Hyperbolic Structures on Surfaces

University of Virginia Distinguished Major Thesis

Trent Lucas
Advised by Professor Sara Maloni

April 2020

Acknowledgements

I would like to thank Sara Maloni for her unwavering dedication and support for this project, and without whom this thesis would not be possible. I would also like to thank Ian Runnels and Alec Traaseth, who helped me learn much of this material in the summer of 2019. Finally, I give thanks to Mikhail Ershov and Thomas Koberda for serving on my thesis committee.

Contents

Introduction

The main objects of study in this thesis will be *surfaces*, i.e. compact, connected, oriented 2-manifolds. One of the crowning achievements of 19th century mathematics was the realization that a surfaces's geometry is constrained by its topology. Perhaps the most classical result which formalizes this notion is the Gauss-Bonnet Theorem, which says that if S is a closed surface endowed with a Riemannian metric, then

$$\int_S K \; dA = 2\pi\chi(S),$$

where K is the Gaussian curvature and $\chi(S)$ is the Euler characteristic. This tells us that if S is the sphere, then S has positive total curvature, if S is a torus, then S has zero total curvature, and if S is any other closed surface, then S has negative total curvature. As a consequence, we have that "most" closed surfaces exhibit negative total curvature. A similar characterization holds for surfaces with boundary. Therefore, if we wish to study the geometry of constant curvature surfaces, we will most often be dealing with hyperbolic geometry.

Hyperbolic geometry is the study of spaces with constant negative curvature. In the two dimensional case, a space has constant negative curvature if it resembles a saddle at every point. In such spaces, Euclid's Fifth Postulate, known as the *Parallel Postulate*, fails to hold, meaning that given a line L and a point p not on L, there are infinitely many lines through p parallel to L. Up to isometry and scaling, there is a unique simply connected two-dimensional Riemannian manifold of constant negative curvature, which we call *the hyperbolic plane*. A theorem of Hilbert says that the hyperbolic plane cannot be isometrically embedded in $\mathbb{R}^2$. Therefore, one must visualize the hyperbolic plane through *models*, i.e. subsets of $\mathbb{R}^2$ with a metric of constant negative curvature which differs greatly from the Euclidean metric.

From the Gauss-Bonnet Theorem, we know that "most" constant curvature surfaces will locally resemble the hyperbolic plane. Our main motivation in this thesis is to understand *how* a surface can resemble the hyperbolic plane. We formalize this with the notion of a *hyperbolic structure* on a surface; given a surface S, a hyperbolic structure on S is an atlas into the hyperbolic plane where the transition maps are isometries. A remarkable fact is that a surface can be given multiple hyperbolic structures which are different, which is to say non-isometric. Therefore, we seek to understand the relationship between different hyperbolic structures on the same surface.

To this end, we will define the *Teichmüller space* of a surface S, denoted $\mathrm{Teich}(S)$. This is essentially the space of all possible *marked* hyperbolic structures on S (up to a natural equivalence, and with some restrictions). A marked hyperbolic structure is simply a surface X with a hyperbolic structure equipped with a homeomorphism $S \to X$, called a *marking*. We endow $\mathrm{Teich}(S)$ with a topology where "similar" hyperbolic structures are "close" in $\mathrm{Teich}(S)$. Therefore, understanding the relationship between different marked hyperbolic structures on S amounts to understanding the topology of $\mathrm{Teich}(S)$. We will see that $\mathrm{Teich}(S)$ can be equivalently viewed as the space of marked complex structures on S, and this viewpoint can be used to define a metric on $\mathrm{Teich}(S)$. Therefore, there is a *geometry* on $\mathrm{Teich}(S)$ which further informs the relationship between hyperbolic structures on S.

We will be particularly interested in the following question: if S is endowed with a hyperbolic structure and we modify S by some topological symmetry, what is the resulting effect on the hyperbolic structure? In the language of Teichmüller theory, we are essentially asking how the symmetry group of S acts on $\mathrm{Teich}(S)$. In order to answer this, we must first clarify what we mean by the "symmetry group" of S. We are really refering to the *mapping class group* of S, denoted $\mathrm{Mod}(S)$. This group is comprised of homotopy classes of orientation-preserving homeomorphisms of S which fix ∂S pointwise. We will formally define how

Mod(S) acts on Teich(S), but it amounts to changing the marking of a hyperbolic structure to a different homeomorphism. Then, we will prove a remarkable theorem of Fricke: the action of Mod(S) on Teich(S) is *properly discontinuous*. We will formally define this property later, but this means in particular that the orbits of this action are discrete. Therefore, this tells us that if we modify S by a non-trivial symmetry, we will get a marked hyperbolic structure which is substantially different.

To develop the theory of hyperbolic structures, we will make use of *Fuchsian groups*, which are discrete groups of isometries of the hyperbolic plane $\mathbb{H}$. This is because Fuchsian groups are symmetry groups of tessellations of $\mathbb{H}$, and it follows that the quotient of $\mathbb{H}$ by the the action of a Fuchsian group is a surface with a hyperbolic structure. In fact, we will see that every (complete) hyperbolic structure on a surface S is given by $\mathbb{H}/G$, where G is some Fuchsian group isomorphic to $\pi_1(S)$. Moreover, we will also see that the group of orientation-preserving isometries of $\mathbb{H}$ is isomorphic to the Lie group $\mathrm{PSL}(2,\mathbb{R})$. Therefore, we can characterize marked hyperbolic structures as injective homomorphisms $\pi_1(S) \to \mathrm{PSL}(2,\mathbb{R})$ with a discrete image, which we called *discrete and faithful representations* of $\pi_1(S)$ into $\mathrm{PSL}(2,\mathbb{R})$.

This allows us to view Teich(S) as living in the much broader space $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))$, which we call the $\mathrm{PSL}(2,\mathbb{R})$-*representation variety* of S. In particular, we will see that Teich(S) can be embedded into the quotient of $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))$ by the conjugation action of $\mathrm{PSL}(2,\mathbb{R})$ (actually, we will restrict our attention to a subspace of this quotient). Fricke's theorem tells us that, dynamically speaking, the action of Mod(S) on Teich(S) is "nice". However, we will see a conjecture of Goldman which says that if we extend this action to the broader representation space, it is "chaotic" elsewhere. Marché and Wolff proved Goldman's conjecture in the case that S is a closed surface of genus 2 by answering an older question of Bowditch. We will conclude the thesis by showing why an affirmative answer to Bowditch's question proves Goldman's conjecture; this connection had previously been folklore, but was first formally established by Marché and Wolff. We hope to generalize this connection between Bowditch's question and Goldman's conjecture to the case of non-orientable surfaces in a future paper.

This thesis is broadly organized as follows. In Chapter 1, we will review the basics of hyperbolic geometry, prove some results about Fuchsian groups, and formally define hyperbolic structures on surfaces. In Chapter 2, we will study mapping class groups and Teichmüller space, and we will prove Fricke's theorem. In Chapter 3, we will discuss the topological structure of the representation variety and its quotient, state Goldman's conjecture, and outline Marché and Wolff's result.

Chapter 1

Hyperbolic Geometry

Our first goal in this chapter is to explore the basics of hyperbolic geometry. In Section 1.1, we will define two models of the hyperbolic plane: the upper half plane model $\mathbb{H}$ and the disk model $\mathbb{D}$. To be acquaint ourselves with these models, we will focus on classifying the geodesics (i.e. "straight lines") in these models as well as finding their isometry groups up to isomorphism. In particular, we will see that $\mathrm{Isom}^+(\mathbb{H}) \cong \mathrm{PSL}(2, \mathbb{R})$ and $\mathrm{Isom}(\mathbb{D}) \cong \mathrm{PSU}(1,1)$. This will require us to use tools from complex analysis. In fact, our main objects of interest will be Möbius transformations, which are rational functions on the Riemann sphere $\widehat{\mathbb{C}}$ and comprise its automorphism group $\mathrm{Aut}(\widehat{\mathbb{C}})$. In Section 1.2, we will examine the actual behavior of hyperbolic isometries. We will see that elements of $\mathrm{Isom}^+(\mathbb{H})$ fall into one of three categories: hyperbolic transformations, which behave like translations; elliptic transformations, which behave like rotations; and parabolic transformations, which behave like "degenerate translations" or "degenerate rotations". We will see that the category to which an isometry belongs is determined by the isometry's trace (if we view it as an element of $\mathrm{PSL}(2, \mathbb{R})$), or equivalently by its fixed points on $\mathbb{H}$.

Our second goal in this chapter is to see how we can endow surfaces with a geometry which locally resembles the hyperbolic plane. In Section 1.3 we will discuss Fuchsian groups, which are discrete subgroups of $\mathrm{Isom}^+(\mathbb{H})$. We will show that a subgroup of $\mathrm{Isom}^+(\mathbb{H})$ is discrete if and only if it acts properly discontinuously on $\mathbb{H}$. For this reason, Fuchsian groups will arise as symmetry groups of tessellations of the hyperbolic plane. In Section 1.4, we'll define and study hyperbolic surfaces. An important class of hyperbolic surfaces are the spaces obtained by quotienting $\mathbb{H}$ be the action of a Fuchsian group. In fact, we will show that any (complete) hyperbolic surface is precisely such a quotient. This will allow us to characterize hyperbolic surfaces via $\mathrm{PSL}(2, \mathbb{R})$-representations of their fundamental group, which motivates what is to come in the next two chapters.

For this chapter, we will follow the notes of Caroline Series [12].

1.1 The Hyperbolic Plane

1.1.1 Two Models of Hyperbolic Geometry

Our primary model of the hyperbolic plane will be the upper half plane.

Definition 1.1.1. The *upper half plane model* is defined by the set $\mathbb{H} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) > 0\}$ endowed with the Riemannian metric

$$ds^2 := \frac{dx^2 + dy^2}{y^2}.$$

Given a path $\gamma : [a, b] \to \mathbb{H}$, we define the *length* of γ to be

$$\ell(\gamma) := \int_a^b ds.$$

Then, given two points P and Q in $\mathbb{H}$, we define the *distance* between them as $d_{\mathbb{H}}(P, Q) = \inf \ell(\gamma)$, where the infimum is taken over all paths γ connecting P and Q.

Recall that a *geodesic* is a path which locally minimizes distance. If we wish to understand the geometry of $\mathbb{H}$, then it is important that we understand the geodesics in this space.

Proposition 1.1.2. *In $\mathbb{H}$, vertical lines are geodesics, and if $x_0, a, b \in \mathbb{R}$ with $a < b$, then*

$$d_{\mathbb{H}}(x_0 + ia, x_0 + ib) = \log\left(\frac{b}{a}\right).$$

Proof. Let $\gamma : [c, d] \to \mathbb{H}$ be a path from $x_0 + ai$ to $x_0 + bi$, and let $\gamma(t) = (x(t), y(t))$. Then,

$$\ell(\gamma) = \int_c^d \frac{\sqrt{x'(t)^2 + y'(t)^2}}{y(t)} dt \geq \int_c^d \frac{y'(t)}{y(t)} dt = \int_a^b \frac{1}{y} dy = \log\left(\frac{b}{a}\right).$$

We have equality if and only if $\frac{dx}{dt} \equiv 0$, which occurs if and only if γ is a vertical line. This means that a vertical line not only locally minimizes distance, but globally minimizes distance. So, we can conclude that a vertical line is a geodesic and $d_{\mathbb{H}}(x_0 + ai, x_0 + bi) = \log\left(\frac{b}{a}\right)$. □

Notice that if we fix b and let $a \to 0$, then $d(x_0 + ia, x_0 + ib) \to \infty$. For this reason, we call $\widehat{\mathbb{R}} = \mathbb{R} \cup \{\infty\} = \{z \in \mathbb{C} \mid \mathrm{Im}(z) = 0\} \cup \{\infty\}$ the *boundary at infinity*, and denote it $\partial\mathbb{H}$. For instance, a *horocycle* is a Euclidean circle in $\mathbb{H}$ tangent to $\mathbb{R}$ or a horizontal straight line (i.e. a circle tangent to ∞); horocycles hence have two ends which asymptotically approach each other but never meet.

Classifying all geodesics in $\mathbb{H}$ will require a bit more work involving automorphisms of the Riemann sphere. Before that, we introduce our second model of hyperbolic geometry.

Definition 1.1.3. The *disk model* is defined by the set $\mathbb{D} = \{z \in \mathbb{C} \mid |z| < 1\}$ endowed with the Riemannian metric

$$ds^2 = \frac{2(dx^2 + dy^2)}{(1 - (x^2 + y^2))^2}.$$

We similarly define the length of paths and the distance $d_{\mathbb{D}}$ between two points in $\mathbb{D}$. If it is clear which model we are using, we will often write $d_{\mathbb{H}}$ or $d_{\mathbb{D}}$ simply as d.

Proposition 1.1.4. *In $\mathbb{D}$, radial lines are geodesics, and if $a \in [0, 1)$, then $d_{\mathbb{D}}(0, a) = \log\left(\frac{1+a}{1-a}\right)$.*

Proof. By switching to polar coordinates (r, θ), we have that $dx^2 + dy^2 = dr^2 + r^2 d\theta^2$. Let $a = r_1 e^{i\phi}$ and $b = r_2 e^{i\phi}$ be two points on the same radial line, and $\gamma = (r(t), \theta(t))$ a path defined on $[a, b]$ connecting them. Then,

$$\begin{aligned} \ell(\gamma) &= \int_a^b \frac{2\sqrt{r'(t)^2 + r(t)^2\theta'(t)^2}}{1 - r(t)^2} dt \\ &\geq \int_a^b \frac{2r'(t)}{1 - r(t)^2} dt = \int_{r_1}^{r_2} \frac{2}{1 - r^2} dr = \int_{r_1}^{r_2} \frac{1}{1 - r} + \frac{1}{1 + r} dr = \log\left(\frac{1 + r}{1 - r}\right)\Bigg|_{r_1}^{r_2}. \end{aligned}$$

This time, we see that the minimal distance is achieved if and only if $\frac{d\theta}{dt} \equiv 0$, or equivalently, if and only if γ is a radial line. The given formula for $d_{\mathbb{D}}(0, a)$ is a special case of our calculation. □

This time, we see that $d_{\mathbb{D}}(0, a) \to \infty$ as $a \to 1$, and so we call $\partial\mathbb{D}$ the *boundary at infinity.*

We will tend to use $\mathbb{H}$ as our primary model of the hyperbolic plane, as it lends itself to easier calculations. However, $\mathbb{D}$ is generally better for visualizations, and there are instances where switching to $\mathbb{D}$ offers a simpler argument. Soon, we will present an isometry $C : \mathbb{H} \to \mathbb{D}$, called the Cayley transform, which allows us to translate statements about one model to the other.

1.1.2 Möbius Transformations and Conformal Maps

There is a reason we have defined our models as subsets of $\mathbb{C}$, rather than just $\mathbb{R}^2$. In this section, we will take a small detour to explore some analytic properties of our models; we can then leverage these properties to study their geometry.

Let $\widehat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ denote the Riemann sphere.

Definition 1.1.5. Let $U, V \subseteq \widehat{\mathbb{C}}$ be open. A *conformal map* is a map $f : U \to V$ which preserves (signed) angles. We let $\mathrm{Aut}(U)$ denote the set of all conformal bijections $f : U \to U$. Note that $\mathrm{Aut}(U)$ forms a group under composition.

Recall from complex analysis that a map f is conformal if and only if it is holomorphic and has a non-vanishing derivative. We will later show that $\mathrm{Aut}(\mathbb{D})$ and $\mathrm{Aut}(\mathbb{H})$ act by isometries on $\mathbb{D}$ and $\mathbb{H}$ respectively. Therefore, our current goal is to understand these groups.

Definition 1.1.6. A *Möbius transformation* is a map $f : \widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$ of the form

$$f(z) = \frac{az+b}{cz+d}$$

where $a, b, c, d \in \mathbb{C}$ such that $ad - bc \neq 0$. In particular, $f(\infty) = \frac{a}{c}$ and $f\left(-\frac{d}{c}\right) = \infty$.

Möbius transformations are realized by the action of $\mathrm{GL}(2, \mathbb{C})$ on $\widehat{\mathbb{C}}$ defined by

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

One can check directly that this is indeed a group action. Note also that for any nonzero $\lambda \in \mathbb{C}$,

$$\begin{pmatrix} \lambda a & \lambda b \\ \lambda c & \lambda d \end{pmatrix} z = \frac{\lambda az + \lambda b}{\lambda cz + \lambda d} = \frac{az+b}{cz+d} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} z.$$

Therefore, if A is a matrix representation of a Möbius transformation, we can can always normalize it to the representation $\frac{1}{\sqrt{\det(A)}} A \in \mathrm{SL}(2, \mathbb{C})$. Note that this representation is still not unique, as multiplication by -1 will not change the transformation. Throughout the rest of this document, we will generally be identifying matrices with Möbius transformations and *not* linear maps.

Proposition 1.1.7. *Any Möbius transformation can be written as the composition of four elementary transformations:*

1. *Translations:* $z \mapsto z + a$ *for* $a \in \mathbb{C}$.
2. *Rotations:* $z \mapsto e^{i\theta} z$ *for* $\theta \in [0, 2\pi)$.
3. *Dilations:* $z \mapsto \lambda z$ *for* $\lambda \in \mathbb{R}_{>0}$.
4. *Inversion:* $z \mapsto \frac{1}{z}$.

Proof. Let f be a Möbius transformation with coefficients a, b, c, and d. By normalizing, we can assume $ad - bc = 1$. Using this, we can rewrite

$$\frac{az+b}{cz+d} = \frac{a}{c} - \frac{1}{c(cz+d)}.$$

Since a map of the form $z \mapsto wz$ for $w \in \mathbb{C}$ is a composition of transformations 2 and 3, we can see that the expression on the right hand side is indeed a composition of our elementary transformations. □

One can directly check that these elementary transformations are conformal bijections. Also, one can check that they map circles to circles, where we consider a straight line as a circle passing through ∞. Therefore, we get the following.

Corollary 1.1.8. *Möbius maps are conformal bijections, and they map circles to circles.*

We can summarize our work so far in group-theoretic terms.

Proposition 1.1.9. *The map* $F : \mathrm{SL}(2,\mathbb{C}) \to \mathrm{Aut}(\widehat{\mathbb{C}})$ *given by*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \left(z \mapsto \frac{az+b}{cz+d} \right)$$

is a group homomorphism with $\mathrm{Ker}(F) = \{\pm I\}$*, where* I *is the identity matrix.*

Proof. Again, one can check directly that F is a homomorphism. To compute the kernel, we have that

$$\frac{az+b}{cz+d} \equiv z \iff cz^2 + (d-a)z - b \equiv 0 \iff b = c = 0,\ a = d.$$

□

In fact, one can show that *every* element of $\mathrm{Aut}(\widehat{\mathbb{C}})$ is a Möbius transformation.

Theorem 1.1.10. *The map* F *in Proposition 1.1.9 is surjective.*

The proof of this fact requires some tools from complex analysis. One shows that any element of $\mathrm{Aut}(\widehat{\mathbb{C}})$ must be a meromorphic function on $\widehat{\mathbb{C}}$, and then one use Liouville's theorem to prove that meromorphic functions on $\widehat{\mathbb{C}}$ are rational functions. Finally, one observes that an injective rational function must have polynomials of degree at most 1 in the numerator and denominator.

Corollary 1.1.11. *There is an isomorphism* $\mathrm{Aut}(\widehat{\mathbb{C}}) \cong \mathrm{PSL}(2,\mathbb{C})$*, where* $\mathrm{PSL}(2,\mathbb{C}) = \mathrm{SL}(2,\mathbb{C})/\{\pm I\}$.

Now that we have identified $\mathrm{Aut}(\widehat{\mathbb{C}})$, we can identify $\mathrm{Aut}(\mathbb{D})$ and $\mathrm{Aut}(\mathbb{H})$ as subgroups. The following characterization of $\mathrm{Aut}(\mathbb{D})$ is a result of complex analysis obtained by the Schwarz lemma.

Theorem 1.1.12. *The group* $\mathrm{Aut}(\mathbb{D})$ *is a subgroup of* $\mathrm{Aut}(\widehat{\mathbb{C}})$ *consisting of Möbius transformations of the form*

$$z \mapsto \frac{az+b}{\bar{b}z+\bar{a}}$$

where $a, b \in \mathbb{C}$ *with* $|a|^2 - |b|^2 = 1$*. This implies that* $\mathrm{Aut}(\mathbb{D}) \cong \mathrm{PSU}(1,1) = \mathrm{SU}(1,1)/\{\pm I\}$.

Now, we introduce an important map $\mathbb{H} \to \mathbb{D}$.

Lemma 1.1.13. *Define the map* $C : \mathbb{H} \to \mathbb{D}$ *by*

$$C(z) = \frac{z-i}{z+i}.$$

Then, C *is a conformal bijection. We call* C *the* Cayley transformation.

Proof. Since C is a Möbius transformation, we know that it is a conformal bijection $\widehat{\mathbb{C}} \to \widehat{\mathbb{C}}$. Since $C(0) = -1$, $C(1) = -i$, and $C(\infty) = 1$, and C must map circles to circles, we know that C maps $\widehat{\mathbb{R}}$ to $\partial\mathbb{D}$. Then, we know that C will map $\mathbb{H}$ to one of the connected components of $\widehat{\mathbb{C}} \backslash \partial\mathbb{D}$. Since $\mathbb{C}(i) = 0$, we can conclude that $C(\mathbb{H}) = \mathbb{D}$. □

By combining Theorem 1.1.12 with Lemma 1.1.13, one gets the following.

Theorem 1.1.14. *The group* $\mathrm{Aut}(\mathbb{H})$ *is a subgroup of* $\mathrm{Aut}(\widehat{\mathbb{C}})$ *consisting of Möbius transformations of the form*

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

where $a, b, c, d \in \mathbb{R}$ *such that* $ad - bc > 0$*. This implies that* $\mathrm{Aut}(\mathbb{H}) \cong \mathrm{PSL}(2,\mathbb{R})$*, where* $\mathrm{PSL}(2,\mathbb{R}) = \mathrm{SL}(2,\mathbb{R})/\{\pm I\}$.

Proof. First, one can check that any map f of the given form is a bijection $\mathbb{H} \to \mathbb{H}$. Indeed, we have that

$$\operatorname{Im}(f(z)) = \operatorname{Im}\left(\frac{(az+b)(c\overline{z}+d)}{|cz+d|^2}\right) = \operatorname{Im}\left(\frac{ac|z|^2 + adz + bc\overline{z} + bd}{|cz+d|^2}\right) = \frac{\operatorname{Im}(z)(ad-bc)}{|cz+d|^2} > 0.$$

Similarly, one can show that $\operatorname{Im}(f^{-1}(z)) > 0$.

Now, take any $F \in \operatorname{Aut}(\mathbb{H})$. Define $h \in \operatorname{Aut}(\mathbb{H})$ by

$$h(z) = \frac{z - \operatorname{Re}(F(i))}{\operatorname{Im}(F(i))}.$$

We know that $h \in \operatorname{Aut}(\mathbb{H})$ since it is of the form described in the theorem. It follows that $h \circ F \in \operatorname{Aut}(\mathbb{H})$ and $(h \circ F)(i) = i$. Now, define $g \in \operatorname{Aut}(\mathbb{H})$ by

$$g(z) = \frac{\cos(\theta)z + \sin(\theta)}{-\sin(\theta)z + \cos(\theta)}$$

where θ is chosen so that if we define $T = g \circ h \circ F$, then $T(i) = i$ and $T'(i) = 1$. Let C denote the Cayley transformation. It follows that $\hat{T} := CTC^{-1} \in \operatorname{Aut}(\mathbb{D})$ with $\hat{T}(0) = 0$ and $\hat{T}'(0) = 1$. Theorem 1.1.12 tells us that $\hat{T} = id$ and hence $T = id$. This means that $F = h^{-1} \circ g^{-1}$, which one can check is of the form described in the theorem. $\square$

1.1.3 Geodesics and Isometries

Now that we understand $\operatorname{Aut}(\mathbb{H})$ and $\operatorname{Aut}(\mathbb{D})$, we can return to studying the geometry of $\mathbb{H}$ and $\mathbb{D}$.

Proposition 1.1.15. *The groups* $\operatorname{Aut}(\mathbb{H})$ *and* $\operatorname{Aut}(\mathbb{D})$ *act on* $\mathbb{H}$ *and* $\mathbb{D}$ *respectively by isometries.*

Proof. Let $T \in \operatorname{Aut}(\mathbb{H})$. Then, there exists $a, b, c, d \in \mathbb{R}$ with $ad - bc > 0$ such that $T(z) = \frac{az+b}{cz+d}$. Recall from the proof of Theorem 1.1.14 that

$$\operatorname{Im}(T(z)) = \frac{\operatorname{Im}(z)(ad-bc)}{|cz+d|^2}.$$

If we let $w = T(z)$, we can compute that $dw = \frac{dz}{(cz+d)^2}$. Then, for any path γ in $\mathbb{H}$, the change of variables theorem tells us that

$$\ell(T(\gamma)) = \int_{T(\gamma)} \frac{1}{\operatorname{Im}(w)}|dw| = \int_\gamma \frac{|cz+d|^2}{\operatorname{Im}(z)} \frac{|dz|}{|cz+d|^2} = \int_\gamma \frac{1}{\operatorname{Im}(z)}|dz| = \ell(\gamma).$$

A similar computation gives the result for $\mathbb{D}$. $\square$

As mentioned before, the Cayley transformation gives us a way to translate statements about $\mathbb{H}$ into statements about $\mathbb{D}$. We formalize this with the following result.

Proposition 1.1.16. *The Cayley transformation* $C : z \mapsto \frac{z-i}{z+i}$ *is an isometry* $\mathbb{H} \to \mathbb{D}$.

