THE MODULAR GROUP AND BUILDING RIEMANN SURFACES

COLLEEN LEACOCK

August 2024

Abstract. In this paper we take a geometric approach to understanding the action of the modular group on the upper half plane, and use this approach to study quotients of the upper half plane by its congruence subgroups. In particular we show that these quotient spaces are Hausdorff and discuss their structure as Riemann surfaces. More so than on the existence of these Riemann surfaces, this paper focuses on leveraging the algebra of the modular group's action to expose some of the rich geometric and combinatorial details baked into them. It concludes with a discussion of two concrete examples.

Contents

1. Introduction 1
2. Formal definition of a Riemann surface 3
3. Automorphisms of $P^1_{\mathbb{C}}$ and $\mathbb{H}$ as projective actions 4
4. The modular group acts on farey edges 7
5. The Dedekind tessellation built from scratch 11
6. From orbits to the quotient space 16
7. Overview of building charts to form a Riemann surface 22
8. Examples and further discussion 23
Acknowledgments 26
References 27
Image Credits 27

1. Introduction

This paper seeks to address a practical question: How can someone new to the theory of Riemann surfaces gain access to concrete, non-trivial examples without much machinery?

As far as background knowledge, we assume the reader has taken a course in basic group theory and basic complex analysis. Familiarity with hyperbolic geometry is also helpful, but not crucial. We notably do not assume any formal background knowledge of Riemann surfaces (or even manifolds).

For context, Riemann surfaces are objects that can arise in several qualitatively different ways, with links that are highly non-obvious. Historically, people were driven to develop the theory of Riemann surfaces out of observations related to analytic continuation, which suggested that the plane is not always the most natural domain for some holomorphic maps [1]. Meanwhile, algebraically, Riemann surfaces arise as the zero sets of complex polynomials in several variables; that all Riemann surfaces (up to isomorphism) can be built this way is a highly deep result that stands as an important pillar of the theory [10]. For this paper, though, we stick more to the complex-analytic side of things and take inspiration from yet another deep result of the theory, the uniformization theorem, which states that the only simply connected Riemann surfaces are the Riemann sphere, the complex plane, and the unit disk (which is isomorphic to the upper half plane) [11]. A corollary of this is that every Riemann surface has one of these three as its universal cover, which points to considering quotients of them as a way to potentially construct more complicated Riemann surfaces.

As far as which surface to start with, one can show using the Riemann-Hurwitz formula that all quotients of the Riemann sphere are topologically 2-spheres (i.e. of genus 0) [1]—hence isomorphic to the Riemann sphere by uniformization—so nothing new there. As regards quotients of the complex plane, there is a rich and important theory of elliptic curves, which arise from quotients of $\mathbb{C}$ by discrete translation subgroups, and are topologically all tori (i.e. of genus 1) [6]. However, the focus of this paper will be on studying quotients of the upper half plane—which historically came into focus as a way of *parametrizing* the space of elliptic curves—with the hope of potentially constructing a surface of genus higher than 1.

First, as background information, we will start with the formal definition of a Riemann surface (Section 2), as well as an introduction to the automorphism groups of the Riemann sphere and upper half plane when viewed as projective actions (Section 3). This will eventually lead us toward studying the group $PSL_2(\mathbb{Z})$ (i.e. the modular group quotiented by $\pm I$).

We then take an in-depth look at the geometric and algebraic properties of the modular group's action on the upper half plane, and how this translates to properties of the associated quotient space (Section 4, Section 5, Section 6). This is where the majority of this paper's theorems lie. For example, one theorem we will prove is that the quotient space we construct from the modular group's action (or from the action of any of its congruence subgroups) is Hausdorff, which is an important underlying topological property of any Riemann surface.

Despite grounding itself in the topic of Riemann surfaces, one limitation of this paper is that it does not include a rigorous proof of the existence of a complex atlas on the quotient space, and hence does not officially show that the quotient carries the necessary complex structure to make it a Riemann surface. Considering this, one may reasonably ask what the approach of this paper serves to contribute, given that there are sources such as [2] which give a concise proof of Hausdorffness and proceed through a rigorous construction of a complex atlas all within roughly the same page count. It comes down to the practical goals of the exposition. What this paper seeks to offer has less to do with concise generality and more to do with a certain kind of clarity in the mind of the reader: a clarity that comes from understanding some of the geometric and combinatorial aspects at play in the construction of this type of quotient space. As such, in the end we focus less on confirming the *existence* of a given Riemann surface, and more on the question of what it *looks* like.

To this end, what the paper does offer next is a way of motivating how each piece of the atlas construction from [2] is carried out (Section 7), setting up intuition and giving a sense for the main considerations involved. (Here we also touch on the compactification process.) Furthermore, the last section of the paper (Section 8) draws on theory and discussion built up over the previous sections to introduce two concrete examples of Riemann surfaces that can be built from quotients of the upper half plane. The second example is a Riemann surface of genus 3, making it beyond what a quotient of the complex plane or the Riemann sphere could yield.

2. Formal definition of a Riemann surface

Riemann surfaces are manifolds with added structure, so we begin with the definition of a manifold. (This section follows definitions from [3] and [1].)

Definition 2.1.

(1) Let X be a topological space. We say that X is *locally Euclidean* if for any point $p \in X$, there is an open neighborhood U of p and a map $h : U \to R^n$, for some $n \in N$, such that $h : U \to h(U)$ is a homeomorphism. Such a map h is called a *chart* of X, and U its *chart domain*.

(2) A *manifold* is a locally Euclidean topological space that is also Hausdorff as well as second-countable (meaning its topology has a countable base).

Remark 2.2. In Definition 2.1, n is called the dimension of X. One can check that n must be consistent across charts [3], so this dimension is well-defined.

2-dimensional manifolds are called surfaces, and Riemann surfaces all have dimension 2 because we want them to locally look like $\mathbb{C}$, which is homeomorphic to $\mathbb{R}^2$. But $\mathbb{C}$ has more structure than $\mathbb{R}^2$, and to be able to do complex analysis on a Riemann surface (e.g. to be able to recognize when a map is holomorphic), we need a consistency across charts that respects this additional structure. In particular, what we end up needing is a way to transition between charts holomorphically.

Definition 2.3. Given a manifold X and any two of its charts $(U, h), (V, k)$, we define the transition map w between them to be the map

$$w := k \circ h^{-1}|_{h(U \cap V)} : h(U \cap V) \to k(U \cap V).$$

Note that w is also a homeomorphism, being the composition of two of them [3].

Definition 2.4.

(1) Let X be a 2-dimensional manifold. A *complex atlas* $\mathcal{A}$ for X is a collection of charts that cover X with the property that the transition map between any pair of charts in $\mathcal{A}$ is holomorphic (thus conformal).

(2) Two complex atlases are considered equivalent if their union is also a complex atlas, and in this way a complex atlas generates a *maximal* complex atlas as the union of all complex atlases that are equivalent to it. A maximal complex atlas is said to give a *complex structure* to X.

Definition 2.5. A Riemann surface is a connected 2-dimensional manifold equipped with a complex structure. (This makes it a 1-dimensional complex manifold.)

Definition 2.6.

(1) A map between Riemann surfaces is *holomorphic* if it is holomorphic in charts. More precisely, if M and N are Riemann surfaces, then $f : M \to N$ is holomorphic at $p \in M$ if there is a chart $h : U_1 \to V_1$ of M with $p \in U_1$ and a chart $k : U_2 \to V_2$ of N with $f(U_1) \subseteq U_2$ for which the map $k \circ f \circ h^{-1} : V_1 \to V_2$ is holomorphic. (One can check that this property is independent of choice of chart because of the holomorphic transition maps.)

(2) Two Riemann surfaces are isomorphic (sometimes called biholomorphic) if they are related by a holomorphic bijection.

Remark 2.7. Since conformal maps are orientation-preserving, Riemann surfaces have an orientation as manifolds (see [7] for a discussion of this). Also, it turns out that among orientable connected 2-dimensional manifolds, the compact ones can be completely characterized by their *genus* [12], which is essentially the number of mug-handles they have (a.k.a. how many doughnut-holes—so for example a 2-sphere has genus 0 and a torus has genus 1). So if a Riemann surface is compact, finding its genus determines its underlying topology.

We end this section with some examples of familiar Riemann surfaces. (By the uniformization theorem, the list below accounts for all the simply connected ones.)

Example 2.8.

(1) $\mathbb{C}$ itself is a Riemann surface with the identity map as one big chart.

(2) Open connected subsets of $\mathbb{C}$ are Riemann surfaces, again with the identity for charts. Of these subsets, the simply connected ones—except for $\mathbb{C}$ itself—are all isomorphic to the unit disk by the Riemann mapping theorem; however, they are not isomorphic to $\mathbb{C}$ (even though homeomorphic to it), as evidenced by Liouville's theorem from complex analysis [4].

(3) The Riemann sphere from complex analysis is a Riemann surface. Topologically it is the 2-sphere, so it is compact (of genus 0) and simply connected. There are several ways of characterizing the Riemann sphere, but for now let us view it as the 1-point compactification of $\mathbb{C}$, meaning $\mathbb{C}$ plus an extra point called ∞, where an open neighborhood of ∞ is any subset containing ∞ whose complement is compact in $\mathbb{C}$ [13].

With this characterization, the identity on $\mathbb{C}$ is automatically a chart for all points of the Riemann sphere except ∞, so we only need one more chart, which we can define on $\mathbb{C} \cup \{\infty\} \setminus \{0\}$ as $z \mapsto \frac{1}{z}$ with $\infty \mapsto 0$. Note that this second chart overlaps with the identity chart on $\mathbb{C} \setminus \{0\}$, with the transition map given by $z \mapsto \frac{1}{z}$, which is indeed holomorphic there.

Often it is also convenient to view the Riemann sphere as a literal geometric sphere, and with this perspective, the above two charts can be obtained through stereographic projection from the north and south pole [1].

3. Automorphisms of $P^1_{\mathbb{C}}$ and $\mathbb{H}$ as projective actions

Recall from complex analysis that the automorphism group of the Riemann sphere consists of all maps of the form

$$z \mapsto \frac{az + b}{cz + d}$$

(where $a, b, c, d \in \mathbb{C}$ with $ad - bc \neq 0$), and that these automorphisms are called Möbius transformations [1].

It will be helpful for this paper to introduce another way of describing the group of Möbius transformations, through what is called the *projective* general linear group. We will see that this group acts on the following space:

Definition 3.1. Given a field F, the *projective line* over F, denoted P^1_F, is the space of lines through the origin in F^2. More precisely, it is a quotient of $F^2 \setminus \{0\}$ by the equivalence relation $\sim$ defined as

$$v_1 \sim v_2 \iff v_1 = \lambda v_2 \text{ for some } \lambda \in F \setminus \{0\},$$

which identifies vectors with the same span.

Example 3.2. Taking $F = \mathbb{C}$, $P^1_\mathbb{C}$ is topologically the 1-point compactification of $\mathbb{C}$, just like the Riemann sphere (see Example 2.8), and indeed one can show that $P^1_\mathbb{C}$ has the structure of a Riemann surface that makes it isomorphic to the Riemann sphere [1]. One chart is given by $[z_1 : z_2] \mapsto \frac{z_1}{z_2}$, defined on all points of $P^1_\mathbb{C}$ except for $[\lambda : 0]$, and another chart is given by $[z_1 : z_2] \mapsto \frac{z_2}{z_1}$, defined everywhere but at $[0 : \lambda]$. Just like in Example 2.8, the transition map is $z \mapsto \frac{1}{z}$. To give an explicit isomorphism φ from $P^1_\mathbb{C}$ to the Riemann sphere, we can just extend the pairing from the first chart (where $[z_1 : z_2] \mapsto \frac{z_1}{z_2}$) by filling in $[\lambda : 0] \mapsto \infty$.

