n$ for some $n \in \mathbf{Z}_{\geq 1}$.*

We bring up this definition because we have already seen some examples of complex-analytic objects like Riemann surfaces. We seek to broaden our horizons and find other complex analytic objects. The case of tori is a potentially easy test case for higher dimensional objects because we would automatically have compactness and the structure of (at least) a topological group.

Example 135. *Suppose that $L \cong \mathbf{Z}^2$ is a group of complex-analytic isomorphisms $\mathbf{C} \to \mathbf{C}$ such that if $\ell \in L$ then ℓ acts nontrivially. Represent L as a discrete subgroup of $\mathbf{C}$. Then the quotient $\mathbf{C}/L$ is a topological torus. Moreover, it is also a Riemann surface of genus one and so there are lots of nonconstant meromorphic functions on $E = \mathbf{C}/L$.*

Definition 136. *An elliptic curve over $\mathbf{C}$ is a pair $E = \{S, P\}$ such that S is a genus one Riemann surface and $P \in S$ is a point, which will serve as the identity for the group law on E.*

Example 137. *Moving the identity point around is a holomorphic map. Namely if $P_1, P_2 \in S$, then the map $S \to S$ given by $Q \mapsto Q + P_2 - P_1$ is holomorphic and induces an isomorphism of elliptic curves $\{S, P_1\} \to \{S, P_2\}$.*

Note that as elliptic curves, these pairs are different, and so we do not consider this an isomorphism $E \to E$, nor even a holomorphic map $E \to E$. If $E = \{S, P\}$ is an elliptic curve, we consider a map of elliptic curves $E \to E$ to be a holomorphic map $S \to S$ fixing P.

Example 138. *If $n \in \mathbf{Z}$ then the multiplication by n map induced by the group law on E is a holomorphic map $[n] : E \to E$.*

Definition 139. *Consider that the set of elliptic curve maps $E \to E$ form a ring with multiplication given by conjugation and addition given by the group law. Call this ring* $\mathrm{End}(E)$.

Consider now how we formed $S = \mathbf{C}/L$. We note that L is a discrete subgroup of $\mathbf{C}$, acting by translation. We call this a lattice. Moreover, to guarantee compactness, we need L to be "big enough," in some sense.

Definition 140. *Let V be a finite dimensional complex vector space. Let V' be V considered as a real vector space. Let L be a discrete*

subgroup of V with a $\mathbf{Z}$-basis $\alpha_1, \ldots, \alpha_n$. Let $W \subset V'$ be the real vector space generated by the α_i. We say the L is a full lattice *of V if $W = V'$.*

Note that necessarily, the $\mathbf{Z}$-rank of a full lattice $L < V$ is $2g$ where g is the $\mathbf{C}$-dimension of V. Note also that V/L is now homeomorphic to $(S^1)^{2g}$. With this in mind we make the following definition.

Definition 141. *If V is a g-dimensional complex vector space and $L < V$ is a full lattice, let $Q = V/L$ and let $P \in T$. We say that the pair $T = \{V/L, P\}$ is a* complex torus *of dimension g.*

It may be tempting to simply think of these tori as $\mathbf{C}^g/\mathbf{Z}^{2g}$, and indeed without much harm we can consider a complex torus as either $\mathbf{C}^g/L$ or $V/\mathbf{Z}^{2g}$ but going further than that shreds all of our complex-analytic information.

Example 142. *Let $L_1 = \mathbf{Z} \oplus i\mathbf{Z}$, $L_2 = \mathbf{Z} \oplus i\sqrt{2}\mathbf{Z}$ and let $L_3 = \mathbf{Z} \oplus i\pi\mathbf{Z}$. Likewise, let $E_i = \mathbf{C}/L_i$. It turns out that $\mathrm{End}(E_3) = \mathbf{Z}$, that is, it is made of only the maps $[n]$. Meanwhile $\mathrm{End}(E_1) = \mathbf{Z}[i]$ and $\mathrm{End}(E_2) = \mathbf{Z}[i\sqrt{2}]$, which have 4 and 2 invertible elements respectively.*

Recall however that our original aim was to create complex-analytic objects. While these tori have the flavor of being complex analytic, they may be somewhat barren in terms of holomorphic or even meromorphic functions. In dimension one, we are saved by the virtue of elliptic curves being given as Riemann surfaces. In higher dimensions we may not be so lucky.

Example 143 (Siegel, Analytic functions of Several Complex Variables, p.104)**.** *Let $V = \mathbf{C}^2$ and consider the lattice L given by $\Pi\mathbf{Z}^4$, where*

$$\Pi = \begin{pmatrix} 1 & 0 & i\sqrt{2} & i\sqrt{5} \\ 0 & 1 & i\sqrt{3} & i\sqrt{7} \end{pmatrix}.$$

It is easy to see that L is a full lattice in $\mathbf{C}^2$, but Siegel showed that the only meromorphic functions on V/L are constant maps. This is to say that the field of meromorphic functions $\mathbf{C}(V/L)$ is none other than $\mathbf{C}$.