Proof. The inverse map of C is given by $C^{-1}(w) = \frac{i(w+1)}{1-w}$. If we write $z = C^{-1}(w)$, then

$$\frac{dz}{dw} = \frac{2i}{(1-w)^2}$$

and

$$\operatorname{Im}(z) = \operatorname{Im}\left(\frac{i(w+1)}{(1-w)}\right) = \operatorname{Im}\left(\frac{i(w+i)(1-\overline{w})}{|1-w|^2}\right) = \frac{1-|w|^2}{|1-w^2|}.$$

Then,

$$\frac{|dz|}{\operatorname{Im}(z)} = \frac{2|dw|}{1-|w|^2}.$$

So, if γ is a path in $\mathbb{D}$, then by changing variables we see that

$$\ell(C^{-1}(\gamma)) = \int_{T(\gamma)} \frac{|dz|}{\mathrm{Im}(z)} = \int_{\gamma} \frac{2|dw|}{1-|w|^2} = \ell(\gamma).$$

□

Now that we understand some hyperbolic isometries, we can classify all the geodesics in our models.

Proposition 1.1.17. *Let $P, Q \in \mathbb{H}$. If P and Q have the same real part, then the vertical segment between them is the unique geodesic connecting them. Otherwise, let C be the circle centered on $\partial\mathbb{H}$ containing both P and Q. Then, the arc on $C \cap \mathbb{H}$ between P and Q is the unique geodesic connecting them.*

Proof. First, suppose $\mathrm{Re}(P) = \mathrm{Re}(Q)$. We showed in Proposition 1.1.2 that the vertical segment between them is a geodesic. In fact, our argument shows that it is the unique geodesic.

Otherwise, let $\eta, \xi \in \mathbb{R}$ with $\eta < \xi$ be the points on $C \cap \mathbb{R}$. Then, let $T \in \mathrm{Aut}(\mathbb{H})$ be the map

$$T(z) = \frac{z-\xi}{z-\eta}.$$

Since $T(\xi) = 0$ and $T(\eta) = \infty$, and Möbius transformations map circles to circles, we can conclude that T maps C to the imaginary axis. If we let A denote the arc on C from P to Q, then $T(A)$ will be the unique geodesic connecting $T(P)$ and $T(Q)$. Since $\mathrm{Aut}(\mathbb{H})$ acts by isometries, we can conclude that A must be the unique geodesic from P to Q. □

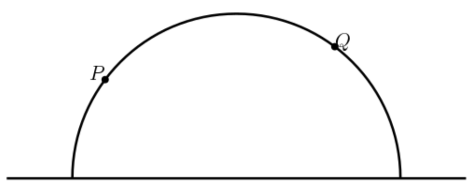

Figure 1.1: The geodesic in $\mathbb{H}$ containing the points P and Q.

Now, one can either use a similar argument or the Cayley transformation to get the following.

Proposition 1.1.18. *Let $P, Q \in \mathbb{D}$. If P and Q lie on a diameter of $\mathbb{D}$, then the segment connecting them is the unique geodesic from P to Q. Otherwise, the arc of the circle orthogonal to S^1 containing P and Q is the unique geodesic connecting them.*

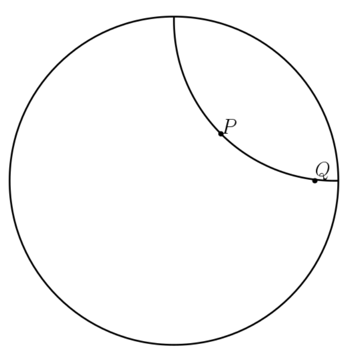

Figure 1.2: The geodesic in $\mathbb{D}$ containing the points P and Q.

As we can see, understanding the isometries of $\mathbb{H}$ and $\mathbb{D}$ is key to understanding their geometries. Our last goal is to understand their full isometry groups. We use the notation Isom and Isom^+ to denote their groups of isometries and orientation-preserving isometries respectively.

Lemma 1.1.19. *The group* $\mathrm{Aut}(\mathbb{H})$ *acts transitively on equidistant pairs of points in* $\mathbb{H}$*. The same is true for* $\mathrm{Aut}(\mathbb{D})$ *and points in* $\mathbb{D}$*.*

Proof. Let $P, P', Q, Q' \in \mathbb{H}$ such that $d(P, P') = d(Q, Q')$. We will show that there exists $T \in \mathrm{Aut}(\mathbb{H})$ such that $T(P) = Q$ and $T(P') = Q'$. It suffices to assume that $Q = i$ and $Q' = ie^{d(P,P')}$.

If $\mathrm{Re}(P) \neq \mathrm{Re}(P')$, then let C be the semi-circle centered on $\mathbb{R}$ containing them, and let $S_1 \in \mathrm{Aut}(\mathbb{H})$ be the transformation from Proposition 1.1.17 which maps C to the imaginary axis. Otherwise, let S_1 be the map $z \mapsto z - \mathrm{Re}(P)$. Now, let $a = |S_1(P)|$, and let S_2 be the map $z \mapsto z/a$, so $S_2S_1(P) = i$. Then, $d(i, S_2S_1(P')) = d(S_2S_1(P), S_2S_1(P')) = d(P, P')$, so $S_2S_1(P') = ie^{\pm d(P,P')}$. If $S_2S_1(P') = ie^{d(P,P')}$, we can take $T = S_2S_1$. Otherwise, we can let S_3 be the map $z \mapsto -1/z$, and let $T = S_3S_2S_1$.

To prove this for $\mathrm{Aut}(\mathbb{D})$, one can adapt the above argument or use the Cayley transformation. □

Lemma 1.1.20. *Circles in* $\mathbb{H}$ *and* $\mathbb{D}$ *are also Euclidean circles (possibly with different centers).*

Proof. Notice that the metric on $\mathbb{D}$ is invariant under rotation about 0. Therefore, hyperbolic circles in $\mathbb{D}$ centered at 0 are also Euclidean circles centered at 0. Then, since $\mathrm{Aut}(\mathbb{D})$ acts transitively on $\mathbb{D}$ and acts by isometries, it acts transivitely on the set of hyperbolic circles of a fixed radius. Since elements of $\mathrm{Aut}(\mathbb{D})$ are Möbius transformations, they carry Euclidean circles to Euclidean circles. Hence every hyperbolic circle in $\mathbb{D}$ is a Euclidean circle. To get the result for $\mathbb{H}$, apply the Cayley transformation. □

Theorem 1.1.21.

(i) $\mathrm{Isom}^+(\mathbb{H}) = \mathrm{Aut}(\mathbb{H}) \cong \mathrm{PSL}(2, \mathbb{R})$.

(ii) $\mathrm{Isom}(\mathbb{D}) = \mathrm{Aut}(\mathbb{D}) \cong \mathrm{PSU}(1, 1)$.

Proof. To prove (i), it remains only to show that $\mathrm{Isom}^+(\mathbb{H}) \subseteq \mathrm{Aut}(\mathbb{H})$ (since we know that $\mathrm{Aut}(\mathbb{H})$ acts by isometries, and conformal maps are orientation-preserving). Let $T \in \mathrm{Isom}^+(\mathbb{H})$. By Lemma 1.1.19, choose $S \in \mathrm{Aut}(\mathbb{H})$ so that $T' := ST \in \mathrm{Isom}^+(\mathbb{H})$ fixes two points P and P' on the imaginary axis. Now, take any $Q \in \mathbb{H}$ which doesn't lie on the imaginary axis. Since $d(P, Q) = d(P, T'(Q))$, we know that $T'(Q)$ lies on the hyperbolic circle centered at P with radius $d(P, Q)$. Similarly, $T'(Q)$ lies on the hyperbolic circle centered at P' with radius $d(P', Q)$. These are also Euclidean circles, and since one can check that $z \mapsto -\overline{z}$ is an isometry, they are symmetric about the imaginary axis. Therefore, these circles intersect at two points on opposite sides of the imaginary axis. Therefore, either $T'(Q) = Q$ or $T'(Q)$ lies on the opposite side of the imaginary axis. However, the latter would contradict that T' is orientation-preserving, so it must be that T' fixes Q. In this next section, we will see that a non-trivial Möbius transformation has at most two fixed points (this is essentially because the equation $\frac{az+b}{cz+d} = z$ is a quadratic in z), and hence $T' = id$. Therefore, $T = S^{-1}$, which means that $T \in \mathrm{Aut}(\mathbb{H})$.

To prove (ii), one can either do a similar argument or use the Cayley transformation.

□

So far, we have only found orientation-preserving isometries of $\mathbb{H}$. It we wish to find the full group $\mathrm{Isom}(\mathbb{H})$, then it suffices to find a single orientation-reversing isometry $S \in \mathrm{Isom}(\mathbb{H})$; this is because any other orientation-reversing isometry is of the form $T \circ S$ for some $T \in \mathrm{Isom}^+(\mathbb{H})$. We do not have to look too hard to find such a map S; we can take the map $S(z) = -\overline{z}$. This is simply a reflection across the imaginary axis, and it is straightforward to check that S is indeed an orientation-reversing isometry. Then, we get the following.

Corollary 1.1.22. *Any orientation-reversing* $A \in \mathrm{Isom}(\mathbb{H})$ *is of the form*

$$A(z) = \frac{a\overline{z} + b}{c\overline{z} + d}$$

where $a, b, c, d \in \mathbb{R}$ *and* $ad - bc = -1$*.*

Proof. We can write $A = T \circ S$, where $T \in \mathrm{Isom}^+(\mathbb{H})$ and $S(z) = -\overline{z}$. Then, T is a Möbius transformation, say with coefficients $a, b, c, d \in \mathbb{R}$ such that $ad - bc = 1$. This tells us that

$$A(z) = \frac{aS(z) + b}{cS(z) + d} = \frac{-a\overline{z} + b}{-c\overline{z} + d}.$$

Since $ad - bc = 1$, it follows that $(-a)d - b(-c) = -(ad - bc) = -1$. □

Now, we can see how this corresponds to the identification $\mathrm{Isom}^+(\mathbb{H}) \cong \mathrm{PSL}(2,\mathbb{R})$. The following result is straightforward to check in light of the work we have done thus far.

Corollary 1.1.23. *The map* $G : \mathrm{GL}(2,\mathbb{R}) \to \mathrm{Isom}(\mathbb{H})$ *given by*

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \mapsto \begin{cases} \left(z \mapsto \frac{az+b}{cz+d}\right) & ad - bc > 0 \\ \left(z \mapsto \frac{a\overline{z}+b}{c\overline{z}+d}\right) & ad - bc < 0 \end{cases}$$

is a surjective homomorphism, and $\mathrm{Ker}(G)$ *is the set of scalar matrices. In particular,* $\mathrm{Isom}(\mathbb{H}) \cong \mathrm{PGL}(2,\mathbb{R})$ *where* $\mathrm{PGL}(2,\mathbb{R}) = \mathrm{GL}(2,\mathbb{R})/\{\lambda I \mid \lambda \in \mathbb{R}\backslash\{0\}\}$.

1.2 Classifying Hyperbolic Isometries

1.2.1 Trace Classification of $\mathrm{SL}(2,\mathbb{C})$

In the last section, we saw that $\mathrm{Isom}^+(\mathbb{H}) = \mathrm{Aut}(\mathbb{H}) \cong \mathrm{PSL}(2,\mathbb{R})$. Now, we wish to study the behavior of different hyperbolic isometries. As before, we start by studying elements of $\mathrm{Aut}(\widehat{\mathbb{C}}) \cong \mathrm{PSL}(2,\mathbb{C})$. To avoid issues that arise from working in quotient groups, we will first study elements of $\mathrm{SL}(2,\mathbb{C})$.

Lemma 1.2.1. *Let $T = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{C})$ such that $T \neq I$. Then, T has 1 or 2 fixed points in $\widehat{\mathbb{C}}$, and T has 1 fixed point if and only if $\mathrm{tr}(T)^2 = 4$.*

Proof. We have that

$$T(z) = z \iff \frac{az+b}{cz+d} = z \iff cz^2 + (d-a)z - b = 0 \iff z = \frac{a-d \pm \sqrt{(d-a)^2+4bc}}{2c}.$$

Then,

$$(d-a)^2 + 4bc = (a+d)^2 - 4ad + 4bc = \mathrm{tr}(T)^2 - 4(ad-bc) = \mathrm{tr}(T)^2 - 4.$$

$\square$

Now, we can study the behavior of elements of $\mathrm{SL}(2,\mathbb{C})$ with particularly nice fixed points. Fix an element $T \in \mathrm{SL}(2,\mathbb{C})$ with $T \neq I$.

First, suppose that T has 1 fixed point z_0. In this case, we call T *parabolic.* Note that since the trace of an element of $\mathrm{SL}(2,\mathbb{C})$ is invariant under conjugation, its number of fixed point is as well. Let $S \in \mathrm{SL}(2,\mathbb{C})$ be the transformation $z \mapsto 1/(z-z_0)$, and let $\widehat{T} = STS^{-1}$. Then, $\widehat{T}$ fixes ∞. Let $\widehat{T} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$. Then, $\widehat{T}(\infty) = \infty$ implies that $c = 0$, and since $\widehat{T} \in \mathrm{SL}(2,\mathbb{C})$, this implies that $d = a^{-1}$. So, we have that $\widehat{T} = \left(\begin{smallmatrix} a & b \\ 0 & a^{-1} \end{smallmatrix}\right)$. Since $a + a^{-1} = \mathrm{tr}(\widehat{T}) = \pm 2$, this tells us that $a = \pm 1$. Therefore, $\widehat{T}(z) = z + \beta$ for some $\beta \in \mathbb{C}$, meaning that $\widehat{T}$ is a translation.

Otherwise, suppose that T has two fixed points $z_0^{\pm}$. This time, let $\widehat{T} = STS^{-1}$ where $S \in \mathrm{SL}(2,\mathbb{C})$ is the map

$$z \mapsto \frac{z - z_0^+}{z - z_0^-}$$

So, the fixed points of $\widehat{T}$ are 0 and ∞. Since $\widehat{T}(\infty) = \infty$, we can again deduce that $\widehat{T}$ is of the form $\left(\begin{smallmatrix} a & b \\ 0 & a^{-1} \end{smallmatrix}\right)$. Moreover, the fact that $\widehat{T}(0) = 0$ implies that $b = 0$, and since $\mathrm{tr}(\widehat{T}) = \mathrm{tr}(T) \neq \pm 2$, we know that $a \neq \pm 1$. Therefore, $\widehat{T}$ is a map of the form $z \mapsto \lambda z$ for $\lambda \in \mathbb{C}$ such that $\lambda \neq 0, 1$. There are three possibilities:

- We call T *hyperbolic* if $\lambda \in \mathbb{R}_{>0}$. In this case, $\widehat{T}$ is a dilation.
- We call T *elliptic* if $|\lambda| = 1$. In this case, $\widehat{T}$ is a rotation.
- Otherwise, we call T *loxodromic.* In this case, $\widehat{T}$ is a simultaneous rotation and dilation.

Theorem 1.2.2. *Let $T \in \mathrm{SL}(2,\mathbb{C})$ such that $T \neq I$.*

(i) T is parabolic if and only if $\mathrm{tr}(T)^2 = 4$.

(ii) T is hyperbolic if and only if $\mathrm{tr}(T) \in \mathbb{R}\backslash[-2,2]$.

(iii) T is elliptic if and only if $\mathrm{tr}(T) \in (-2,2)$.

(iv) T is loxodromic if and only if $\mathrm{tr}(T) \notin \mathbb{R}$.

Proof. We have already proved (i), so suppose T is not parabolic. Let $\widehat{T}$ be its conjugate which fixes 0 and ∞ as before. Then, we know that $\widehat{T}(z) = \lambda z$ for some $\lambda \in \mathbb{C}$ with $\lambda \neq 0, 1$, and hence

$$\widehat{T} = \begin{pmatrix} \sqrt{\lambda} & 0 \\ 0 & \sqrt{\lambda^{-1}} \end{pmatrix}.$$

If we let $\ell = \log(\lambda) = \ln|\lambda| + i\mathrm{Arg}(\lambda)$ (with $\mathrm{Arg}(\lambda) \in [0, 2\pi)$), then $\mathrm{tr}(T) = \mathrm{tr}(\widehat{T}) = e^{\ell/2} + e^{-\ell/2} = 2\cosh(\ell/2)$. Then,

$$\lambda \in \mathbb{R}_{>0} \iff \mathrm{Im}(\ell) = 0 \iff \mathrm{tr}(T) = \pm 2\cosh(x) \text{ for } x \in \mathbb{R}\backslash\{0\} \iff \mathrm{tr}(T) \in \mathbb{R}\backslash[-2, 2].$$

Also,

$$|\lambda| = 1 \iff \mathrm{Re}(\ell) = 0 \iff \mathrm{tr}(T) = 2\cos\left(\frac{\mathrm{Im}(\ell)}{2}\right) \iff \mathrm{tr}(T) \in (-2, 2).$$

This proves (ii) and (iii), and (iv) follows automatically. □

In $\mathrm{PSL}(2,\mathbb{C})$ and $\mathrm{PSL}(2,\mathbb{R})$, the ordinary trace is not well-defined, and so we have to either speak of the square of the trace or the trace up to multiplication by -1. However, we can see that conditions (i)-(iv) are invariant under multiplication by -1, and so this classification descends to our quotient groups.

1.2.2 Behavior of Hyperbolic Isometries

We can use our classification of elements of $\mathrm{Aut}(\widehat{\mathbb{C}})$ to study the different types of isometries of $\mathbb{H}$ and $\mathbb{D}$.

If $T \in \mathrm{SL}(2,\mathbb{C})$ represents an element of $\mathrm{Aut}(\mathbb{H})$, then we know that $T \in \mathrm{SL}(2,\mathbb{R})$ and hence $\mathrm{tr}(T) \in \mathbb{R}$. If $T \in \mathrm{SL}(2,\mathbb{C})$ represents an element of $\mathrm{Aut}(\mathbb{D})$, then the Cayley transformation tells us that T is conjugate to an element of $\mathrm{SL}(2,\mathbb{R})$, and since trace is conjugation invariant, $\mathrm{tr}(T) \in \mathbb{R}$. Therefore, elements of $\mathrm{Aut}(\mathbb{H})$ and $\mathrm{Aut}(\mathbb{D})$ fall into three categories: parabolic, hyperbolic, and elliptic. Since we can always move from $\mathbb{H}$ to $\mathbb{D}$ via the Cayley transformation, we will restrict our attention to elements of $\mathrm{Aut}(\mathbb{H})$.

Let $T = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}(2,\mathbb{R})$ represent an element of $\mathrm{Aut}(\mathbb{H})$. Then, the fixed point(s) of T are given by the equation

$$z_0^{\pm} = \frac{a - d \pm \sqrt{\mathrm{tr}(T)^2 - 4}}{2c}.$$

If T is parabolic, then T has a single fixed point z_0, and it will belong to $\widehat{\mathbb{R}}$. Recall from earlier in this section that if we conjugate T to a transformation $\widehat{T}$ which fixes ∞, then $\widehat{T}(z) = z + \alpha$ for some $\alpha \in \mathbb{R}$. This is a horizontal translation which moves vertical geodesics to other vertical geodesics and preserves horizontal Euclidean lines, which are horocycles tangent to ∞. Conjugating back to T, we see that T preserves horocycles tangent to z_0 and moves geodesics intersecting at z_0 to one another. We think of parabolic transformations as translations with one fixed point at infinity.

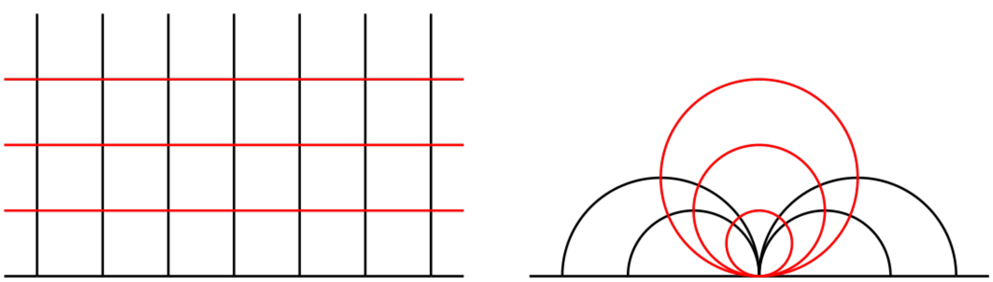

Figure 1.3: On the left is a parabolic transformation with a fixed point at ∞, and on the right is a parabolic transformation with a fixed point on $\mathbb{R}$. The black geodesics are mapped to one another, and the red curves are mapped to themselves.

If T is hyperbolic, then T has two fixed points $z_0^{\pm}$ on $\widehat{\mathbb{R}}$. If we conjugate T to a transformation $\widehat{T}$ which fixes 0 and ∞, then $\widehat{T}(z) = \lambda z$ for some $\lambda \in \mathbb{R}_{>0}$. This preserves Euclidean lines through the origin and maps non-vertical geodesics centered at 0 to each other. If we conjugate back to T, we see that T preserves arcs from z_0^- to z_0^+ and maps non-vertical geodesics centered at $z_0^{\pm}$ to one another. We think of hyperbolic transformations as translations with two fixed points at infinity. In terms of dynamics, one fixed point will be a source and the other will be a sink. To determine which is which, one can either compute the image of a non-fixed point or compute the derivative of T at the fixed points.

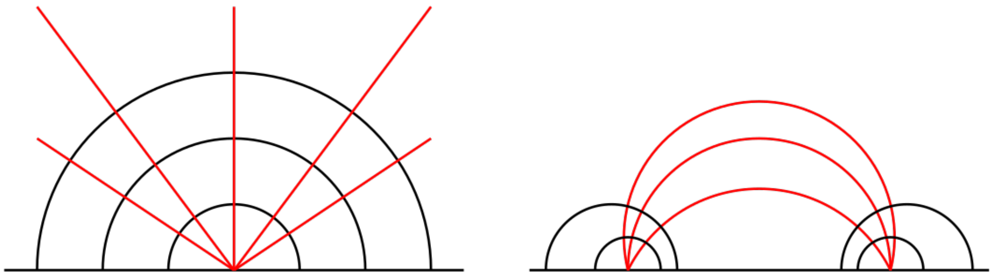

Figure 1.4: On the left is a hyperbolic transformation with a fixed points at 0 and ∞, and on the right is a hyperbolic transformation with a fixed points on $\mathbb{R}$. The black geodesics are mapped to one another, and the red curves are mapped to themselves.

We can see that a hyperbolic transformation preserves the unique geodesic connecting its fixed points. This geodesic is called the *axis* of T, denoted $\mathrm{Ax}(T)$. In fact, all points on $\mathrm{Ax}(T)$ move the same distance under T. To prove this, we can assume without loss of generality that the fixed points of T are 0 and ∞. Then, $T(z) = \lambda z$ for some $\lambda \in \mathbb{R}$, and $\mathrm{Ax}(T)$ is the imaginary axis. Then for any $t \in \mathbb{R}_{>0}$,

$$d(it, T(it)) = d(it, i\lambda t) = \log(\lambda).$$

This shows that the distance moved by points on $\mathrm{Ax}(T)$ is in fact related to $\mathrm{tr}(T)$, and we call λ the *translation length* of T. Moreover, one can show that $\log(\lambda)$ is the smallest distance traveled by any point under T. This is in contrast to parabolic elements, where there is no positive lower bound on the distance a point travels.

If T is elliptic, then T has one fixed point z_0 in $\mathbb{H}$; the other fixed point will be its complex conjugate in the lower half plane. Since T fixes z_0 and preserves distances, it must preserve hyperbolic circles centered at z_0. Moreover, it will map geodesics containing z_0 into one another. We think of elliptic elements as rotations.

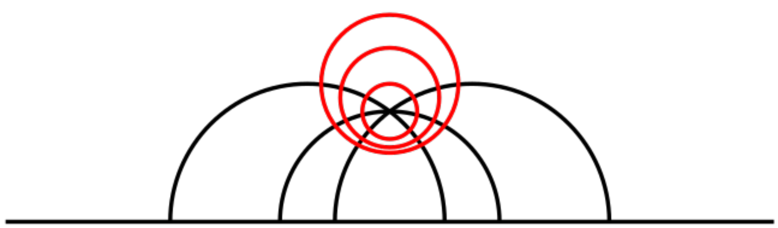

Figure 1.5: An elliptic transformation. The black geodesics are mapped to one another, and the red curves are mapped to themselves.

We can in fact find the angle of rotation of T (i.e. the angle between a geodesic containing z_0 and its image) just by $\mathrm{tr}(T)$. Since $\mathrm{SL}(2, \mathbb{R})$ acts transitively on $\mathbb{H}$, we can conjugate T by an element of $\mathrm{SL}(2, \mathbb{R})$ to assume that T has fixed points $\pm i$. Then, $T(i) = i$ implies that $i = \frac{ai+b}{ci+d}$, and hence $a = d$, $b = -c$, and $a^2 + b^2 = 1$. This implies that

$$T = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

for some $\theta \in [0, 2\pi)$. We can then compute that $T'(i) = e^{2i\theta}$. Therefore, if T has rotation angle ψ, then $\mathrm{tr}(T) = 2\cos(\psi/2)$.

1.3 Fuchsian Groups

1.3.1 Discreteness and Proper Discontinuity

Our goal in this section is to study discrete groups of hyperbolic isometries. These groups are of interest as they are symmetry groups of tessellations of the hyperbolic plane. Later, we will see that we can quotient the hyperbolic plane by the action of these groups to get hyperbolic surfaces, and consequently they will have discrete isometry groups as their fundamental groups.

However, we will first need to determine what "discrete" actually means in this context. Recall that a *topological group* is a group equiped with a topology such that the multiplication and inversion maps are continuous (and hence homeomorphisms). We first give $\mathrm{SL}(2,\mathbb{R})$ the subspace topology from $\mathbb{R}^4$ by identifying with the set

$$X = \{(a,b,c,d) \in \mathbb{R}^4 \mid ad - bc = 1\}.$$

Then, the homeomorphism $\delta : X \to X$ given by $(a,b,c,d) \mapsto (-a,-b,-c,-d)$ generates an order two group acting on X, and we give $\mathrm{PSL}(2,\mathbb{R})$ the quotient topology by identifying it with $X/\langle\delta\rangle$. The multiplication and inversion maps in $\mathrm{SL}(2,\mathbb{R})$ and $\mathrm{PSL}(2,\mathbb{R})$ are polynomials and hence continuous, so these are both topological groups.