We first observe that the general linear group $GL_2(F)$ acts on P^1_F by matrix multiplication. That is, one can check directly that if

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in GL_2(F)$$

then the map $\hat{\mu}_A : P^1_F \to P^1_F$ defined by

$$\hat{\mu}_A \left(\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \right) = A \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} az_1 + bz_2 \\ cz_1 + dz_2 \end{bmatrix} \in P^1_F$$

is well-defined (i.e. unaffected by scaling z_1 and z_2 by the same scalar), and yields the group homomorphism

$$\hat{\mu}_F : GL_2(F) \to \mathrm{Aut}(P^1_F)$$
$$A \mapsto \hat{\mu}_A$$

whereby $GL_2(F)$ acts on P^1_F. However, this action $\hat{\mu}_F$ is not faithful (i.e. not injective). It comes close, but intuitively speaking it has inherited a blindness to overall scalar multiple from the way we built P^1_F, and indeed one can check that $\ker(\hat{\mu}_F) = \{\lambda I \mid \lambda \in \mathbb{F} \setminus \{0\}\}$. This brings us to the projective general linear group.

Definition 3.3. The quotient space $GL_2(F)/\{\lambda I \mid \lambda \in \mathbb{F} \setminus \{0\}\}$ is called $PGL_2(F)$ (where P is for projective).

Shifting our domain to $PGL_2(\mathbb{F})$ gives a now-injective-by-design homomorphism

$$\mu_F : PGL_2(F) \to \mathrm{Aut}(P^1_F)$$
$$A \mapsto \mu_A$$

that is well-defined due to the kernel of $\hat{\mu}_F$, whereby $PGL_2(F)$ acts faithfully on P^1_F.

Now consider the case where $F = \mathbb{C}$ again. With $P^1_\mathbb{C}$ being isomorphic to the Riemann sphere, we know $\mathrm{Aut} P^1_\mathbb{C}$ must be isomorphic to the group of Möbius transformations. Specifically this isomorphism can be obtained by conjugating by φ from Example 3.2. For example, one can verify directly that given $A \in PGL_2(\mathbb{C})$, taking $\mu_A \in \mathrm{Aut}P^1_{\mathbb{C}}$ and sending it to $\varphi\mu_A\varphi^{-1}$ recovers an ordinary Möbius transformation whose coefficients come straight from the entries of A. Noting that any Möbius transformation has coefficients from some element of $PGL_2(\mathbb{C})$, one can then show that $\mu_{\mathbb{C}}$ is not only injective but also *surjective* (exercise), hence is a group isomorphism.

In addition, $PGL_2(\mathbb{C})$ is isomorphic to the group $PSL_2(\mathbb{C}) := SL_2(\mathbb{C})/\{\pm I\}$, due to $\mathbb{C}$ being algebraically closed [14]. (The same is not true of $PGL_2(\mathbb{R})$ versus $PSL_2(\mathbb{R})$; in this case the latter is isomorphic to an index-2 subgroup of $PGL_2(\mathbb{R})$ due to all squares being non-negative, which prevents the ability to swap the sign of the determinant by scaling.) But at least when $F = \mathbb{C}$, combining the above two isomorphisms yields a group isomorphism $\mu : PSL_2(\mathbb{C}) \to \mathrm{Aut}(P^1_{\mathbb{C}})$, and in this way $PSL_2(\mathbb{C})$ acts faithfully on $P^1_{\mathbb{C}}$.

We now turn our attention to the upper half plane $\mathbb{H} \subseteq \mathbb{C}$ (consisting of all points strictly above the real axis), which is a model for hyperbolic space. It is often convenient to view $\mathbb{H}$ as being embedded in $P^1_{\mathbb{C}}$ as a hemisphere, with boundary $P^1_{\mathbb{R}}$ (which embeds as a great circle). Recall that the automorphisms of $\mathbb{H}$ are also Möbius transformations, so when $\mathbb{H}$ is viewed in $P^1_{\mathbb{C}}$,

$$\mu^{-1}(\mathrm{Aut}(\mathbb{H})) \leq \mu^{-1}(\mathrm{Aut}(P^1_{\mathbb{C}})) = PSL_2(\mathbb{C}).$$

From the observation that automorphisms of $\mathbb{H}$ preserve its boundary $P^1_{\mathbb{R}}$, a reasonable (though incorrect) first guess is that $\mathrm{Aut}(\mathbb{H})$ might turn out to be isomorphic to $PGL_2(\mathbb{R})$; however, using the further observation that its automorphisms must *not* flip $\mathbb{H}$ to the lower half plane (which translates to a restriction on the sign of the determinant), one can show that in fact

$$\mu^{-1}(\mathrm{Aut}(\mathbb{H})) = PSL_2(\mathbb{R}).$$

Since μ is a group isomorphism, so is

$$\mu\mid_{PSL_2(\mathbb{R})} : PSL_2(\mathbb{R}) \to \mathrm{Aut}(\mathbb{H}),$$

and thus $PSL_2(\mathbb{R})$ acts faithfully on $\mathbb{H}$ (and on its boundary $P^1_{\mathbb{R}}$).

Observation 3.4. In hyperbolic space, lines are in bijection with pairs of distinct ideal points (a.k.a. points at infinity), and also automorphisms preserve lines. We can directly observe this in the upper half plane model: Ideal points are points on the boundary $P^1_{\mathbb{R}}$, and hyperbolic lines are circles in $P^1_{\mathbb{C}}$ which intersect the circle $P^1_{\mathbb{R}}$ at right angles. Equivalently they are circles centered on $P^1_{\mathbb{R}}$, thus specifiable by their two points of intersection with $P^1_{\mathbb{R}}$. Furthermore, Möbius transformations are conformal and take circles in $P^1_{\mathbb{C}}$ to circles in $P^1_{\mathbb{C}}$, and hence those which fix $P^1_{\mathbb{R}}$—in particular all automorphisms of $\mathbb{H}$—preserve the set of hyperbolic lines.

Notation 3.5. We will denote the set of hyperbolic lines in $\mathbb{H}$ by L.

Sometimes we care about the orientation of a hyperbolic line, sometimes not. In either situation, it is notationally convenient to specify an edge by its endpoints in $P^1_{\mathbb{R}}$ as an ordered pair (which is safe to do in light of Observation 3.4). So to help disambiguate when order matters with this notation, we will also distinguish between L and the related set $L_{\to}$, where members of $L_{\to}$ are viewed with definite orientation.

One consequence of Observation 3.4 is that μ induces an action of $PSL_2(\mathbb{R})$ on $L_{\to}$. Also, like any group, $PSL_2(\mathbb{R})$ also acts on itself, by (let's say left) multiplication. In the next lemma we introduce a map from $PSL_2(\mathbb{R})$ to $L_{\to}$, and so a natural question is whether it is coordinated with the respective actions of $PSL_2(\mathbb{R})$ on its domain and codomain.

Lemma 3.6. *The map*

$$\hat{\theta} : PSL_2(\mathbb{R}) \to L_{\to}$$
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mapsto \left(\begin{bmatrix} a \\ c \end{bmatrix}, \begin{bmatrix} b \\ d \end{bmatrix} \right)$$

(where $[a : c]$ *and* $[b : d]$ *are viewed in* $P^1_{\mathbb{R}}$*) is a* $PSL_2(\mathbb{R})$*-equivariant map. That is, for all* $A, B \in PSL_2(\mathbb{R})$*,*

$$A \cdot \hat{\theta}(B) = \hat{\theta}(AB). \tag{3.7}$$

Proof. We leave this as an exercise to the reader. [Hint: Consider how *Observation* 3.4 might allow for viewing θ as factoring through the space $(P^1_{\mathbb{R}})^2$ with the diagonal excluded. How does $PSL_2(\mathbb{R})$ act on this intermediate space?] □

4. The modular group acts on farey edges

Eventually, we would like to build a new space by quotienting $\mathbb{H}$ by a group action, which essentially means passing to the space of orbits of the action. What happens if we quotient by the full automorphism group of $\mathbb{H}$, from the action of $PSL_2(\mathbb{R})$? Well this action can send any point to any other point (exercise), so only has one orbit; thus we get a single point, and that's boring. So then we can ask about subgroups, and in search of something more discrete, hopefully it seems non-ridiculous to at least consider the subgroup $PSL_2(\mathbb{Z})$.

Convention 4.1. Technically the modular group refers to $SL_2(\mathbb{Z})$, which acts on $\mathbb{H}$ just like $PSL_2(\mathbb{Z})$ does—though not quite faithfully—by pre-composing μ with the natural projection that quotients out by $\pm I$. In practice, which group you prefer to have in mind as acting is typically not of major importance, since the image of both actions is the same. For the purposes of this paper we really mean the action of $PSL_2(\mathbb{Z})$ even when referring to the modular group action.

We saw last section that $PSL_2(\mathbb{R})$ acts on L, so to try to understand its subgroup $PSL_2(\mathbb{Z})$ better, a natural question to ask is whether $PSL_2(\mathbb{Z})$ preserves anything more specific than $PSL_2(\mathbb{R})$ does, such as a subset of L. In fact this is the case.

Notation 4.2. We will write $\mathbb{Q}^*$ to denote $\mathbb{Q} \cup \{\infty\}$ as a subset of the Riemann sphere (or sometimes to mean the corresponding subset of $P^1_{\mathbb{C}}$, depending on context).

Observation 4.3. Each element of $\mathbb{Q}^*$ has two lowest-terms representations, related by $\frac{n}{d} = \frac{-n}{-d}$. (For ∞, it will be convenient to consider $\frac{\pm 1}{0}$ as lowest terms.)

Definition 4.4.

(1) We say that a hyperbolic line $(r_1, r_2) \in L$ is a *farey edge* if its endpoints r_1, r_2 are in $\mathbb{Q}^*$ with the property that, when written in lowest terms as $r_1 = \frac{a}{c}$ and $r_2 = \frac{b}{d}$ (with either sign option from Observation 4.3), they satisfy $ad - bc = \pm 1$.

(2) In the case where $ad - bc = \pm 2$, we say that (r_1, r_2) defines a *trident edge*.

Notation 4.5. The set of farey edges is denoted E (or $E_{\to}$ if oriented), and the set of trident edges is denoted E^* (or $E^*_{\to}$ if oriented).

Lemma 4.6. *The restriction of the map* $\hat{\theta}$ *from Lemma 3.6 to*

$$\theta : PSL_2(\mathbb{Z}) \to E_{\to}$$

is invertible.

Proof. To show that the codomain of θ is well-defined, let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in PSL_2(\mathbb{Z}).$$

Then $ad - bc = 1$, and one can check that this requires the columns of A to be coprime so that $\frac{a}{c}$ and $\frac{b}{d}$ are already lowest-terms representations of the endpoints of $\theta(A)$. Therefore these endpoints define a farey edge, so $\theta(A) \in E_{\to}$, and we conclude that $\theta(PSL_2(\mathbb{Z})) \subseteq E_{\to}$.

To show that θ is surjective, let $e = (r_1, r_2) \in E_{\to}$. First write r_1 in lowest terms as $\frac{a}{c}$, say choosing sign such that $c \geq 0$. (If $r_1 = \infty$, write $r_1 = \frac{1}{0}$.) Next write r_2 in lowest terms as $\frac{b}{d}$, choosing sign such that $ad - bc > 0$. (This is always possible since the determinant of a matrix switches sign under negating both entries of one column.) Then $ad - bc = 1$ because $e \in E_{\to}$, so the matrix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

defines an element of $PSL_2(\mathbb{Z})$ satisfying $\theta(A) = e$.