In fact, we would really like a complex torus V/L of dimension g to have a large number of meromorphic functions. We should in fact have the transcendence degree of $\mathbf{C}(V/L)$ equal to g. To fix this, we introduce the following.

Definition 144. *A* Hermitian form *H on a complex vector space V is a complex bilinear map*

$$\overline{V} \times V \to \mathbf{C}.$$

This is to say that if $x, y, x_1, x_2, y_1, y_2 \in V$, $a, b \in \mathbf{C}$ then H is given as a map $V \times V \to \mathbf{C}$ such that

$$H(x_1 + x_2, y) = H(x_1, y) + H(x_2, y),$$

$$H(x, y_1 + y_2) = H(x, y_1) + H(x, y_2),$$

$$H(ax, by) = \overline{a}bH(x, y),$$

and

$$\overline{H(x, y)} = H(y, x).$$

Example 145. *On the elliptic curve given by the lattice $\mathbf{Z} \oplus i\mathbf{Z} < \mathbf{C}$, we have the Hermitian form $H(z, z') = \overline{z}z'$.*

Definition 146. *If M is an R-module for a commutative ring R, we say that a bilinear form $B : M \times M \to R$ is skew-symmetric if $B(x, y) = -B(y, x)$. We say that B is alternating if $B(x, x) = 0$ for all $x \in M$.*

Lemma 147 (Milne, Abelian Varieties, Lemma 1.2.4)**.** *There is a one-to-one correspondence between Hermitian forms H on V and $\mathbf{R}$-valued skew-symmetric forms E on V such that $E(iv, iw) = E(v, w)$. This correspondence is given by*

$$E(v, w) = \Im(H(v, w)),$$

$$H(v, w) = E(iv, w) + iE(v, w).$$

The idea is to use our Hermitian form H to construct a holomorphic function $f_v = H(v, \cdot)$ for all $v \in V$. If this is going to work, the Hermitian form H had better respect the lattice $L < V$.

Definition 148. *Let V be a finite-dimensional complex vector space and let L be a full lattice in V. If the imaginary part of a positive-definite Hermitian form $E : V \times V \to \mathbf{R}$ takes integer values on L, we say that E is a Riemann Form.*

An equivalent definition to the one above is that if V' is V thought of as an $\mathbf{R}$-vector space, then a Riemann form is a skew-symmetric bilinear form $E : L \times L \to \mathbf{Z}$ such that

- $E_{\mathbf{R}} : V' \times V' \to \mathbf{R}$ satisfies $E_{\mathbf{R}}(iv, iw) = E_{\mathbf{R}}(v, w)$
- The Hermitian form associated to $E_{\mathbf{R}}$ is positive definite.

Definition 149. *A (polarized) abelian variety over $\mathbf{C}$ is a complex torus V/L equipped with a Riemann form $E : L \times L \to \mathbf{Z}$.*

We note here that abelian varieties are *projective* in the following sense.

Definition 150. *We define projective n-space over $\mathbf{C}$ to be*

$$\mathbb{P}^n(\mathbf{C}) = \left\{ \begin{array}{c} x_i \in \mathbf{C}, \forall \lambda \in \mathbf{C}^\times \\ [x_0 : \cdots : x_n] : [x_0 : \cdots : x_n] = [\lambda x_0 : \cdots : \lambda x_n] \end{array} \right\}.$$

Theorem 151. *For all abelian varieties $A = \{V/L, L, E\}$ over $\mathbf{C}$, there is a positive integer n and an embedding $V/L \hookrightarrow \mathbb{P}^n(\mathbf{C})$ such that V/L is the zero set of a finite number of homogeneous polynomials in $n+1$ variables.*

Idea. This follows from the Appell-Humbert Theorem and a theorem of Poincare on line bundles. For details, please see *Complex Abelian Varieties* by Birkenhake and Lange. □

The following is a key example.

Example 152. *Let $\tau \in \mathcal{H}$, $B/\mathbf{Q}$ be a quaternion algebra, $\psi : B \hookrightarrow M_2(\mathbf{R})$, and $\mathcal{O}$ a maximal order in B. It follows that $\psi(\mathcal{O})\binom{\tau}{1}$ is a full lattice in $\mathbf{C}^2$. By Eichler's Embedding Theorem, we can pick some $\mu \in \mathcal{O}$ such that $\mu^2 = -\operatorname{disc}(\mathcal{O})$. We may endow $A = \mathbf{C}/\iota(\mathcal{O})\binom{\tau}{1}$ with a Riemann form $E : \mathcal{O} \times \mathcal{O} \to \mathbf{Z}$ by $E(x, y) = t(\mu x \overline{y})$. It follows that (A, E) forms a polarized abelian variety over $\mathbf{C}$.*

11. Lecture 11: Shimura Curves and Abelian varieties

Let's sum up what we know about abelian varieties:

- A Complex Torus is a finite-dimensional complex vector space and $L < V$ is a full lattice.
- A Riemann Form for a full lattice $L < V$ is an alternating form $E : L \times L \to \mathbf{Z}$ which extends to a positive-definite Hermitian form on V.
- An Abelian variety is a complex torus which admits a Riemann form. A polarized abelian variety is an abelian variety with a choice of Riemann form.