In this section, we are concerned with two important properties of subgroups of $\mathrm{PSL}(2,\mathbb{R})$.

Definition 1.3.1. A subgroup $G \subseteq \mathrm{PSL}(2,\mathbb{R})$ is called *discrete* if it has no accumulation points in $\mathrm{PSL}(2,\mathbb{R})$.

Definition 1.3.2. A group G acts *properly discontinuously* on $\mathbb{H}$ if for any compact subset $K \subseteq \mathbb{H}$, the set $\{g \in G \mid gK \cap K \neq \varnothing\}$ is finite.

Our goal is to show that for subgroups of $\mathrm{PSL}(2,\mathbb{R})$, these two properties are equivalent. First, we develop some equivalent characterizations of discreteness.

Lemma 1.3.3. *Let $G \subseteq \mathrm{PSL}(2,\mathbb{R})$ be a subgroup. The following are equivalent:*

(i) G has no accumulation points in itself.

(ii) G has no accumulation points in $\mathrm{PSL}(2,\mathbb{R})$.

(iii) The identity map id is an isolated point in G.

Proof. (i) $\implies$ (ii): Suppose that (ii) fails. Then, there exists a sequence of distinct points $\{g_n\}$ in G such that $g_n \to h \in \mathrm{PSL}(2,\mathbb{R})$. It follows that $g_{n+1}^{-1} \to h^{-1}$ and therefore $g_n g_{n+1}^{-1} \to hh^{-1} = id$ where each $g_n g_{n+1}^{-1}$ is distinct. Since G is a subgroup, we have that $id \in G$, and hence (i) fails.

(ii) $\implies$ (iii): If (iii) fails, then (ii) immediately fails.

(iii) $\implies$ (i): Suppose (iii) holds. Let $U \subseteq \mathrm{PSL}(2,\mathbb{R})$ be a neighborhood of id such that $G \cap U = \{id\}$. Take any $g \in G$. Then, since multiplication by g is a homeomorphism of G, it follows that $g(G \cap U) = G \cap gU = \{g\}$ is a neighborhood of g in G. Since this holds for any $g \in G$, we can conclude that (i) holds. □

Note that in general topological spaces, (ii) is a strictly stronger statement than (i). However, we can see from the proof that they are equivalent for any topological group.

Next, we provide some equivalent characterizations of properly discontinuous actions.

Lemma 1.3.4. *Let $G \subseteq \mathrm{PSL}(2,\mathbb{R})$ be a subgroup. The following are equivalent:*

(i) G does not act properly discontinuously on $\mathbb{H}$.

(ii) Some G-orbit in $\mathbb{H}$ has an accumulation point.

(iii) Every G-orbit in $\mathbb{H}$ has an accumulation point, except possibly one which is a single point fixed by every element of G.

Proof. (i) $\implies$ (ii): Suppose that (i) holds. So, there exists a compact subset $K \subseteq \mathbb{H}$ such that $gK \cap K \neq \varnothing$ for infinitely-many distinct $g_n \in G$. This means that there exist points $z_n \in K$ such that $g_n z_n \in K$ for all $n \in \mathbb{N}$. Since K is compact, we assume without loss of generality that $z_n \to w \in K$. Since K is compact, there exists $R > 0$ such that $K \subseteq \overline{B_R(w)}$. For sufficiently large n, we have that $d(z_n, w) < 1$, and hence $d(g_n z_n, g_n w) < 1$. Since $g_n z_n \in K$, it follows that $g_n w \in \overline{B_{R+1}(w)}$ for sufficiently large n.

So, if infinitely-many $g_n w$ are distinct, then the compactness of $\overline{B_{R+1}(w)}$ will give us a convergent subsequence, and in particular this means that the G-orbit of w has an accumulation point. Otherwise, suppose that $g_n w = g_m w$ for infinitely-many $n \neq m$. Since $w \notin \partial\mathbb{H}$, it follows that we have an infinite family $g_m^{-1} g_n$ of elliptic transformations with fixed point w. So, if we take any other point $w' \in \mathbb{H}$, its orbit will contain infinitely-many points on a circle, which will give us a convergent subsequence and hence an accumulation point.

(ii) $\implies$ (iii): Suppose that (ii) holds, so there exists $z_0 \in \mathbb{H}$ which has an accumulation point w_0 in its G-orbit. In particular, this means there exists infinitely-many distinct $g_n \in G$ such that $g_n z_0 \to w_0$. Now, take any $z \in \mathbb{H}$. We have that

$$d(g_n z, z) \leq d(g_n z, g_n z_0) + d(g_n z_0, w_0) + d(w_0, z) = d(z, z_0) + d(g_n z_0, w_0) + d(w_0, z).$$

Since $g_n z_0 \to w_0$, we have in particular that $d(g_n z, z)$ is bounded independently of n, and hence the set of all $g_n z$ is contained in some closed ball. If $g_n z$ is distinct for infinitely-many n, we again get an accumulation point via a convergent subsequence. Otherwise, we again have infinitely-many distinct elliptic transformations which fix z, and hence the orbit of any other point in $\mathbb{H}$ will have infinitely-many points on a circle and thus will have an accumulation point.

(iii) $\implies$ (i): Suppose that (iii) holds. Choose any $z \in \mathbb{H}$ which as an accumulation point w in its orbit. Choose $R > 0$ so that $z \in \overline{B_R(w)}$. It follows that $g\overline{B_R(w)} \cap \overline{B_R(w)} \neq \varnothing$ for infinitely-many $g \in G$. ☐

Lemmas 1.3.3 and 1.3.4 give us one half of our main result.

Corollary 1.3.5. *If a subgroup $G \subseteq \mathrm{PSL}(2, \mathbb{R})$ acts properly discontinuously on $\mathbb{H}$, then G is discrete.*

Proof. Suppose that G is not discrete. From characterization (iii) in lemma 1.3.3, this means that there exists a sequence of distinct nontrivial points $\{g_n\}$ in G such that $g_n \to id$. So, for any $z \in \mathbb{H}$, $g_n z \to z$. Choose $z_0 \in \mathbb{H}$ so that $g_n z_0 \neq g_m z_0$ for all $n \neq m$; such a z_0 exists by a cardinality argument, since $g_n z = g_m z$ if and only if z is a fixed point of $g_m g_n^{-1}$, and any non-trivial element of G has at most two fixed points. Then, each $g_n z_0$ is distinct and $g_n z_0 \to z_0$. By characterization (ii) of lemma 1.3.4, we have that G does not act properly discontinuously. ☐

Now, we can prove the converse to Corollary 1.3.5.

Lemma 1.3.6. *Take any $z_0 \in \mathbb{H}$ and any compact subset $K \subseteq \mathbb{H}$. Then, the set $E = \{T \in \mathrm{PSL}(2, \mathbb{R}) \mid T(z_0) \in K\}$ is compact.*

Proof. It suffices to show that E is closed and bounded, since $\mathrm{PSL}(2, \mathbb{R})$ inherits its topology from $\mathbb{R}^4$. To see that E is closed, consider that the map $\beta : \mathrm{PSL}(2, \mathbb{R}) \to \mathbb{H}$ given by $\beta(T) = T(z_0)$ is continuous, and $E = \beta^{-1}(K)$.

So, it remains to show that E is bounded. Since K is compact we know that it is bounded away from $\partial\mathbb{H}$, and hence there exists $M_1 > 0$ such that $\mathrm{Im}(w) > M_1$ for all $w \in K$. Moreover, since K is compact there exists $M_2 > 0$ such that $|w| < M_2$ for any $w \in K$. Now, take any $T = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{PSL}(2, \mathbb{R})$. Then,

$$\frac{\mathrm{Im}(z_0)}{|cz_0 + d|^2} = \frac{(ad - bc)\mathrm{Im}(z_0)}{|cz_0 + d|^2} = \mathrm{Im}\left(\frac{az_0 + b}{cz_0 + d}\right) > M_1,$$

so

$$|cz_0 + d| < \sqrt{\frac{\mathrm{Im}(z_0)}{M_1}}.$$

Then since

$$|T(z_0)| = \left|\frac{az_0 + b}{cz_0 + d}\right| < M_2,$$

it follows that

$$|az_0 + b| < M_2|cz_0 + d| < M_2\sqrt{\frac{\mathrm{Im}(z_0)}{M_1}}.$$

Therefore, the quantities $|az_0 + b|$ and $|cz_0 + d|$ are uniformly bounded for all $T \in E$. This implies that the entries of any $T \in E$ are bounded, and hence E is bounded. $\square$

Theorem 1.3.7. *A subgroup $G \subseteq \mathrm{PSL}(2, \mathbb{R})$ is discrete if and only if it acts properly discontinuously on $\mathbb{H}$.*

Proof. By Corollary 1.3.5, it remains to show the forward implication. So, suppose that G is discrete. Let $K \subseteq \mathbb{H}$ be compact, and without loss of generality, assume that $K = \overline{B_R(i)}$ (the closed ball around i of radius R) for some $R > 0$. So for any $T \in \mathrm{PSL}(2, \mathbb{R})$, if $T(K) \cap K \neq \varnothing$, then $T(i) \in \overline{B_{2R}(i)}$. Let

$$E = \{T \in \mathrm{PSL}(2, \mathbb{R}) \mid T(i) \in \overline{B_{2R}(i)}\},$$

and define $E_G = E \cap G$. From Lemma 1.3.6, we know that E is compact. Since G is discrete, it follows that E_G is discrete. Therefore, E_G is a discrete subset of a compact set, and thus E_G is finite. Since

$$\{T \in G \mid T(K) \cap K \neq \varnothing\} \subseteq E_G,$$

we can conclude that G acts properly discontinuously. $\square$

We give subgroups of this type a special name.

Definition 1.3.8. A subgroup of $\mathrm{PSL}(2, \mathbb{R})$ is called *Fuchsian* if it is discrete, or equivalently, if it acts properly discontinuously on $\mathbb{H}$.

1.3.2 Fundamental Domains

As mentioned before, Fuchsian groups are in fact symmetry groups of tessellations of $\mathbb{H}$. We will now formalize this idea.

Definition 1.3.9. Let G be a Fuchsian group. A subset $R \subseteq \mathbb{H}$ is called a *fundamental domain* for G if

(i) $gR \cap R = \varnothing$ for all $g \in G \backslash \{id\}$.

(ii) $\bigcup_{g \in G} \overline{gR} = \mathbb{H}$.

(iii) R is the interior of a convex geodesic polygon in $\mathbb{H}$. That is, R is the interior of a convex set (a set where any two points can be connected by a geodesic in that set) and $\partial R \cap \mathbb{H}$ is a countable union of geodesic segments of positive length, only finitely many of which meet in any compact set.

(iv) $\bigcup_{g \in G} gR$ is locally finite. That is, for any compact $K \subseteq \mathbb{H}$, $|\{g \in G \mid gR \cap K \neq \varnothing\}| < \infty$.

One can think of a fundamental domain of a group G as a polygon whose G-translates tessellate $\mathbb{H}$, similar to how the unit square tessellates $\mathbb{R}^2$ under the action of $\mathbb{Z}^2$.

Theorem 1.3.10. *Any Fuchsian group admits a fundamental domain.*

Note that if R is a fundamental domain, then R many have infinitely-many sides. One can prove that a Fuchsian group G admits fundamental domain with finitely many sides if and only if G is finitely generated, although this requires quite a bit of work. One can prove Theorem 1.3.10 with the following construction.

Definition 1.3.11. Let G be a Fuchsian group. Choose $z_0 \in \mathbb{H}$ such that if $gz_0 = z_0$, then $g = id$; such a z_0 exists since G is discrete. Define $H_g(z_0) = \{z \in \mathbb{H} : d(z, z_0) < d(z, gz_0)\}$. Then the *Dirichlet domain of G centered at z_0* is given by

$$R = R_{z_0} = \bigcap_{g \in G \backslash \{id\}} H_g(z_0).$$

Theorem 1.3.12 ([12], Theorem 5.3)**.** *Let G be a Fuchsian group and R a Dirichlet domain of G centered at z_0. Then, R is a fundamental domain for G.*

1.4 Hyperbolic Structures

1.4.1 Surfaces

Moving forward, our primary objects of study will be surfaces.

Definition 1.4.1. A *surface* is a compact, connected, oriented two-dimensional topological manifold, possibly with boundary. We call a surface *closed* if it is has no boundary.

In particular, we do not assume surfaces to have a smooth structure. However, much of the upcoming theory can be adapted to the category of smooth manifolds.

For the sake of completeness, we recall some classical facts about surfaces (see [10]).

Theorem 1.4.2 (Classification of Surfaces)**.** *Any closed surface is homeomorphic to the connected sum of S^2 with g tori for some $g \geq 0$. The value g is called the* genus *of the surface. Any surface is obtained from a closed surface by removing $n \geq 0$ open disks with disjoint closures.*

Throughout this document, we will let $S_{g,n}$ denote a surface of genus g with n boundary components.

Theorem 1.4.3 (Fundamental Group of a Surface)**.** *The fundamental group of the surface $S_{g,n}$ is given by the presentation*

$$\pi_1(S_{g,n}) = \langle X_1, Y_1, \ldots, X_g, Y_g, C_1, \ldots, C_n \mid [X_1, Y_1] \cdots [X_g, Y_g] C_1 \cdots C_n \rangle.$$

The generators X_i and Y_i correspond to a meridian and a longitude of each torus in the connected sum that define $S_{g,n}$. The generators C_i correspond to the boundary components of $S_{g,n}$.

We will often make use of the following topological invariant.

Definition 1.4.4. Let $S = S_{g,n}$ be a surface. The *Euler characteristic* of S is defined to be

$$\chi(S) := 2 - 2g - n.$$

1.4.2 Hyperbolic Surfaces

In this section, we will explore how we can move hyperbolic geometry from the plane onto different surfaces. In doing so, we will make use of our work from Section 1.3. We start with a local definition for hyperbolic surfaces, which is perhaps the most natural definition.

Definition 1.4.5. Let S be a surface. A *hyperbolic structure* on S is an atlas of charts $\{\varphi_\alpha : U_\alpha \to \mathbb{H}\}$ such that if $V \subseteq U_\alpha \cap U_\beta$ is non-empty and connected, then the transition map $\varphi_\alpha \circ \varphi_\beta^{-1} : \varphi_\beta(V) \to \varphi_\alpha(V)$ is an orientation-preserving isometry (with respect to the hyperbolic metric). A *hyperbolic surface* is a surface equipped with a hyperbolic structure.

Since any surface S is assumed to be oriented, we will assume that any hyperbolic structure on S respects this orientation, hence why we require the transition maps to lie in $\mathrm{Isom}^+(\mathbb{H})$. Generally, the letter S will denote an ordinary surface (i.e. just a topological space), while the letter X will denote a hyperbolic surface.

Note that a hyperbolic surface automatically comes with a metric obtained by pulling back the metric from $\mathbb{H}$. This metric will have constant negative curvature, so by the Gauss-Bonnet Theorem, a closed surface S admits a hyperbolic structure only if $\chi(S) < 0$. The converse is true too.

Theorem 1.4.6 ([2], Theorem 1.2)**.** *Let S be a surface. If $\chi(S) < 0$, then S admits a hyperbolic structure.*

One of the simplest ways to build a hyperbolic surface uses fundamental domains of Fuchsian groups.

Definition 1.4.7. Suppose G is a Fuchsian group and R is a fundamental domain for G. A *side* of R is a segment $s \subseteq \partial R$ of positive length where there exists $g \in G$ such that $s \subseteq \overline{R} \cap g\overline{R}$. A *side pairing* of R is an element $g \in G$ such that $\overline{R} \cap g\overline{R}$ is a side.

Suppose we have a Fuchsian group G and a fundamental domain R. Define an equivalence relation $\sim$ on $\overline{R}$ by $x \sim y$ if and only if there exists $g \in G$ such that $x = gy$. Our hyperbolic surface will be $\overline{R}/\sim$. Let $\pi : R \to \overline{R}/\sim$ be the projection map. Given $[z] \in \overline{R}/\sim$, we define a chart around $[z]$ as follows:

- If $z \in \text{Int}(R)$, then there exists a neighborhood $U \subseteq R/\sim$ of $[z]$ such that $[w] = \{w\}$ for each $[w] \in U$. In this case, we take the chart $(U, \pi^{-1}|_U)$.
- Suppose $z \in s$ for some side s, but z is not a vertex. Choose $g \in G$ such that $s \subseteq g\overline{R} \cap \overline{R}$. Choose $r > 0$ such that $B_r(z) \cap R$ and $B_r(g^{-1}z) \cap R$ are disjoint open half-disks. Let $A = B_r(z) \cap \overline{R}$ and $B = B_r(z) \cap g\overline{R}$. Then, let $U = \pi(A \cup g^{-1}B)$. Define $\varphi : U \to \mathbb{H}$ by

$$\varphi([x]) = \begin{cases} x & \text{if } x \in A \\ gx & \text{if } x \in B \end{cases}.$$

  Notice that φ is well-defined if $x \in A \cap B$. We take the chart (U, φ).
- Suppose z is a vertex of $\overline{R}$. Then, do the analogous construction to the previous case, except with all the G-translates of R around z.

See figure 1.6 for an illustration of these neighborhoods. It is straightforward to verify that the transition maps are hyperbolic isometries, so we have indeed constructed a hyperbolic structure.

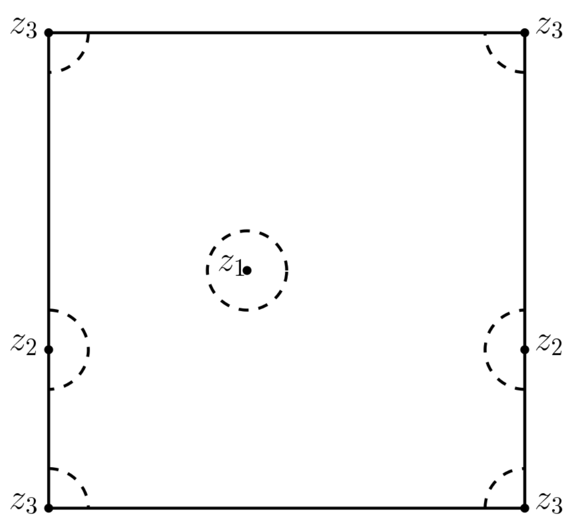

Figure 1.6: We can picture $\overline{R}$ as the unit square in $\mathbb{R}^2$ under the action of $\mathbb{Z}^2$. Here we see how to choose neighborhoods in $\overline{R}/\sim$ around an interior point z_1, a non-vertex side point z_2, and a vertex z_3.

Now suppose that G acts freely on $\mathbb{H}$, meaning that any nontrivial $g \in G$ has no fixed points in $\mathbb{H}$. This is equivalent to G being torsion-free, and also equivalent to G having no elliptic elements (since G is discrete, any elliptic elements must have finite order). Then, we can put a hyperbolic structure on $\mathbb{H}/G$. We define our charts as follows. Take any $[z] \in \mathbb{H}/G$, and choose a representative z. No element of $g \in G$ fixes z, and since G acts properly distcontinuously, no orbit has an accumulation point. So, we can choose $r > 0$ such that $gB_r(z) \cap B_r(z) = \varnothing$ for all $g \in G$. Then, we take the chart (U, φ) where U is the projection of $B_r(z)$ to $\mathbb{H}/G$, and φ is the inverse of this projection.

One can check that both of these constructions in fact yield the same surface. However, neither of these constructions actually require both a Fuchsian group and fundamental domain. One only needs a polygon with side-pairing isometries to carry out the first construction, and one only needs a Fuchsian group acting freely to carry out the second construction.

Furthermore, if G is a Fuchsian group that does not act freely (equivalently, if G contains elliptic elements), then we can still carry out the second construction. However, if $z \in \mathbb{H}$ is the fixed point of an elliptic element $T \in G$, then in order to make our charts bijective, we will have to assign z a chart into the space

$\mathbb{H}/\langle T\rangle$. We call this construction an *orbifold*. If the order of T is n, then $\mathbb{H}/\langle T\rangle$ looks like a "wedge" of $\mathbb{H}$ with angle $2\pi/n$, and $[z] \in \mathbb{H}/G$ will resember a cone of angle $2\pi/n$; therefore, we call the elliptic fixed points *cone points*.

1.4.3 Global Characterization of Hyperbolic Structures

We have seen that given a torsion-free Fuchsian group G, we can always obtain a hyperbolic surface via the quotient $\mathbb{H}/G$. Moreover, it follows from the theory of covering spaces that $\pi_1(\mathbb{H}/G) \cong G$. Now, our goal is to prove a partial converse to this example; namely, if X is any hyperbolic surface with a complete metric, then there exists a torsion-free Fuchsian group G such that $\pi_1(X) \cong G$ and X is isometric to $\mathbb{H}/G$. This means that rather define (complete) hyperbolic structures locally using charts, we can define them globally as pairs (S, φ) where S is a surface and $\varphi : \pi_1(S) \to \mathrm{Isom}^+(\mathbb{H}) \cong \mathrm{PSL}(2, \mathbb{R})$ is an injective homomorphism with a discrete image.

To accomplish this, we will define the developing map. The idea behind this map is that since a hyperbolic surface locally looks like $\mathbb{H}$, we should be able to somehow "unroll" our surface onto the hyperbolic plane, much like how one may unroll a torus or cone onto the Euclidean plane.

We define the developing map as follows. Let X be a hyperbolic surface, and $\tilde{X}$ its universal cover. We will view $\tilde{X}$ as the set of homotopy classes (relative to endpoints) of curves in X based at $x_0 \in X$. Take any $[\alpha] \in \tilde{X}$, where α is a representative path in X. We can cover α with a finite sequence of charts, starting with (U_0, φ_0) around $\alpha(0) = x_0$ and ending with (U_n, φ_n) about $\alpha(1)$, such that sequential charts have a connected intersection and non-sequential charts have an empty intersection. For $0 \leq i < n$, we have some $\gamma_i \in \mathrm{Isom}^+(\mathbb{H})$ such that $\varphi_i(U_i \cap U_{i+1}) = \gamma_i \varphi_{i+1}(U_i \cap U_{i+1})$ (in particular, γ_i is the isometry extending $\varphi_i \circ \varphi_{i+1}^{-1}$). This means that we can adjust the chart (U_{i+1}, φ_{i+1}) so that it "lines up" with (U_i, φ_i). Starting from the last chart, we adjust (U_n, φ_n) to line up with (U_{n-1}, φ_{n-1}), then adjust both of these charts to align with (U_{n-2}, φ_{n-2}), and continue this process until everything is lined up with (U_0, φ_0). In the end, we have that the image of α starts at $\varphi(\alpha(0)) = \varphi(x_0) \in \varphi_0(U_0)$ and ends at $\gamma_0 \cdots \gamma_{n-1}\varphi_n(\alpha(1)) \in \gamma_0 \cdots \gamma_{n-1}\varphi_n(U_n)$. See figure 1.7 for an illustration.

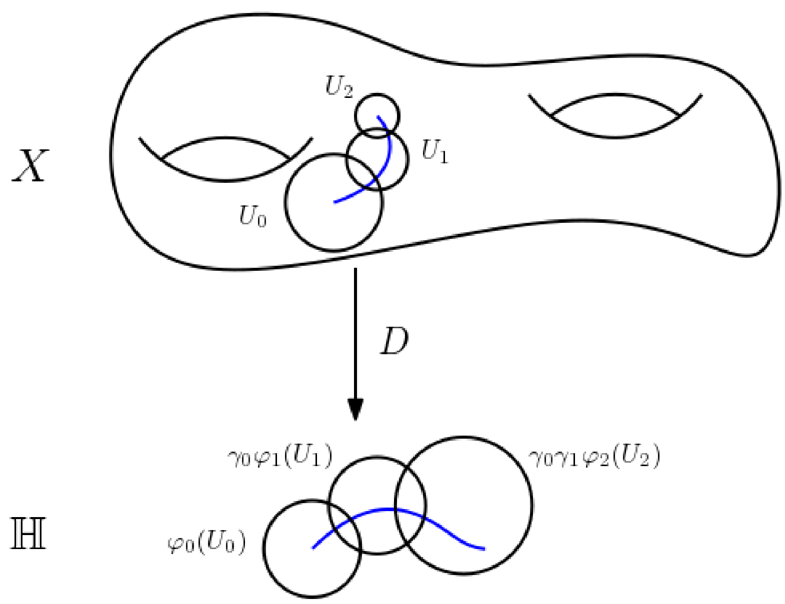

Figure 1.7: The image of a path on $S_{2,0}$ (in blue) under the developing map.

Definition 1.4.8. In the setup of the previous paragraph, we define the *developing map* based at (U_0, φ_0) to be the map $D : \tilde{X} \to \mathbb{H}$ given by $D([\alpha]) = \gamma_0 \cdots \gamma_{n-1}\varphi_n(\alpha(1))$. More generally, we define the *developing image* of α to be the curve $\beta(t) := \gamma_0 \cdots \gamma_{i-1}\varphi_i(\alpha(t))$, where $\alpha(t) \in U_i$.

Certainly, one must check that the developing map is well-defined. First, one shows that if we fix a representative path α, then the definition is independent of the charts we choose (other than the initial chart). One can show inductively that two different coverings of α give the same developing image; the base case that we have two coverings of the form $\{(U_0, \varphi_0), (U_1, \varphi_1)\}$ and $\{(U_0, \varphi_0), (U_1', \varphi_0')\}$ is not hard to show. Next, one shows that $D([\alpha])$ is independent of the representative α. To do this, take any two representatives α, α' of $[\alpha]$. Then, we can produce a finite "grid" of charts covering α and α'; given a homotopy H from α to α', we partition the domain of H into rectangles $[t_i, t_{i+1}] \times [s_i, s_{i+1}]$ each contained in a single chart. We can deform α to α' in finitely many steps where at each step, we only deform a segment of α which lies entirely in one of the charts in our grid. One checks that deforming within a single chart does not affect $D([\alpha])$.