To see that θ is also injective, suppose

$$A_1 = \begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix}, A_2 = \begin{bmatrix} a_2 & b_2 \\ c_2 & d_2 \end{bmatrix} \in PSL_2(\mathbb{Z})$$

are such that $\theta(A_1) = \theta(A_2) = (r_1, r_2)$. As mentioned above, the columns of A_1, A_2 must be coprime due to their determinant of 1, so $\frac{a_1}{c_1} = \frac{a_2}{c_2}$ are both lowest terms representations of r_1, and likewise for $\frac{b_1}{d_1} = \frac{b_2}{d_2}$ representing r_2. In light of Observation 4.3, it follows that corresponding columns of A_1 and A_2 are related by an overall scalar multiple of ± 1. But if one column pair is equal and the other is related by negation, then $\det(A_1) = -\det(A_2)$, contradicting that both matrices have determinant 1. Therefore the entries of A_1 and A_2 are either all the same or all negated, and either way $A_1 = A_2$ in $PSL_2(\mathbb{Z})$.

Therefore θ is a well-defined bijection.

□

Example 4.7. An important reference-point edge that will come up in the proofs to follow is the farey edge

$$e_0 := \theta(I) = (\infty, 0) \in E_{\to}.$$

Before stating the next result, it will be useful call attention to a certain nice property that group actions can have called *simple transitivity*, which we now define.

Definition 4.8. Given an action of a group G on a set S, the action is *transitive* if for all $(x, y) \in S \times S$, there exists $\gamma \in G$ for which $\gamma \cdot x = y$. If in addition this γ is *unique* to each $(x, y) \in S \times S$, then we say that G acts *simply transitively* on S.

Theorem 4.9. *The action of* $PSL_2(\mathbb{Z})$ *takes farey edges to farey edges. Furthermore,* $PSL_2(\mathbb{Z})$ *acts simply transitively on* $E_{\to}$*.*

Proof. For the first statement, let $e \in E_{\rightarrow}$ and $A \in PSL_2(\mathbb{Z})$, and consider $A \cdot e$. By Lemma 4.6, $\theta^{-1}(e) \in PSL_2(\mathbb{Z})$, so

$$\theta^{-1}(A \cdot e) = A \circ \theta^{-1}(e) \in PSL_2(\mathbb{Z})$$

due to Lemma 3.6. Hence $A \cdot e \in E_{\rightarrow}$ (by Lemma 4.6 again).

For the second statement, we first show that there exists a unique element in $PSL_2(\mathbb{Z})$ taking e_0 (as defined in Example 4.7) to e. Since $e_0 = \theta(I)$, Lemma 3.6 gives

$$A \cdot e_0 = A \cdot \theta(I) = \theta(A) \tag{4.10}$$

for all $A \in PSL_2(\mathbb{Z})$. Thus the condition $A \cdot e_0 = e$ is equivalent to $\theta(A) = e$, so $A = \theta^{-1}(e)$ gives the unique solution in $PSL_2(\mathbb{Z})$.

So to show transitivity, suppose $e_1, e_2 \in E_{\rightarrow}$, and set $A_1 = \theta^{-1}(e_1), A_2 = \theta^{-1}(e_2)$ so that $A_1 \cdot e_0 = e_1$ and $A_2 \cdot e_0 = e_2$. Then $A_2 \circ A_1^{-1}$ takes e_1 to e_2.

To show that the transitivity is simple, suppose $B, C \in PSL_2(\mathbb{Z})$ both take e_1 to e_2. Then since $A_1 \cdot e_0 = e_1$, we have

$$BA_1 \cdot e_0 = e_2 = CA_1 \cdot e_0.$$

From here, the uniqueness shown above implies that $BA_1 = CA_1$, which in turn implies $B = C$. This completes the proof. □

Proposition 4.11.

(1) Each farey edge $e \in E$ has a natural dual trident edge $e^* \in E^*$ (and vice versa).

(2) If a hyperbolic line ℓ shares one endpoint with a farey edge e and one endpoint with its dual e^*, then ℓ is a farey edge.

Proof. To show the first statement, we will prove that the operation of putting a pair of rationals in lowest terms and then applying the correspondence

$$\left(\frac{a}{c}, \frac{b}{d}\right) \mapsto \left(\frac{a+b}{c+d}, \frac{a-b}{c-d}\right)$$

induces an involution $* : E \cup E^* \to E \cup E^*$ that takes farey edges to trident edges and vice versa.

Note that to be well-defined, this correspondence does rely on putting the fractions in lowest terms first, but one can verify that it is unaffected by the sign ambiguity noted in Observation 4.3.

Suppose $e \in E$, and let $(\frac{a}{c}, \frac{b}{d})$ be a lowest terms representation of e, so $ad - bc = \pm 1$. One can check directly that, in general,

$$(a-b)(c+d) - (a+b)(c-d) = 2(ad - bc), \tag{4.12}$$

so in this case $(a+b)(c-d) - (a-b)(c+d) = \pm 2$. Therefore to show that the edge e^* is a trident edge it suffices to show that

$$e^* = \left(\frac{a+b}{c+d}, \frac{a-b}{c-d}\right) \tag{4.13}$$

is already a lowest-terms representation of it. Note that the greatest common factor of either numerator-denominator pair involved in (4.13) divides the left-hand side of (4.12), so is at most 2 (in absolute value). Furthermore, if it *is* 2, then a and b must have the same parity, and likewise for c and d, which would force $ad - bc$ to be even (exercise), thereby contradicting our choice of $e \in E$. Therefore, $\frac{a+b}{c+d}$ and $\frac{a-b}{c-d}$ are both already as reduced as they can be, verifying that $e^* \in E^*$.

Let's see what happens when we apply this same operation to e^*. Since we just saw that (4.13) is still in lowest terms, we know its dual is given by

$$(e^*)^* = \left(\frac{(a+b)+(a-b)}{(c+d)+(c-d)}, \frac{(a+b)-(a-b)}{(c+d)-(c-d)}\right) = \left(\frac{2a}{2b}, \frac{2c}{2d}\right) = e. \tag{4.14}$$

One may yet worry that there are trident edges which do not arise as the dual to any farey edge. So now switch perspectives and suppose that $t = (\frac{a}{c}, \frac{b}{d}) \in E^*$ is a lowest-terms representation of a trident edge, so $ad - bc = \pm 2$. Since $\frac{a}{c}$ and $\frac{b}{d}$ are in lowest terms, one can check that this can only happen if a and b have the same parity and c and d also have the same parity (exercise). [Hint: Break this into cases.] Thus $a+b, a-b, c+d, c-d$ are all even, and by (4.12), we have

$$(a+b)(c-d) - (a-b)(c+d) = \pm 4,$$

so

$$\frac{a+b}{2} \cdot \frac{c-d}{2} - \frac{a-b}{2} \cdot \frac{c+d}{2} = \pm 1.$$

Hence

$$t^* = \left(\frac{a+b}{c+d}, \frac{a-b}{c-d}\right) = \left(\frac{\frac{a+b}{2}}{\frac{c+d}{2}}, \frac{\frac{a-b}{2}}{\frac{c-d}{2}}\right) \in E$$

is a farey edge, and a computation similar to (4.14) confirms that $(t^*)^* = t$ in this case as well.

From here the second statement follows directly from the identity

$$ad-bc = a(c+d)-(a+b)c = (a+b)d-b(c+d) = (a-b)c-a(c-d) = b(c-d)-(a-b)d,$$

which shows that if $(\frac{a}{c}, \frac{b}{d})$ is a farey edge then

$$\left(\frac{a}{c}, \frac{a+b}{c+d}\right), \left(\frac{b}{d}, \frac{a+b}{c+d}\right), \left(\frac{a}{c}, \frac{a-b}{c-d}\right), \left(\frac{b}{d}, \frac{a-b}{c-d}\right)$$

are all farey edges too.

□

Example 4.15. The reader is encouraged to verify that the dual of $e_0 \in E$ is

$$e_0^* = (1, -1) \in E^*.$$

Lemma 4.16. *The action of $PSL_2(\mathbb{Z})$ preserves dual pairings. That is, for all $A \in PSL_2(\mathbb{Z})$ and for all $e \in E \cup E^*$,*

$$A \cdot e^* = (A \cdot e)^*.$$

Proof. Let

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in PSL_2(\mathbb{Z}).$$

We first show that $A \cdot e_0^* = (A \cdot e_0)^*$. In light of (4.10),

$$(A \cdot e_0)^* = (\theta(A))^* = \left(\frac{a}{c}, \frac{b}{d}\right)^* = \left(\frac{a+b}{c+d}, \frac{a-b}{c-d}\right).$$

Also, observe that

$$\theta\left(\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\right) = (1, -1) = e_0^*.$$

Thus using Lemma 3.6 we obtain

$$A \cdot e_0^* = A \cdot \theta \left(\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \right) = \theta \left(\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \right)$$

$$= \theta \left(\begin{bmatrix} a+b & -a+b \\ c+d & -c+d \end{bmatrix} \right) = \left(\frac{a+b}{c+d}, \frac{a-b}{c-d} \right) = (A \cdot e_0)^*.$$

So now let $e \in E$ be any farey edge, and set $B = \theta^{-1}(e)$ so that $B \cdot e_0 = e$ (see Theorem 4.9). Applying the above argument to the action of B, we have

$$A \cdot e^* = A(B \cdot e_0)^* = AB \cdot e_0^*,$$

and applying it to the action of AB then gives

$$AB \cdot e_0^* = (AB \cdot e_0)^* = (A \cdot e)^*.$$

Therefore, $A \cdot e^* = (A \cdot e)^*$ as desired.

□

Corollary 4.17. *The action of $PSL_2(\mathbb{Z})$ takes trident edges to trident edges and is simply transitive on $E^*_{\rightarrow}$.*

Proof. This is more or less immediate from the combination of Theorem 4.9 and Lemma 4.16. The details are left as an exercise. □

5. The Dedekind tessellation built from scratch

The results of the last section suggest that the action of $PSL_2(\mathbb{Z})$ somehow *knows* about farey and trident edges at a fundamental level. So what does this pattern of edges actually *look* like in $\mathbb{H}$? To answer this, we begin by studying how these edges intersect each other.

Observation 5.1. In general, if two hyperbolic lines $(a_1, a_2), (b_1, b_2) \in L$ intersect in $\mathbb{H}$, then all four of $a_1, a_2, b_1, b_2 \in P^1_{\mathbb{R}}$ must be distinct, and their cyclic order in $P^1_{\mathbb{R}}$ must be alternating. That is, up to cyclic reordering it must be (a_1, b_1, a_2, b_2) rather than (a_1, a_2, b_1, b_2). In Proposition 5.2 below, this observation will help us impose betweenness conditions on *finite* endpoints when we view them affinely (i.e. back in $\mathbb{R} \subseteq \mathbb{C}$ rather than in $P^1_{\mathbb{R}} \subseteq P^1_{\mathbb{C}}$).

Proposition 5.2. *The intersection properties of farey and trident edges in $\mathbb{H}$ (so excluding at ideal points) are as follows:*

1. *(1) No farey edge intersects any other farey edge.*
2. *(2) Each farey edge intersects exactly one trident edge, which is its dual; these meet at right angles.*
3. *(3) Each trident edge intersects a total of 4 other trident edges, a pair at each of two points; the angles of intersection at these points are multiples of $\pi/3$.*

Proof. First we show that statements 1 and 2 apply to e_0.