Example 153. *Let $V = \mathbf{C}$ and let $L = \mathbf{Z} \oplus i\mathbf{Z}$. Then*

$$E(x + iy, x' + iy') = xy' = x'y$$

extends to the positive definite Hermitian form $H(z, z') = \overline{z}z'$ and is thus a Riemann form. Every Riemann form on V/L is equivalent to this one in some way.

We also note that by the Appell-Humbert theorem, Riemann forms are in correspondence with embeddings into projective space. In particular, they can be embedded into projective space as the combined

zero sets of complex homogeneous polynomials. This makse them of interest for us.

We have thus described the objects of the category of abelian varieties of dimension g. Let us describe the morphisms. Let $A_1 = \mathbf{C}^g/L_1$, $A_2 = \mathbf{C}^g/L_2$ be complex tori. Note that if $M \in M_g(\mathbf{C})$ is such that $ML_1 \subset L_2$ then the map induced by M is holomorphic. In fact the converse is true. If we have a holomorphic map $\phi : A_1 \to A_2$ then we must fix the identity point and we obtain M as a lift of ϕ to the universal covers of A_1 and A_2, which are both $\mathbf{C}^g$. To get a morphism of abelian varieties, we need to make sure this linear map is compatible with the Riemann form.

Definition 154. *A morphism $(\mathbf{C}^g/L_1, E_1) \to (\mathbf{C}^g/L_2, E_2)$ is one induced by a linear map $M \in M_g(\mathbf{C})$ such that $ML_1 \subset L_2$ and such that for $v, w \in \mathbf{C}^g$, $E_1(Mv, Mw) = \det(M)E_2(v, w)$.*

The other reason we study complex tori and abelian varieties are for the study of Theta functions.

Definition 155. *Let V be a complex vector space. A theta function on V is a meromorphic function $f : V \to \mathbf{C}$ such that there is a full lattice L, a function $\phi : L \to \mathbf{C}$, and a bilinear form $B : L \times L \to \mathbf{C}$ which becomes $\mathbf{C}$-linear in the first factor when extended from L to $\mathbf{C}$ such that for $v \in V$, $\ell \in L$,*

$$f(v + \ell) = \phi(\ell)f(v)e^{2\pi i B(v,\ell)}.$$

Example 156. *If $q : \mathbf{C}^g \to \mathbf{C}$ is a quadratic form, let $B : \mathbf{C}^g \times \mathbf{C}^g \to \mathbf{C}$ be given by $B(x, y) = q(x + y) - q(x) - q(y)$. Then B restricts to a bilinear form $L \times L \to \mathbf{Z}$. Then $f(\vec{x}) = e^{2\pi i q(\vec{x})}$ is a theta function.*

The theta functions form a multiplicative group. We say that a theta function $\theta(\vec{x})$ is normalized if it is entire and unique up to multiplication by a quadratic form theta function. The number of different normalized theta functions is related to the Riemann forms on L.

We now return to the example from the end of the last class. Let $B/\mathbf{Q}$ be a quaternion algebra, $\mathcal{O} \subset B$ be a maximal order and $\psi : B \hookrightarrow M_2(\mathbf{R})$.

Lemma 157. *Let $\tau \in \mathcal{H}$ and let $V = \mathbf{C}^2$. Then $\psi(\mathcal{O})\binom{\tau}{1}$ is a full lattice in V.*

Proof. Since $\mathcal{O}$ is an order, there exist $\alpha, \beta, \gamma, \delta \in \mathcal{O}$ which are linearly independent over $\mathbf{Q}$ and $\mathcal{O} = \alpha\mathbf{Z} \oplus \beta\mathbf{Z} \oplus \gamma\mathbf{Z} \oplus \delta\mathbf{Z}$. Since they are linearly independent over $\mathbf{Q}$, $\psi(\alpha), \psi(\beta), \psi(\gamma)$, and $\psi(\delta)$ are linearly independent over $\mathbf{R}$ and thus generate $M_2(\mathbf{R})$ as a vector space.