It is important to note that the developing map does in fact depend on the choice of the initial chart (U_0, φ_0) (for this reason, it is perhaps more appropriate to say *a* developing map). However, this only changes the developing map up to composition by a hyperbolic isometry which carries one initial chart to another.

There is a special scenario where $\alpha \in \tilde{X}$ is a loop in X.

Definition 1.4.9. Let $D : \tilde{X} \to \mathbb{H}$ be the developing map based at (U_0, φ_0). Take any $[\alpha] \in \pi_1(X, x_0)$. Then, $D([\alpha]) = h\varphi_0(x_0)$ for some $h \in \mathrm{Isom}^+(\mathbb{H})$. We call h the *holonomy* of $[\alpha]$.

Since the developing map is well-defined, it follows that we have a well-defined (again, up to the choice of our initial chart) map $\mathrm{hol} : \pi_1(X, x_0) \to \mathrm{Isom}^+(\mathbb{H}) \cong \mathrm{PSL}(2, \mathbb{R})$ which sends $[\alpha]$ to its holonomy. Then, the next result follows directly from the definition of the developing map and multiplication in $\pi_1(X)$.

Proposition 1.4.10. *Let* $\mathrm{hol} : \pi_1(X) \to \mathrm{Isom}^+(\mathbb{H})$ *map* $[\alpha] \in \pi_1(X)$ *to its holonomy. Then,* hol *is a group homomorphism. We call* hol *the* holonomy homomorphism*.*

The holonomy homomorphism will be the homomorphism in our global definition of a hyperbolic structure mentioned earlier. Therefore, it is important to note the following.

Proposition 1.4.11. *The holonomy homomorphism* $\mathrm{hol} : \pi_1(X) \to \mathrm{Isom}^+(\mathbb{H})$ *is injective.*

Proof. Suppose $\mathrm{hol}([\alpha]) = id$. This means that the developing image of α is a loop entirely contained in the image of the base chart $\varphi_0(U_0)$. Since $\mathbb{H}$ is simply connected, we know the developing image of α is null-homotopic, and applying φ_0^{-1}, we get that α is null-homotopic. ☐

Recall that we put a metric on a hyperbolic surface X by pulling back the metric from $\mathbb{H}$. Then, we can put a metric on $\tilde{X}$ by pulling back the metric on X along the covering map. Then, we have the following crucial characterization.

Theorem 1.4.12. *Let X be a hyperbolic surface and $p : \tilde{X} \to X$ its universal cover. The following are equivalent:*

(i) The developing map $D : \tilde{X} \to \mathbb{H}$ is a covering map.

(ii) X is metrically complete.

(iii) $\tilde{X}$ is metrically complete.

A detailed proof theorem 1.4.12 is a bit long. However, the high-level ideas involved are relatively simple.

Sketch of proof.

- $(i) \implies (ii)$: Since D is a covering map between two simply connected spaces, D is in fact a homeomorphism. Moreover, it follows by definition of the metrics on X and $\tilde{X}$ that D is a local isometry. Thus, D is an isometry. This implies that closed balls in $\tilde{X}$ are compact. Then, this implies closed balls in X are compact, and it follows that X is complete.

- $(ii) \implies (iii)$: The projection map p does not increase distances, so a Cauchy sequence x_n in $\tilde{X}$ projects to a Cauchy sequence $p(x_n)$ in X. By assumption, $p(x_n)$ has a limit w. Since p is a local isometry, there exists $r > 0$ such that $p^{-1}(B_r(w)) = \bigsqcup_i V_i$, where each V_i is a ball of radius r. Then, x_n must eventually stay in one V_i, and hence must converge to $p^{-1}(w) \cap V_i$.

- $(iii) \implies (i)$: We can show that D satisfies the path lifting property. Let $\alpha : [0,1] \to \mathbb{H}$ be a path. If we can lift $\alpha|_{[0,t_0]}$ to $\tilde{S}$ for some t_0, then since D is a local isometry, we can extend this lift to a lift of $\alpha|_{[0,t_0+\varepsilon)}$. Also, if we can lift $\alpha|_{[0,t_0)}$ to $\tilde{S}$ for some t_0, then since D is a local isometry and $\tilde{S}$ is complete, we can complete this to a lift of $\alpha|_{[0,t_0]}$. It follows that we can lift all of α. Since $\mathbb{H}$ is path connected, it follows that D is surjective. Then, any local homeomorphism satisfying the path lifting property is necessarily a covering map.

□

Corollary 1.4.13. *Let X be a complete hyperbolic surface. Then, X is isometric to $\mathbb{H}/G$ for some torsion-free Fuchsian group G.*

Proof. Let $D : \tilde{X} \to \mathbb{H}$ be a developing map and $\mathrm{hol} : \pi_1(X) \to \mathrm{Isom}^+(\mathbb{H})$ the holonomy homomorphism. We let $G = \mathrm{hol}(\pi_1(X))$. Let $\pi : \mathbb{H} \to \mathbb{H}/G$ be the projection map. We define a map $\bar{D} : X \to \mathbb{H}/G$ as follows. Fix a base point $x_0 \in X$. For any $x \in X$, choose a path α from x_0 to x. Then, define $\bar{D}(x) := \pi(D([\alpha]))$.

We claim that $\bar{D}$ is well-defined. Suppose α and β are two paths in X from x_0 to x. Choose a covering of α and β as in the definition of the developing map, such that both coverings have the same final chart (U, φ) around x. Then, there exists isometries $g_\alpha, g_\beta \in \mathrm{Isom}^+(\mathbb{H})$ such that $D([\alpha]) = g_\alpha\varphi(x)$ and $D([\beta]) = g_\beta\varphi(x)$. Now, let $\bar{\beta}$ denote the path obtained by reverserving β (that is, $\bar{\beta}(t) = \beta(1-t)$), and let $\bar{\beta} \cdot \alpha$ be the concatenated path obtained first by following α and then following $\bar{\beta}$. In particular, $\bar{\beta} \cdot \alpha$ is a loop based at x_0. If we take the developing image of $\bar{\beta} \cdot \alpha$ by concatenating our covers of α and β, then the final chart will be $g_\beta^{-1} g_\alpha \varphi(x)$. Since $\bar{\beta} \cdot \alpha$ is a loop based at x_0, it follows that $g_\beta^{-1} g_\alpha \in G$. Thus, $D([\alpha])$ and $D([\beta])$ are in the same equivalence class in $\mathbb{H}/G$, meaning $\pi(D[\alpha]) = \pi(D[\beta])$.

Now, it follows by definition of hol that D is G-equivariant, meaning that for any $\beta \in \pi_1(X)$ and $[\alpha] \in \tilde{X}$, $D(\beta \cdot [\alpha]) = \mathrm{hol}(\beta) \cdot D([\alpha])$. Therefore, it follows from the theory of covering spaces that $\bar{D}$ is a covering map (essentially, $\bar{D}$ is obtained by quotienting the domain and codomain of D by the same group action). Then, if we let $\bar{D}_*$ denote the induced homomorphism $\pi_1(X) \to \pi_1(\mathbb{H}/G)$, we know that $\bar{D}_*$ is injective and the number of sheets of $\bar{D}$ is $[\pi_1(\mathbb{H}/G) : \bar{D}_*(\pi_1(X))]$. But since $\pi_1(X) \cong \pi_1(\mathbb{H}/G)$, this implies that $\bar{D}$ has one sheet, and hence is injective. Also, by definition of the metrics on X and $\tilde{X}$, $\bar{D}$ is a local isometry, and since it's injective, it is an isometry.

Since X is a compact, connected, and oriented manifold, we know that $\pi_1(X)$, and hence G, is torsion-free. So, it remains to check that G is a Fuchsian group. This is a consequence of the following fact: if a group H acts freely on a topological space Y in a way such that Y/H is Hausdorff and the projection map $p : Y \to Y/H$ is a covering map, then H acts properly discontinuously on Y.

To prove this, we first take any $x, y \in Y$ and show that we can choose neighborhoods U and V of x and y such that $hU \cap V \neq \varnothing$ for at most one $h \in H$. If $p(x) \neq p(y)$, then $p(x)$ and $p(y)$ have disjoint neighborhoods in Y/H which we can lift to obtain U and V. Otherwise, $y = hx$ for some $h \in H$. Since H acts freely, h is the only element of H satisfying this equation. So, we can take U to be a neighborhood of x such that $p|_U$ is a homeomorphism, and let $V = hU$. Now, we can show that for any compact subset $K \subseteq Y$, $\{h \in H \mid hK \cap K \neq \varnothing\}$ is finite. Take any compact set $K \subseteq Y$. For any $(x, y) \in K \times K$ we have a pair of sets (U_x, V_y) as above. The collection $(U_x \times V_y)_{(x,y) \in K \times K}$ is an open cover of $K \times K$ and hence has a finite subcover $\{U_1 \times V_1, \dots, U_n \times V_n\}$. Suppose that $h \in H$ such that $hK \cap K \neq \varnothing$. This means that there exists $x \in K$ such that $(x, hx) \in K \times K$. The element (x, hx) must lie in one of finitely many sets $U_1 \times V_1, \dots, U_n \times V_n$. Moreover, if $(x, hx) \in U_i \times V_i$, then $hU_i \cap V_i \neq \varnothing$, and hence the value of h is determined. So, there are only finitely-many possible values of h. Thus, H acts properly discontinuously on Y. □

Chapter 2

Mapping Class Groups and Teichmüller Space

We saw in the last chapter that we can endow a surface S with a geometry which locally resembles $\mathbb{H}$. Our goal now is to understand how this hyperbolic structure changes if we modify S by some non-trivial symmetry. To achieve this goal, we have to answer two questions: what are the non-trivial symmetries of S, and how can we describe the relationship between two different hyperbolic structures?

In Section 2.1, we will answer the first question by defining the mapping class group $\mathrm{Mod}(S)$ of S. The group $\mathrm{Mod}(S)$ will consist of homotopy classes of orientation-preserving homeomorphisms of S which fix boundary components. We will use simple closed curves on S to study the action of $\mathrm{Mod}(S)$ on S. Then, we will prove an important structural theorem: $\mathrm{Mod}(S)$ is generated by particularly simple elements called Dehn twists. In Section 2.2, we will answer the second question by defining the Teichmüller space $\mathrm{Teich}(S)$ of S. This is essentially the collection of all possible (complete) hyperbolic structures on S (up to a natural equivalence). We will endow $\mathrm{Teich}(S)$ with a topology that describes when two hyperbolic structures are "similar". Finally, in Section 2.3, we will achieve our goal by studying the relationship between these two objects. Namely, we will see that $\mathrm{Mod}(S)$ naturally acts on $\mathrm{Teich}(S)$, and this action is properly discontinuous, which makes it "nice".

For this chapter, we will follow Benson Farb and Dan Margalit's book [2], namely Chapters 1-4, 10, and 12.

2.1 Mapping Class Groups

2.1.1 Simple Closed Curves

Recall that surfaces are assumed to be compact, connected, oriented, and possibly with boudary components. The theory of mapping class groups can be extended to punctured surfaces (i.e. surfaces with finitely many points removed), but we will not discuss such surfaces.

Given a surface S, we will soon define the mapping class group of S to be the group of boundary-fixing homeomorphisms of S up to homotopy. However, if we wish to study the behavior of a homeomorphism of S, we need to know how it moves points around within S. Therefore, we will study the mapping class group by examining its action on curves that lie on S.

Definition 2.1.1. Let S be a surface. A *closed curve* on S is a continuous map $S^1 \to S$ (we often identify a closed curve with its image). A closed curve is called *simple* if it is injective. A closed curve is called *essential* if it is not null-homotopic or homotopic into a neighborhood of a boundary component.

There are two main types of simple closed curves.

Definition 2.1.2. Let α be a closed curve on a surface S. We say that α is *non-separating* is $S\backslash\alpha$ is connected. Otherwise, we say α is *separating*.

Since the mapping class group is comprised of homotopy classes of maps, we will focus our attention on homotopy classes of curves. In particular, we study *free* homotopy classes of curves, meaning that we do not require our homotopies to fix a particular base point. Since we assume our surfaces to be connected (and hence path-connected, since surfaces are manifolds), we have a bijective correspondence between conjugacy classes of $\pi_1(S)$ and free homotopy classes of closed curves on S. Unless stated otherwise, homotopy classes of simple closed curves will refer to free homotopy classes.

The following fact, which we will use later, connects homotopy classes of curves with the geometry of a surface.

Proposition 2.1.3. *Let X be a complete hyperbolic surface and α an essential simple closed curve on S. Then, α is homotopic to a unique geodesic closed curve.*

Proof. Let $\mathrm{hol} : \pi_1(X) \to \mathrm{PSL}(2, \mathbb{R})$ be the holonomy homomorphism of the hyperbolic structure on X with some choice of an initial chart (U, φ), and let $\phi = \mathrm{hol}([\alpha])$. Since we know the image of the holonomy homomorphism acts freely and properly discontinuously on $\mathbb{H}$, we know that ϕ cannot be an elliptic element. Moreover, we claim that ϕ is a cannot be parabolic. Let $\tau(\phi) = \inf\{d(x, \phi(x)) \in \mathbb{R}_{>0} \mid x \in \mathbb{H}\}$. If ϕ is parabolic, then $\tau(\phi) = 0$. This implies that, up to free homotopy, the length of α can be made arbitrarily small, which contradicts that α is essential. So, ϕ must be hyperbolic.

Recall that (U, φ) is our initial chart about $\alpha(0)$. If we let A denote the axis of ϕ, then the orbit points of $\varphi(\alpha(0))$ under ϕ tend towards the endpoints of A on $\partial\mathbb{H}$. So, take a lift $\widetilde{\alpha}$ of α to $\mathbb{H}$. In this context, if we view α as a map $\alpha : S^1 \to S$, then by taking a lift we really mean a lift of the map $\alpha \circ \pi$ where $\pi : \mathbb{R} \to S^1$ is the usual covering map. In other words, we take $\widetilde{\alpha}$ to have $\mathbb{R}$ as its domain and intersect every orbit point of $\varphi(\alpha(0))$ under ϕ. Then, $\widetilde{\alpha}$ has the same endpoints on $\partial\mathbb{H}$ as A, and hence is homotopic to A. This descends to a homotopy from α to the image of A under the covering map $\mathbb{H} \to X$. But since A is a geodesic, its image is also a geodesic (as the covering map is a local isometry). See Figure 2.1 for an illustration.

So, this shows the existence of a geodesic representative of α. To show uniqueness, we can play a similar game. Namely, if α is homotopic to some geodesic γ on X, then γ lifts to some geodesic $\widetilde{\gamma}$ in $\mathbb{H}$, and we get a homotopy from $\widetilde{\alpha}$ to $\widetilde{\gamma}$. By compactness, homotopies can move points in $\mathbb{H}$ only a bounded distance, and hence $\widetilde{\gamma}$ and $\widetilde{\alpha}$ must have the same endpoints. But we also know A and $\widetilde{\alpha}$ have the same endpoints, and a geodesic is uniquely determined by its endpoints, so it must be that $\widetilde{\gamma} = A$. It follows that γ must be the image of A under the covering map $\mathbb{H} \to X$, which proves uniqueness. $\square$

In the above proof, we discussed why the holonomy of an essential simple closed curve cannot be parabolic. In general, parabolic elements only arise in the holonomy groups of punctured surfaces, which are surfaces with finitely-many points removed. This is because a puncture is the only place where a simple closed curve can be nontrivial but still be homotoped to have arbitrarily small length (and hence its holonomy can move points an arbitrarily small distance).

Note that the existence part of this proposition is true for any Riemannian manifold; you can always take a curve, lift it to the universal cover, homotope it to a geodesic, and project it back down to the manifold. In the case of closed surfaces, the uniqueness part holds only for hyperbolic surfaces.

Our next goal is to understand how homotopy classes of curves interact with each other.

Definition 2.1.4. Let S be a surface, and let a and b be homotopy classes of closed curves on S. The *geometric intersection number of a and b* is defined to be

$$i(a, b) := \min\{|\alpha \cap \beta| \in \mathbb{N} \mid \alpha \in a, \beta \in b\}.$$

That is, $i(a, b)$ is the minimal number of times any two representatives of a and b must intersect. If $\alpha \in a$ and $\beta \in b$, we say that α and β are in *minimal position* if $|\alpha \cap \beta| = i(a, b)$.

Example 2.1.5.

- On the torus T^2, the two standard generators of $\pi_1(T^2)$ have geometric intersection number 1.
- Let S be any surface, and let a and b be homotopy classes of closed curves on S such that a is separating (note that the property of being separating or non-separating is homotopy invariant). Then, $i(a, b)$ must be even. To see this, take any $\alpha \in a$ and $\beta \in b$. If β were to cross α an odd number of times, it would start and end in different connected components of $S \backslash \alpha$, which is not possible.

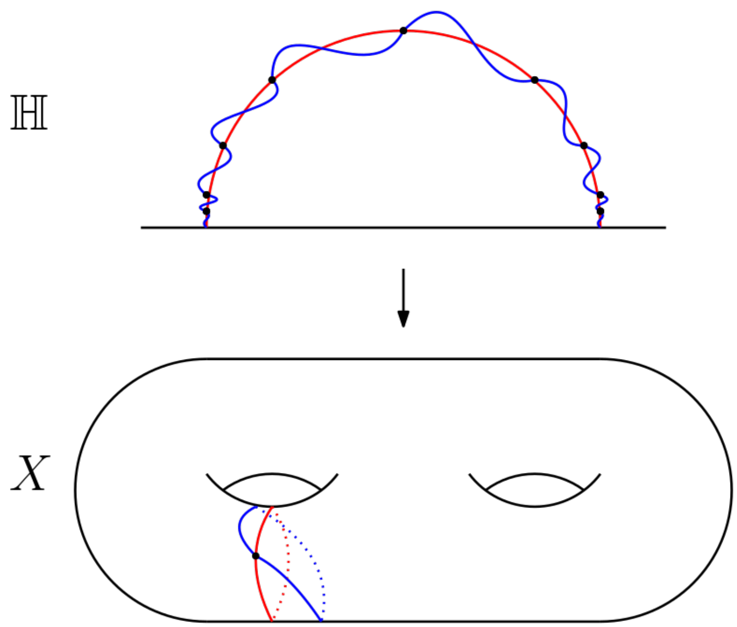

Figure 2.1: The curves α and $\widetilde{\alpha}$ are drawn in blue, A and its image are drawn in red, and the black dots denote the orbit of $\varphi(\alpha(0))$.

An important concept in the study of closed curves is the *change of coordinates principle*. This principle is not a precise statement, but rather the rough idea that if we have two simple closed curves α and β which divide up a surface S in the same way, then there will be a homeomorphism of S which maps α to β. To illustrate this idea, we can first show that if α and β are non-separating, then there is a homeomorphism of S mapping α to β.

Definition 2.1.6. Let S be a surface and α a simple closed curve on S. The surface obtained by *cutting S along α* is a surface S_α with two distinguished boundary components B_1 and B_2 and a homeomorphism $h : B_1 \to B_2$ satisfying the following:

- There is a homeomorphism $\phi : S_\alpha/(x \sim h(x)) \to S$.
- The homeomorphism ϕ satisfies $\phi(B_1 \cup B_2) = \alpha$.

Suppose α and β are two non-separating simple closed curves on a surface S. Then, the cut surfaces S_α and S_β are connected and have the same Euler characteristic and number of boundary components. Therefore, it follows from the classification of surfaces that S_α is homeomorphic to S_β. If we choose a homeomorphism which respects the equivalence relations on the distinguished boundary components of S_α and S_β, then we obtain a homeomorphism of S which maps α to β.

This argument works for any two non-separating simple closed curves. If α and β are instead separating curves, then we can say the same thing if the individual connected components of S_α and S_β are homeomorphic. We summarize this as follows.

Proposition 2.1.7. *Let α and β be simple closed curves on a surface S. Then, there exists a homeomorphism of S mapping α to β if and only if the surfaces obtained by cutting S along α and β are homeomorphic.*

The idea behind the change of coordinates principle is that we can state Proposition 2.1.7 for more general collections of curves. For instance, we can make an analogous statement for pairs of simple closed curves which intersect once. The idea behind the proof is always the same: cut your surface along your collection of curves and apply the classification of surfaces.

2.1.2 Curve Graphs

There is an important tool in the study of simple closed curves called the curve graph.

Definition 2.1.8. Let S be a surface. We define the *curve graph of* S to be the graph $\mathcal{C}(S)$ where the vertex set V is the set of homotopy classes of essential simple closed curves on S, and the edge set E is the set of pairs $(a,b) \in V^2$ such that $i(a,b) = 0$. That is, there is an edge connecting a and b if and only if a and b are disjoint.

Theorem 2.1.9. *If $3g+n \geq 5$, then $\mathcal{C}(S_{g,n})$ is connected.*

Proof. We want to show that for two vertices a and b of $\mathcal{C}(S_{g,n})$, there exists a sequence of vertices $a = c_0, c_1, \ldots, c_m = b$ such that $i(c_i, c_{i+1}) = 0$ for all i. We can do this by induction on $i(a,b)$.

If $i(a,b) = 0$, then there is an edge from a straight to b. If $i(a,b) = 1$, then choose representatives α and β in minimal position. Then, a neighborhood of $\alpha \cup \beta$ is a one-holed torus; let c be the vertex corresponding to the boundary of this one-holed torus. Then, c must be non-essential, since otherwise we would have that $(g,n) = (1,0)$ or $(g,n) = (1,1)$, which contradicts our assumption that $3g+n \geq 5$. So, a, c, b is the desired path.

Now, suppose that $i(a,b) \geq 2$. By induction, it suffices to find a vertex c such that $i(a,c), i(b,c) < i(a,b)$. So, choose $\alpha \in a$ and $\beta \in b$ in minimal position and orient them arbitrarily. Figure 2.2 indicates the two possibilities for two consecutive intersections of α and β. If the consecutive intersections have the same sign, we choose γ as in the left-hand side of Figure 2.2 (we assume it continues to follow α everywhere else on S). We see that γ is non-essential since $|\gamma \cap \alpha| = 1$. If the consecutive intersections have opposite signs, let γ_1 and γ_2 be as in the right-hand side of Figure 2.2. We know that these curves cannot be null-homotopic since this would imply that α and β are not in minimal position. If they are both homotopic into a neighborhood of a boundary component, then the side of α containing these curves is a two-holed disk. We can draw analogous curves γ_3 and γ_4 on the other side of α, which similarly cannot be null-homotopic. If γ_3 and γ_4 are also homotopic into neighborhoods of boundary components, this would imply that $(g,n) = (0,4)$, contradicting our assumption. So, it must be that some γ_i is essential, so let $\gamma = \gamma_i$. In any case, the homotopy class of γ is the desired vertex c. $\square$

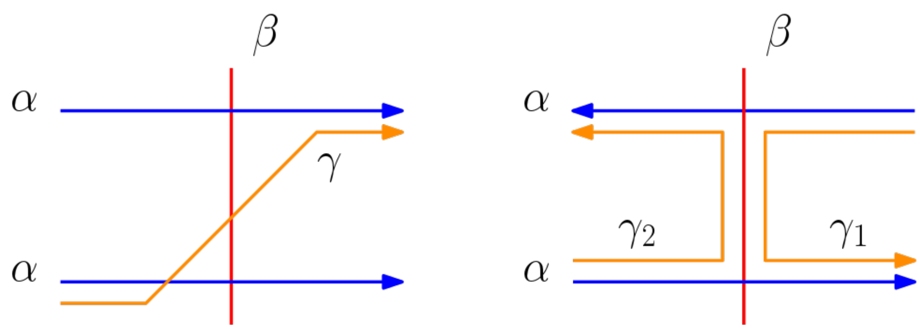

Figure 2.2: How to choose γ (orange) depending on the configuration of α (blue) and β (red). We assume γ continues to follow α outside of this area.

There are two related graphs we are interested in.

Definition 2.1.10. Let S be a surface.

(i) We let $\mathcal{N}(S)$ denote the subgraph of $\mathcal{C}(S)$ comprised of verticies corresponding to non-separating simple closed curves.

(ii) We let $\mathcal{NI}(S)$ denote the graph with the same vertex set at $\mathcal{N}(S)$, but where two verticies a and b are joined by an edge if and only if $i(a,b) = 1$.

Proposition 2.1.11. *If $g \geq 2$, then $\mathcal{N}(S_{g,n})$ is connected.*

Proof. First, assume $g \geq 2$ and $n \leq 1$; we will induct on n. Take any verticies a and b of $\mathcal{N}(S_{g,n})$. Since $3g+n \geq 5$, we know that $\mathcal{C}(S_{g,n})$ is connected, so choose a path $c_0 = a, c_1, \ldots, c_{m-1}, c_m = b$ in $\mathcal{C}$. We want to modify this path so that it consists only of non-separating curves. Suppose some c_i is separating. Since

$g \geq 2$ and $n \leq 1$, the components of $S_{g,n} \backslash c_i$ must have positive genus (otherwise, c_i would be homotopic into a neighborhood of a boundary component and hence would not be essential). If c_{i-1} and c_{i+1} lie on different components of $S_{g,n} \backslash c_i$, then c_{i-1} and c_{i+1} are already disjoint and so we can just remove c_i from our path. If c_{i-1} and c_{i+1} lie on the same component of $S_{g,n} \backslash c_i$, then just replace c_i with any non-separating curve on the other component (we know we can do this since the other component has positive genus). We can repeat this process for any separating c_i to get a path from a to b in $\mathcal{N}(S_{g,n})$.