Suppose $(r_1, r_2) \in E \cup E^*$ is a farey or trident edge that crosses $e_0 = (\infty, 0)$ somewhere in $\mathbb{H}$. Then in light of Observation 5.1, r_1, r_2 are finite and nonzero, and of opposite signs, so without loss of generality assume $r_1 > 0 > r_2$. Write r_1, r_2 in lowest terms as $r_1 = \frac{a}{c}, r_2 = \frac{b}{d}$, such that $c, d \geq 0$. Then multiplying through by cd we have

$$ad > 0 > bc. \tag{5.3}$$

Since $ad, bc \in \mathbb{Z}$, this implies that $ad - bc \geq 2$. Hence (r_1, r_2) is not a farey edge. If it is a trident edge, then $ad - bc = 2$, so (5.3) implies $ad = 1$ and $bc = -1$. It follows that $|r_1| = |r_2| = 1$ and $r_2 = -r_1$. Thus $(r_1, r_2) = (1, -1) = e_0^*$.

Therefore e_0 intersects no farey or trident edges except possibly e_0^*. Using basic Euclidean geometry in the affine version of $\mathbb{H}$, we see immediately that e_0 and e_0^* indeed intersect at $i \in \mathbb{C}$ with perpendicular tangent lines (exercise).

Next we show that statement 3 applies to e_0^*.

Suppose $(r_1, r_2) \in E^*$ is a trident edge that intersects $e_0^* = (1, -1)$ in $\mathbb{H}$. Write r_1, r_2 in lowest terms as $r_1 = \frac{a}{c}, r_2 = \frac{b}{d}$ such that $c, d \geq 0$ (where ∞ is expressed as $\frac{1}{0}$), so that

$$ad - bc = \pm 2. \tag{5.4}$$

We consider two cases, based on whether or not an endpoint is infinite.

Case 1: Suppose $r_1, r_2 \neq \infty$. Then $c, d > 0$. To get an intersection, one of r_1, r_2 must be between the endpoints of e_0 and the other not between them (in the affine sense), so without loss of generality suppose that

$$|r_1| > 1 > |r_2|. \tag{5.5}$$

Multiplying through by $cd > 0$ gives

$$|a|d > cd > |b|c,$$

so (5.4) implies $(|a| - c)d = 1$ and $(d - |b|)c = 1$. Thus

$$|c| = |d| = \big||a| - c\big| = \big|d - |b|\big| = 1,$$

so $a, b \in \{-2, 0, 2\}$, and from here (5.5) further implies that $|a| = 2$ and $b = 0$. Hence in this case $(r_1, r_2) = (\pm 2, 0)$.

Case 2: Without loss of generality, suppose $r_1 = \infty$, so $a = 1$ and $c = 0$. Therefore (5.4) gives

$$1 \cdot d - 0 \cdot b = \pm 2,$$

so $|d| = 2$. With r_1 being infinite, to get an intersection we need r_2 between the endpoints of e_0 (in the affine sense), so $|r_2| < 1$. Since $|d| = 2$, the only way this can happen is if $|b| < 2$, and if $b = 0$ then $\frac{b}{d}$ is not in lowest terms, so $|b| = 1$. Thus in this case $(r_1, r_2) = (\infty, \pm\frac{1}{2})$.

Therefore, the set of trident edges that intersect e^* is some subset of

$$T = \left\{(0, 2), (0, -2), \left(\infty, \frac{1}{2}\right), \left(\infty, \frac{-1}{2}\right)\right\} \subseteq E^*.$$

Once again using Euclidean geometry, the reader is encouraged to verify that e_0^* does intersect all four members of T, and furthermore that these intersections occur in two triplets at $e^{i\pi/3}$ and $e^{i2\pi/3}$ with tangent lines meeting at angles of $\pi/3$.

Now to complete the proof, let $e \in E$ and $e^* \in E^*$ be an arbitrary dual edge pair, and set $A = \theta^{-1}(e)$ so that $A \cdot e_0 = e$ and $A \cdot e_0^* = e^*$ (see Theorem 4.9 and Lemma 4.16).

Note that $PSL_2(\mathbb{R})$ acts on $\mathbb{H}$ by conformal bijections, preserving incidence and angles of incidence. So in particular, for all $\ell_1, \ell_2 \in L$, $A \cdot \ell_1$ intersects $A \cdot \ell_2$ at a given angle if and only if ℓ_1 intersects ℓ_2 at that angle. Also, A preserves the sets E and E^*. Hence ℓ is a farey or trident edge that intersects e if and only if $A^{-1} \cdot \ell$ is a farey or trident edge that intersects $A^{-1} \cdot e = e_0$. As we have seen, this is true if and only if $A^{-1} \cdot \ell = e_0^*$, which by choice of A is equivalent to the condition $\ell = e^*$.

The same argument can be used to show that ℓ is a trident edge that intersects e_0^* if and only if $\ell = A \cdot t$ for some $t \in T$.

As mentioned, the remaining details about angle follow straight from conformality, and this completes the proof.

□

Definition 5.6. A point in $\mathbb{H}$ where a farey and trident edge meet is called a *dual point*, and where three trident edges meet is a *trident point*.

We will soon use Proposition 5.2 to deduce exactly how the farey and trident edges cut up the upper half plane. First, though, it will be helpful to consider how the farey edges break things up on their own, and in particular to verify that there are enough of them to essentially triangulate all of $\mathbb{H}$.

Lemma 5.7. *The edge set E yields a tessellation (called the Farey tessellation) that covers $\mathbb{H}$ by hyperbolic ideal triangles. That is, every point in $\mathbb{H}$ is either part of a farey edge or enclosed by 3 farey edges that meet pairwise at ideal points.*

Proof. Let $p = x + iy \in \mathbb{H}$. So $y > 0$. If $x \in \mathbb{Z}$, then p is on the edge $(\infty, x) = (\frac{1}{0}, \frac{x}{1})$, which is a farey edge, in which case we are done. Otherwise, if $x \notin \mathbb{Z}$, let $m = \lfloor x \rfloor, n = \lceil x \rceil$, so that $m < x < n$ and $n - m = 1$. Then $(n, m) = (\frac{n}{1}, \frac{m}{1})$ is also a farey edge, so we are done if p lies on (n, m). Moreover, we are also done if p lies outside the semi-disk enclosed between (n, m) and $\mathbb{R}$, since in that case p is in the ideal triangle with vertices n, m, ∞. Thus to cover all remaining cases, it is safe to assume p lies inside the semi-disk under (n, m).

More generally, suppose p lies in the semi-disk under any farey edge $(r_1, r_2) \in E$ that satisfies $r_1, r_2 \neq \infty$. (If an endpoint were infinite the edge would be a Euclidean ray rather than a Euclidean semi-circle.) Consider the trident edge $t \in E^*$ that is dual to (r_1, r_2). Since t intersects (r_1, r_2) inside $\mathbb{H}$ by Proposition 5.2, it must have an endpoint $r_3 \in \mathbb{Q}^*$ that satisfies $r_1 < r_3 < r_2$ (see Observation 5.1). Furthermore, by Proposition 4.11, (r_1, r_3) and (r_3, r_2) are both farey edges.

In general for a farey edge with finite endpoints $\frac{a}{c}$ and $\frac{b}{d}$, being a Euclidean semi-circle, its diameter is the distance between its endpoints:

$$\Big|\frac{a}{c} - \frac{b}{d}\Big| = \Big|\frac{ad - bc}{cd}\Big| = \Big|\frac{1}{cd}\Big|.$$

In particular, every farey edge with finite endpoints has diameter $\frac{1}{n}$ for some $n \in \mathbb{N}$. So find $k \in \mathbb{N}$ such that $|r_1 - r_2| = \frac{1}{k}$. Then since $r_1 < r_3 < r_2$, we have $|r_1 - r_3| < |r_1 - r_2|$ and similarly $|r_3 - r_2| < |r_1 - r_2|$, so both new farey edges must have diameter at most $\frac{1}{k+1}$.

We now ask: Does p lie in either of the two semicircles defined by (r_1, r_3) and (r_3, r_2)? If not, then p is either on a farey edge or enclosed in the ideal triangle with endpoints r_1, r_2, r_3, in which case we are done. Otherwise, repeat the same process as above, with (r_1, r_2) replaced by the new edge that p is enclosed under, to obtain a new pair of semi-circles with radius at most $\frac{1}{k+2}$; then $\frac{1}{k+3}$, and so on. At the end of each stage, ask the same question to see if it is necessary to continue.

Since the sequence $(\frac{1}{k})_{k \in \mathbb{N}}$ converges to zero, after some finite number of iterations, y will exceed the radius of all new semi-circles obtained in this process, at which point p must no longer lie under any of them. Thus the algorithm necessarily ends in victory, and we conclude that p is either on a farey edge or in an ideal triangle with farey edges for sides.

□

Theorem 5.8. *The edge set $E \cup E^*$ yields a tessellation of $\mathbb{H}$ by congruent hyperbolic ideal triangles that each have interior angles of $\pi/2$ (at a dual point) and $\pi/3$ (at a trident point), and an ideal vertex at a point in $\mathbb{Q}^*$. This tessellation is called the Dedekind tessellation (henceforth "the tessellation").*

Proof. The reader is encouraged to prove this by drawing! Below we present guiding steps for one possible strategy—but this is by no means the only one.

Preliminary note: To draw the hemisphere $\mathbb{H} \subseteq P^1_{\mathbb{C}}$, there are a couple of natural options. One option is to work in the upper half plane model by stereographically projecting from the north pole $[1 : 0] \in P^1_{\mathbb{C}}$ as usual. Another option that may be helpful for seeing the overall symmetry is to draw $\mathbb{H}$ as a disk, by stereographically projecting from the point $[-i : 1]$ onto a copy of $\mathbb{C}$ placed so that it passes through the great circle $P^1_{\mathbb{R}} \subseteq P^1_{\mathbb{C}}$. (Intuitively, one can imagine peering at the hemisphere from the point on the sphere directly across from it.)

Guiding steps: Start by drawing e_0 and its dual e_0^*. What other farey edges can be filled in? [Hint: See statement 2 of Proposition 4.11.] Well now those farey edges need dual trident edges too. Can we deduce anything about the endpoints of these trident edges? What about where the they intersect each other? [Hint: Statement 2 of Proposition 5.2 should help, and also statement 3.] Okay once those are in place, fill in more new farey edges (again from Proposition 4.11), and then new trident edges (again from Proposition 5.2), and so on. Can you see how to continue this process indefinitely? Will it account for all farey and trident edges in the long run? (Spoiler: Yes it will.)

As a final note, Lemma 5.7 shows that the Dedekind tessellation fills all of $\mathbb{H}$, since it is just a refinement of the Farey tessellation.

□

Definition 5.9. The hyperbolic triangles from Theorem 5.8 are called *faces* of the tessellation. Each face is considered to include its whole boundary in $\mathbb{H}$ (i.e. all points on its sides and vertices except for its ideal vertex in $\mathbb{Q}^*$), so is closed in $\mathbb{H}$ by definition (where $\mathbb{H}$ has the standard subspace topology inherited from $\mathbb{C}$).

Observation 5.10. Since every face of the tessellation has one one dual point, one trident point, and one ideal point for vertices (see Theorem 5.8), they each have a definite *handedness* based on the cyclic ordering of their vertices. It is a nice exercise to verify that adjacent faces have opposite handedness. (Feel free to make your own rule about which ordering is for which hand; Figure 1 and Figure 2 show handedness with color.)