The result follows since

$$M_2(\mathbf{R})\begin{pmatrix}\tau\\1\end{pmatrix} = \begin{pmatrix}\tau\\0\end{pmatrix}\mathbf{R}\oplus\begin{pmatrix}1\\0\end{pmatrix}\mathbf{R}\oplus\begin{pmatrix}0\\\tau\end{pmatrix}\mathbf{R}\oplus\begin{pmatrix}0\\1\end{pmatrix}\mathbf{R} = \mathbf{C}^2.$$

□

It follows that $\mathbf{C}^2/\psi(\mathcal{O})\binom{\tau}{1}$ is a complex torus. Now let $D = \mathrm{disc}(\mathcal{O}) \in \mathbf{Z}_{>0}$ and pick $\mu \in \mathcal{O}$ such that $\mu^2 + D = 0$. Note that if $v \in \mathbf{C}^2$ then we need to contruct an alternating form $E_{\mathbf{R}} : V \times V \to \mathbf{R}$ such that $E(iv, iw) = E(v, w)$. Therefore since v can be written as $r\psi(x)\binom{\tau}{1}$ for $r \in \mathbf{R}$, $x \in \mathcal{O}$, iv can be written as $r\psi(x)\binom{i\tau}{i}$ and we need to find $r' \in \mathbf{R}$, $x' \in \mathcal{O}$ $r'\psi(x')\binom{\tau}{1} = \binom{i\tau}{i}$. We will choose $r' = \dfrac{1}{\sqrt{D}}$ and $x' = \mu$, and to that end we define

$$E : \psi(\mathcal{O})\begin{pmatrix}\tau\\1\end{pmatrix} \times \psi(\mathcal{O})\begin{pmatrix}\tau\\1\end{pmatrix} \to \mathbf{Z} : E(\psi(x)\begin{pmatrix}\tau\\1\end{pmatrix}, \psi(y)\begin{pmatrix}\tau\\1\end{pmatrix}) = t(\mu x\overline{y}).$$

Lemma 158. *The function E described above is a Riemann form on $\psi(\mathcal{O})\binom{\tau}{1}$.*

Proof. First we check that E is an alternating bilinear pairing over $\mathbf{Z}$. It is very easy to check that E is bilinear over $\mathbf{Z}$. To see that E is alternating, recall that if $x \in \mathcal{O}$ then $n(x) \in \mathbf{Z}$ and thus

$$\begin{aligned} E\left(\psi(x)\begin{pmatrix}\tau\\1\end{pmatrix}, \psi(x)\begin{pmatrix}\tau\\1\end{pmatrix}\right) &= t(\mu x\overline{x}) \\ &= t(\mu n(x)) \\ &= n(x)t(\mu) = n(x)0 = 0. \end{aligned}$$

Now suppose that $v, w \in V$. Let $r, s \in \mathbf{R}$ and $x, y \in \mathcal{O}$ be such that $v = r\psi(x)\binom{\tau}{1}$ and $w = s\psi(y)\binom{\tau}{1}$. Therefore $iv = r\psi(x)\dfrac{\psi(\mu)}{\sqrt{D}}\binom{\tau}{1}$ and $iw = s\psi(y)\dfrac{\psi(\mu)}{\sqrt{D}}\binom{\tau}{1}$. We find that

$$\begin{aligned} E(iv, iw) &= \frac{r}{\sqrt{D}}\frac{s}{\sqrt{D}}t(\mu x\mu\overline{y}\overline{\mu}) \\ &= \frac{rs}{D}t(\mu x\mu\overline{\mu}\,\overline{y}) \\ &= \frac{rs}{D}t(\mu xD\overline{y}) \\ &= rst(\mu x\overline{y}) \\ &= E(v, w) \end{aligned}$$

Therefore E extends to a Hermitian form. To check that E extends to a positive definite Hermitian form, we cite Lang's Introduction to

Algebraic and Abelian Functions, Chapter IX. It's not that it's hard to prove, but we'd have to expend some time studying positive involutions. □

Consider now that the quotient $\mathbf{C}^2/\psi(\mathcal{O})\binom{\tau}{1}$ has even more structure than that of an abelian variety. Consider that if $\alpha \in \mathcal{O}$ then $\alpha\mathcal{O} \subset \mathcal{O}$ with equality if and only if $\alpha \in \mathcal{O}^\times$. Therefore $\psi(\alpha\mathcal{O})\binom{\tau}{1} \subset \psi(\mathcal{O})\binom{\tau}{1}$. It follows that $\psi(\alpha) \in M_2(\mathbf{R})$ defines a linear map $\mathbf{C}^2 \to \mathbf{C}^2$ which induces an endomorphism of $\mathbf{C}^2/\psi(\mathcal{O})\binom{\tau}{1}$. Since this holds for all $\alpha \in \mathcal{O}$, we have a homomorphism $\iota : \mathcal{O} \to \mathrm{End}(\mathbf{C}/\psi(\mathcal{O})\binom{\tau}{1})$. If $n(\alpha) \neq 0$ and B is division, then $\alpha \neq 0$ and so $n(\alpha)E$ is another polarization on $\mathbf{C}/\psi(\mathcal{O}^1)\binom{\tau}{1}$.