Now, suppose $n \geq 2$. Apply the same procedure as before. This time, it may be that if c_i is separating, then c_{i-1} and c_{i+1} lie on the same component S' of $S_{g,n} \backslash c_i$ and the other component has genus 0. However, we can then apply our inductive hypothesis to get a path in $\mathcal{N}(S')$ from c_{i-1} to c_i. □

Proposition 2.1.12. *If $g \geq 2$, then $\mathcal{NI}(S_{g,n})$ is connected.*

Proof. Take any verticies a and b of $\mathcal{NI}(S_{g,n})$. Since $g \geq 2$, we can choose a path $a = c_0, c_1, \ldots, c_{m-1}, c_m = b$ in $\mathcal{N}(S_{g,n})$. Then, for any consecutive vertices c_i and c_{i+1}, there exists some vertex d_i such that $i(c_i, d_i) = i(c_{i+1}, d_i) = 1$. To see this, apply the change of coordinates principle to map c_i and c_{i+1} to a pair of curves as in Figure 2.3. □

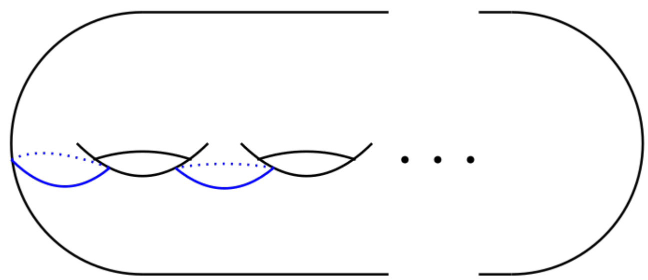

Figure 2.3: We map c_i and c_{i+1} to the blue curves.

Proposition 2.1.11 is not true for any surface of genus 1. However, it turns out that Propsition 2.1.12 is.

Proposition 2.1.13. *For any $n \geq 0$, $\mathcal{NI}(S_{1,n})$ is connected.*

Proof. One can prove this by induction on n. The base cases are $S_{1,0} = T^2$ and $S_{1,1}$, which one can check manually. Otherwise, suppose $n \geq 2$. Take any verticies a and b of $\mathcal{NI}(S_{1,n})$. Since $3g + n \geq 5$, the curve graph is connected, so choose a path $a = c_0, c_1, \ldots, c_m = b$ in $\mathcal{C}(S_{1,n})$. Suppose some c_i is separating. If c_{i-1} and c_{i+1} live on different components of $S_{1,n} \backslash c_i$, remove c_i. If they live on the same component, then by induction, there exists a path $c_{i-1} = d_0, d_1, \ldots, d_k = c_{i+1}$ such that each d_j is non-separating and $i(d_j, d_{j+1}) = 1$ for all j. So, remove c_i and add each d_j to our path. By repeating this process, we can assume that each c_i is non-separating and $i(c_i, c_{i+1}) = 1$ or $i(c_i, c_{i+1}) = 0$ for all i. Then, arguing as in Proposition 2.1.12, we can modify this path so that $i(c_i, c_{i+1}) = 1$ for all i. □

The important take-away from this discussion of curve graphs is the following, which is just a reformulation of Propositions 2.1.12 and 2.1.13.

Corollary 2.1.14. *Suppose $g \geq 1$, $n \geq 0$, and a and b are non-separating essential simple closed curves on $S_{g,n}$. Then, there exists a sequence of non-separating essential simple closed curves $a = c_0, c_1, \ldots, c_m = b$ such that $i(c_i, c_{i+1}) = 1$ for all i.*

2.1.3 Mapping Class Groups

Now that we are familiar with curves on surfaces, we are ready to define the mapping class group.

Definition 2.1.15. Let S be a surface. We let $\mathrm{Homeo}^+(S, \partial S)$ denote the group of orientation-preserving homeomorphisms of S which fix ∂S pointwise (if S is closed, then this is simply the group of orientation-preserving homeomorphisms). Given $f, g \in \mathrm{Homeo}^+(S, \partial S)$, we write $f \sim g$ if f and g are homotopic. Then, we define the *mapping class group of S* to be

$$\mathrm{Mod}(S) := \mathrm{Homeo}^+(S, \partial S)/\sim .$$

That is, $\mathrm{Mod}(S)$ is the group of homotopy classes of orientation-preserving pointwise-boundary-fixing homeomorphisms of S. A *mapping class of* S is an element of $\mathrm{Mod}(S)$.

Example 2.1.16. Suppose we embed S_g in $\mathbb{R}^3$ in a "star" configuration, as illustrated for S_4 in Figure 2.4. Then, we get a homeomorphism $f : S_g \to S_g$ by rotating S_g by an angle of $\frac{2\pi}{g}$, so that each handle is sent to its neighbor. Then, $[f]$ is an order g element of $\mathrm{Mod}(S)$. To see this, let α be the blue curve in Figure 2.4, and we see that $\alpha, f(\alpha), f^2(\alpha), \ldots, f^{g-1}(\alpha)$ are pairwise non-homotopic.

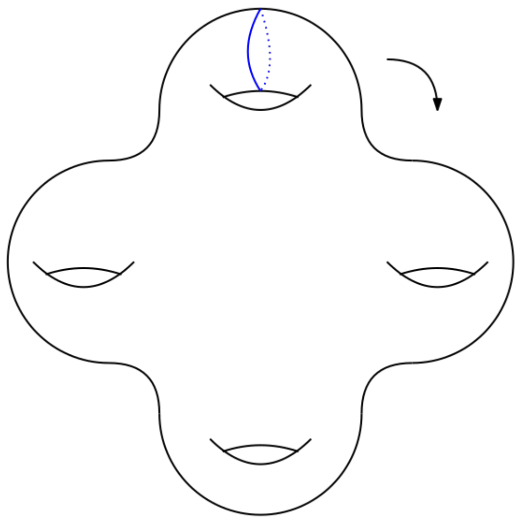

Figure 2.4: The blue curve α travels to each of the four handles under f.

Example 2.1.17. Suppose we view S_g as a regular $4g$-gon with its sides identified. We can rotate this $4g$-gon to get non-trivial mapping classes which permute the standard generators of $\pi_1(S_g)$.

A crucial example of a mapping class is called a Dehn twist. The idea behind a Dehn twist is simple: take a simple closed curve α on a surface S, cut S along α, twist along the cut, and glue S back together. We formalize this as follows.

Definition 2.1.18. Let $A = S^1 \times [0,1] \simeq S_{0,2}$ denote the annulus, and let $T : A \to A$ be the orientation-preserving homeomorphism $T(\theta, t) = (\theta + 2\pi t, t)$. Notice that T fixes ∂A pointwise. Now, let S be a surface and α a simple closed curve on S. Then, there exists a closed neighborhood $N \subseteq S$ of α with a homeomorphism $\phi : A \to N$. Define the map $T_\alpha : S \to S$ by

$$T_\alpha(x) = \begin{cases} (\phi \circ T \circ \phi^{-1})(x) & \text{if } x \in N \\ x & \text{if } x \notin N. \end{cases}$$

We call T_α the *Dehn twist of* S *about* α. See Figure 2.5 for an illustration.

Techinically, the definition of T_α depends on the choice of N and ϕ. However, T_α is well-defined up to homotopy. Moreover, one can show that for two simple closed curves on S, T_α is homotopic to T_β if and only if α is homotopic to β. So, given a homotopy class a of a simple closed curve on S, we get a well-defined mapping class T_a. Moving forward, we will identify a Dehn twist with its homotopy class without comment.

Now that we have seen some examples of individual mapping classes, we can see some mapping class groups. For instance, there are two mapping class groups which are easy to compute.

Proposition 2.1.19. *The group* $\mathrm{Mod}(S^2) = \mathrm{Mod}(S_{0,0})$ *is trivial.*

Proof. Let f be any homeomorphism of S^2. Up to homotopy, we may assume that f has a fixed point x_0 (for instance, one can choose x_0 arbitrarily and rotate S^2 along the great circle between x_0 and $f(x_0)$). We can identify $S^2 \setminus \{x_0\}$ with $\mathbb{R}^2$ via the stereographic projection. Then, any homeomorphism of $\mathbb{R}^2$ is homotopic to the identity map via the straight-line homotopy. It follows that f must be homotopic to id_{S^2}. $\square$

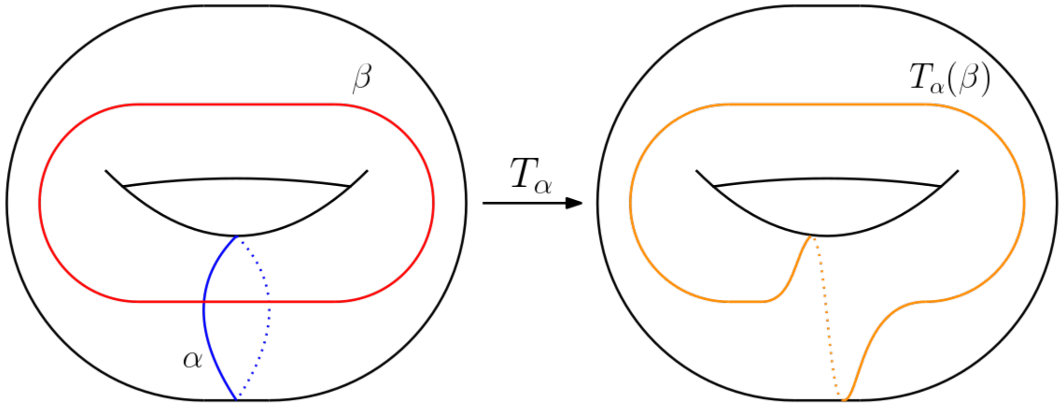

Figure 2.5: The orange curve is the image of β (red) under the Dehn twist about α (blue).

Proposition 2.1.20 (The Alexander Trick)**.** *Let D^2 denote the closed disk. Then,* $\mathrm{Mod}(D^2) = \mathrm{Mod}(S_{0,1})$ *is trivial.*

Proof. Identify D^2 with the closed unit disk in $\mathbb{R}^2$. Let $\phi : D^2 \to D^2$ be any homeomorphism of which fixes ∂D^2 pointwise. Then, define a homotopy $H : D^2 \times [0,1] \to D^2$ as follows. For $t \in [0,1)$, define

$$H(x,t) = \begin{cases} (1-t)\phi\left(\frac{x}{1-t}\right) & 0 \le |x| \le 1-t \\ x & 1-t \le |x| \le 1. \end{cases}$$

Then, define $H(x,1) = x$. Intuitively, at time t, H applies ϕ to the disk of radius $1-t$ and applies id_{D^2} outside this disk. This is a homotopy from ϕ to id_{D^2} which always fixes ∂D^2. □

As mentioned before, simple closed curves are a vital tool for studying the behavior of mapping classes. One particular instance of this notion is the Alexander Method. This is a relatively general theorem which allows us to reduce claims about mapping classes to combinatorial statements. We will not require the full generality of the theorem, so we state a special case here.

Theorem 2.1.21. *Let S be a closed surface. Suppose γ_1 and γ_2 are two curves in minimal position which fill S, meaning that $S \backslash (\gamma_1 \cup \gamma_2)$ is a union of disjoint union of disks. Then, there are only finitely many $f \in \mathrm{Mod}(S)$ which fix the set $\{\gamma_1, \gamma_2\}$ up to homotopy.*

Sketch of Proof. The full proof requires some careful consideration, but the main idea is as follows. We can view $\gamma_1 \cup \gamma_2$ as an embedded graph $\Gamma \subseteq S$ where the vertices are the intersection points of γ_1 and γ_2. Suppose that f fixes $\{\gamma_1, \gamma_2\}$ up to homotopy. One can show that f necessarily fixes Γ up to homotopy. It follows that, up to homotopy, f simply permutes the complementary regions of Γ. If $g \in \mathrm{Mod}(S)$ permutes the complementary regions of Γ in the same way, then since the mapping class groups on the disk is trivial, it must be that $f = g$. Since there are only finitely many complementary regions of Γ, there are only finitely many possibilities for f. □

2.1.4 Generating Mapping Class Groups

We mentioned before that Dehn twists are crucial examples of mapping classes. This is because of the following theorem, which we will prove shortly.

Theorem 2.1.22. *The group* $\mathrm{Mod}(S_{g,n})$ *is generated by Dehn twists.*

The idea behind the proof is to doubly induct on g and n. Therefore, there are a couple of base cases to establish first.

Proposition 2.1.23. *Let $A = S^1 \times [0,1] \simeq S_{0,2}$ be the annulus. Then,* $\mathrm{Mod}(A) \cong \mathbb{Z}$*. In particular,* $\mathrm{Mod}(A)$ *is generated by the Dehn twist about the central circle of A.*

The idea to prove this is to take any simple arc α (i.e. an injective continuous map $[0,1] \to A$) which connects the two boundary components of A. One shows that the homotopy class of α is determined by its winding number around the central boundary component. One formalizes and proves this claim using the oriented self-intersection number of α. Then, we let β be a straight arc in A of the form $\beta(t) = (x_0, t)$, so β has winding number 0. We define $\varphi : \mathrm{Mod}(A) \to \mathbb{Z}$ by mapping $[f]$ to the winding number of $f(\beta)$. The surjectivity of this map comes from taking powers of the Dehn twist about the central circle of A. The injectivity comes from the fact that if $\varphi([f]) = 0$, then up to homotopy f fixes β pointwise, so f is determined by its action on the surface obtained by cutting A along β. But this surface is a homeomorphic to a disk, which we know has a trivial mapping class group, and so f must be trivial.

Proposition 2.1.24. *Let $P = S_{0,3}$; we call P the* pair of pants*. Then,* $\mathrm{Mod}(P) \cong \mathbb{Z}^3$*. In particular,* $\mathrm{Mod}(P)$ *is generated by the three Dehn twists about each boundary component of P.*

This is proved similarly to the case of the annulus. First, one shows that the homotopy class of an arc connecting two boundary components is determined by its winding number around each boundary component. Then, choose a straight arc β connecting a pair of boundary components. We define $\varphi : \mathrm{Mod}(P) \to \mathbb{Z}^3$ by mapping $[f]$ to $(w_1(f(\beta)), w_2(f(\beta)), w_3(f(\beta)))$, where $w_i(\gamma)$ denotes the winding number of γ around the ith boundary component of P. Surjectivity is proved using Dehn twists, and injectivity is proved by cutting along β and getting an annulus.

We now have our base cases for the proof of Theorem 2.1.22. We just need one additional lemma.

Lemma 2.1.25. *Suppose a and b are two essential simple closed curves such that $i(a, b) = 1$. Then, $T_a T_b(a) = b$.*

Proof. The equality $T_a T_b(a) = b$ can be equivalently written as $T_b(a) = T_a^{-1}(b)$. If a and b are as in Figure 2.6, then this is immediate. For the general case, we can apply the change of coordinates principle to get a and b as in the special case. □

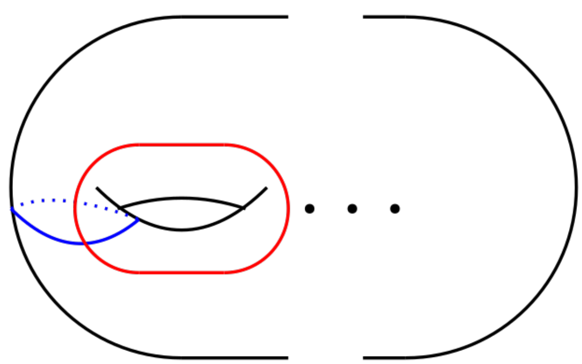

Figure 2.6: We map a and b to the red and blue curves.

Proof of Theorem 2.1.22. First, we will prove by induction on n that for all $n \geq 0$, $\mathrm{Mod}(S_{0,n})$ is generated by Dehn twists. The cases $n = 0$, 1, 2, and 3 are already taken care of. So, suppose $S = S_{0,n}$ for some $n \geq 4$. Take any $F = [f] \in \mathrm{Mod}(S)$. Note that every simple closed curve on a surface of genus 0 is separating. So, choose a simple closed curve α such that the cut surface S_α is a pair of pants and a copy of $S_{0,n-1}$. Let a denote the homotopy class of α. Then, we claim there exists $G = [g] \in \mathrm{Mod}(S)$ such that g is a product of Dehn twists and $i(a, G(F(a))) \leq 2$.

To see this, suppose that $i(a, F(a)) \geq 3$, and we will show that there exists a Dehn twist T such that $i(a, T(F(a))) < i(a, F(a))$. Since $i(a, F(a)) \geq 3$ and α is separating, we may assume that α and $f(\alpha)$ are positioned with three consecutive intersections as in the left side of Figure 2.7. We take T to be the Dehn twist around the red curve β, where β continues to follow $f(\alpha)$ outside of the image. The middle of Figure 2.7 depicts $T(f(\alpha))$. This is homotopic to the curve γ on the right side of Figure 2.7, which we assume follows $f(\alpha)$ outside the image. We see that γ intersects α two fewer times than $f(\alpha)$ within the image, and intersects α as many times as $f(\alpha)$ outside the image.

Thus, there exists a product of Dehn twists $G = [g] \in \mathrm{Mod}(S)$ such that $i(a, G(F(a))) \leq 2$. Since all simple closed curves on S are separating, this means $i(a, G(F(a))) = 0$ or $i(a, G(F(a))) = 2$. With the change of coordinates principle, one can show that if c and d are homotopy classes of simple closed curves on S such that $i(c, d) \in \{0, 2\}$ and $c \neq d$, then c and d surround different sets of boundary components. Since elements of $\mathrm{Mod}(S)$ fix boundary components, we know that a and $G(F(a))$ surround the same set of boundary components, and hence it must be that $a = G(F(a))$. Therefore, $G \circ F$ is determined by its action on $S \backslash \alpha$, which is a pair of pants and a copy of $S_{0,n-1}$. By induction, this means that $G \circ F$ is equal to a product of Dehn twists H, and hence $F = G^{-1} \circ H$. Thus, F is a product of Dehn twists.

Now, suppose that $S = S_{g,n}$ for $g \geq 1$. Choose any $F = [f] \in \mathrm{Mod}(S_{g,n})$, and let α be any non-separating essential simple closed curve. Then, $f(\alpha)$ is also a non-separating essential simple closed curve. Corollary 2.1.14 and Lemma 2.1.25 tell us that we can compose f with a product of Dehn twists to assume that f fixes α. Therefore, F is determined by its action on the cut surface S_α, which is a copy of $S_{g-1,n+2}$. By induction (i.e. with the inductive hypothesis that $\mathrm{Mod}(S_{g-1,m})$ is generated by Dehn twists for all m), we can conclude that F is a product of Dehn twists. $\square$

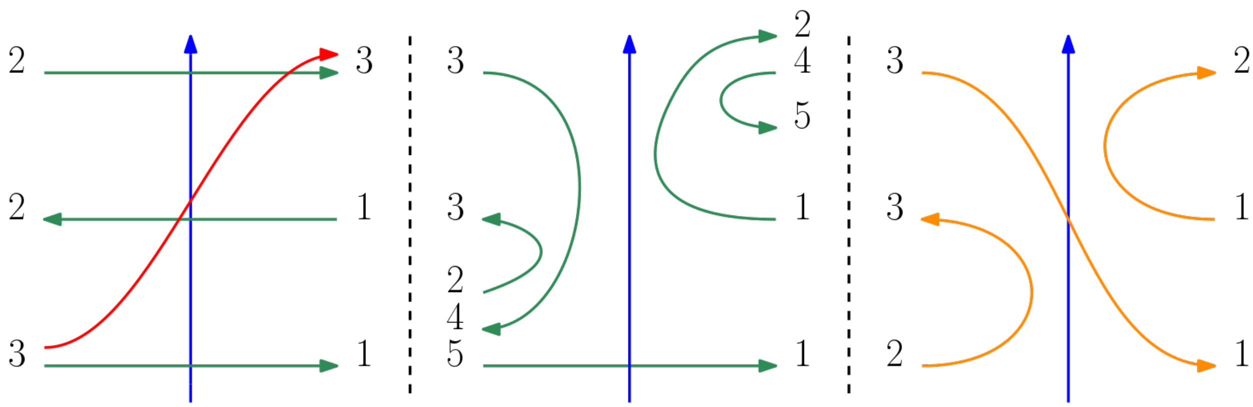

Figure 2.7: The left shows the initial configuration of α (blue), $f(\alpha)$ (green), and β (red). The middle shows $T(f(\alpha))$ (green) relative to α (blue). The right shows γ (orange) relative to α (blue). The numbers indicate the order that these curves enter and leave the image.

In the inductive step of this proof, we used our result about curve graphs. This was not strictly necessary; one can do the inductive step with a hands-on argument similar to how we did the base case. However, curve graphs are worth discussing as interesting objects in their own right. Moreover, curve graphs play a vital role in proving the following stronger theorem, which tells us that $\mathrm{Mod}(S)$ is finitely generated for closed surfaces.

Theorem 2.1.26 (Dehn-Lickorish). *For any $g \geq 0$, $\mathrm{Mod}(S_g)$ is generated by finitely many Dehn twists about non-separating curves, as illustrated in Figure 2.8.*

For a proof of the Dehn-Lickorish Theorem, see [2] Chapter 4.

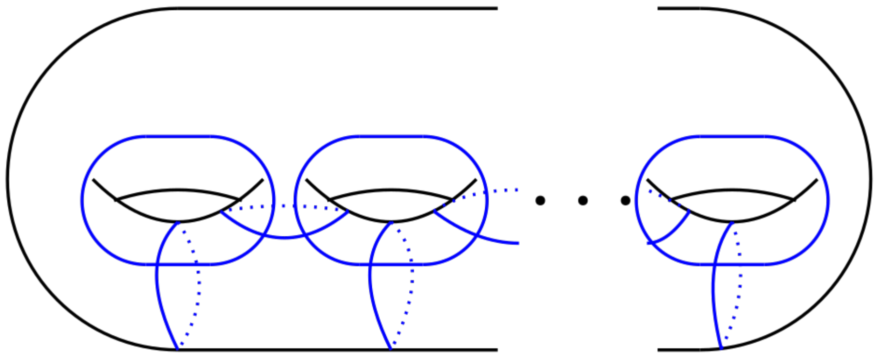

Figure 2.8: The Dehn-Lickorish generating set.

2.2 Teichmüller Space

2.2.1 Defining Teichmüller Space

Throughout the remainder of this chapter, we will assume all surfaces to be closed. Given a surface S, we will want to consider hyperbolic structures on S up to homotopy. That is, if we equip S with two different hyperbolic structures, we will want to call them equivalent if there is an isometry between these two "versions" of S which is homotopic to the identity map. We formalize this as follows.

Definition 2.2.1. Let S be a surface. A *marked hyperbolic structure on* S is a pair (X, ϕ) where X is a complete hyperbolic surface and $\phi : S \to X$ is a homeomorphism. The map ϕ is called a *marking*.

If we have a marked hyperbolic structure (X, ϕ), then we can think of X just as S endowed with a hyperbolic structure. A marked hyperbolic structure on S yields a hyperbolic structure directly on S by pulling back along the marking.

Definition 2.2.2. Let (X_1, ϕ_1) and (X_2, ϕ_2) be two marked hyperbolic structures on a surface S. We call (X_1, ϕ_1) and (X_2, ϕ_2) *equivalent* if there exists an orientation-preserving isometry $I : X_1 \to X_2$ such that $I \circ \phi_1$ and ϕ_2 are homotopic.

One reason we wish to consider marked hyperbolic structures is that an arbitrary hyperbolic surface X does not come with a canonical homeomorphism $S \to X$. However, a pair of marked hyperbolic structures always comes with a canonical homeomorphism between them.

Definition 2.2.3. Let (X_1, ϕ_1) and (X_2, ϕ_2) be two marked hyperbolic structures on a surface S. The *change of marking homeomorphism* is the map $\phi_2 \circ \phi_1^{-1} : X_1 \to X_2$.

Recall that a closed surface S admits a hyperbolic structure if and only if $\chi(S) < 0$. We will only define the Teichmüller space of S when $\chi(S) < 0$. However, one can reasonably adapt this definition to the torus by defining *flat structures*.

Definition 2.2.4. Let S be a surface with $\chi(S) < 0$. The *Teichmüller space of* S, denoted $\mathrm{Teich}(S)$, the set of equivalence classes $[(X, \phi)]$ of marked hyperbolic structures on S.

Of course, if we wish to refer to $\mathrm{Teich}(S)$ as a space, we will need to topologize it somehow. We will do this by identifying $\mathrm{Teich}(S)$ with a space of representations of $\pi_1(S)$.

For a surface S, we define $\mathrm{DF}(\pi_1(S), \mathrm{PSL}(2, \mathbb{R}))$ to be the set of discrete and faithful $\mathrm{PSL}(2, \mathbb{R})$-representations of $\pi_1(S)$, i.e. the set of injective homomorphisms $\pi_1(S) \to \mathrm{PSL}(2, \mathbb{R})$ with a discrete image. Then, $\mathrm{PGL}(2, \mathbb{R})$ acts on $\mathrm{DF}(\pi_1(S), \mathrm{PSL}(2, \mathbb{R}))$ by conjugation; namely, for a representation ρ and $A \in \mathrm{PGL}(2, \mathbb{R})$, we define $A \cdot \rho$ by $(A \cdot \rho)(g) := A\rho(g)A^{-1}$ for all $g \in \pi_1(S)$.

For a surface S with $\chi(S) < 0$, there is a natural bijective correspondence

$$\mathrm{Teich}(S) \longleftrightarrow \mathrm{DF}(\pi_1(S), \mathrm{PSL}(2, \mathbb{R}))/\mathrm{PGL}(2, \mathbb{R}).$$

For the map $\mathrm{Teich}(S) \to \mathrm{DF}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PGL}(2,\mathbb{R})$, we map $[(X,\phi)]$ to the conjugacy class of $\mathrm{hol}_X \circ \phi_*$, where ϕ_* is the induced map $\pi_1(S) \to \pi_1(X)$ and $\mathrm{hol}_X : \pi_1(X) \to \mathrm{PSL}(2,\mathbb{R})$ is the holonomy homomorphism of the hyperbolic structure on X. Of course, we have to check that this is well-defined independent of our choice of hol_X and independent of our representative of $[(X,\phi)]$. It follows from our definition of the holonomy that any two holonomy homomorphisms of the same hyperbolic structure differ by conjugation (more precisely, choosing a different initial chart of the developing map amounts to post-composing the initial chart by an element of $\mathrm{Isom}(\mathbb{H})$, which conjugates the holonomy homomorphism). So, it doesn't matter how we pick hol_X. Suppose we pick a different representative (Y,ψ) of $[(X,\phi)]$. Then, there exists an isometry $I : X \to Y$ such that $\psi \sim I \circ \phi$. Therefore, $\mathrm{hol}_Y \circ \psi_* = \mathrm{hol}_Y \circ I_* \circ \phi_*$. Then, $\mathrm{hol}_Y \circ I_*$ and hol_X are conjugate, since developing a curve α on X versus developing $I \circ \alpha$ on Y amounts to post-composing the charts on X by an isometry.