Theorem 5.11. *The action of $PSL_2(\mathbb{Z})$ takes left-hand faces to left-hand faces and right-hand faces to right-hand faces. Furthermore, it acts simply transitively on each of these two sets.*

Proof. As regards the first statement, we have already seen that $PSL_2(\mathbb{Z})$ acts on E and on E^*, which together form the edge set of the tessellation. Therefore, if $A \in PSl_2(\mathbb{Z})$, then A acts bijectively on the union of all edges in the tessellation, hence also bijectively on their complement in $\mathbb{H}$. This complement is the union of all face interiors, each of which is its own connected component. Given that A is a homeomorphism, preserving connectedness in both directions, it follows that A must map each face interior surjectively onto another, while also taking

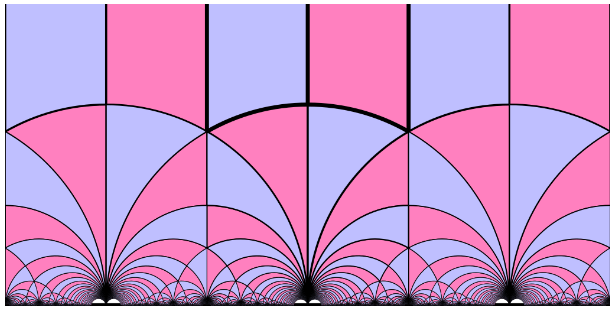

FIGURE 1. The Dedekind tessellation, shown here in the upper half plane model of the hyperbolic plane (achieved by stereographically projecting $\mathbb{H} \subseteq P^1_{\mathbb{C}}$ from the north pole) [22].

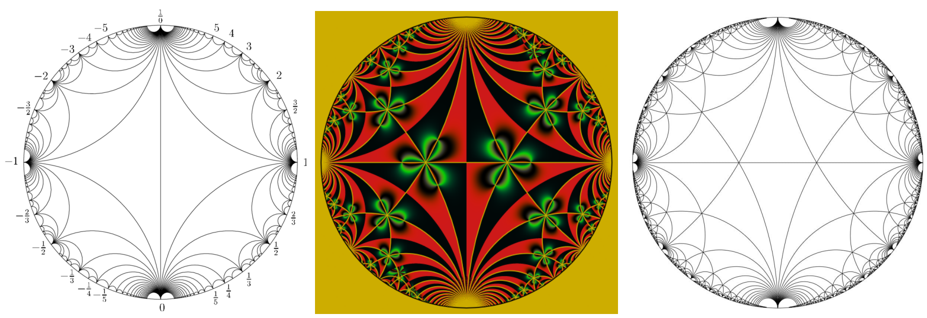

FIGURE 2. At left: farey edges [20]; at right: trident edges [21]. Here both are shown in the Poincaré disk model (achieved by stereographically projecting from the point $[-i : 1] \in P^1_{\mathbb{C}}$). They are combined at center to construct the Dedekind tessellation (with added clovers for good luck) [19].

their boundaries along (exercise). Hence, A takes faces to faces. It preserves their handedness because all conformal maps are orientation-preserving (see [1]).

Showing simple transitivity we leave as an exercise to the reader. [Hint: Find a nice bijective correspondence between $E_{\rightarrow}$ and the set of left-hand faces (or right-hand, say whichever you write with). In particular, make sure your correspondence is nice enough that the action of $PSL_2(\mathbb{Z})$ respects it—i.e, that each element of $PSL_2(\mathbb{Z})$ sends associated pairs to other associated pairs. Then apply Theorem 4.9. Related exercise: Can you find any other sets on which the action is also simply transitive using a similar tactic?]

□

Remark 5.12. One can also show (along similar lines as for Theorem 5.11) that $PSL_2(\mathbb{Z})$ acts transitively on the set of dual points and on the set of trident points. However, the reader may have already noticed that the transitivity is in this case *not* simple. (More on this later.)

6. From orbits to the quotient space

Using the geometric understanding we built up in the previous section, we now seek to understand how properties of the action of $PSL_2(\mathbb{Z})$ translate to properties of the new quotient space we are about to build. But $PSL_2(\mathbb{Z})$ itself is by no means the only group we can quotient by—many of its subgroups work as well!

Definition 6.1.

(1) For any $n \in \mathbb{N}$ (excluding 0), the *principle congruence subgroup* of level n of the modular group is the kernel of the natural projection $\pi_n : SL_2(\mathbb{Z}) \to SL_2(\mathbb{Z}/n\mathbb{Z})$ that takes each entry of the given matrix to its residue mod n. Hence it has the form

$$\Gamma(n) := \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in SL_2(\mathbb{Z}) \;\middle|\; a, d \equiv 1 \pmod{n} \text{ and } b, c \equiv 0 \pmod{n} \right\}$$

for $n \geq 2$. (For $n = 1$, we take $\Gamma(1)$ to be all of $SL_2(\mathbb{Z})$ [2].)

(2) Let π_s denote the projection $SL_2(\mathbb{Z}) \to PSL_2(\mathbb{Z})$ that identifies $\pm A$ for all $A \in SL_2(\mathbb{Z})$. Then the principle congruence subgroup of level n of $PSL_2(\mathbb{Z})$ is $\pi_s(\Gamma(n)) \leq PSL_2(\mathbb{Z})$ [8].

Definition 6.2. A *congruence subgroup* is any subgroup which contains a principle congruence subgroup. (Its level is the minimal n for which it contains the principle congruence subgroup of level n.)

Remark 6.3. The reader is encouraged to check using Definition 6.1 that principle congruence subgroups are normal and of finite index. Definition 6.2 then implies that *all* congruence subgroups are of finite index (though not necessarily normal like the principle ones).

Definition 6.4. Given a congruence subgroup Γ of $PSL_2(\mathbb{Z})$, we form the topological space $\mathbb{H}/\Gamma$ as the set of Γ-orbits of $\mathbb{H}$ equipped with the quotient topology, from the projection

$$\pi : \mathbb{H} \to \mathbb{H}/\Gamma$$
$$h \mapsto \Gamma h.$$

(The quotient topology means that a subset $U \subseteq \mathbb{H}/\Gamma$ is defined to be open in $\mathbb{H}/\Gamma$ if and only if $\pi^{-1}(U)$ is open in $\mathbb{H}$.)

To get a sense for the quotient space in Definition 6.4, it is natural to look for a subset of $\mathbb{H}$ that has one point from each orbit to serve as representatives. The following result is helpful to have in mind when considering how to do this, and will also be of immense use for several proofs later this section.

Proposition 6.5. *Each face of the Dedekind tessellation contains at most one point per* Γ*-orbit.*

Proof. Let F be a face of the Dedekind tessellation, and suppose $h_1, h_2 \in F$ are in the same Γ-orbit. Then $h_2 = \gamma(h_1)$ for some $\gamma \in \Gamma$, so $h_2 \in \gamma F \cap F$. By Theorem 5.11, γF is a face of the same handedness as F, distinct from F unless γ is the identity. If γ is the identity then we are done because $h_2 = h_1$. Otherwise, F and γF are not adjacent (see Observation 5.10). This implies they intersect at at most one vertex in $\mathbb{H}$, which means $h_2 \in \gamma F \cap F$ must be a dual or trident point, and h_1 the same kind of point (since Γ preserves both types). But F has only one dual point and only one trident point, so we conclude that $h_1 = h_2$.

□

Definition 6.6. A *fundamental domain* for a space with respect to a group action is a subset that contains a point from every orbit such that its interior contains no more than one point per orbit.

Example 6.7. The union of any pair of adjacent faces in the Dedekind tessellation is a fundamental domain for the action of $PSL_2(\mathbb{Z})$ on $\mathbb{H}$. The reader is encouraged to verify this as an exercise. [Hint: Consider Observation 5.10 and how Theorem 5.11 might apply; Proposition 6.5 may also come in handy.]

Note that uniqueness of the orbit representatives does break down on the boundary, for any such choice of fundamental domain. The way in which it breaks down informs the side identifications needed to glue the fundamental domain into the right underlying topological space to represent the quotient.

For example, a conventional choice of fundamental domain for this action is the hyperbolic triangle $\Delta_F \subseteq \mathbb{H}$ with vertices $e^{i\pi/3}$ and $e^{i2\pi/3}$ and the ideal point ∞, joined along the trident edges

$$(1, -1), \left(\infty, \frac{1}{2}\right), \left(\infty, \frac{-1}{2}\right)$$

(old friends from the proof of Proposition 5.2). To glue Δ_F into a space homeomorphic to $\mathbb{H}/PSL_2(\mathbb{Z})$, we need to identify points on its boundary that are in the same orbit, and one can check that in this case this amounts to gluing together points related by a reflection over the imaginary axis. (This effectively zips Δ_0 into a dumpling with a leak at ∞, and we get something homeomorphic to the 2-sphere minus a point, or equivalently to the plane.)

As helpful as fundamental domains can be for visualizing quotient spaces, even without directly finding one for a given congruence subgroup, we can still gain a substantial amount of insight about the quotient space by studying certain properties of the action back in $\mathbb{H}$. To this end, it will be helpful to first introduce the following terminology:

Definition 6.8. Given a point $h \in \mathbb{H}$, its *face cluster* $\Delta(h)$ is the union of all faces containing h.

Observation 6.9. The reader may have noticed that if $h \in \mathbb{H}$, then $\Delta(h)$ includes

- only 1 face if h is not on an edge
- 1 face of each handedness if h is on an edge but not a vertex
- 2 of each handedness if h is a dual point
- 3 of each handedness if h is a trident point.

Lemma 6.10. *For all $h \in \mathbb{H}$, $\Delta(h)$ is a neighborhood of h.*

Proof. This is true for points in Δ_F from Example 6.7 more or less by inspection, and we leave the details to the reader. [Hint: Break this into cases based on Observation 6.9.]

One may yet worry about points which are close to the boundary of $\mathbb{H}$, but we can dispel remaining doubt by moving them over to Δ_F and back again, as follows:

Let $h \in \mathbb{H}$. Example 6.7 tells us that we can find $B \in PSL_2(\mathbb{Z})$ such that $B \cdot h \in \Delta_F$, so that, as noted, $\Delta(B \cdot h)$ is a neighborhood of $B \cdot h$.

Note that for all $A \in PSL_2(\mathbb{Z})$,

$$A \cdot \Delta(h) = \Delta(A \cdot h). \tag{6.11}$$

This is because the action of $PSL_2(\mathbb{Z})$ preserves faces (see Theorem 5.11); in particular, a face F of the tessellation contains h if and only if $A \cdot F$ is a face containing $A \cdot h$.

Hence, $B \cdot \Delta(h) = \Delta(B \cdot h)$ is a neighborhood of $B \cdot h$, meaning there is an open set $U \subseteq B \cdot \Delta(h)$ that contains $B \cdot h$. Since $PSL_2(\mathbb{Z})$ acts by homeomorphisms, it follows that $B^{-1} \cdot U \subseteq \Delta(h)$ is an open set containing h, and we are done. □

We now use face clusters to prove a group-action-related property that will set us up for proving that the quotient space is Hausdorff.