Definition 159. *We say that a two-dimensional abelian variety is an abelian surface. We say that an abelian surface A admitting an embedding $\iota : \mathcal{O} \to \mathrm{End}(A)$ admits quaternionic multiplication or QM. We will say that a (polarized) QM abelian surface is the data of $(A = \mathbf{C}^2/L, \iota : \mathcal{O} \hookrightarrow \mathrm{End}(A))$. The category of QM abelian surfaces has morphisms $\phi : A_1 \to A_2$ such that for all $\alpha \in \mathcal{O}$, $\phi\iota(\alpha) = \iota(\alpha)\phi$.*

Lemma 160. *For all $\tau \in \mathcal{H}$, the endomorphisms of $(\mathbf{C}^2/\psi(\mathcal{O})\binom{\tau}{1})$ as a QM abelian surface are the complex scalar maps $g : \mathbf{C}^2 \to \mathbf{C}^2$ for $g \in \mathbf{C}$ such that $\mathbf{Z}[g] \hookrightarrow \mathcal{O}$.*

Proof. When we ask about the morphisms which commute with ι given by ψ, we are essentially asking about the matrices in $M_2(\mathbf{C})$ which commute with the ring $\psi(\mathcal{O})$. Since $\psi(\mathcal{O})$ generates $M_2(\mathbf{R})$ as a vector space over $\mathbf{R}$, $\psi(\mathcal{O})$ generates $M_2(\mathbf{C})$ over $\mathbf{C}$. Therefore the matrices which commute with $\psi(\mathcal{O})$ in $M_2(\mathbf{C})$ are contained in the elements of the center of $M_2(\mathbf{C})$, i.e. the complex scalar matrices.

The only other restriction we have to make is that for $g \in \mathbf{C}$, $M = \begin{pmatrix} g & 0 \\ 0 & g \end{pmatrix}$ takes $\psi(\mathcal{O})\binom{\tau}{1}$ into itself. Therefore, there exists some $x \in \mathcal{O}$ such that $\psi(x)\binom{\tau}{1} = \binom{g\tau}{g} = g\binom{\tau}{1}$. It follows that $g^2\binom{\tau}{1} = \psi(x^2)\binom{\tau}{1}$ and thus

$$(g^2 - t(x)g + n(x))\begin{pmatrix}\tau\\1\end{pmatrix} = \psi(x^2 - t(x)x + n(x))\begin{pmatrix}\tau\\1\end{pmatrix}.$$

By linear independence, g embeds into $\mathcal{O}$. □

Definition 161. *We say that a QM abelian surface (A, ι) is special or CM if $\mathrm{End}(A, \iota) \supsetneq \mathbf{Z}$.*

We can thus see, using the following theorem, that in this case that $\mathrm{End}(A, \iota)$ will have to be a quadratic subring of $\mathcal{O}$.

Theorem 162. *If (A, E, ι) is a QM abelian surface, then there is some $\tau \in \mathcal{H}$ such that $A = \mathbf{C}^2/\psi(\mathcal{O})\binom{\tau}{1}$, $E(\psi(x)\binom{\tau}{1}, \psi(y)\binom{\tau}{1}) = t(\mu x\overline{y})$ and ι is induced by ψ. Moreover, if there is some $u \in \mathcal{O}^1$ such that $\psi(u)\tau_1 = \tau_2$ then u defines an isomorphism of QM abelian surfaces $\mathbf{C}^2/\psi(\mathcal{O})\binom{\tau_1}{1} \to \mathbf{C}^2/\psi(\mathcal{O})\binom{\tau_2}{1}$.*

Proof. Note that if $u \in \mathcal{O}^1$, replacing $\binom{\tau}{1}$ with $\left[\psi(u)\binom{\tau}{1}\right]$

- Does not change $\psi(\mathcal{O})\binom{\tau}{1}$ since $\psi(\mathcal{O})\psi(u) = \psi(\mathcal{O}u) = \psi(\mathcal{O})$.
- Does not change E because

$$\begin{aligned} E(\psi(x)\left[\psi(u)\binom{\tau}{1}\right], \psi(y)\left[\psi(u)\binom{\tau}{1}\right]) &= E(\psi(xu)\binom{\tau}{1}, \psi(yu)\binom{\tau}{1}) \\ &= t(\mu xu\overline{u}\,\overline{y}) \\ &= t(\mu x1\overline{y}) \\ &= E(\psi(x)\binom{\tau}{1}, \psi(y)\binom{\tau}{1}). \end{aligned}$$

Therefore we need only show that there is some $g \in \mathbf{C}^\times$ such that

$$g\binom{\psi(u)\tau}{1} = \psi(u)\binom{\tau}{1}.$$

Let $\psi(u) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{R})$, so that $\psi(u)\binom{\tau}{1} = \binom{a\tau+b}{c\tau+d}$. Meanwhile $\psi(u)\tau = \dfrac{a\tau+b}{c\tau+d}$, and thus if we set $g = c\tau + d$ then

$$g\binom{\psi(u)\tau}{1} = g\binom{\frac{a\tau+b}{c\tau+d}}{1} = \binom{a\tau+b}{c\tau+d} = \psi(u)\binom{\tau}{1}.$$

For the first part of the theorem, see section IX.5 of Lang's Introduction to Algebraic and Abelian Functions. □

We therefore have a bijection between the set of abelian surfaces with QM by $\mathcal{O}$ up to isomorphism and the set $Y_B(\mathbf{C})$. Let us see how Atkin-Lehner involutions interact with this interpretation of $Y_B(\mathbf{C})$.