For the map $\mathrm{DF}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PGL}(2,\mathbb{R}) \to \mathrm{Teich}(S)$, take any $[\rho] \in \mathrm{DF}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PGL}(2,\mathbb{R})$ and let $G = \mathrm{Im}(\rho)$. Then, $\mathbb{H}/G$ is a hyperbolic surface with fundamental group G, and hence $\mathbb{H}/G$ is homeomorphic to S. So, we map $[\rho]$ to $[(\mathbb{H}/G, \phi)]$, where ϕ is chosen such that $\rho = \mathrm{hol}_{\mathbb{H}/g} \circ \phi_*$. One can show that such a map ϕ exists and that this map $\mathrm{DF}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PGL}(2,\mathbb{R}) \to \mathrm{Teich}(S)$ is well-defined using tools from algebraic topology (see [2], Proposition 10.2).

Perhaps it may seem unusual to quotient by $\mathrm{PGL}(2,\mathbb{R})$ instead of $\mathrm{PSL}(2,\mathbb{R})$. This is because we want to be able to conjugate a representation by any isometry, not just orientation-preserving ones. If we instead quotient $\mathrm{DF}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))$ by $\mathrm{PSL}(2,\mathbb{R})$, we instead get the space $\mathrm{Teich}(S) \sqcup \mathrm{Teich}(\bar{S})$, where $\bar{S}$ denotes S with the opposite orientation. We will return to this matter in the next chapter.

We give $\pi_1(S)$ the discrete topology and $\mathrm{PSL}(2,\mathbb{R})$ its usual topology (as a quotient of a subspace of $\mathbb{R}^4$). Since a homomorphism $\pi_1(S) \to \mathrm{PSL}(2,\mathbb{R})$ is determined by the images of its $2g$ generators, we can view $\mathrm{Hom}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))$ as a subset of $\mathrm{PSL}(2,\mathbb{R})^{2g}$ and give it the subspace topology. One can show that the resulting topology is independent of the chosen set of generators, and agrees with the compact-open topology. Then, we give $\mathrm{DF}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))$ the subspace topology and given $\mathrm{DF}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PGL}(2,\mathbb{R})$ the quotient topology. Finally, we topologize $\mathrm{Teich}(S)$ by declaring the bijection above to be a homeomorphism.

As mentioned before, one of the motivations of Teichmüller theory is to relate the topology of $\mathrm{Teich}(S)$ to the geometry of hyperbolic surfaces. A basic yet important example of this notion is the following.

Definition 2.2.5. Let S be a surface with $\chi(S) < 0$, and let $\mathcal{S} = \mathcal{S}(S)$ denote the set of (free) homotopy classes of essential simple closed curves on S. For a point $\mathcal{X} = [(X,\phi)] \in \mathrm{Teich}(S)$, we define the *length function of* $\mathcal{X}$ to be the map $\ell_{\mathcal{X}} : \mathcal{S} \to \mathbb{R}_+$ which maps $c \in \mathcal{S}$ to the length of the unique geodesic on X in the homotopy class of $\phi(c)$.

Recall that we proved the existence and uniqueness of geodesic representatives of simple closed curves in Proposition 2.1.3.

Proposition 2.2.6. *Let S be a surface with $\chi(S) < 0$ and let c be a homotopy class of an essential simple closed curve on S. Then, the function $\mathrm{Teich}(S) \to \mathbb{R}_+$ given by $\mathcal{X} \mapsto \ell_{\mathcal{X}}(c)$ is continuous.*

Proof. Let γ be the geodesic representative of c. For $\mathcal{X} = [(X,\phi)] \in \mathrm{Teich}(S)$, let $[\mathrm{hol}_X]$ be the corresponding element of $\mathrm{DF}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PGL}(2,\mathbb{R})$. Note that $\mathrm{hol}_X(\gamma)$ is hyperbolic since γ is essential. Since γ is a geodesic, it follows that if we lift γ to a curve $\widetilde{\gamma}$ in $\mathbb{H}$, then $\widetilde{\gamma}$ is in fact the axis of $\mathrm{hol}_X(\gamma)$. Therefore, the length of γ is equal to the translation length of $\mathrm{hol}_X(\gamma)$. So,

$$\ell_{\mathcal{X}}(c) = 2\cosh^{-1}\left(\frac{\mathrm{tr}(\mathrm{hol}_X(\gamma))}{2}\right).$$

Since the map $\mathrm{DF}(\pi_1(S), \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PGL}(2,\mathbb{R}) \to \mathbb{R}$ given by $[\rho] \mapsto \mathrm{tr}(\rho(\gamma))$ is continuous, the proposition follows. $\square$

2.2.2 Teichmüller's Theorem and the Teichmüller Metric

The Teichmüller space of a surface can be equivalently defined in terms of complex structures on the surface. Suppose S is a *Riemann surface*, meaning S is a one-dimensional complex manifold. If S has genus $g \geq 2$, then the Uniformization Theorem says that S is isomorphic to a quotient of $\mathbb{H}$ by some $G \subseteq \mathrm{Aut}(\mathbb{H})$. But as we know, $\mathrm{Aut}(\mathbb{H})$ is precisely the isometry group of $\mathbb{H}$ equipped with the hyperbolic metric. Hence, G induces a hyperbolic structure on S. Conversely, a hyperbolic structure on S induces a complex structure on S. It shoulded be noted this correspondence is non-constructive, so if you have particular hyperbolic structure, it will not tell you much about the corresponding complex structure.

We defined $\mathrm{Aut}(\mathbb{D})$ and $\mathrm{Aut}(\mathbb{H})$ to be groups of conformal (i.e. angle-preserving maps), but we also mentioned that a map is conformal if and only if it is holomorphic and has a non-vanishing derivative. For this reason, conformal maps are intimately linked with complex analysis. Indeed, the complex analysis viewpoint of Teichmüller space starts with the following.

Let $U, V \subseteq \mathbb{C}$ be open and let $f : U \to V$ be a homeomorphism which is smooth (as a map $\mathbb{R}^2 \to \mathbb{R}^2$) at all but finite-many points. We can switch between real coordinates (x, y) and a complex coordinate z by letting $z = x + iy$. If $f = u(x, y) + iv(x, y)$, then we can write

$$df = \begin{pmatrix} u_x & u_y \\ v_x & v_y \end{pmatrix}.$$

In real notation, we write $df = (u_x, v_x)dx + (u_y, v_y)dy$. In complex notation, we can write $df = f_z dz + f_{\bar{z}} d\bar{z}$ where

$$f_z = \frac{1}{2}((u_x + iv_x) - i(u_y + iv_y))$$
$$f_{\bar{z}} = \frac{1}{2}((u_x + iv_x) + i(u_y + iv_y)).$$

Definition 2.2.7. In the setup of the paragraph above, the quantity

$$\mu_f := \frac{f_{\bar{z}}}{f_z}$$

is called the *complex dilation of f*.

Note that the condition $f_{\bar{z}} \equiv 0$ is equivalent to the Cauchy-Riemann equations, so f is holomorphic if and only if $\mu_f \equiv 0$.

Definition 2.2.8. Let $U, V \subseteq \mathbb{C}$ be open, and let $f : U \to V$ be an orientation-preserving homeomorphism which is smooth at all but finitely many points. If f is differentiable at $p \in U$, we define the *dialation of f at p* to be the quantity

$$K_f(p) := \frac{1 + |\mu_f(p)|}{1 - |\mu_f(p)|}.$$

Then, we define the *dilation of f* to be the quantity

$$K(f) := \sup\{K_f(p) \in \mathbb{R} \mid f \text{ is differentiable at } p\}.$$

In particular, $K(f) \in [0, \infty]$. If $K(f) < \infty$, we call f a *$K(f)$-quasiconformal map.*

This formula may feel a bit arbitrary, but it has a rather geometric interpretation. At $p \in U$, the linear map df_p maps the unit circle to some ellipse $E_p \subseteq \mathbb{C}$ centered at the origin. We see that

$$|df_p(e^{i\theta})| = |f_z(p)e^{i\theta} + f_{\bar{z}}(p)e^{-i\theta}| = |f_z(p)||1 + \mu_f(p)e^{-i2\theta}|.$$

So, the maximum modulus of $|df_p(e^{i\theta})|$ as we range over θ is $|f_z(p)|(1+|\mu_f(p)|)$, while the minimum modulus is $|f_z(p)|(1 - |\mu_f(p)|)$. Thus, $K_f(p)$ measures the ratio of the major axis of E_p to the minor axis of E_p, and $K(f)$ take the supremum of this ratio as we range over all $p \in U$. So, we can think of a K-quasiconformal map as "preserving angles up to a factor of K". One can check that a map is 1-quasiconformal if and only if it is conformal.

If f is a map between two Riemann surfaces rather than just open subsets of $\mathbb{C}$, we can define the dilation of f using charts. In this case, how does the dilation of f change with homotopy? If we look at the homotopy class of f, is there a member of this class which minimizes the dilation? In the case of negative Euler characteristic, Teichmüller himself proved that the answer to this question is yes.

Theorem 2.2.9 (Teichmüller)**.** *Let X and Y be two Riemann surfaces of genus $g \geq 2$, and let $f : X \to Y$ be a homeomorphism. Let $\mathcal{F}$ be the set of quasiconformal maps homotopic to f, and let $c = \inf\{K(g) \in \mathbb{R}_+ \mid g \in \mathcal{F}\}$. Then, there exists a unique map $h \in \mathcal{F}$ such that $K(h) = c$.*

A similar but simpler result is known as Grötzsch's Problem.

Theorem 2.2.10 (Grötzsch's Problem)**.** *Suppose $X, Y \subseteq \mathbb{R}^2$ are the rectangles $[0, a] \times [0, 1]$ and $[0, Ka] \times [0, 1]$ respectively, for some $K \geq 1$. Suppose $f : X \to Y$ is an orientation-preserving homeomorphism which is smooth away from a finite number of points and takes horizontal and vertical sides to horizontal and vertical sides respectively. Then, $K_f \geq K$, with equality if and only if f is affine.*

The proof of Grötzsch's Problem, while not immediate, follows rather directly from the definition of K_f. Then, the main strategy in proving Teichmüller's Theorem is to reduce to Grötzsch's problem. For both of these proofs, see [2], Chapter 11.

Teichmüller's theorem allows us to define a metric on Teichmüller space. Let S be a surface with genus $g \geq 2$, and let $(X, \phi), (Y, \psi)$ represent two points $\mathcal{X}, \mathcal{Y} \in \text{Teich}(S)$. Let $f : X \to Y$ be the change of marking homeomorphism $f := \psi \circ \phi^{-1}$. Let h be the unique quasiconformal map of minimal dilation homotopic to f.

Definition 2.2.11. In the the setup of the above paragraph, we define the *Teichmüller metric* on Teich(S) by

$$d_{\text{Teich}}(\mathcal{X}, \mathcal{Y}) := \frac{1}{2} \log(K(h)).$$

One can show that the Teichmüller metric is compatible with the topology we have already put on Teich(S). As discussed before, one can extract geometric information about Riemann surfaces from geometric information about their Teichmüller spaces. This particular metric plays a role in proving the Nielsen-Thurston classification theorem, a major result which classifies elements of mapping class groups. However, this is not to suggest that d_{Teich} is the "canonical" metric on Teich(S); there are other, equally informative metrics one can use, including ones called the Weil-Petersson metric and the Thurston metric.

2.3 The Action of Mapping Class Groups on Teichmüller Space

2.3.1 Defining the Action and the Statement of Fricke's Theorem

In the last two sections, we studied seemingly disparate subjects; now, we can look at a bridge between mapping class groups and Teichmüller space. Indeed, if S is a surface equipped with a hyperbolic structure, then the action of $\mathrm{Mod}(S)$ on S should affect this structure. In this way, we should be able to extend the action of $\mathrm{Mod}(S)$ on S to an action on $\mathrm{Teich}(S)$. It is natural to wonder about the behavior of this action.

We define this action formally as follows. Let S be a surface of genus $g \geq 2$, let $\mathcal{X} = [(X, \phi)] \in \mathrm{Teich}(S)$, and let $F = [f] \in \mathrm{Mod}(S)$. Then we define

$$F \cdot \mathcal{X} := [(X, \phi \circ f^{-1})].$$

The use of f^{-1} ensures that we satisfy the axioms of a left action. Note that if $f_1, f_2 \in \mathrm{Homeo}^+(S)$ are homotopic and (X, ϕ) is a hyperbolic structure on S, then $(X, \phi \circ f_1^{-1})$ and $(X, \phi \circ f_2^{-1})$ are homotopic hyperbolic strucutres. Similarly, if (X_1, ϕ_1) and (X_2, ϕ_2) are homotopic hyperbolic structures and $f \in \mathrm{Homeo}^+(S)$, then $(X_1, \phi_1 \circ f^{-1})$ and $(X_2, \phi_2 \circ f^{-1})$ are homotopic. So, this action is well-defined.

Note that $\mathrm{Mod}(S)$ acts by isometries with respect to d_{Teich}. This is because if (X, ϕ) and (Y, ψ) are hyperbolic structures on S and $f \in \mathrm{Homeo}^+(S)$, then the changing of marking map from $(X, \phi \circ f^{-1})$ to $(Y, \psi \circ f^{-1})$ is $\psi \circ \phi^{-1}$. In other words, the action of $\mathrm{Mod}(S)$ does not affect the change of marking map. Similarly, we can see that the orbit of $[(X, \phi)] \in \mathrm{Teich}(S)$ is

$$\mathrm{Mod}(S) \cdot [(X, \phi)] = \{[(X, \phi \circ f^{-1})] \in \mathrm{Teich}(S) \mid f \in \mathrm{Homeo}^+(S)\}.$$

So, the orbit of $[(X, \phi)]$ is obtained by ranging over all possible markings of X.

Definition 2.3.1. Let S be a surface of genus $g \geq 2$. The *moduli space of* S, denoted $\mathcal{M}(S)$, is the quotient space

$$\mathcal{M}(S) := \mathrm{Teich}(S)/\mathrm{Mod}(S).$$

Since the orbit of a point in $\mathrm{Teich}(S)$ is obtained by ranging over all possible markings, we can think of $\mathcal{M}(S)$ as the space of oriented isometry classes of unmarked hyperbolic surfaces which are homeomorphic to S.

Now, it turns out that this action is, in some sense, "nice". More formally, recall that if a group G acts on a space X, we say that G acts properly discontinuously if for every compact subset $K \subseteq X$, the set $\{g \in G \mid g \cdot K \cap K\}$ is finite.

Theorem 2.3.2 (Fricke)**.** *Let S be a surface of genus $g \geq 2$. Then, the action of* $\mathrm{Mod}(S)$ *on* $\mathrm{Teich}(S)$ *is properly discontinuous.*

Here is one reason why we say that proper discontinuity makes this action "nice". Suppose that G is a group acting isometrically on a metric space (X, d). Then, the metric on X descends to a *pseudometric* $\bar{d}$ on X/G. Essentially, for $[x], [y] \in X/G$, we define

$$\bar{d}([x], [y]) := \inf_{\substack{x' \in [x] \\ y' \in [y]}} d(x', y').$$

We call this a pseudometric because in general, it satisfies all but one of the metric space axioms; namely, there may exist $[x], [y] \in X/G$ such that $[x] \neq [y]$ but $\bar{d}([x], [y]) = 0$. However, if G acts properly discontinuously, then this problem does not occur, and $\bar{d}$ will be a true metric. Therefore, Fricke's Theorem tells us that d_{Teich} descends to a metric on $\mathcal{M}(S)$.

2.3.2 Proof of Fricke's Theorem

The proof of Fricke's Theorem will require two main lemmas.

Lemma 2.3.3. *Let X be a complete hyperbolic surface, and let $\mathcal{S}$ denote the set of homotopy classes of essential simple closed curves in X. For $c \in \mathcal{S}$, let $\ell_X(c)$ denote the length of the unique geodesic in c. Then, for any $L \in \mathbb{R}$, the set $A(L) := \{c \in \mathcal{S} \mid \ell_X(c) \leq L\}$ is finite.*

Proof. Note that if $L \leq 0$, then $A(L) = \varnothing$. So, take any $L > 0$. We can view X as $\mathbb{H}/G$ where G is a Fuchsian group acting freely on $\mathbb{H}$. Since X is closed, we can choose a compact fundamental domain K for this action. Let K_L be the closed L-neighborhood of K. Any geodesic γ on X corresponds to a unique $g_\gamma \in G$ (i.e. its holonomy). Since X is closed, we know that every element of G is hyperbolic. Moreover, for any geodesic γ, its length $\ell_X(\gamma)$ is precisely the translation length λ_{g_γ} of g_γ. Therefore, the set $A(L)$ is in bijective correspondence with $\{g \in G \mid \lambda_g \leq L\}$. But this set is equal to $\{g \in G \mid g \cdot K_L \cap K_L \neq \varnothing\}$. Since G acts properly discontinuously on $\mathbb{H}$, we know that this last set is finite. $\square$

Lemma 2.3.4 (Wolpert's Lemma)**.** *Let X and Y be complete hyperbolic surfaces and $\phi : X \to Y$ a K-quasiconformal homeomorphism. Let c be a homotopy class of simple closed curves on X. Then,*

$$\frac{1}{K}\ell_X(c) \leq \ell_Y(\phi(c)) \leq K\ell_X(c).$$

It requires a bit of work to go through all the details of Wolpert's Lemma (see [2] Lemma 12.5 for the full proof), but the main idea is as follows. Let $\psi_1, \psi_2 \in \mathrm{Isom}(\mathbb{H})$ denote the holonomies of c and $\phi(c)$ respectively. Then, ψ_1 and ψ_2 are hyperbolic transformations, and hence $\mathbb{H}/\langle\psi_1\rangle$ and $\mathbb{H}/\langle\psi_2\rangle$ are annuli. One can show that these quotients are conformally equivalent to annuli A_1 and A_2 with circumference 1 whose (Euclidean) heights are $m_1 = \pi/\ell_X(c)$ and $m_2 = \pi/\ell_Y(\phi(c))$ respectively. Since ϕ lifts to a K-quasiconformal map $A_1 \to A_2$, one can modify the proof of Grötzsch's Problem to show that

$$\frac{1}{K}m_2 \leq m_1 \leq Km_2.$$

Thus, we get the inequality in the lemma.

The following corollary of Wolpert's Lemma follows directly from the definition of the Teichmüller metric.

Corollary 2.3.5. *Let S be a surface of genus $g \geq 2$. Suppose $\mathcal{X}, \mathcal{Y} \in \mathrm{Teich}(S)$ such that $d_{\mathrm{Teich}}(\mathcal{X}, \mathcal{Y}) \leq \frac{1}{2}\log(K)$ for some $K > 0$. Then, if c is a homotopy class of an essential simple closed curve in S, then*

$$\frac{1}{K}\ell_{\mathcal{X}}(c) \leq \ell_{\mathcal{Y}}(c) \leq K\ell_{\mathcal{X}}(c).$$

Armed with everything we've discussed in this chapter, the proof of Fricke's theorem is rather quick.

Proof of Fricke's Theorem. Let $B \subseteq \mathrm{Teich}(S)$ be compact, and let D denote the diameter of B (with respect to d_{Teich}). Take any $f \in \mathrm{Mod}(S)$ such that $f \cdot B \cap B \neq \varnothing$. Let $\mathcal{X} \in B$. So, $d_{\mathrm{Teich}}(\mathcal{X}, f \cdot \mathcal{X}) \leq 2D$. Choose homotopy classes c_1 and c_2 of essential simple closed curves in S which fill S, and let $L = \max\{\ell_{\mathcal{X}}(c_1), \ell_{\mathcal{X}}(c_1)\}$. By Corollary 2.3.5,

$$\ell_{\mathcal{X}}(f^{-1}(c_i)) = \ell_{f\cdot\mathcal{X}}(c_i) \leq KL$$

where $K = e^{4D}$ and $i \in \{1, 2\}$. By Lemma 2.3.3, this means there are only finitely many possibilities for $f^{-1}(c_i)$. Moreover, for a particular choice of $f^{-1}(c_i)$, Theorem 2.1.21 says that there are only finitely many possibilities for f^{-1}, and hence for f. Thus, there are only finitely many f such that $f \cdot B \cap B \neq \varnothing$. $\square$

Chapter 3

Representation Spaces of Surfaces

In the last chapter, we saw that the Teichmüller space of a surface S can be viewed as the set of $\mathrm{PGL}(2,\mathbb{R})$-conjugacy classes of discrete and faithful representations $\pi_1(S) \to \mathrm{PSL}(2,\mathbb{R})$. However, this means that $\mathrm{Teich}(S)$ lives inside the much bigger world of $\mathrm{PSL}(2,\mathbb{R})$-conjugacy classes of *all* representations $\pi_1(S) \to \mathrm{PSL}(2,\mathbb{R})$. Our first goal in this chapter is to figure out how exactly it sits within this space. In Section 3.1, we will see that this larger space has finitely-many connected components, one of which corresponds to $\mathrm{Teich}(S)$. To distinguish these components, we will define an invariant on the space $\mathrm{Hom}(\pi_1(S),\mathrm{PSL}(2,\mathbb{R}))$ called the Euler class.

We also saw in the last chapter that the mapping class group $\mathrm{Mod}(S)$ acts "nicely" on $\mathrm{Teich}(S)$. In Section 3.2, we will see how this action extends to the larger representation space containing $\mathrm{Teich}(S)$. A conjecture of Goldman says that the action is "chaotic" on the rest of the space. As discussed in the introduction, Marché and Wolff proved Goldman's conjecture in the case that S has genus 2 by answering an older question of Bowditch. We will end this chapter by outlining the connection between Bowditch's question and Goldman's conjecture.

For more information on the Euler class and the connected components of $\mathrm{Hom}(\pi_1(S),\mathrm{PSL}(2,\mathbb{R}))$, our main reference is Goldman's paper [5]. Palesi's paper [11] provides a more introductory exposition to the subject, and discusses how Goldman's results can be extended to non-orientable surfaces. Marché and Wolff's result can be found in their paper [9].

3.1 $\mathrm{PSL}(2,\mathbb{R})$-Representation Spaces

3.1.1 The Representation Variety and its Quotients

Given a closed surface S of genus $g \geq 2$, we saw in the last chapter that there is a natural bijective correspondence

$$\mathrm{Teich}(S) \longleftrightarrow \mathrm{DF}(\pi_1(S),\mathrm{PSL}(2,\mathbb{R}))/\mathrm{PGL}(2,\mathbb{R}).$$

This is because $\mathrm{PSL}(2,\mathbb{R}) \cong \mathrm{Isom}^+(\mathbb{H})$ and $\mathrm{PGL}(2,\mathbb{R}) \cong \mathrm{Isom}(\mathbb{H})$. Recall that the forward direction of this correspondence comes from taking the holonomy of a hyperbolic structure. Moreover, we have a proper embedding

$$\mathrm{Teich}(S) \hookrightarrow \mathrm{DF}(\pi_1(S),\mathrm{PSL}(2,\mathbb{R}))/\mathrm{PSL}(2,\mathbb{R}).$$

In fact, as we mentioned before, $\mathrm{DF}(\pi_1(S),\mathrm{PSL}(2,\mathbb{R}))/\mathrm{PSL}(2,\mathbb{R})$ is the space $\mathrm{Teich}(S) \sqcup \mathrm{Teich}(\bar{S})$, where $\bar{S}$ is S with the opposite orientation. This means we also have an embedding of $\mathrm{Teich}(S)$ into a much larger space:

$$\mathrm{Teich}(S) \hookrightarrow \mathrm{Hom}(\pi_1(S),\mathrm{PSL}(2,\mathbb{R}))/\mathrm{PSL}(2,\mathbb{R}).$$

Our goal in this section is to see how exactly $\mathrm{Teich}(S)$ fits into this larger world. We will start by taking a bit of time to understand the space $\mathrm{Hom}(\pi_1(S),\mathrm{PSL}(2,\mathbb{R}))/\mathrm{PSL}(2,\mathbb{R})$.

Definition 3.1.1. Let Γ be a finitely-generated group and let G be a Lie group. The *G-representation variety of* Γ is the space $\mathcal{R}(\Gamma, G) := \mathrm{Hom}(\Gamma, G)$. If $\Gamma = \pi_1(S)$ for some surface S, then we define the *G-representation variety of S* as $\mathcal{R}(S,G) := \mathcal{R}(\pi_1(S), G)$.

If one is interested in the more general theory of representation spaces, then one needs to place additional restrictions on G (namely, G should be something called an *algebraic reductive Lie group*), but any groups we discuss will satisfy these restrictions.

If Γ is generated by $X_1, \ldots, X_n$, then we can identify $\rho \in \mathcal{R}(\Gamma, G)$ with the n-tuple $(\rho(X_1), \ldots, \rho(X_n))$. With this correspondence, we can view the representation variety as the subset of G^n consisting of n-tuples which satisfy the relations of Γ. These relations can be viewed as polynomials in n variables, which explains the usage of the word "variety." We topologize the representation variety in the same way we did Teichmüller space, which again agrees with the compact-open topology.