Proposition 6.12. *Let Γ denote any congruence subgroup of $PSL_2(\mathbb{Z})$. If $h_1, h_2 \in \mathbb{H}$ such that $\Gamma h_1 \neq \Gamma h_2$, then there exist neighborhoods U_1 of h_1 and U_2 of h_2 such that $\Gamma U_1 \cap \Gamma U_2 = \emptyset$.*

Proof. Suppose $h_1, h_2 \in \mathbb{H}$ such that $\Gamma h_1 \neq \Gamma h_2$. Consider the set

$$S := \Delta(h_1) \cap \Gamma h_2.$$

Each face included in $\Delta(h_1)$ contains at most one point of Γh_2 by Proposition 6.5, so S is finite given Observation 6.9. Also, $h_1 \notin \Gamma h_2 \supseteq S$ because that would mean $\Gamma h_1 = \Gamma h_2$, contradicting our choice of h_1, h_2. Therefore, so long as S is non-empty, we can set

$$r_1 := \frac{1}{2} \min_{x \in S} d(h_1, x) > 0$$

so that $B_{r_1}(h_1)$ is a neighborhood of h_1. So now if S *is* empty, set $N_1 = \Delta(h_1)$, and otherwise, set $N_1 = \Delta(h_1) \cap B_{r_1}(h_1)$. Then either way, N_1 is a neighborhood of h_1 due to Lemma 6.10, and $\overline{N_1} \cap S$ is empty by choice of r_1. Also, since $\Delta(h)$ is closed in $\mathbb{H}$ (being a finite union of closed faces), we have $\overline{N_1} \subseteq \overline{\Delta(h_1)} = \Delta(h_1)$. Therefore,

$$\overline{N_1} \cap \Gamma h_2 = \left(\overline{N_1} \cap \Delta(h_1)\right) \cap \Gamma h_2 = \overline{N_1} \cap \left(\Delta(h_1) \cap \Gamma h_2\right) = \overline{N} \cap S = \emptyset.$$

We now show that Γh_2 is disjoint from *all* of $\Gamma \overline{N_1}$. Suppose for contradiction that $p \in \Gamma \overline{N_1} \cap \Gamma h_2$. Then $p \in \gamma(\overline{N_1}) \cap \Gamma h_2$ for some $\gamma \in \Gamma$, meaning $\gamma^{-1}(p) \in \overline{N_1} \cap \Gamma h_2$, but this is empty. So $\Gamma \overline{N_1} \cap \Gamma h_2 = \emptyset$ as hoped.

Since Γ acts by homeomorphisms, we have $\Gamma \overline{N_1} = \overline{\Gamma N_1}$ (exercise). This implies in particular that h_2 itself is not in the closure of ΓN_1, so has positive distance from ΓN_1. Letting $r_2 > 0$ be half this distance, then, we obtain a neighborhood $N_2 := B_{r_2}(h_2)$ of h_2 that is disjoint form $\Gamma(N_1)$.

We now finish the proof by showing that ΓN_1 is in fact disjoint from all of ΓN_2. Similar to earlier, suppose $p \in \Gamma N_1 \cap \Gamma N_2$. Then there is some $\gamma \in \Gamma$ for which $p \in \Gamma N_1 \cap \gamma(N_2)$, so $\gamma^{-1}(p) \in \Gamma N_1 \cap N_2 = \emptyset$, a contradiction. Thus $\Gamma N_1 \cap \Gamma N_2 = \emptyset$, and we are done. □

To translate Proposition 6.12 into Hausdorffness, we just need one more lemma first.

Lemma 6.13. *Given any congruence subgroup Γ of $PSL_2(\mathbb{Z})$, the projection π from Definition 6.4 is an open map.*

Proof. Recall that $\pi(h) = \Gamma h$ for all $h \in \mathbb{H}$ (by definition), so given $p \in \mathbb{H}$, one can check that $\pi(h) = \pi(p)$ if and only if $h \in \Gamma p$ (exercise). Thus $\pi^{-1}(\pi(p)) = \Gamma p$ for all $p \in \mathbb{H}$, from which it follows that $\pi^{-1}(\pi(S)) = \Gamma S$ for all $S \subseteq \mathbb{H}$.

In particular, if $U \subseteq \mathbb{H}$ is open, then $\pi^{-1}(\pi(U)) = \Gamma U$ is a union of open sets (since Γ acts by homeomorphisms), and hence is open. Therefore $\pi(U)$ is open in $\mathbb{H}/\Gamma$ with the quotient topology, which verifies that π is an open map.

□

Theorem 6.14. *Let Γ be any congruence subgroup of $PSL_2(\mathbb{Z})$. Then $\mathbb{H}/\Gamma$ is Hausdorff.*

Proof. Let $p_1, p_2 \in \mathbb{H}/\Gamma$ with $p_1 \neq p_2$. Since π is surjective, find $h_1 \in \pi^{-1}(p_1)$ and $h_2 \in \pi^{-1}(p_2)$. Then $\pi^{-1}(p_1) = \pi^{-1}(\pi(h_1)) = \Gamma h_1$ (see the proof of Lemma 6.13), and similarly $\pi^{-1}(p_2) = \pi^{-1}(\pi(h_2)) = \Gamma h_2$. Note that, being distinct, p_1 and p_2 have different pre-images; hence $\Gamma h_1 \neq \Gamma h_2$. Therefore by Proposition 6.12, there exist neighborhoods N_1 and h_1 and N_2 of h_2 such that $\Gamma N_1 \cap \Gamma N_2 = \emptyset$. Since $\pi^{-1}(\pi(N_1)) = \Gamma N_1$ and $\pi^{-1}(\pi(N_2)) = \Gamma N_2$, it follows that

$$\pi^{-1}\left(\pi(N_1) \cap \pi(N_2)\right) = \pi^{-1}(\pi(N_1)) \cap \pi^{-1}(\pi(N_2)) = \emptyset,$$

so $\pi(N_1) \cap \pi(N_2) = \emptyset$ by the surjectivity of π. Finally, by Lemma 6.13, $\pi(N_1)$ and $\pi(N_2)$ are open and hence neighborhoods of p_1 and p_2, respectively, and the proof is complete.

□

It is not hard to show that $\mathbb{H}/\Gamma$ also inherits second-countability straight from $\mathbb{H}$ [2], so given Theorem 6.14 we now know that $\mathbb{H}/\Gamma$ at least has hope of turning out to be a manifold—and hopefully even a Riemann surface—if we can find suitable charts for it. Excellent.

To start down this road of charts, then, we once again look for a group-action-related property that can translate to a topological property—in this case to the existence of a local homeomorphism.

Proposition 6.15. *Let Γ be a congruence subgroup of $PSL_2(\mathbb{Z})$. Given any $h \in \mathbb{H}$, there exists a neighborhood N of h such that for all $\gamma \in \Gamma$, $\gamma(N) \cap N \neq 0$ only if γ stabilizes h (i.e. only if $\gamma(h) = h$). In particular, int$(\Delta(h))$ is such a neighborhood.*

Proof. Let $h \in \mathbb{H}$, and $\gamma \in \Gamma$. Set $N = \text{int}(\Delta(h))$—a neighborhood of h due to Lemma 6.10—and suppose that $\gamma(N) \cap N \neq \emptyset$. Then since $\gamma(N) \cap N$ is open in $\mathbb{H}$, it is not a subset of the union of edges of the tessellation, so it must contain a point p in the *interior* of some face F in the tessellation. Note that

$$\gamma(N) \cap N \subseteq \gamma(\Delta(h)) \cap \Delta(h) = \Delta(\gamma(h)) \cap \Delta(h)$$

by (6.11), so p is in the face cluster of both $\gamma(h)$ and h. Furthermore, p is not in any other face besides F, since distinct faces overlap at most along boundaries. Hence, with face clusters being unions of faces, we deduce that $F \subseteq \Delta(\gamma(h)) \cap \Delta(h)$. Therefore F must contain both $\gamma(h)$ and h, which share an orbit. So by Proposition 6.5, $\gamma(h) = h$.

□

Theorem 6.16. *Let Γ be any congruence subgroup of $PSL_2(\mathbb{Z})$. If $h \in \mathbb{H}$ is such that Γ is free on h (i.e. if h has trivial stabilizer in Γ), then the projection map π from Definition 6.4 is a local homeomorphism at h.*

Proof. Suppose $h \in \mathbb{H}$ is such that Γ is free on h. By *Proposition* 6.15, there is a neighborhood N of h with the property that $\gamma(N) \cap N$ is empty unless γ stabilizes h. But only the identity stabilizes h because Γ is free on h. So if $p \in N$, then $\gamma(p) \in N$ implies γ is the identity. Thus $\Gamma p \cap N = \{p\}$ for all $p \in N$, meaning N includes at most one point from every Γ-orbit. Therefore, π is 1-to-1 on N, so

$$\pi|_N : N \to \pi(N)$$

is invertible. Furthermore, $\pi|_N$ is continuous (by definition of the quotient topology), and since π is open by Lemma 6.13, it follows that $\pi|_N^{-1}$ is also continuous. Hence $\pi|_N$ is homeomorphic onto its image (which is a neighborhood of $\pi(h)$), and we conclude that π is a local homeomorphism at h.

□

Remark 6.17. We won't go into the theory of covering maps here (see [5] for a definition), but for the reader familiar with them, it is nice to note that if S is the set of points on which Γ acts freely, then $\pi|_S$ is in fact a covering map of its image $\pi(S)$. (This can be shown using the proof of Theorem 6.16.)

For the points where Γ acts freely, Theorem 6.16 gives us a direct way of building charts: namely, by taking local inverses of π. This is almost great; the only problem is that we still need to figure out when and where the action is free (if anywhere)—not to mention what to do in cases where it's not.

Proposition 6.18. *With respect to the action of $PSL_2(\mathbb{Z})$, the size of the stabilizer subgroup of a point $h \in \mathbb{H}$ is equal to*

- 3 *if h is a trident point*
- 2 *if h is a dual point*
- 1 *in all other cases.*

Proof. Let $h \in h$, and consider any face $F \subseteq \Delta(h)$. Let n be the number of faces with the same handedness as F which are included in $\Delta(h)$. Theorem 5.11 tells us that there is a unique element of $PSL_2(\mathbb{Z})$ sending F to each one of these n faces, so exactly n total elements which send F to a subset of $\Delta(h)$. Also, from Observation 6.9 we know that n is 3 for trident points, 2 for dual points, and 1 for all other points. Therefore, to prove the result, all that remains to show is that $A \in PSL_2(\mathbb{Z})$ stabilizes h if and only if $A \cdot F \subseteq \Delta(h)$. In fact, this remaining piece holds for any congruence subgroup, as we now show.

Let Γ be a congruence subgroup of $PSL_2(\mathbb{Z})$. Note that $h \in F$ because we chose $F \subseteq \Delta(h)$, so $\gamma(h) \in \gamma F$ (which is a face by Theorem 5.11) for all $\gamma \in \Gamma$.

Hence if γ stabilizes h, then γF contains $h = \gamma(h)$, so $\gamma F \subseteq \Delta(h)$.

Conversely, if $\gamma \in \Gamma$ is such that $\gamma F \subseteq \Delta(h)$, then γF contains not only $\gamma(h)$ but also h, implying that $\gamma(h) = h$ by Proposition 6.5. This completes the proof.

□

Remark 6.19. With at most 3 elements, the stabilizer subgroup in $PSL_2(\mathbb{Z})$ of any $h \in \mathbb{H}$ is small enough that it must be cyclic, and so we often describe stabilizer subgroups by just their order. (This bound on size applies to congruence subgroups of $PSL_2(\mathbb{Z})$ as well, since the stabilizer with respect to the action of a subgroup is a subgroup of the original stabilizer.)

Definition 6.20. Given a congruence subgroup Γ of $PSL_2(\mathbb{Z})$,

(1) The *period* of a point $h \in \mathbb{H}$ is the size of its stabilizer subgroup in Γ.

(2) Points with period *higher* than 1 are called *elliptic points* for Γ. (That leaves the non-elliptic points as the ones that Γ acts freely on.)