Lemma 163. *Let β_m be an element of reduced norm $m|D$ in $\mathcal{O}$ such that $\beta_m\mathcal{O} = \mathcal{O}\beta_m$ (i.e. $\beta_m \in N\mathcal{O}$). The action of w_m on $(A, \iota : \mathcal{O} \hookrightarrow \mathrm{End}(A))$ fixes A but sends ι to the map ι_m where for all $\alpha \in \mathcal{O}$, $\iota_m(\alpha) = \iota(\beta_m)\iota(\alpha)\iota(\beta_m)^{-1}$. Moreover, if (A, ι) is w_m-fixed, then we have an explicit embedding of $\mathbf{Z}[\sqrt{-m}]$ into $\mathrm{End}(A, \iota)$ (or possibly $\mathbf{Z}[\sqrt{-1}]$ if $m = 2$).*

What about when $B = M_2(\mathbf{Q})$ and there are no Atkin-Lehner involutions?

Lemma 164. *If A is an abelian surface and $\iota : M_2(\mathbf{Z}) \hookrightarrow \mathrm{End}(A)$ then there is an elliptic curve E such that $A \cong E \times E$, in particular, $Y_{M_2(\mathbf{Q})}(\mathbf{C})$ is in bijection with the set of elliptic curves over $\mathbf{C}$.*

12. Lecture 12: Complex analytic wrapup

Last time: if $B/\mathbf{Q}$ is an indefinite quaternion algebra and $\mathcal{O}$ is a maximal order, then $Y_B = \psi(\mathcal{O}^1) \setminus \mathcal{H}$ is in bijection with the set of isomorphism classes of pairs (A, ι) where $A/\mathbf{C}$ is a two-dimensional abelian variety and $\iota : \mathcal{O} \hookrightarrow \mathrm{End}(A)$.

We also claimed that $Y_{M_2(\mathbf{Q})}$ is in one-to-one correspondence with the set of isomorphism classes of elliptic curves E over $\mathbf{C}$. As we will see, this is because a pair (A, ι) where $\iota : M_2(\mathbf{Z}) \hookrightarrow \mathrm{End}(A)$ is simply an abelian surface with an *idempotent decomposition*.

Definition 165. *An idempotent e of a ring R is an element such that $e^2 = e$. Every ring contains the* trivial idempotents *$e = 0$ and $e = 1$.*

Lemma 166. *If R has nontrivial idempotents, then R is not a division ring.*

Definition 167. *In $M_2(\mathbf{Z})$, let $e = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and let $I = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.*

Note that $I = I^{-1}$ and moreover that $IeI = 1 - e$.

Theorem 168. *If $A/\mathbf{C}$ is an abelian surface, $\iota : M_2(\mathbf{Z}) \hookrightarrow \mathrm{End}(A)$ then the pair (A, ι) uniquely determines an elliptic curve E and similarly any elliptic curve determines a pair (A, ι).*

Proof. Given any abelian variety A, there is a unique embedding $\varepsilon : \mathbf{Z} \hookrightarrow \mathrm{End}(A)$ by $n \mapsto [n] : A \to A$. Therefore there is a unique embedding $M_n(\mathbf{Z}) \hookrightarrow \mathrm{End}(A^n) \cong M_n(\mathrm{End}(A))$ extending ε. We thus obtain a pair $(A = E \times E, \iota)$ from an elliptic curve E.

Meanwhile from (A, ι) we determine an embedding

$$(\ker(\iota(e)) \times \ker(\iota(1-e)), \iota \mid_{\ker \iota(e)} \times \iota \mid_{\ker \iota(1-e)}) \hookrightarrow (A, \iota).$$

Note however that $\ker \iota(e) \cong \ker \iota(1-e)$ because if $\iota(e)P = 0$ then $\iota(I)P$ satisfies $\iota(1-e)\iota(I)P = 0$. Therefore $\iota(I)$ is an automorphism of A interchanging $\ker \iota(e)$ and $\ker \iota(1-e)$. There is a natural isomorphism $A \to \ker \iota(1-e) \times \ker \iota(e) \cong \ker \iota(e) \times \ker \iota(e)$ by $v \mapsto (\iota(e)v, \iota(1-e)v)$. The embedding $\iota \mid_{\ker \iota(e)}$ has image $\iota(1-e)\iota(M_2(\mathbf{Z}))\iota(1-e) \cong \mathbf{Z}$ and factors through ε and is thus uniquely determined by ε. □

We now study holomorphic 1-forms on $X(\Gamma)$ for $\Gamma < \mathrm{SL}_2(\mathbf{R})$ a finite covolume Fuchsian group.