It seems that we would like to study the space $\mathcal{R}(\Gamma, G)/G$, where G acts on $\mathcal{R}(\Gamma, G)$ by conjugation. Unfortunately, this space is not always well-behaved. In particular, it need not be Hausdorff. To see this, suppose that $\Gamma = \mathbb{Z}$ and $G = \mathrm{SL}(2, \mathbb{R})$, and define the representation $\rho : \mathbb{Z} \to \mathrm{SL}(2, \mathbb{R})$ by

$$1 \mapsto \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Notice that for any nonzero $t \in \mathbb{R}$, we have

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

This shows that the trivial representation is an accumulation point of the $\mathrm{SL}(2, \mathbb{R})$-conjugation orbit of ρ in $\mathcal{R}(\mathbb{Z}, \mathrm{SL}(2, \mathbb{R}))$, but it does not belong to the orbit since the identity matrix is the only member of its conjugacy class. So, $[\rho]$ cannot be separated from the trivial representation in $\mathcal{R}(\mathbb{Z}, \mathrm{SL}(2, \mathbb{R}))/\mathrm{SL}(2, \mathbb{R})$. The same issue arises in $\mathcal{R}(\mathbb{Z}, \mathrm{PSL}(2, \mathbb{R}))$. Another issue arises in $\mathcal{R}(S, \mathrm{PSL}(2, \mathbb{R}))/\mathrm{PSL}(2, \mathbb{R})$ with representations of Euler class 0 in the case that S is a closed surface of genus 2, which we will discuss in the next section. To fix these issues, we will need to work with a slightly different space.

Definition 3.1.2. We call a subgroup $G \subseteq \mathrm{PSL}(2, \mathbb{R})$ *elementary* if its action on $\mathbb{H} \cup \partial\mathbb{H}$ has a finite orbit. We call a representation $\rho : \Gamma \to \mathrm{PSL}(2, \mathbb{R})$ *elementary* if $\rho(\Gamma)$ is elementary.

Definition 3.1.3. Let Γ be a finitely-generated group, and let $G = \mathrm{PSL}(2, \mathbb{R})$. Define $\mathcal{R}_{ne}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))$ to be the subset of $\mathcal{R}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))$ consisting of non-elementary representations. We define the space

$$\mathfrak{X}(\Gamma) := \mathcal{R}_{ne}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))/\mathrm{PSL}(2, \mathbb{R}),$$

where $\mathrm{PSL}(2, \mathbb{R})$ acts on $\mathcal{R}_{ne}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))$ by conjugation. If $\Gamma = \pi_1(S)$ for some surface S, then we define $\mathfrak{X}(S) := \mathfrak{X}(\pi_1(S))$.

In particular, we have that $\mathfrak{X}(\Gamma)$ is an open subset of $\mathcal{R}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))/\mathrm{PSL}(2, \mathbb{R})$ [7]. In fact, if S is a surface of genus $g \geq 2$, we have that $\mathrm{Teich}(S) \hookrightarrow \mathfrak{X}(S)$, since discrete and faithful representations $\pi_1(S) \to \mathrm{PSL}(2, \mathbb{R})$ are necessarily non-elementary. Notice that $\mathfrak{X}(\Gamma)$ solves the issue above, since the trivial representation is elementary. In fact, one can show that the space $\mathfrak{X}(\Gamma)$ is Hausdorff.

In some sense, the space $\mathfrak{X}(\Gamma)$ may seem like an ad hoc fix, but there is a much deeper theory lurking in the background. The space $\mathfrak{X}(\Gamma)$ sits inside a space called the *polystable quotient* of $\mathcal{R}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))$, which is related to an object called the $\mathrm{PSL}(2, \mathbb{R})$*-character variety of* Γ. The $\mathrm{PSL}(2, \mathbb{R})$-character variety is obtained as the *geometric invariant theory quotient*, or *GIT quotient*, denoted $\mathcal{R}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))//\mathrm{PSL}(2, \mathbb{R})$. These other objects are defined with the language of representation theory and algebraic geometry, and concern the structure of $\mathcal{R}(\Gamma, \mathrm{PSL}(2, \mathbb{R}))$ as a variety. For a survey of these spaces, see for example [13].

3.1.2 The Euler Class

Recall once more that for us, surfaces are compact, connected, and oriented. In this chapter, we are no longer assuming that all surfaces are closed.

To help get a handle on the space $\mathfrak{X}(S)$, we will define a useful invariant called the Euler class. There are a couple different ways to define this invariant with varying levels of abstraction; we will opt for the most "concrete" definition. In preparation, we will need to recall some facts about covering spaces of groups.

Proposition 3.1.4. *Suppose G is a path-connected and locally path-connected topological group. Suppose $p : \widetilde{G} \to G$ is a path-connected and locally path-connected covering space. Let $e \in G$ be the identity element and choose $\widetilde{e} \in p^{-1}(e)$. Then, there is a unique topological group structure on $\widetilde{G}$ such that $\widetilde{e}$ is the identity element and p is a homomorphism.*

Proof. Let $m : G \times G \to G$ be the multiplication map, and let $p \times p : \widetilde{G} \times \widetilde{G} \to G \times G$ be the product map (so $p \times p$ is a covering map). Let $\varphi : \widetilde{G} \times \widetilde{G} \to G$ be the composition $m \circ (p \times p)$. Let $p_* : \pi_1(\widetilde{G}) \to \pi_1(G)$ be the induced homomorphism. Note that $\pi_1(\widetilde{G} \times \widetilde{G}) = \pi_1(\widetilde{G} \times \{\widetilde{e}\}) \times \pi_1(\{\widetilde{e}\} \times \widetilde{G})$, and each factor (and hence the whole group) lands in $p_*(\pi_1(\widetilde{G}))$ under the induced homomorphism $\varphi_* : \pi_1(\widetilde{G} \times \widetilde{G}) \to \pi_1(G)$. So, there is a unique lift $\widetilde{m} : \widetilde{G} \times \widetilde{G} \to \widetilde{G}$ of φ such that $\widetilde{m}(\widetilde{e}, \widetilde{e}) = \widetilde{e}$. One can show that $\widetilde{m}$ satisfies the group axioms using uniqueness of path lifts. The fact that p is a homomorphism follows from the commutativity of the following diagram:

$$\begin{tikzcd}
\widetilde{G} \times \widetilde{G} \arrow[r, "\widetilde{m}"] \arrow[d, "p \times p"] & \widetilde{G} \arrow[d, "p"] \\
G \times G \arrow[r, "m"] & G
\end{tikzcd}$$

□

Proposition 3.1.5. *If G is a connected topological group and $H \subseteq G$ is a discrete normal subgroup, then $H \subseteq Z(G)$.*

Proof. Take any $h \in H$, and define the map $\varphi_h : G \to H$ by $\varphi_h(g) = ghg^{-1}$. Since H is normal, the codomain of φ_h is indeed H. Since φ_h is continuous and G is connected, $\varphi_h(G)$ is connected. Since H is discrete, this means $\varphi_h(G)$ must be a single point. Since $\varphi_h(e) = h$, we can conclude that $\varphi_h(G) = \{h\}$. This means that $ghg^{-1} = h$ for all $g \in G$, and hence $h \in Z(G)$. □

Now, let $p : \widetilde{\mathrm{PSL}}(2, \mathbb{R}) \to \mathrm{PSL}(2, \mathbb{R})$ denote the universal cover. We equip $\widetilde{\mathrm{PSL}}(2, \mathbb{R})$ with a group structure as above, and let Z denote its center. Then, we have the following key fact.

Proposition 3.1.6. *Let $p : \widetilde{\mathrm{PSL}}(2, \mathbb{R}) \to \mathrm{PSL}(2, \mathbb{R})$ be the universal cover of $\mathrm{PSL}(2, \mathbb{R})$. Then, $Z = \mathrm{Ker}(p)$ and $Z \cong \mathbb{Z}$.*

Proof. Since $\mathrm{Ker}(p)$ is a discrete normal subgroup of $\widetilde{\mathrm{PSL}}(2, \mathbb{R})$, this tells us that $\mathrm{Ker}(p) \subseteq Z$. On the other hand, $p(Z) \subseteq Z(\mathrm{PSL}(2, \mathbb{R}))$ since p is a homomorphism, and since $Z(\mathrm{PSL}(2, \mathbb{R}))$ is trivial, this implies that $Z \subseteq \mathrm{Ker}(p)$.

Now, we can show that $\mathrm{Ker}(p)$ is isomorphic to the group D of deck transformations of $\widetilde{\mathrm{PSL}}(2, \mathbb{R})$. For $g \in \mathrm{PSL}(2, \mathbb{R})$ and $\widetilde{g} \in \widetilde{\mathrm{PSL}}(2, \mathbb{R})$, let L_g and $L_{\widetilde{g}}$ denote the respective left multiplication maps. So, if $\widetilde{g} \in p^{-1}(g)$, then $L_{\widetilde{g}}$ is a lift of $L_g \circ p$. That means that if $z \in \mathrm{Ker}(p)$, then $L_z \in D$. Furthermore, if $\varphi \in D$ maps $\widetilde{e}$ to $z \in \mathrm{Ker}(p)$, then $L_{z^{-1}} \circ \varphi$ is a deck transformation which fixes $\widetilde{e}$, and therefore must be trivial, so it must be that $\varphi = L_z$.

So, we have that $Z = \mathrm{Ker}(p) \cong D$. Since $\widetilde{\mathrm{PSL}}(2, \mathbb{R})$ is the universal cover, we know that $D \cong \pi_1(\mathrm{PSL}(2, \mathbb{R}))$. Then, any $A \in \mathrm{SL}(2, \mathbb{R})$ has a polar decomposition RP where $R \in \mathrm{SO}(2)$ and $P \in \mathrm{SL}(2, \mathbb{R})$ is symmetric and positive-definite. One can show that the set of symmetric positive-definite elements of $\mathrm{SL}(2, \mathbb{R})$ is contractible, which yields a deformation retract of $\mathrm{PSL}(2, \mathbb{R})$ onto $\mathrm{PSO}(2)$, which is homotopy equivalent to S^1. So, $\pi_1(\mathrm{PSL}(2, \mathbb{R})) \cong \pi_1(\mathrm{PSO}(2)) \cong \mathbb{Z}$, as desired. □

So, we can fix a generator z of Z. Fix a closed surface S of genus $g \geq 2$. We can write

$$\pi_1(S) = \left\langle X_1, Y_1, \ldots, X_g, Y_g \mid \prod_{i=1}^{g} [X_i, Y_i] \right\rangle.$$

We define the map $R_g : \mathrm{PSL}(2, \mathbb{R})^{2g} \to \widetilde{\mathrm{PSL}}(2, \mathbb{R})$ by

$$R_g(A_1, B_1, \ldots, A_g, B_g) = \prod_{i=1}^{g} \left[\widetilde{A}_i, \widetilde{B}_i \right]$$

where $\widetilde{A}_i$ and $\widetilde{B}_i$ denote arbitrary lifts of A_i and B_i respectively. This map is well-defined, since any two lifts of A_i or B_i will differ by a central element of $\widetilde{\mathrm{PSL}}(2,\mathbb{R})$, which will vanish in the commutators. More specifically, suppose we have another lift $\widetilde{A}'_i$ of A_i. Then, there exists a deck transformation φ such that $\varphi(\widetilde{A}_i) = \widetilde{A}'_i$. As we saw in the proof of Proposition 3.1.6, φ corresponds to left multiplication by some $g \in \mathrm{Ker}(p) = Z$. So,

$$\left[\widetilde{A}'_i, \widetilde{B}_i\right] = \widetilde{A}'_i\widetilde{B}_i\widetilde{A}'^{-1}_i\widetilde{B}^{-1}_i = g\widetilde{A}_i\widetilde{B}_i\widetilde{A}^{-1}_i g^{-1}\widetilde{B}^{-1}_i = gg^{-1}\widetilde{A}_i\widetilde{B}_i\widetilde{A}^{-1}_i\widetilde{B}^{-1}_i = \left[\widetilde{A}_i, \widetilde{B}_i\right].$$

We use the map R_g to define the Euler class.

Definition 3.1.7. Let S be a closed surface of genus g. The *Euler class* is a map

$$e : \mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R})) \to \mathbb{Z}$$

defined as follows. Let $X_1, Y_1, \ldots, X_g, Y_g$ denote the generators of $\pi_1(S)$ as above. Then, for any $\rho \in \mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R}))$, we have that

$$R_g(\rho(X_1), \rho(Y_1), \ldots, \rho(X_g), \rho(Y_g)) = z^k$$

for some $k \in \mathbb{Z}$. We define $e(\rho) := k$.

The following classical result gives a nice bound on the Euler class.

Theorem 3.1.8 (Milnor-Wood Inequality)**.** *Let S be a closed surface of genus $g \geq 2$. Then, for any $\rho \in \mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R}))$, $|e(\rho)| \leq |\chi(S)| = 2g-2$.*

The idea to prove this result is to decompose S into pairs of pants, prove the inequality on these subsurfaces, and observe how the bound changes when we glue the pants back together. In order to do this, we need to define the Euler class for surfaces with boundary.

Suppose now $S = S_{g,n}$ and $\chi(S) < 0$. We can write

$$\pi_1(S) = \left\langle X_1, Y_1, \ldots, X_g, Y_g, C_1, \ldots, C_n \mid \left(\prod_{i=1}^{g}[X_i, Y_i]\right) C_1 \cdots C_n \right\rangle.$$

It seems like we could define the Euler class in the same way that we defined it in the closed case; namely, take a lift of our relation and see which central element we get in $\widetilde{\mathrm{PSL}}(2,\mathbb{R})$. In the closed case, however, we were only lifting commutators; this allowed us to take arbitrary lifts, since any differences would get cancelled out. If we try to play the same game in this case, then the result depends on how we choose the lifts of the boundary elements C_i. So, we have to be able to choose a canonical lift of these elements.

The action of $\mathrm{PSL}(2,\mathbb{R})$ on $\mathbb{H}^2$ yields an action on $\partial\mathbb{H}^2 \simeq S^1$. This lifts to an action of $\widetilde{\mathrm{PSL}}(2,\mathbb{R})$ on the universal cover $\widetilde{\partial\mathbb{H}^2} \simeq \mathbb{R}$. So, if $T \in \mathrm{PSL}(2,\mathbb{R})$ has a fixed point on $\partial\mathbb{H}^2$, there is a unique lift $\widetilde{T}$ with a fixed point on $\widetilde{\partial\mathbb{H}}$. We call $\widetilde{T}$ the *canonical lift* of T. In particular, T has a canonical lift if and only if T is not elliptic. Note that $(\widetilde{T})^{-1} = \widetilde{(T^{-1})}$.

So, let $\mathcal{NE} \subseteq \mathrm{PSL}(2,\mathbb{R})$ denote the set of non-elliptic elements. We define a map

$$R_{g,n} : \mathrm{PSL}(2,\mathbb{R})^{2g} \times \mathcal{NE}^n \to \widetilde{\mathrm{PSL}}(2,\mathbb{R})$$

by setting

$$R_{g,n}(A_1, B_1, \ldots, A_g, B_g, C_1, \ldots, C_n) = \left(\prod_{i=1}^{g}\left[\widetilde{A}_i, \widetilde{B}_i\right]\right)\widetilde{C}_1 \cdots \widetilde{C}_n$$

where $\widetilde{A}_i$ and $\widetilde{B}_i$ are arbitrary lifts of A_i and B_i and $\widetilde{C}_i$ are the canonical lifts of C_i.

Definition 3.1.9. Let $S = S_{g,n}$ with $\chi(S) < 0$. We define the set

$$W(S) = \{\rho \in \mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R})) \mid \rho(C_i) \text{ is not elliptic for all } i\}.$$

The *relative Euler class* is a map $e : W(S) \to \mathbb{Z}$ defined as follows. Let $X_1, Y_1, \ldots, X_g, Y_g, C_1, \ldots, C_n$ denote the generators of $\pi_1(S)$ as above. For any $\rho \in W(S)$, we have that

$$R_{g,n}(\rho(X_1), \rho(Y_1), \ldots, \rho(X_g), \rho(Y_g), \rho(C_1), \ldots, \rho(C_n)) = z^k$$

for some $k \in \mathbb{Z}$. We define $e(\rho) := k$.

Note that the relative Euler class agrees with the Euler class in the case of closed surfaces. To prove the Milnor-Wood inequality, one has to start with the following base case.

Theorem 3.1.10 ([5]). *Let $P = S_{0,3}$ be a pair of pants. Then, for any $\rho \in W(P)$, $|e(\rho)| \leq 1 = |\chi(P)|$.*

Next, we can prove that the Euler class is additive when we glue pants together.

Proposition 3.1.11. *Let P_1 and P_2 be pairs of pants with boundary components C_1, C_2, C_3 and D_1, D_2, D_3 respectively. Suppose S is obtained by gluing C_3 and D_1 together, and let C be the curve on S corresponding to C_3 and D_1. Then, if $\rho \in W(S)$ such that $\rho(C)$ is not elliptic, then $e(\rho) = e(\rho|_{\pi_1(P_1)}) + e(\rho|_{\pi_1(P_2)})$.*

Proof. Take any $\rho \in W(S)$ such that $\rho(C)$ is not elliptic. We can write

$$\pi_1(S) = \langle C_1, C_2, D_2, D_3 \mid C_1 C_2 D_2 D_3 \rangle,$$

where C_i and D_i are the boundary components coming from P_1 and P_2 respectively. Assume C is oriented such that $C = C_1 C_2$. Note that $\pi_1(S)$ is obtained as the amalgamated free product of $\pi_1(P_1)$ and $\pi_1(P_2)$. In particular, $\pi_1(P_1) = \langle C_1, C_2, C^{-1} \rangle$ and $\pi_1(P_2) = \langle C, D_2, D_3 \rangle$. Then, we have that

$$z^{e(\rho)} = \widetilde{C_1}\widetilde{C_2}\widetilde{D_2}\widetilde{D_3} = (\widetilde{C_1}\widetilde{C_2}\widetilde{C^{-1}})(\widetilde{C}\widetilde{D_2}\widetilde{D_3}) = z^{e(\rho|_{\pi_1(P_1)})} z^{e(\rho|_{\pi_1(P_2)})}.$$

Thus, we have that $e(\rho) = e(\rho|_{\pi_1(P_1)}) + e(\rho|_{\pi_1(P_2)})$. □

Then, we can check that the Euler class is invariant when we glue two boundary components of the same surface.

Proposition 3.1.12. *Let $S = S_{g,n}$ with $\chi(S) < 0$ and $n \geq 2$. Let S' be obtained by gluing together two boundary components C_1 and C_2 of S, let $f : S \to S'$ be the gluing map, and let $f_* : \pi_1(S) \to \pi_1(S')$ be the induced map. If $\rho \in W(S')$ such that $\rho(D)$ is not elliptic, where D is the curve on S' obtained from C_1 and C_2, then $e(\rho) = e(\rho \circ f_*)$.*

Proof. The fundamental group $\pi_1(S')$ is obtained from $\pi_1(S)$ via an HNN extension. That is, if we write

$$\pi_1(S) = \left\langle X_1, Y_1, \ldots, X_g, Y_g, C_1, \ldots, C_n \mid \left(\prod_{i=1}^{g} [X_i, Y_i] \right) C_1 \cdots C_n \right\rangle$$

$$\pi_1(S') = \left\langle X_1, Y_1, \ldots, X_g, Y_g, X_{g+1}, Y_{g+1}, C_3, \ldots, C_n \mid \left(\prod_{i=1}^{g+1} [X_i, Y_i] \right) C_3 \cdots C_n \right\rangle,$$

then induced map f_* is given by maps C_1 to X_{g+1}, C_2 to $Y_{g+1} X_{g+1}^{-1} Y_{g+1}^{-1}$, and preserves the rest of the generators. It follows that

$$\begin{aligned} z^{e(\rho)} &= \left[\widetilde{\rho(X_1)}, \widetilde{\rho(Y_1)}\right] \cdots \left[\widetilde{\rho(X_g)}, \widetilde{\rho(Y_g)}\right] \widetilde{\rho(X_{g+1})} \left(\widetilde{\rho(Y_{g+1})\rho(X_{g+1}^{-1})\rho(Y_{g+1}^{-1})} \right) \widetilde{\rho(C_3)} \cdots \widetilde{\rho(C_n)} \\ &= \left[\rho(\widetilde{f_*(X_1)}), \rho(\widetilde{f_*(Y_1)})\right] \cdots \left[\rho(\widetilde{f_*(X_g)}), \rho(\widetilde{f_*(Y_g)})\right] \rho(\widetilde{f_*(C_1)}) \rho(\widetilde{f_*(C_2)}) \rho(\widetilde{f_*(C_3)}) \cdots \rho(\widetilde{f_*(C_n)}) \\ &= z^{e(\rho \circ f_*)} \end{aligned}$$

□

Any surface S with $\chi(S) < 0$ can be obtained by gluing together pairs of pants in the ways described by Propositions 3.1.11 and 3.1.12. Therefore, to prove the Milnor-Wood inequality, it only remains to show that any representation $\rho \in W(S)$ can be modified without changing the Euler class so that it maps the curves defining this pair of pants decomposition to non-elliptic elements (see [5]).

3.1.3 Classification of Connected Components

Let S be a surface with $\chi(S) < 0$. We have seen that Euler class is a map $e : W(S) \to \mathbb{Z}$ which is bounded above and below by $-\chi(S)$ and $\chi(S)$ respectively. Now, we can see how the Euler class is a useful topological invariant. We can make the following immediate observation.

Proposition 3.1.13. *Let $S = S_{g,n}$ with $\chi(S) < 0$. The Euler class is constant on path-components of $W(S)$.*

Proof. Suppose ρ_t for $t \in [0,1]$ is a path in $W(S)$. Then,

$$\widetilde{\gamma}(t) := R_{g,n}(\rho_t(X_1), \dots, \rho_t(Y_g), \rho(C_1), \dots, \rho(C_n))$$

is a path in $\widetilde{\mathrm{PSL}}(2,\mathbb{R})$. In particular, $\widetilde{\gamma}(t)$ is a lift of the constant path

$$\gamma(t) := \left(\prod_{i=1}^{g} [\rho_t(X_i), \rho_t(Y_i)] \right) \rho(C_1) \cdots \rho(C_n)$$

in $\mathrm{PSL}(2,\mathbb{R})$, and hence $\widetilde{\gamma}(t)$ must be constant. This implies that $e(\rho_t)$ is constant. $\square$

In fact, there is a much stronger statement one can make relating the Euler class to path components.

Theorem 3.1.14 (Goldman, [5]). *Let $S = S_{g,n}$ with $\chi(S) < 0$.*

(i) For $k \in \{\chi(S), \dots, -\chi(S)\}$, $e^{-1}(k)$ is non-empty and connected.

(ii) The path-components of $W(S)$ are precisely the pre-images $e^{-1}(k)$ for $\chi(S) \leq k \leq -\chi(S)$. In particular, $W(S)$ has $2|\chi(S)| + 1$ connected components.

(iii) A representation $\rho \in W(S)$ satisfies $|e(\rho)| = |\chi(S)|$ if and only if ρ is discrete and faithful.

The proof of (i) is similar in spirit to the proof the Milnor-Wood inequality. Namely, one first proves the result for the case that S is a pair of pants. For the general case, one decomposes S into pairs of pants and studies the behavior of representations as the pants are glued together.

All of our results thus far have been stated for the ordinary representation variety $\mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R}))$. Now, we can see that they descened to $\mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PSL}(2,\mathbb{R})$.

Proposition 3.1.15. *Let S be a closed surface of genus $g \geq 2$, and let $\rho \in \mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R}))$. Then, for any $h \in \mathrm{PSL}(2,\mathbb{R})$, $e(h \cdot \rho) = e(\rho)$, where $h \cdot \rho$ denotes the conjugation action.*

To prove this, we need a quick lemma regarding lifts of elements of $\mathrm{PSL}(2,\mathbb{R})$.