Remark 6.21. One can directly check that if $H \leq \Gamma$ is the stabilizer of $h \in \mathbb{H}$, then the stabilizer of $\gamma(h)$ is $\gamma H \gamma^{-1}$, which has the same order. Thus period is consistent across all points of a given orbit, so is well-defined on points of the quotient space.

Observation 6.22. From Proposition 6.18, we see that the elliptic points for the action of $PSL_2(\mathbb{Z})$ are precisely the dual and trident points. (Together these make up two orbits given Remark 5.12, and hence account for two points in $\mathbb{H}/PSL_2(\mathbb{Z})$.) Additionally, this implies that dual and trident points are the only points in $\mathbb{H}$ that can ever be elliptic—regardless of the congruence subgroup—since, as noted earlier, passing to a subgroup keeps the stabilizer of a point at least as small.

Elliptic points complicate the chart-building process. (More on this in the next section.) But before stressing out too much, it is comforting to note that for many congruence subgroups they never show up in the first place! We include the next result to demonstrate this.

Proposition 6.23. *$\Gamma(n)$ is free on $\mathbb{H}$ for all $n \geq 2$*

Proof. Let $n \geq 2$, and suppose $\gamma \in \Gamma(n)$ is such that $\gamma(h) = h$ for some $h \in \mathbb{H}$.

If h is *not* a dual or trident point, then we know automatically that γ is the identity because h is non-elliptic (see Observation 6.22).

So now assume that h *is* a dual or trident point. Then it shares an orbit with either the dual point i or the trident points $e^{i\pi/3}$ and $e^{i2\pi/3}$, meaning there exists $\alpha \in \Gamma$ such that

$$\alpha(h) \in \{i, e^{i\pi/3}, e^{i2\pi/3}\}.$$

Note that

$$\alpha\gamma\alpha^{-1}(\alpha(h)) = \alpha(\gamma(h)) = \alpha(h),$$

i.e. $\alpha\gamma\alpha^{-1}$ stabilizes $\alpha(h)$. Also, $\alpha\gamma\alpha^{-1}$ is in $\Gamma(n)$ because $\Gamma(n)$ is normal in $PSL_2(\mathbb{Z})$ (being the kernel of a group homomorphism).

We will now explicitly narrow down what $\alpha\gamma\alpha^{-1}$ could be. From Proposition 6.18 we know that exactly 2 elements of $PSL_2(\mathbb{Z})$ stabilize i and exactly 3 stabilize each of $e^{i\pi/3}$ and $e^{i2\pi/3}$. One can check that, in $P^1_{\mathbb{C}}$,

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} i \\ 1 \end{bmatrix} \equiv \begin{bmatrix} i \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} e^{i\pi/3} \\ 1 \end{bmatrix} \equiv \begin{bmatrix} e^{i\pi/3} \\ 1 \end{bmatrix} \equiv \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} e^{i\pi/3} \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} e^{i2\pi/3} \\ 1 \end{bmatrix} \equiv \begin{bmatrix} e^{i2\pi/3} \\ 1 \end{bmatrix} \equiv \begin{bmatrix} 0 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} e^{i\pi/3} \\ 1 \end{bmatrix},$$

which therefore means that the identity matrix plus the above five matrices are the *only* elements of $PSL_2(\mathbb{Z})$ that can possibly stabilize $\alpha(h)$. Therefore, $\alpha\gamma\alpha^{-1}$ is one of them. On the other hand, none of these matrices besides I is in $\Gamma(n)$, since ± 1 is not congruent to 0 mod n for any $n \geq 2$. Therefore, $\alpha\gamma\alpha^{-1}$ must be the identity, implying that $\gamma = \alpha^{-1}(\alpha\gamma\alpha^{-1})\alpha$ is the identity as well. This proves $\Gamma(n)$ is free on $\mathbb{H}$.

□

7. Overview of building charts to form a Riemann surface

For a congruence subgroup Γ, in order to turn $\mathbb{H}/\Gamma$ into a Riemann surface it remains to find charts for its points and show that their transition maps are conformal. As noted in the previous section, Theorem 6.16 has already given us a way to construct charts for non-elliptic points, by taking local inverses of π. Note that due to Proposition 6.23, this means $\mathbb{H}/\Gamma(n)$ is locally Euclidean for all $n \geq 2$ because every point is non-elliptic so already has a chart. Thus if Γ is a *principle* congruence subgroup other than $PSL_2(\mathbb{Z})$ itself (which is $\Gamma(1)$), then we at least know the quotient is a manifold. However, when elliptic points are present, building charts for them takes more care, and for this paper we will only give a brief overview of the process; we refer the reader to [2] for a much more in-depth exposition.

The starting point for handling elliptic points is essentially to find a holomorphic map that mimics how π locally glues points together. We already know that the stabilizer group of an elliptic point $p \in \mathbb{H}$ is cyclic, and its elements are called hyperbolic rotations since they are automorphisms of $\mathbb{H}$ that fix a non-ideal point. Hyperbolic rotations behave like Euclidean rotations locally, and if they were actually Euclidean rotations, then the points that π identifies near p would be precisely those points with equal distance from p that are spaced out evenly around p, i.e., at angles of $2\pi/n$ and its multiples, where n is the period of p. This would be great, since then we could simply use the holomorphic map $(z-p)^n$ to identify them!

Okay but the rotations are not Euclidean. Except. In the disk model, a hyperbolic rotation about the *origin* really is a Euclidean rotation, so mapping the upper half plane to the unit disk and then moving our elliptic point to the origin—both of which can be accomplished with Möbius transformations—solves the issue. We can think of this initial adjustment as the "straightening map", and then composing with z^n achieves the "wrapping action" that completes the identification to simulate π.

Indeed π is locally $n - to - 1$ about p, so to get a chart for the point $\pi(p)$ in the quotient, we can restrict π to a suitably small neighborhood N of p and then define a chart on $\pi(N)$ (which is a neighborhood of $\pi(p)$ by Lemma 6.13) as follows: For any point $a \in \pi(N)$, first pull back to the n pre-images of a in N, and then re-identify them with the straightening-wrapping combo map. All told, this process gives a bijection from $\pi(N)$ to a neighborhood of the origin in $\mathbb{C}$, which one can check is continuous in both directions, and thus has created a valid chart for $\pi(p) \in \mathbb{H}/\Gamma$.

This process of giving charts to the images of elliptic points is what finishes the work of building an atlas for $\mathbb{H}/\Gamma$, and it turns out that with suitably small chart domains, the transition maps between its charts are indeed conformal, making it a complex atlas. (The exposition in [2] includes a rigorous proof of this conformality, which we will also not go into here.) In the end though, the punchline is that $\mathbb{H}/\Gamma$ indeed carries a natural complex structure (generated from the atlas) that turns it into a Riemann surface. To recognize this additional structure, we introduce the following terminology and notation:

Notation 7.1. With the added structure of a Riemann surface, the space $\mathbb{H}/\Gamma$ is called a *modular curve* and is often denoted $Y(\Gamma)$.

Note that $Y(\Gamma)$ is in general not compact, due to having punctures called *cusps* as a relic of the ideal points of the tessellation. However, for any congruence subgroup

Γ it turns out that there is always a natural way to compactify $Y(\Gamma)$, and we again refer the reader to [2] for a rigorous explanation of how to do this. The details take work, but the general idea is as follows:

Since the set of ideal points of the tessellation is $\mathbb{Q}^*$, to fill in the cusps it makes sense to extend our scope from $\mathbb{H}$ to $\mathbb{H} \cup \mathbb{Q}^*$ (called the *extended upper half plane*). Since $PSL_2(\mathbb{Z})$ also acts on $\mathbb{Q}^*$, the domain of π can easily be extended to $\mathbb{H} \cup \mathbb{Q}^*$ with the same rule $a \mapsto \Gamma(a)$.

Topologically some care is needed, because the subspace topology on $\mathbb{H} \cup \mathbb{Q}^*$ gets problematic at the boundary when passing to the quotient. But by adjusting what counts as a neighborhood for the points of $\mathbb{Q}^*$, one can put a new topology on $\mathbb{H} \cup \mathbb{Q}^*$ that is able to make the quotient compact while keeping it Hausdorff.

Then for charts, the story has some parallels with the case of the elliptic points: it boils down to a straightening action and a wrapping action that makes use of periodicity. Compare: While elliptic points are fixed by hyperbolic rotations, which can be straightened into Euclidean rotations by sending the fixed point to 0, ideal points are fixed by transformations which can be made into Euclidean *translations* by sending the fixed point to ∞ (these are *parabolic* transformations). Also, where elliptic points have a period, cusps each have a similar kind of cyclic parameter associated to them, known as their width. After being straightened, while z^n does the wrapping for elliptic points of period $n \in \mathbb{N}$, the map $e^{i2\pi/k}$ is what does the wrapping for cusps of width $k \in \mathbb{N}$, by identifying points that are related via a translation by any multiple of k.

Again, this is by no means the full story; we only hope it gives a general sense of the considerations involved in the compactification process. We end this section with the following standard notation:

Notation 7.2. The compactified version of a modular curve $Y(\Gamma)$ is typically denoted $X(\Gamma)$. It is also called a modular curve.

8. Examples and further discussion

Example 8.1 ($\Gamma(2)$ and the little Picard theorem.)**.** Consider the action of $\Gamma(2)$ on $\mathbb{H}$. From the first isomorphism theorem, we know the 6 cosets of $\Gamma(2)$ in $PSL_2(\mathbb{Z})$ form a group isomorphic to $PSL_2(\mathbb{Z}/2\mathbb{Z})$, which is isomorphic to S_3 [14]. Now consider the Farey tessellation from Lemma 5.7. It is a nice exercise to go through and color its endpoints according to the types $\frac{odd}{odd}$, $\frac{odd}{even}$, and $\frac{even}{odd}$, from which it emerges that the farey edges each have endpoints of two *differing* types from this list, allowing them to be colored systematically by the leftover type that they miss (see Figure 3). One can then show that the action of $\Gamma(2)$ preserves this coloring of the Farey tessellation and acts simply transitively on the set of *unoriented* farey edges of any fixed color. [Hint: Consider where e_0 can go by an element of $\Gamma(2)$, in the spirit of the proof of Theorem 4.9. Could e_0 ever land there with the endpoints switched?] In fact, each coset of $\Gamma(2)$ in $PSL_2(\mathbb{Z})$ induces its own unique permutation of the three colors—a direct way of seeing the isomorphism with S_3.

From here one can then show that the union of any two adjacent faces in the Farey tessellation forms a fundamental domain for $\mathbb{H}/\Gamma(2)$, which we leave as an exercise to the reader (c.f. Example 6.7). (Note that back in the Dedekind tessellation, this fundamental domain includes 6 times as many faces as the fundamental domain for $\mathbb{H}/PSL_2(\mathbb{Z})$, one for each coset of $\Gamma(2)$. No accident.) In terms of side-identifications for these fundamental domains, they are already given to us by the

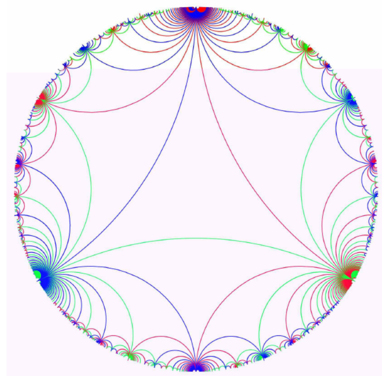

FIGURE 3. A 3-coloring of the edges of the Farey tessellation [23], based on the parity of the numerator and denominator of their endpoints. (Types of endpoints: $\frac{odd}{odd}, \frac{odd}{even}, \frac{even}{odd}$; each edge connects two different types and misses one.) The group $\Gamma(2)$ preserves this 3-coloring, and is simply transitive on the set of *unoriented* edges of any given color.

color pairings of the 4 boundary sides, and gluing these again gives a dumpling shape but this time with 3 punctures. So when compactifying $Y(\Gamma(2))$ into $X(\Gamma(2))$ (by filling in these punctures), topologically we still end up with the 2-sphere, and from the uniformization theorem this means that even as a Riemann surface, $X(\Gamma(2))$ is isomorphic to the Riemann sphere. Hence $Y(\Gamma(2))$ is isomorphic to the Riemann sphere with 3 punctures, or equivalently to $\mathbb{C}$ with 2 punctures. In fact, there is a well-known analytic map called the *modular λ-function* that goes from $\mathbb{H}$ to $\mathbb{C}$ with image $\mathbb{C} \setminus \{0, 1\}$—while being well-defined and injective on $\mathbb{H}/\Gamma(2)$—that makes this isomorphism explicit. ([15] has more details.)