- A holomorphic 1-form on $\mathcal{H}$ is of the form $f(z)\mathrm{d}z$.
- Holomorphic 1-forms ω on $Y(\Gamma)$ are thus of the form $f(z)\mathrm{d}z$ on some small enough subset $U \cong \mathbb{D}$.
- We can extend out from U to all of $\mathcal{H}$ if and only if ω is holomorphic on $\mathrm{cusp}(\Gamma)$ and for all $\gamma \in \Gamma$, $f(\gamma z)\mathrm{d}\gamma z = f(z)\mathrm{d}z$.
- By the quotient rule, if $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then

$$\begin{aligned} \mathrm{d}\left(\frac{az+b}{cz+d}\right) &= \frac{a\mathrm{d}z(cz+d) - c\mathrm{d}z(az+b)}{(cz+d)^2} \\ &= \frac{(ad-bc)\mathrm{d}z}{(cz+d)^2} = \frac{\mathrm{d}z}{(cz+d)^2}. \end{aligned}$$

- Therefore the holomorphic 1-forms on $X(\Gamma)$ are given by the holomorphic maps $f : \mathcal{H} \to \mathbf{C}$ such that $f\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f(z)$ and which has a holomorphic extension to $\Lambda(\Gamma)$.

Definition 169. *We say that a modular form for Γ (or more precisely a cusp form) of weight k is a holomorphic map $f : \mathcal{H} \to \mathbf{C}$ such that $f\left(\frac{az+b}{cz+d}\right) = (cz+d)^2 f(z)$ and which has a holomorphic extension to $\Lambda(\Gamma)$.*

Even better: let $\Gamma < \mathrm{SL}_2(\mathbf{Z})$ be a *congruence subgroup*.

Definition 170. *If $B/\mathbf{Q}$ is a quaternion algebra, $\psi : B \hookrightarrow M_2(\mathbf{R})$, and $\mathcal{O}$ is a maximal order in B then $\Gamma < \psi(\mathcal{O}^1)$ is a congruence subgroup if for some integer N coprime to $\mathrm{disc}(\mathcal{O})$, $\Gamma > \{\psi(\gamma) : \gamma \in \mathcal{O}^1, \gamma \equiv 1 \bmod N\mathcal{O}\}$.*

Congruence subgroups for quaternion algebras over some totally real field are the absolute widest class of Fuchsian groups that one could consider to produce Shimura curves.

If $\Gamma < \mathrm{SL}_2(\mathbf{Z})$ is congruence, let $q = e^{2\pi i z}$ so that $q = 0$ if and only if $z = i\infty$ as in our toy model. A modular form of weight k for Γ has a fourier expansion $f(q) = \sum_{n \geq 0} a_n q^n$, so that f is a cusp form if and only if $a_0 = 0$.

Example 171. *We may identify $\mathrm{SL}_2(\mathbf{Z})\backslash\mathcal{H}$ with $\mathbf{C}$ via the holomorphic j-function*

$$j(q) = 1/q + 744 + 196884q + \dots.$$

McKay noted that the coefficients of j are the sums of the lowest-dimensional irreducible representations of the Monster Simple group. This area of research is sometimes known as "Moonshine."

We have a problem though. The genus of $X_{M_2(\mathbf{Q})}$ is zero and thus there are no nonzero holomorphic 1-forms.

Definition 172. *Let $B/\mathbf{Q}$ be a quaternion algebra. An Eichler order is the intersection of two maximal orders.*

Lemma 173. *If $\mathcal{O} \subset M_2(\mathbf{Q})$ is an Eichler order, there is a positive integer N such that $\mathcal{O}$ is conjugate to*

$$\mathcal{O}_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in M_2(\mathbf{Z}) : N \mid c \right\}.$$

Definition 174. *Let $\Gamma_0(N) = \mathcal{O}_0(N)^1$ and $X_0(N) = X(\Gamma_0(N))$.*

Lemma 175. *The holomorphic 1-forms on $X_0(14)$ are the complex multiples of $f(q)\dfrac{\mathrm{d}q}{q}$ where*

$$f(q) = q - q^2 - 2q^3 + q^4 + 2q^6 + q^7 - q^8 + q^9 + \dots.$$

Note how slowly the coefficients grow in comparison to say the j -function. This is a general phenomenon.

Theorem 176 (Ramanujan-Petersson Conjecture)**.** *If f is a "normalized" cusp form of weight k for $\Gamma_0(N)$ and $p \nmid N$ then $|a_p| \leq 2p^{\frac{k-1}{2}}$.*

This theorem was proved by Deligne via his proof of the Weil Conjectures. The question is what does it mean to be "normalized?" The idea is that we can generate lots of commuting operators T_{p^n}, and that if f is an eigenvector for each of these, we can find the normalized coefficients a_p as the eigenvalues of T_p.