Lemma 3.1.16. *Let $g, h \in \mathrm{PSL}(2,\mathbb{R})$. Fix a lift $\widetilde{g}$ of g. Then, for any lift $\widetilde{ghg^{-1}}$ of ghg^{-1} to $\widetilde{PSL}(2,\mathbb{R})$, there exist a lift $\widetilde{h}'$ of h such that $\widetilde{ghg^{-1}} = \widetilde{g}\widetilde{h}'(\widetilde{g})^{-1}$.*

Proof. Fix any lift $\widetilde{ghg^{-1}}$ of ghg^{-1}. Also, choose any lift $\widetilde{h}$ of h. First, note that $p((\widetilde{g})^{-1}) = p(\widetilde{g})^{-1} = g^{-1}$, and hence $(\widetilde{g})^{-1} = \widetilde{g^{-1}}$ for some lift $\widetilde{g^{-1}}$ of g^{-1}. Then, by definition we have that $\widetilde{g}\widetilde{h} = \widetilde{gh}$ for some lift $\widetilde{gh}$ of gh. Finally, we have by definition that $(\widetilde{gh})\widetilde{g^{-1}}$ is a lift of ghg^{-1}, and so there exists $k \in \mathbb{Z}$ such that $(\widetilde{gh})\widetilde{g^{-1}} = z^k \widetilde{ghg^{-1}}$. So, we have that

$$\widetilde{g}(z^{-k}\widetilde{h})(\widetilde{g})^{-1} = z^{-k}\widetilde{g}\widetilde{h}\widetilde{g^{-1}} = z^{-k}(\widetilde{gh})\widetilde{g^{-1}} = \widetilde{ghg^{-1}}.$$

So, take $\widetilde{h}' = z^{-k}\widetilde{h}$. $\square$

Proof of Proposition 3.1.15. Take any lift $\widetilde{h}$ of h. Then, when we compute R_{2g}, we can take lifts of $h\rho(X_i)h^{-1}$ and $h\rho(Y_i)h^{-1}$ to be elements of the form $\widetilde{h}\widetilde{\rho(X_i)}(\widetilde{h})^{-1}$ and $\widetilde{h}\widetilde{\rho(Y_i)}(\widetilde{h})^{-1}$, where $\widetilde{\rho(X_i)}$ and $\widetilde{\rho(Y_i)}$ are lifts of $\rho(X_i)$ and $\rho(Y_i)$. Then,

$$\begin{aligned}
z^{e(h\cdot\rho)} &= R_{2g}(h\rho(X_1)h^{-1}, h\rho(Y_1)h^{-1}, \ldots, h\rho(X_g)h^{-1}, h\rho(Y_g)h^{-1}) \\
&= \prod_{i=1}^{g} \left[\widetilde{h}\widetilde{\rho(X_i)}(\widetilde{h})^{-1}, \widetilde{h}\widetilde{\rho(Y_i)}(\widetilde{h})^{-1}\right] \\
&= \widetilde{h}\left(\prod_{i=1}^{g} \left[\widetilde{\rho(X_i)}, \widetilde{\rho(Y_i)}\right]\right)(\widetilde{h})^{-1} \\
&= \widetilde{h}z^{e(\rho)}(\widetilde{h})^{-1} \\
&= z^{e(\rho)}.
\end{aligned}$$

□

Corollary 3.1.17. *Let S be a closed surface of genus $g \geq 2$. Then, the space $\mathcal{R}' = \mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PSL}(2,\mathbb{R})$ has $2|\chi(S)|+1$ connected components. Moreover, two elements $[\rho_1], [\rho_2] \in \mathcal{R}'$ lie on the same connected component if and only if $e(\rho_1) = e(\rho_2)$.*

Proof. Consider that the map $e : \mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R})) \to \mathbb{Z}$ is continuous, since it is locally constant. Proposition 3.1.15 tells us that if $\rho_1, \rho_2 \in \mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R}))$ are conjugate, then $e(\rho_1) = e(\rho_2)$. Hence e factors through a continuous map $e' : \mathcal{R}' \to \mathbb{Z}$. Since the map e' is continuous and its image contains $2|\chi(S)|+1$ points, we know that $\mathcal{R}'$ has at least $2|\chi(S)|+1$ connected components. On the other hand, since $\mathcal{R}'$ is the continuous image of $\mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R}))$, it has at most $2|\chi(S)|+1$ components. So, $\mathcal{R}'$ has $2|\chi(S)|+1$ connected components. Then, each component of $\mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R}))$ must map to a unique component of $\mathfrak{X}$ under the natural projection, which implies the "moreover" statement. □

Thus, we can finally understand how $\mathrm{Teich}(S)$ fits into $\mathfrak{X}(S)$. Theorem 3.1.14 tells us that the discrete and faithful representations of $\pi_1(S)$ are precisely those of extreme Euler class, and they comprise two connected components of $\mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R}))$. Then, Corollary 3.1.17 tells us that this characterization descends to the quotient $\mathcal{R}(S, \mathrm{PSL}(2,\mathbb{R}))/\mathrm{PSL}(2,\mathbb{R})$, and so $\mathrm{Teich}(S) \sqcup \mathrm{Teich}(\bar{S})$ comprises two connected components of this space. Since discrete and faithful representations are non-elementary, this means that $\mathrm{Teich}(S) \sqcup \mathrm{Teich}(\bar{S})$ comprises two connected components of $\mathfrak{X}(S)$.

3.2 The Mapping Class Group Action on Representation Spaces

3.2.1 Goldman's Conjecture

Throughout this section, fix a closed surface S of genus $g \geq 2$. In Chapter 2, we saw that the mapping class group acts on $\mathrm{Teich}(S)$ as follows: given $[f] \in \mathrm{Mod}(S)$ and $[(X, \phi)] \in \mathrm{Teich}(S)$, we define

$$[f] \cdot [(X, \phi)] = [(X, \phi \circ f^{-1})].$$

However, we are now more interested in viewing $\mathrm{Teich}(S)$ as a subset of $\mathfrak{X} = \mathfrak{X}(S)$. So, we want to extend the action of $\mathrm{Mod}(S)$ onto all of $\mathfrak{X}$. Given $[\rho] \in \mathfrak{X}$ and $[f] \in \mathrm{Mod}(S)$, we define the action of $\mathrm{Mod}(S)$ on $\mathfrak{X}$ by

$$[f] \cdot [\rho] = [\rho \circ f_*^{-1}],$$

where f_* is the induced automorphism of $\pi_1(S)$. Since homotopy classes of maps $S \to S$ correspond to conjugacy classes of homomorphisms $\pi_1(S) \to \pi_1(S)$, this action is well-defined. This action agrees with our original action of $\mathrm{Mod}(S)$ on $\mathrm{Teich}(S)$; namely, given a marked hyperbolic structure $[(X, \phi)]$, the marking $\phi \circ f^{-1}$ corresponds to the representation $\mathrm{hol}_X \circ (\phi \circ f_*^{-1}) = (\mathrm{hol}_X \circ \phi_*) \circ f_*^{-1}$.

Fricke's theorem tells us that the action of $\mathrm{Mod}(S)$ on $\mathrm{Teich}(S)$ is "nice", as it is properly discontinuous. When we proved this, we were viewing $\mathrm{Teich}(S)$ as the space of hyperbolic structures on S, not as a space of representations of $\pi_1(S)$. Since arbitrary representations of $\pi_1(S)$ cannot be interpreted as hyperbolic structures on S, it is not clear that $\mathrm{Mod}(S)$ should act "nicely" on the rest of the space $\mathfrak{X}$. In fact, Goldman conjectured in 2006 that $\mathrm{Mod}(S)$ acts "chaotically" on the rest of $\mathfrak{X}$ [3].

To formally state Goldman's conjecture, we need to clarify what a "chaotic" action is. We are really refering to an ergodic action.

Definition 3.2.1. Suppose a group G acts on a measure space $(X, \mathcal{A}, \mu)$ via measure-preserving transformations (i.e. $\mu(g \cdot A) = \mu(A)$ for all $A \in \mathcal{A}$). We say the action is *ergodic* if $g \cdot A = A$ implies that $\mu(A) = 0$ or $\mu(A^c) = 0$.

Notice that to prove G acts ergodically on X, it suffices to show that G acts ergodically on a set of full measure in X. We will also make use of an alternate characterization of ergodic actions.

Proposition 3.2.2. *A measure-preserving action of a group G on a measure space $(X, \mathcal{A}, \mu)$ is ergodic if and only if any G-invariant measurable function $f : X \to \mathbb{R}$ is almost-everywhere constant.*

Ergodicity is a broad concept from probability theory that arises in many different areas. We mainly think of ergodic actions as those which are the opposite of properly discontinuous actions. This is because properly discontinuous actions have discrete orbits, while ergodic actions have dense orbits (hence why we think of them as "chaotic").

Perhaps ergodicity may seem like a strange notion to introduce at this point. After all, we have only discussed $\mathfrak{X}$ as a topological space; we have yet to give it a measure. Our measure comes from another strong result from Goldman on the topological structure of $\mathfrak{X}$.

Theorem 3.2.3 (Goldman [4])**.** *The space $\mathfrak{X}$ is a smooth symplectic manifold manifold of dimension $6g-6$.*

Here, a *symplectic manifold* refers to a manifold equipped with a closed non-degenerate differential 2-form, called a *symplectic form.* A symplectic form gives rise to a volume form on the manifold, which in turn yields a measure.

Now, we can return to Goldman's conjecture. We partition $\mathfrak{X}$ into subsets $\mathfrak{X}^{2g-2}, \ldots, \mathfrak{X}^{2-2g}$, where $\mathfrak{X}^k$ consists of representations of Euler class k. Recall that the extremal components are those consisting of discrete faithful representations (i.e. Teichmüller spaces). The Euler class is invariant under the action of $\mathrm{Mod}(S)$, so we can restrict the action to each of these subsets.

Conjecture 3.2.4 (Goldman, [3])**.** *The group $\mathrm{Mod}(S)$ acts ergodically on the components $\mathfrak{X}^k$ for nonzero and non-extremal k.*

In 2015, Marché and Wolff proved one case of Goldman's conjecture.

Theorem 3.2.5 (Marché-Wolff, [9])**.** *In the case $g = 2$,* $\mathrm{Mod}(S)$ *acts ergodically on $\mathfrak{X}^{-1}$ and $\mathfrak{X}^1$, and* $\mathrm{Mod}(S)$ *does not act ergodically on $\mathfrak{X}^0$.*

There is a link between Goldman's conjecture and an older question of Bowditch.

Question 3.2.6 (Bowditch, [1])**.** If a $\mathrm{PSL}(2,\mathbb{R})$-representation of $\pi_1(S)$ is not discrete and faithful, does it necessarily map a simple closed curve to a non-hyperbolic element?

In particular, Marché and Wolff answered Bowditch's question affirmatively in the genus 2 case, and formalized the following link between Bowditch's question and Goldman's conjecture (which had previously been well known, but never formally stated).

Theorem 3.2.7 (Marché-Wolff, [9])**.** *Let $\mathcal{NH}^k$ denote the subset of $\mathfrak{X}^k$ consisting of representations which map a simple closed curve to a non-hyperbolic element. If $(g,k) \neq (2,0)$, then the action of* $\mathrm{Mod}(S)$ *on $\mathcal{NH}^k$ is ergodic.*

In other words, Marché and Wolff proved that if the set $\mathcal{NH}^k$ has full measure in $\mathfrak{X}^k$ for non-zero and non-extremal k, then Goldman's conjecture is true.

The case $(g,k) = (2,0)$ is rather exceptional. Marché and Wolff showed that $\mathfrak{X}^0$ is the disjoint union of two non-empty $\mathrm{Mod}(S)$-invariant open sets $\mathfrak{X}^0_{\pm}$; in particular, this means the action on $\mathfrak{X}^0$ is not ergodic. This comes from the behavior of a particular mapping class φ called a *hyperelliptic involution.* If we view a surface of genus g as embedded in $\mathbb{R}^3$ in the standard way, φ is the mapping class which rotates the surface 180° around the axis in Figure 3.1. When $g = 2$, this mapping class has the peculiar property that it preserves every simple closed curve [6]. Marché and Wolff use this fact, along with the fact that representations of Euler class 0 lift to $\mathrm{SL}(2,\mathbb{R})$-representations, to build a continuous $\mathrm{Mod}(S)$-invariant map $\mathfrak{X}^0 \to \{1,-1\}$. They were able to show that elements of $\mathfrak{X}^0_-$ map a simple closed curve to a non-hyperbolic element, and hence the action on $\mathfrak{X}^0_-$ is ergodic. In another paper [8], they used different techniques to show that elements of $\mathfrak{X}^0_+$ also must map a simple closed curve to a non-hyperbolic element. Thus, their results can be summarized as follows.

Theorem 3.2.8 ([9], [8])**.** *In the case $g = 2$, any $[\rho] \in \mathfrak{X}^1 \cup \mathfrak{X}^{-1} \cup \mathfrak{X}^0_+ \cup \mathfrak{X}^0_-$ maps a simple closed curve to a non-hyperbolic element.*

Corollary 3.2.9. *In the case $g = 2$,* $\mathrm{Mod}(S)$ *acts ergodically on $\mathfrak{X}^1$, $\mathfrak{X}^{-1}$, $\mathfrak{X}^0_+$, and $\mathfrak{X}^0_-$.*

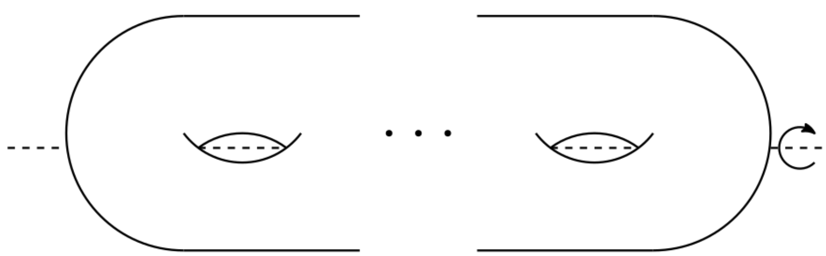

Figure 3.1: The hyperelliptic involution rotates S_g around the axis indicated by the dashed line.

3.2.2 Outline of the Proof of Theorem 3.2.7

Fix any $k \in \{3-2g, \ldots, 2g-3\}$. Let $\mathcal{S}$ denote the set of simple closed curves on S, viewed as a set of conjugacy classes of $\pi_1(S)$, and let $\mathcal{S}^{ns}$ denote the subset of $\mathcal{S}$ consisting of non-separating simple closed curves. To prove Theorem 3.2.7, one wants to show that the action of $\mathrm{Mod}(S)$ on the set

$$\mathcal{NH}^k = \{[\rho] \in \mathfrak{X}^k \mid \text{there exists } [\gamma] \in \mathcal{S} \text{ such that } \rho(\gamma) \text{ is not hyperbolic}\}$$

is ergodic. It suffices to prove that the action is ergodic on a subset of full measure, so the first step is to reduce to a set of representations we can say more about. Namely, we define the set consisting of representations sending a *non-separating* simple closed curve to an elliptic element:

$$\mathcal{E}^k := \{[\rho] \in \mathfrak{X}^k \mid \text{there exists } [\gamma] \in \mathcal{S}^{ns} \text{ such that } \rho(\gamma) \text{ is elliptic}\}.$$

Notice that $\mathcal{E}^k$ is a $\mathrm{Mod}(S)$-invariant subset of $\mathcal{NH}^k$. One shows the following.

Proposition 3.2.10. *If $(g,k) \neq (2,0)$, then $\mathcal{E}^k$ has full measure in $\mathcal{NH}^k$.*

To prove this, we further reduce to the set

$$\mathcal{EI}^k := \{[\rho] \in \mathfrak{X}^k \mid \text{there exists } [\gamma] \in \mathcal{S}^{ns} \text{ such that } \rho(\gamma) \text{ is elliptic of infinite order}\}.$$

Since there are only countably-many conjugacy classes of finite-order elliptic isometries, and only countably-many simple closed curves on S, the set $\mathcal{EI}^k$ has full measure in $\mathcal{E}^k$. So, it's enough to show that $\mathcal{EI}^k$ has full measure in $\mathcal{NH}^k$. We consider another set:

$$\mathcal{N} := \{[\rho] \in \mathfrak{X} \mid \text{for all } [\gamma] \in \mathcal{S},\ \rho(\gamma) \text{ is not parabolic or elliptic of finite order}\}.$$

For a given $\gamma \in \mathcal{S}$, the set of $[\rho]$ for which $\rho(\gamma)$ is parabolic is the zero set of the non-constant algebraic function $[\rho] \mapsto \mathrm{tr}([\rho])^2 - 4$. It follows that $\mathcal{N}$ has full measure in $\mathfrak{X}$. So, to prove Proposition 3.2.10, it suffices to show that $\mathcal{EI}^k \cap \mathcal{N}$ has full measure in $\mathcal{NH}^k \cap \mathcal{N}$. In fact, one can show that these two sets are simply equal.

To prove that $\mathcal{EI}^k \cap \mathcal{N} = \mathcal{NH}^k \cap \mathcal{N}$, one has to show that if if $[\rho] \in \mathfrak{X}^k$ maps a separating simple closed curve to an infinite-order elliptic element, then $[\rho]$ must also map some non-separating simple closed curve to an infinite-order elliptic. To show this, one starts in the case $g = 2$. Let X_1, X_2, Y_1, and Y_2 denote the standard generators of $\pi_1(S)$, and let $A_i = \rho(X_i)$ and $B_i = \rho(Y_i)$. We can assume that the elements A_i and B_i are hyperbolic (or else we are done, since X_i and Y_i are non-separating). Then, up to the action of the mapping class group, $[X_1, Y_1]$ is the only separating curve on S, so we can assume $C = [A_1, B_1]$ is elliptic. Then, the main idea is that one can choose N so that the axes of A_1 and $C^N A_2 C^{-N}$ are positioned so that $A_1 C^N A_2 C^{-N}$ is elliptic. So, $X_1[X_1, Y_1]^N X_2 [X_2, Y_2]^{-N}$ is the desired non-separating curve. The case that $g \geq 3$ follows similarly, but it is not as straightforward to choose curves playing the same role as X_1 and X_2 in the $g = 2$ case.

So, it remains for one to show the following.

Theorem 3.2.11. *The action of $\mathrm{Mod}(S)$ on $\mathcal{E}^k$ is ergodic.*

To do this, we'll introduce one more set into the fray. Given $[\gamma] \in \mathcal{S}$, we define the function $f_\gamma : \mathfrak{X}^k \to \mathbb{R}_+$ by $f_\gamma([\rho]) = \mathrm{tr}(\rho(\gamma))^2$ (we take the trace squared since $\mathrm{tr}(\rho(\gamma))$ is not well-defined). Then, we define $\mathcal{U}^k$ to be the set of $[\rho] \in \mathcal{E}^k$ where there exists curves $\gamma_1, \ldots, \gamma_{6g-6} \in \mathcal{S}$ such that:

(i) The elements $\rho(\gamma_i)$ are all elliptic.

(ii) The differentials df_{γ_i} generate $T^*_{[\rho]}\mathfrak{X}^k$.

We care about the set $\mathcal{U}^k$ because of the following fact, which is the key reason why the action on $\mathfrak{X}^k$ is ergodic.

Proposition 3.2.12. *Suppose $f : \mathcal{U}^k \to \mathbb{R}$ is a $\mathrm{Mod}(S)$-invariant measurable function. Then, for every $[\rho] \in \mathcal{U}^k$, there exists a neighborhood $V_{[\rho]}$ of $[\rho]$ on which f is almost everywhere constant.*

Proof Sketch. The proof of this fact takes advantage of the symplectic form ω on $\mathfrak{X}^k$. Fix $[\rho] \in \mathcal{U}^k$. For each function f_{γ_i}, we define the function $h_i : \mathfrak{X}^k \to \mathbb{R}$ by

$$h_i = \arccos\left(\frac{\sqrt{f_i}}{2}\right).$$

In particular, if ϕ is a representation such that $\phi([\gamma_i])$ is elliptic, then $h_i([\phi])$ is the rotation angle of $\phi([\gamma_i])$. For each h_i, the non-degeneracy of ω gives rise to a vector field X_i called the *Hamiltonian vector field* of h_i; this is the unique vector field satisfying $dh_i(Y) = \omega(X_i, Y)$ for all vector fields Y on $\mathfrak{X}^k$. Let Φ_i^t denote the time t flow of X_i. One can show that Φ_i^t is 2π-periodic in t, and if τ_i is the Dehn twist along γ_i, then

$$\tau_i \cdot [\phi] = \Phi_i^{h_i([\phi])}([\phi]) \tag{$\star$}$$

for all $[\phi] \in \mathfrak{X}^k$. So, if $h_i([\phi]) \notin 2\pi\mathbb{Q}$, then f is almost everywhere constant on the $\mathrm{Mod}(S)$-orbit of $[\phi]$. Now, we can choose a neighborhood $V_{[\rho]}$ of $[\rho]$ defined by the following property: if $[\phi] \in V_{[\rho]}$, then $\phi(\gamma_i)$ is elliptic for all i and the vectors $X_i([\phi])$ span $T_{[\phi]}\mathfrak{X}^k$. A consequence of the latter condition is that if one takes $V_{[\rho]}$ small enough, the flows Φ_i act transitively on $V_{[\rho]}$. By $(\star)$, f is almost everywhere constant on almost every orbit of $\mathrm{Mod}(S)$. It follows that f is almost everywhere constant on $V_{[\rho]}$. $\square$

There are two more facts which one uses to prove Theorem 3.2.11.

Proposition 3.2.13. *The space $\mathcal{EI}^k$ is connected.*

Proof Sketch. To prove this, one considers the set $\mathcal{C}$ of pairs $(a, b) \in \mathcal{S}$ such that $i(a, b) = 1$. We define the set $\mathcal{EI}^k_{(a,b)}$ to consist of representations $[\rho] \in \mathcal{EI}^k$ such that $\rho(a)$ and $\rho(b)$ are non-commuting elliptic elements, at least one of which has infinite order. One can check that $\mathcal{EI}^k$ is the union of all these sets $\mathcal{EI}^k_{(a,b)}$. Moreover, these restrictions are enough to directly parametrize $\mathcal{EI}^k_{(a,b)}$ as a subset of $\mathbb{R}^3$, which makes it easy to prove that this set is connected. So, we define an equivalence relation $\sim$ on $\mathcal{C}$ generated by the relation that $(a, b) \sim (a', b')$ if $\mathcal{EI}^k_{(a,b)} \cap \mathcal{EI}^k_{(a',b')} \neq \varnothing$. It then suffices to prove that $\sim$ has a single equivalence class, which is a combinatorial proof similar in flavor to the proof that the curve graph is connected. $\square$

Proposition 3.2.14. *The set $\mathcal{U}^k$ is an open subset of $\mathfrak{X}^k$ containing $\mathcal{EI}^k$.*

Proof Sketch. To prove this, take any $[\rho] \in \mathcal{EI}^k$, and let $[\gamma] \in \mathcal{S}^{ns}$ such that $\rho(\gamma)$ is elliptic of infinite order. Let $D_{[\rho]} \subseteq T^*_{[\rho]}\mathfrak{X}^k$ be the subspace generated by differentials df_δ of traces of curves δ where $\rho(\delta)$ is elliptic. Then, one wants to prove that $D_{[\rho]} = T^*_{[\rho]}\mathfrak{X}^k$. The strategy is to take any vector $\xi \in T_{[\rho]}\mathfrak{X}^k$ which is mapped to zero by all elements of $D_{[\rho]}$, and show that $\xi = 0$.

Now, we know $df_\gamma(\xi) = 0$, which implies that ξ is tangent to the subspace $\mathfrak{X}_\theta := f_\gamma^{-1}(4\cos(\theta))$, where θ is the rotation angle of $\rho(\gamma)$. We let $r : \mathfrak{X}_\theta(S) \to \mathfrak{X}_\theta(S\backslash\gamma)$ be the restriction map. One can show that the cotangent space of $\mathfrak{X}(S\backslash\gamma)$ is generated by differentials df_δ where δ is disjoint from γ. Moreover, one can show that such differentials lie in $D_{[\rho]}$. Hence $\xi \in \mathrm{Ker}(dr)$. On the other hand, it follows from Equation $(\star)$ that $\mathrm{Ker}(dr)$ is the span of X_γ, where X_γ is the Hamiltonian vector field of f_γ. So, we can write $\xi = \lambda X_\gamma$. Finally, one can show that there exists a curve δ such that $df_\delta(\xi) = 0$ but $df_\delta(X_\gamma) \neq 0$. Thus, it must be that $\xi = 0$. $\square$

Now, armed with Propositions 3.2.12, 3.2.13, and 3.2.14, the proof of Theorem 3.2.11 follows rather quickly.

Proof of Theorem 3.2.11. Suppose $f : \mathcal{E}^k \to \mathbb{R}$ is a $\mathrm{Mod}(S)$-invariant measurable function. We want to show that f is almost everywhere constant. By Proposition 3.2.14, $\mathcal{EI}^k \subseteq \mathcal{U}^k \subseteq \mathcal{E}^k$. Since $\mathcal{EI}^k$ has full measure in $\mathcal{E}^k$, the same must be true for $\mathcal{U}^k$, so it suffices to show that the restriction $f : \mathcal{U}^k \to \mathbb{R}$ is almost everywhere constant. By Proposition 3.2.12, we can define a function $g : \mathcal{U}^k \to \mathbb{R}$ which is locally constant and agrees with f almost everywhere (namely, let $g([\rho])$ be the constant to which f is almost everywhere equal on $V_{[\rho]}$). By Propositions 3.2.13 and 3.2.14, $\mathcal{U}^k$ is connected and hence g is constant. So, f is almost everywhere constant. $\square$

Bibliography

[1] B.H. Bowditch. “Markoff Triples and Quasifuchsian Groups”. In: *Proceedings of the London Mathematical Society* 77 (1998), pp. 697–736. DOI: 10.1112/s0024611598000604.

[2] Benson Farb and Dan Margalit. *A Primer on Mapping Class Groups*. Princeton University Press, 2012.

[3] William Goldman. “Mapping Class Group Dynamics on Surface Group Representations”. In: *Proceedings of Symposia in Pure Mathematics Problems on Mapping Class Groups and Related Topics* (2006), pp. 189–214. DOI: 10.1090/pspum/074/2264541.

[4] William Goldman. “The Symplectic Nature of Fundamental Groups of Surfaces”. In: *Advances in Mathematics* 54.2 (1984), pp. 200–225. DOI: 10.1016/0001-8708(84)90040-9.

[5] William Goldman. “Topological Components of Spaces of Representations”. In: *Inventiones Mathematicae* 93 (1988), pp. 557–607. DOI: 10.1007/bf01410200.

[6] Andrew Haas and Perry Susskind. “The Geometry of the Hyperelliptic Involution in Genus Two”. In: *Proceedings of the American Mathematical Society* 105 (Jan. 1989), pp. 159–159. DOI: 10.1090/s0002-9939-1989-0930247-2.

[7] Sean Lawton and Adam Sikora. “Varieties of Characters”. In: *Algebras and Representation Theory* 20 (Mar. 2017), pp. 1133–1141. DOI: 10.1007/s10468-017-9679-y.

[8] Julien Marché and Maxime Wolff. “Six-point Configurations in the Hyperbolic plane and Ergodicity of the Mapping Class Group”. In: *Groups, Geometry, and Dynamics* 13.2 (June 2019), pp. 731–766. DOI: 10.4171/ggd/503.

[9] Julien Marché and Maxime Wolff. “The modular action on $\mathrm{PSL}_2(\mathbb{R})$ -characters in genus 2”. In: *Duke Mathematical Journal* 165 (2016), pp. 371–412. DOI: 10.1215/00127094-3166522.

[10] William S. Massey. *Algebraic Topology: An Introduction*. Springer, 1996.

[11] Frédéric Palesi. “Connected Components Of PGL(2,$\mathbb{R}$)-Representation Spaces Of Non-Orientable Surfaces”. In: *Geometry, Topology and Dynamics of Character Varieties Lecture Notes Series, Institute for Mathematical Sciences, National University of Singapore* (2012), pp. 281–295. DOI: 10.1142/9789814401364_0008.

[12] Caroline Series. *Hyperbolic Geometry MA 448*. 2010. URL: http://homepages.warwick.ac.uk/~masbb/Papers/MA448.pdf.

[13] Adam Sikora. “Character Varieties”. In: *Transactions of the American Mathematics Society* 364 (2012), pp. 5173–5208.