This isomorphism is often used to prove the little Picard theorem, which states that the image of any nonconstant holomorphic map $\mathbb{C} \to \mathbb{C}$ misses at most one point. We won't go into complete details of the proof since it relies on the theory of covering spaces and the lifting criterion (which [5] has more details on), but for the reader familiar with some topology, the basic idea is as follows:

With Remark 6.17 in mind we see that $\mathbb{H}$ is actually a cover for $\mathbb{H}/\Gamma(2)$, due to $\Gamma(2)$ being free on $\mathbb{H}$ (see Proposition 6.23). So by the above isomorphism $\mathbb{H}$ is also a cover of the twice punctured complex plane. Consequently, if a holomorphic map defined on all of $\mathbb{C}$ misses two distinct points $z_1, z_2 \in \mathbb{C}$, then with $\mathbb{C}$ being simply connected there is no obstruction to lifting this map $\mathbb{C} \to \mathbb{C} \setminus \{z_1, z_2\}$ to a map $\mathbb{C} \to \mathbb{H}$. From here, composing with an isomorphism from $\mathbb{H}$ to the unit disk (e.g. the Möbius transformation $z \mapsto \frac{z-i}{z+i}$) gives us a map from $\mathbb{C}$ to the unit disk, which must be constant by Liouville's theorem [4]. It then follows that the map $\mathbb{C} \to \mathbb{H}$ must have been constant, and in turn that the original map $\mathbb{C} \to \mathbb{C} \setminus \{z_1, z_2\}$ was constant all along; that is, constant maps are the only holomorphic maps on $\mathbb{C}$ that can miss more than one point of $\mathbb{C}$ in their image.

Remark 8.2. An important characteristic of a compact Riemann surface is its genus, which as mentioned in Remark 2.7 characterizes it topologically (even though it by no means determines it completely if the genus is nonzero). Suppose we have found an explicit fundamental domain (and directions for side-identifications) for a modular curve, as in the case of Example 8.1. Then the triangulation from the tessellation allows us to compute its Euler characteristic

$$\chi = V - E + F$$

(the alternating sum of its vertex, edge, and face counts) by simply gluing sides, filling in cusps, and tallying components directly. Then we can apply the relation

$$\chi = 2 - 2g$$

to find the genus g of the surface (see [16]).

For a general congruence subgroup, though, it can be daunting to try to figure out a fundamental domain by hand. Still, there is a relation called the Riemann-Hurwitz formula that can be used to compute the genus of $X(\Gamma)$ by studying the projection $X(\Gamma) \to X(PSL_2(\mathbb{Z}))$ that maps $\Gamma h \to PSL_2(\mathbb{Z})h$ and ends up giving what is called a *branched cover* of the sphere; we refer the reader to [1] for an introduction to branched covers and Riemann-Hurwitz, and to [2] for further details on applying this to congruence subgroups.

It turns out that for the first 5 levels of congruence subgroups, every compactified modular curve stubbornly remains of genus 0, so is isomorphic to the Riemann sphere by the uniformization theorem, and even at level 6 we still only achieve genus 1, the same as what a quotient of $\mathbb{C}$ could yield. However, at level 7 things start to get exciting (see Figure 4 and Example 8.3 below). For more examples and data on genus, [8] lists an extensive collection of congruence subgroups organized by the genus of their associated modular curves.

Example 8.3 ($\Gamma(7)$ and Klein's quartic)**.** In 1879 Klein published a paper on the Riemann surface that is now known as Klein's quartic, which he described both analytically and algebraically, connecting the two representations [9].

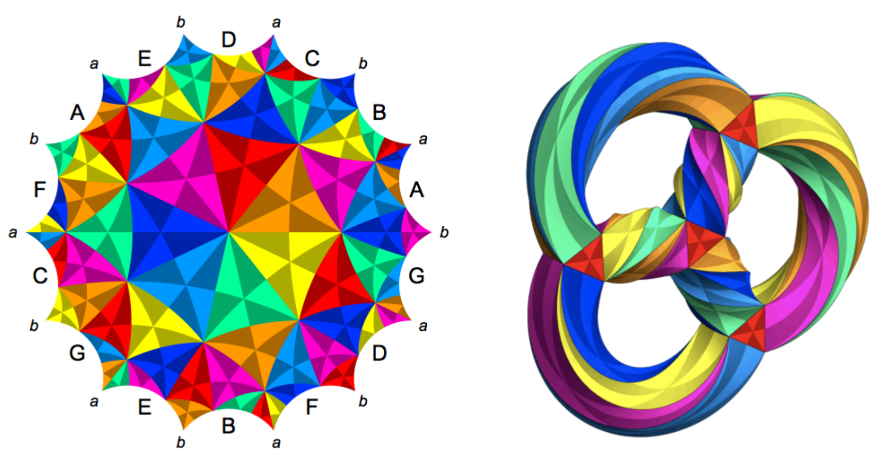

FIGURE 4. The Riemann surface of Klein's quartic can be built from quotienting $\mathbb{H}$ by $\Gamma(7)$ and compactifying. At left: the fundamental domain, where pairs of sides with the same letter are glued; at right: an embedding of the surface into $\mathbb{R}^3$ [24].

The algebraic description of Klein's quartic is where the term "quartic" comes from, since it arises as the space of solutions to the equation

$$x^3y + y^3z + z^3x = 0$$

viewed as a subset of $P^2_{\mathbb{C}}$ (meaning that x, y, z are allowed to range over $\mathbb{C}$ and ordered triples are considered the same if they differ only by overall scalar multiple).

Analytically, Klein also showed that his quartic curve can be built through quotienting, and is in fact $X(\Gamma(7))$. He even found and drew an explicit fundamental domain for the space (along with its side identifications), and Figure 4 shows a similar rendering, where Klein's original 14-sided polygon built of 24 heptagons has been subdivided further into $336 = 2 \cdot |PSL_2(\mathbb{Z}/7\mathbb{Z})|$ faces. Here the side identifications are given by the letter pairs shown along the boundary, and the faces have also been colored to call out the 56 triangles that are dual to the 24 heptagons.

One note is that this cuspless depiction of the fundamental domain applies specifically to the already-compactified space $X(\Gamma(7))$. By contrast, so far in this paper we have built fundamental domains slightly differently, starting from the noncompact space and considering faces straight from the Dedekind tessellation. It is also possible to take this approach for $Y(\Gamma(7))$, though it ends up requiring some more side-identifications to the 24 cusps involved, and can be a bit more unwieldy to draw and parse; hence the compactified version is more commonly depicted.

Either way, though, having an explicit fundamental domain to work with allows us to directly compute genus, as noted in *Remark* 8.2, and it is a nice combinatorial exercise to use the left panel of Figure 4 to verify that in this case the genus is indeed 3 [18], as the right panel suggests.

Riemann surfaces of genus higher than 1 are called *hyperelliptic curves*, and unlike the Riemann sphere (of genus 0) and elliptic curves (of genus 1), hyperelliptic curves turn out to have only *finite* automorphism groups, bounded above by the Hurwitz bound of $84(g-1)$, where g denotes genus [17]. In fact Klein's quartic has automorphism group isomorphic to $PSL_2(\mathbb{Z}/7\mathbb{Z})$, with $168 = 84(3-1)$ elements, making it the first example of what is known as a *Hurwitz surface*, a Riemann surface whose automorphism group actually attains the Hurwitz bound [18].

There is much more to say about Klein's quartic, and needless to say, a full account of its properties would be well beyond the scope of this paper. However, there are many other sources covering this topic in greater detail, and in particular, the interested reader may wish to have a look at [9], which offers an accessible, lively tour of several more properties and visualizations of the surface.

This concludes our discussion of examples.

Acknowledgments

First of all, I would like to say thank you to my mentor Michael Barz, not only for his enthusiasm and guidance in our meetings each week but also for his patient reassurance through the process of narrowing in on a topic, which I found stressful at first. Many thanks also to Peter May for allowing me to participate in the program, and for taking the time to make a personal connection, as well as for the opportunity to give a talk at the end of the program. Finally, thanks to Shanna Goldman and Annika Larson for making the REU a much more fun experience.

References

[1] Wilhelm Schlag. A Concise Course in Complex Analysis and Riemann Surfaces. American Mathematical Society, 2014.
[2] Fred Diamond & Jerry Shurman. A First Course in Modular Forms. Springer, 2005.
[3] Klaus Jänich. Vector Analysis. Springer, 2001.
[4] Joseph Bak & Donald J. Newman. Complex Analysis. Springer, 2010.
[5] Christopher Stith. The Fundamental Group and Connections to Covering Spaces. https://math.uchicago.edu/~may/REU2016/REUPapers/Stith.pdf
[6] Thomas Krämer. Lectures on Elliptic Curves. https://www.mathematik.hu-berlin.de/~kraemeth/old-stuff/elliptic-curves/Elliptic-Curves.pdf
[7] Peter Kronheimer (lecturer) and Maxim Jeffs (transcriber). Riemann Surfaces. https://people.math.harvard.edu/~jeffs/Complex_Analysis_Class_Notes.pdf
[8] Chris Cummins & Sebastian Pauli. Congruence Subgroups of $PSL_2(\mathbb{Z})$. https://mathstats.uncg.edu/sites/pauli/congruence/
[9] John Baez. Klein's Quartic Curve. https://math.ucr.edu/home/baez/klein.html
[10] https://en.wikipedia.org/wiki/Riemann_surface
[11] https://en.wikipedia.org/wiki/Uniformization_theorem
[12] https://en.wikipedia.org/wiki/Surface_(topology)
[13] https://en.wikipedia.org/wiki/Alexandroff_extension
[14] https://en.wikipedia.org/wiki/Projective_linear_group
[15] https://en.wikipedia.org/wiki/Modular_lambda_function
[16] https://en.wikipedia.org/wiki/Euler_characteristic
[17] https://en.wikipedia.org/wiki/Hurwitz%27s_automorphisms_theorem
[18] https://en.wikipedia.org/wiki/Klein_quartic

Image Credits

[19] https://cran.r-project.org/web/packages/RcppColors/readme/README.html
[20] https://www.researchgate.net/publication/2113389_Transcendental_ending_laminations/figures
[21] https://web.sas.upenn.edu/miguellopez/2021/07/11/farey/
[22] https://commons.wikimedia.org/wiki/File:The_modular_group_PSL_2%28Z%29.svg
[23] https://www.researchgate.net/publication/2208524_Wall_Crossing_Discrete_Attractor_Flow_and_Borcherds_Algebra
[24] https://www.gregegan.net/SCIENCE/KleinQuartic/KleinQuartic.html