Theorem 177. *If $\Gamma = \Gamma_0(N)$, then $N\Gamma/\Gamma \cong (\mathbf{Z}/2\mathbf{Z})^{\{p|N : p \text{ is prime}\}} = W$.*

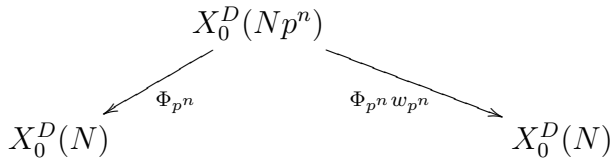
Theorem 178 (Ogg)**.** *For all but finitely many N, $\mathrm{Aut}(X_0(N)) = W$.*

If X, Y, Z are Riemann surfaces, $f : X \to Y$, $g : Y \to Z$ are holomorphic and ω is a holomorphic 1-form, then $f^*\omega$ is a holomorphic 1-form on X and $f_*\omega$ is a holomorphic 1-form on Z.

Definition 179. *If $p \nmid N$, let $w_{p^n} : X_0(Np^n) \to X_0(Np^n)$ be the corresponding involution. If $M \mid N$, let $\Phi_{N/M} : X_0(N) \to X_0(M)$ be the map induced by $\Gamma_0(N) \setminus \mathcal{H} \to \Gamma_0(M) \setminus \mathcal{H}$.*

We use these maps and the pushforwards and pullbacks of holomorphic 1-forms to define operators of cusp forms.

Definition 180. *The Hecke operator T_{p^n} is given by the correspondence of curves*



$$\omega \longmapsto T_{p^n}\omega = (\Phi_{p^n} w_{p^n})_* \Phi_{p^n}^* \omega$$

Theorem 181. *The collection of operators $\{T_p : p \nmid N\}$ all mutually commute and act on modular forms of any weight for $\Gamma_0(N)$.*

There is also a theory of Hecke Operators for other Shimura curves. Recall that if $B/\mathbf{Q}$ is a quaternion algebra, any maximal order has the same discriminant D. In fact this integer determines the quaternion algebra B.

Definition 182. *If $D \in \mathbf{Z}$ such that a maximal order of B has discriminant D, then let $X^D = X_B$.*

Theorem 183 (Jacquet-Langlands, first approximation)**.** *Let $p \neq q$ be primes. There is a Hecke-equivariant bijection between certain modular forms on $X_0(pq)$ and X^{pq}.*

This is an important theorem because $X^{pq} = Y^{pq}$ and its geometry is generally much easier to work with, but there are no cusps and therefore no natural Fourier expansions. On the other hand, there are Fourier expansions on $X_0(pq)$ even though the geometry is harder. This interplay has been a part of countless results over the past century, especially in Wiles' proof of Fermat's last theorem. We finish our discussion of the complex-analytic aspects of Shimura curves by briefly discussing higher-dimensional Shimura varieties. We will define these as higher-dimensional quotients by "Fuchsian groups."

Example 184. *Let $J_g = \begin{pmatrix} 0 & -I_g \\ I_g & 0 \end{pmatrix}$ and $Sp_{2g}(\mathbf{R}) = \{M \in M_{2g}(\mathbf{R}) : MJ_gM^\top = J_g\}$. If we identify $\mathbf{C}^g$ with $\mathbf{R}^{2g}$, then $U_g(\mathbf{C}) = \mathrm{Sp}_{2g}(\mathbf{R}) \cap \mathrm{GL}_g(\mathbf{C})$ is a compact subgroup of the locally compact group $\mathrm{Sp}_{2g}(\mathbf{R})$.*

Theorem 185. *There is a homeomorphism between $\mathrm{Sp}_{2g}(\mathbf{R})/U_g(\mathbf{C})$ and $\mathcal{H}^n$, the set of $g \times g$ complex symmetric matrices π such that the quadratic form $q(\vec{x}) = \vec{x}\Im(\pi)\vec{x}^\top$ is positive definite.*

Recall here that if $\mathbf{C}^g/L$ is an abelian variety which admits a principal polarization E, then $L = \Pi\mathbf{Z}^{2g}$ where $\Pi = (1|\pi)$ where $\pi \in \mathcal{H}^n$.

Theorem 186. *The matrices $M \in M_g(\mathbf{C}) \cong M_{2g}(\mathbf{R})$ which send $\mathbf{C}^g/L \to \mathbf{C}^g/L$ and preserve $E : L \times L \to \mathbf{Z}$ form the group $\mathrm{Sp}_{2g}(\mathbf{Z})$, which is discrete in $\mathrm{Sp}_{2g}(\mathbf{R})$.*

Corollary 187. *The space $\mathrm{Sp}_{2g}(\mathbf{Z})\backslash\mathcal{H}^n$ is Hausdorff and its points are in 1-1 correspondence with g-dimensional principally polarized abelian varietes.*

Definition 188. *A PE Shimura variety is a quotient space $\Gamma \setminus G/K$ whose points are in bijection with triples (A, E, ι) where A is an abelian variety, E is a polarization, and ι is an embedding of a ring into $\mathrm{End}(A, E)$.*